OULU 2015
C 537

ACTA UNIVERSITATIS OULUENSIS

Jukka Komulainen

SOFTWARE-BASED COUNTERMEASURES TO 2D FACIAL SPOOFING ATTACKS

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING,
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING;
INFOTECH OULU

ACTA UNIVERSITATIS OULUENSIS
C Technica 537

JUKKA KOMULAINEN

SOFTWARE-BASED COUNTERMEASURES TO 2D FACIAL SPOOFING ATTACKS

Academic dissertation to be presented, with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu, for public defence in the Wetteri auditorium (IT115), Linnanmaa, on 21 August 2015, at 12 noon

UNIVERSITY OF OULU, OULU 2015

Acta Univ. Oul. C 537, 2015

Supervised by
Professor Matti Pietikäinen
Docent Abdenour Hadid

Reviewed by
Associate Professor Arun Ross
Associate Professor Julian Fierrez

Opponent
Associate Professor Joni Kämäräinen

ISBN 978-952-62-0872-5 (Paperback)
ISBN 978-952-62-0873-2 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2015

Komulainen, Jukka, Software-based countermeasures to 2D facial spoofing attacks.
University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering, Department of Computer Science and Engineering; Infotech Oulu
Acta Univ. Oul. C 537, 2015
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Because of its natural and non-intrusive interaction, identity verification and recognition using facial information is among the most active areas in computer vision research. Unfortunately, it has been shown that conventional 2D face recognition techniques are vulnerable to spoofing attacks, where a person tries to masquerade as another one by falsifying biometric data and thereby gaining an illegitimate advantage.

This thesis explores different directions for software-based face anti-spoofing. The proposed approaches are divided into two categories: first, low-level feature descriptors are applied for describing the static and dynamic characteristic differences between genuine faces and fake ones in general, and second, complementary attack-specific countermeasures are investigated in order to overcome the limitations of generic spoof detection schemes.

The static face representation is based on a set of well-known feature descriptors, including local binary patterns, Gabor wavelet features and histogram of oriented gradients. The key idea is to capture the differences in quality, light reflection and shading by analysing the texture and gradient structure of the input face images. The approach is then extended to the spatiotemporal domain when both facial appearance and dynamics are exploited for spoof detection using local binary patterns from three orthogonal planes.

It is reasonable to assume that no generic spoof detection scheme is able to detect all known, let alone unseen, attacks scenarios. In order to find out well-generalizing countermeasures, the problem of anti-spoofing is broken into two attack-specific sub-problems based on whether the spoofing medium can be detected in the provided view or not. The spoofing medium detection is performed by describing the discontinuities in the gradient structures around the detected face. If the display medium is concealed outside the view, a combination of face and background motion correlation measurement and texture analysis is applied. Furthermore, an open-source anti-spoofing fusion framework is introduced and its system-level performance is investigated more closely in order to gain insight on how to combine different anti-spoofing modules.

The proposed spoof detection schemes are evaluated on the latest benchmark datasets. The main findings of the experiments are discussed in the thesis.

Keywords: anti-spoofing, biometrics, computer vision, countermeasure, face recognition, liveness detection, presentation attack, spoofing

Komulainen, Jukka, Ohjelmistopohjaisia keinoja estää näkyvän valon alueella toimivien kasvontunnistusjärjestelmien huijaaminen.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Tieto- ja sähkötekniikan tiedekunta, Tietotekniikan osasto; Infotech Oulu
Acta Univ. Oul. C 537, 2015
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Kasvokuvaan perustuvan henkilöllisyyden tunnistamisen etuja ovat luonnollinen vuorovaikutus ja etätunnistus, minkä takia aihe on ollut erittäin aktiivinen tutkimusalue konenäön tutkimuksessa. Valitettavasti tavanomaiset kasvontunnistustekniikat ovat osoittautuneet haavoittuvaisiksi hyökkäyksille, joissa kameralle esitetään jäljennös kohdehenkilön kasvoista positiivisen tunnistuksen toivossa.

Tässä väitöskirjassa tutkitaan erilaisia ohjelmistopohjaisia ratkaisuja keinotekoisten kasvojen ilmaisuun petkuttamisen estämiseksi. Työn ensimmäisessä osassa käytetään erilaisia matalan tason piirteitä kuvaamaan aitojen ja keinotekoisten kasvojen luontaisia staattisia ja dynaamisia eroavaisuuksia. Työn toisessa osassa esitetään toisiaan täydentäviä hyökkäystyyppikohtaisia vastakeinoja, jotta yleispätevien menetelmien puutteet voitaisiin ratkaista ongelmaa rajaamalla.

Kasvojen staattisten ominaisuuksien esitys perustuu yleisesti tunnettuihin matalan tason piirteisiin, kuten paikallisiin binäärikuvioihin, Gabor-tekstuureihin ja suunnattujen gradienttien histogrammeihin. Pääajatuksena on kuvata aitojen ja keinotekoisten kasvojen laadun, heijastumisen ja varjostumisen eroavaisuuksia tekstuuria ja gradienttirakenteita analysoimalla. Lähestymistapaa laajennetaan myös tila-aika-avaruuteen, jolloin hyödynnetään samanaikaisesti sekä kasvojen ulkonäköä ja dynamiikkaa irroittamalla paikallisia binäärikuvioita tila-aika-avaruuden kolmelta ortogonaaliselta tasolta.

Voidaan olettaa, ettei ole olemassa yksittäistä yleispätevää vastakeinoa, joka kykenee ilmaisemaan jokaisen tunnetun hyökkäystyypin, saati tuntemattoman. Näin ollen työssä keskitytään tarkemmin kahteen hyökkäystilanteeseen. Ensimmäisessä tapauksessa huijausapuvälineen reunoja ilmaistaan analysoimalla gradienttirakenteiden epäjatkuvuuksia havaittujen kasvojen ympäristössä. Jos apuvälineen reunat on piilotettu kameran näkymän ulkopuolelle, petkuttamisen ilmaisu toteutetaan yhdistämällä kasvojen ja taustan liikkeen korrelaation mittausta ja kasvojen tekstuurianalyysiä. Lisäksi työssä esitellään vastakeinojen yhdistämiseen avoimen lähdekoodin ohjelmisto, jonka avulla tutkitaan lähemmin menetelmien fuusion vaikutuksia.

Tutkimuksessa esitetyt menetelmät on kokeellisesti vahvistettu alan viimeisimmillä julkisesti saatavilla olevilla tietokannoilla. Tässä väitöskirjassa käydään läpi kokeiden päähavainnot.

Asiasanat: biometriikka, eloisuuden ilmaisu, huijaamisen esto, hyökkäys, kasvontunnistus, konenäkö, petkuttaminen, vastakeino

Preface

The research work of this thesis was carried out in the Center for Machine Vision Research of the Department of Computer Science and Engineering at the University of Oulu, Finland, between 2010 and 2015.

First of all, I would like to express my gratitude to my supervisors, Prof. Matti Pietikäinen and Adjunct Prof. Abdenour Hadid, for their support and guidance. I am especially grateful for their trust and patience giving me the freedom to try out my own ideas and gradually become an independent researcher.

I would also like to thank Dr. Sébastien Marcel for hosting my research visit to Idiap Research Institute in Martigny. Further, I want to acknowledge the co-authors of the papers, Tiago de Freitas Pereira, Dr. André Anjos and Prof. José Mario De Martino, for their hard work and fruitful collaboration. I especially want to thank Tiago, André and other members of Biometrics group at Idiap for making my stay at Martigny not only work oriented.

The Center for Machine Vision Research has been an excellent place for doing research, not only because of the talented and helpful personnel, but also because of the very pleasant and easy-going atmosphere. For that, I want to thank the whole group, both research and support staff. However, all the members of kahvilinjasto deserve honourable mention for all the different activities and profound discussions which were also occasionally related to the matters computer vision.

I would like to gratefully acknowledge the official reviewers, Associate Prof. Arun Ross and Associate Prof. Julian Fierrez for their constructive comments and feedback. I also want to thank Dr. Pertti Väyrynen for helping with the language revision on such a short notice.

I am grateful to Infotech Oulu Doctoral Program and the EC project TABULA RASA for the financial support that has given me the possibility to focus on the same research topic throughout my doctoral studies and to compile a unified thesis. Also the financial support I have gotten from HPY Research Foundation, Nokia Foundation, Walter Ahlström Foundation, Tauno Tönning Foundation, Finnish Foundation for Technology Promotion (TES), UniOGS and the Academy of Finland is gratefully acknowledged.

I want to thank Ulla Eerikäinen, Juha Lamminaho, Juha Jylhänlahti, Marko Matero and other specialists who have fixed body every now and then and kept me in working order.

Finally, I want to express my deepest gratitude to my family and friends. My parents, Paula and Mauri, have always supported me and encouraged me to try out many different things in my life. Most of all, I would like to thank my wife Tiina for her love, support, understanding, and especially for taking care of our daughter Immi during the most intensive working hours past year. And Immi, despite the sleepless nights, you have been the perfect distraction for getting my thoughts off the thesis.

Oulu, June 2015

Abbreviations

3DMAD	*3D Mask Attack Database*
AUC	*Area Under Curve*
CNN	*Convolutional Neural Network*
CRF	*Conditional Random Field*
DoG	*Difference of Gaussian*
FASD	*Face Anti-Spoofing Database*
EER	*Equal Error Rate*
FAR	*False Accept Rate*
FFR	*False Fake Rate*
FLR	*False Living Rate*
FRR	*False Rejection Rate*
GLCM	*Gray-Level Co-occurrence Matrix*
HOG	*Histogram of Oriented Gradients*
HOOF	*Histogram of Oriented Optical Flow*
IBG	*International Biometric Group*
ICAO	*International Civil Aviation Organization*
IR	*Thermal Infrared*
LBP	*Local Binary Pattern*
LBP^{u2}	*Uniform local binary pattern operator*
LBP-TOP	*Local Binary Patterns from Three Orthogonal Planes*
LDA	*Linear Discriminant Analysis*
LLR	*Linear Logistic Regression*
MFSD	*Mobile Face Spoofing Database*
MLP	*Multi-Layer Perceptron*
NIR	*Near-Infrared*
PID	*Photograph Imposter Database*
SFAR	*Spoofing False Acceptance Rate*
SIFT	*Scale-Invariant Feature Transform*
SVM	*Support Vector Machine*

List of original publications

This dissertation is based on the following articles, which are referred to in the text by their Roman numerals (I–VII):

I Anjos A, Komulainen J, Marcel S, Hadid A & Pietikäinen M (2014) Face anti-spoofing: Visual approach. In: S. Marcel, M.S. Nixon & S.Z. Li, (eds) Handbook of Biometric Anti-Spoofing, Springer Verlag, 65-82.

II Määttä J, Hadid A & Pietikäinen M (2011) Face spoofing detection from single images using micro-texture analysis. Proc. International Joint Conference on Biometrics (IJCB).

III Määttä J, Hadid A & Pietikäinen M (2012) Face spoofing detection from single images using texture and local shape analysis. IET Biometrics, 1(1):3-10.

IV Komulainen J, Hadid A & Pietikäinen M (2013) Face spoofing detection using dynamic texture. In: ACCV 2012 Workshops, Part I (LBP 2012), Lecture Notes in Computer Science, 7728:146-157.

V de Freitas Pereira T, Komulainen J, Anjos A, De Martino JM, Hadid A, Pietikäinen M & Marcel S (2014) Face liveness detection using dynamic texture. EURASIP Journal on Image and Video Processing, 2014:2.

VI Komulainen J, Hadid A & Pietikäinen M (2013) Context based face anti-spoofing. Proc. the IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS).

VII Komulainen J, Anjos A, Hadid A, Marcel S & Pietikäinen M (2013) Complementary countermeasures for detecting scenic face spoofing attacks. Proc. IAPR International Conference on Biometrics (ICB).

The author of the dissertation is the first author in articles II-IV, VI and VII. The main responsibility for writing and experiments was carried out by the present author, while valuable comments and suggestions were given by the co-authors. Dr. André Anjos was closely involved in the designing and discussing the experiments conducted in Papers V and VII. The author wrote the literature review and discussion presented in Paper I. Paper V was a joint work with the first author Tiago de Freitas Pereira with whom the present author shared an equal role in preparing the article.

Contents

1 Introduction

1.1 Background and motivation

There is a need for large-scale identity management solutions that are able to accurately determine the identity of an individual in the context of several different applications, including online banking, border and access control, just to name a few. Biometrics aims at uniquely recognizing individuals based on their physiological, behavioural and chemical attributes, such as face, fingerprint, iris, voice, gait and DNA. Biometric authentication offers several advantages over traditional schemes, i.e. knowledge-based (e.g. passwords) and token-based (e.g. ID cards) mechanisms. For instance, people tend to use simple passwords because complex passwords are hard to remember, and, even worse, the same password is usually utilized across different applications for the same reason. Where passwords and ID cards can be easily lost, shared, manipulated or stolen, by using biometrics, it is possible to confirm or establish identity based on who the individual is rather than what the individual possesses or what the individual remembers (Jain *et al.* 2008).

Identity management for persons using biometrics has indeed become a reality not only because of the biometric passport (e-passport), but also because of the presence of more and more biometric-enabled applications for personal computers and mobile phones. For instance, Apple released very recently its mobile payment and digital wallet service, Apple Pay, that relies on Touch ID fingerprint recognition feature included in the latest iPhones. Furthermore, a growing number of developing countries are using biometric technologies to create national identification programs, e.g. for voter registration or to bring benefits to the poor. The most ambitious large-scale project was initiated by Unique Identification Authority of India that is implementing the scheme of providing a unique ID (AADHAAR[1] number) for every Indian resident.

Unfortunately, it has been shown that conventional biometric techniques, such as fingerprint and face recognition, are vulnerable to spoofing attacks, where a person tries to masquerade as another one by falsifying a biometric trait of the targeted user and presenting it to the input sensor, and thereby, gaining an illegitimate advantage. Among the different vulnerabilities, spoofing is purely a biometric vulnerability because,

[1] `http://uidai.gov.in/`

e.g. unlike passwords, our biometric traits, e.g. face, iris, fingerprints and even DNA, are widely available and some of them are easy to sample, which is probably the most crucial drawback of biometric authentication (Galbally *et al.* 2014). Since spoofing attacks are performed at the sensor level, i.e. outside the control of the biometric system manufacturer, no digital protection mechanisms can be used against them. This type of attack is therefore very easy to reproduce and has great potential to succeed, thus every user can be seen as a potential attacker.

Without dedicated countermeasures, most of the existing biometric systems are vulnerable to spoofing because they try to maximize the discriminability between identities without regards to whether the presented trait originates from a living legitimate client or not. The latest alarming example of the vulnerability to presentation attacks was exposed by a group of hackers who demonstrated that a fingerprint of the phone user, photographed from a glass surface, was enough to manufacture a gummy finger that could unlock an iPhone 5s secured with TouchID. According to mobile security firm Lookout[2], the same vulnerability exists in the newest iPhone models (6 and 6 Plus), which made the release of Apple Pay a bit awkward.

Currently, spoofing attacks are one of the main problems for companies willing to market identity management solutions based on biometric technologies. The existing biometrics based authentication systems are beginning to be mature enough for consumer-level applications but the lack of public confidence due to vulnerabilities to presentation attacks prevents biometric authentication from making its breakthrough. Thus, there is an urgent need to find efficient and reliable solutions for detecting and circumventing these direct sensor-level attacks and to bring the technology into practical use.

1.2 Contributions of the thesis

This thesis focuses on exploring software-based approaches for improving the robustness of 2D face authentication systems to spoofing attacks. All proposed methods are based on analysis of single image or short video sequences captured with conventional cameras, i.e. the sensor embedded in the face verification system that acquires the samples for the actual recognition. The main contributions of the thesis are listed below:

[2] `https://blog.lookout.com/blog/2014/09/23/iphone-6-touchid-hack/`

- Static facial texture and local shape analysis based methods are proposed for describing the inherent differences between genuine faces and fake ones, including shading, reflectance and texture (quality).
- The texture based spoof detection is extended also to the spatiotemporal domain in order to exploit the dynamics of facial texture information in addition to the appearance.
- A context based method is presented for detecting the boundaries of the used display medium by describing the discontinuities in the gradient structures around the observed face.
- An open-source fusion framework is introduced for combining different countermeasures.
- The fusion framework is used for investigating more closely the effects on combining countermeasures under scenic attacks, i.e. the display medium is concealed outside the provided view, in order to gain insight into how different anti-spoofing modules could be coupled.
- An extensive literature review on software-based spoof detection schemes is presented and future directions are discussed.

It is also worth mentioning that the texture based face anti-spoofing has been adopted by several other researchers[3]. Although the local binary pattern (LBP) based facial representation proposed in Paper II was originally developed for detecting print and video replay attacks, it has later shown to be useful also in 3D mask attack detection (Erdogmus & Marcel 2013b, Kose & Dugelay 2014). Furthermore, some kind of LBP based face representation was used as a component in all best-performing algorithms in the 2nd competition on counter measures to 2D facial spoofing attacks (Chingovska *et al.* 2013b). In 2013, Paper III was recognized with IET Biometrics Premium Award which is given to the best research paper published within the last two years in IET Biometrics journal.

1.3 Summary of the original articles

Paper I is a book chapter that gives an overall introduction to the problem of spoofing 2D face authentication systems. More specifically, the different vulnerabilities are introduced and an extensive literature review on the state-of-the-art of software-based

[3] At the time of writing, Google Scholar lists 68 citations of Paper II.

countermeasures to spoofing is given, including discussion on the current state of face anti-spoofing and its future directions.

Papers II-V describe the application of capturing the static and dynamic characteristic disparities between genuine faces and fake ones. Paper II introduces a micro-texture analysis based face representation for print attack detection. Supplementary texture and local shape based features are included to the face description in Paper III and experiments on two additional benchmark datasets are conducted. The key idea of the static face representation is to capture the inherent differences in quality, light reflection and shading by analysing the texture and gradient structure of the given face samples using a set of well-known features descriptors, including LBP, Gabor wavelet features and histogram of oriented gradients (HOG).

Paper IV extends the micro-texture based face description into the spatiotemporal domain. In addition to analysing the facial appearance based cues, dynamic texture analysis is applied for exploiting specific temporal cues, including facial movements (liveness) and dynamic motion and quality artefacts originated from the display medium, for anti-spoofing. Dynamic texture based face spoof detection was simultaneously proposed by two independent research groups. The work in Paper IV and by de Freitas Pereira *et al.* (2013a) is consolidated in Paper V which provides an in-depth analysis on the use of dynamic texture based face liveness description using a unified experimental setup and evaluation methodology.

Since it is reasonable to assume that no superior spoof detection technique is able to cope with all known, let alone unseen, attack scenarios, Paper VI and Paper VII approach the problem of face spoofing from the attack-specific point of view. In other words, the two complementary attack-specific sub-problems are considered based on whether the boundaries of the used display medium can be detected in the view and different spoof detection schemes are proposed accordingly for each scenario. Paper VI describes a method exploiting contextual information for detecting the display medium in the provided scene, whereas Paper VII concentrates on scenarios where the support medium is concealed outside the view by applying complementary countermeasures, face and background motion correlation and texture, for detecting these scenic attacks. In addition, an open-source fusion framework is released in Paper VII and the fusion of countermeasures is studied more closely in order to gain insight into how different anti-spoofing modules could be coupled.

1.4 Outline of the thesis

This thesis is divided into two parts. The first part consisting of five chapters gives an overview of the state of the art in software-based anti-spoofing and summarizes the main ideas and findings of the original papers. The original papers form the other part of this thesis.

Chapter 1 is an introduction to the thesis presenting the background of the topic, scope and contributions of this thesis, as well as a brief summary of the original papers. Chapter 2 gives an overview on face anti-spoofing, including an introduction to the vulnerabilities of face authentication systems, a literature review of the state-of-the-art techniques and description of the publicly available databases used in the experiments of this thesis.

Chapters 3 and 4 present the new approaches proposed in the original papers. Chapter 3 describes first the key ideas behind texture based face anti-spoofing and, then, summarizes the main findings in print and video attack detection. Chapter 4 presents the rationale behind the proposed attack-specific categorization and approaches for exploiting the visual cues in each attack scenario. Finally, experimental results are shown for each scenario, including a study on combining complementary countermeasures.

Chapter 5 concludes the thesis by discussing directions for future research in face anti-spoofing. Finally, Chapter 6 summarizes the thesis.

2 Overview on face anti-spoofing

This chapter gives first a general introduction to the vulnerabilities of biometric systems and anti-spoofing. Then, the scope is narrowed down to 2D face modality and various aspects of face anti-spoofing are introduced, including different attack scenarios and the state of the art in face spoof detection with particular focus on software-based countermeasures. Finally, face spoofing related benchmark datasets are introduced which have been the basis of the work in this thesis.

2.1 Introduction to biometric anti-spoofing

Between presenting the biometric trait and the final decision, there are various points in the information flow (see Figure 1) where the security of a biometric authentication system can be compromised (Nixon *et al.* 2008). The different attacks can be broadly divided into two classes: indirect attacks and direct attacks (Ratha *et al.* 2001). Indirect attacks are performed at inside the biometric system by intruders, such as cyber-criminal hackers, who attempt to bypass the feature extractor or the matcher (referred to as 3 and 5 in Figure 1), or tamper the templates or models in the database (referred to as 6 in Figure 1, or exploit the possible weaknesses in the communication channels (referred to as 2, 4, 7 and 8 in Figure 1). System security against this kind of attacks can be increased using the conventional digital protection mechanisms, such as firewalls, anti-virus software, intrusion detection and encryption. Indirect attacks fall out of the scope of the present thesis, thus readers are referred to work by Buhan & Hartel (2005) and Roberts (2007), for instance.

Direct attacks (also referred to as presentation attacks) are performed outside the biometric system simply by presenting a biometric trait to the input sensor (referred to as 1 in Figure 1). The nature of presentation attacks can have huge variations from human traits to synthetic artefacts, e.g. a gummy finger, a cosmetic contact lens or a video screen. In the case of fingerprints, the direct attacks can be performed by breathing on a sensor when a latent print of the previous user might be reactivated, or by using something as sinister as a dismembered finger (Nixon *et al.* 2008).

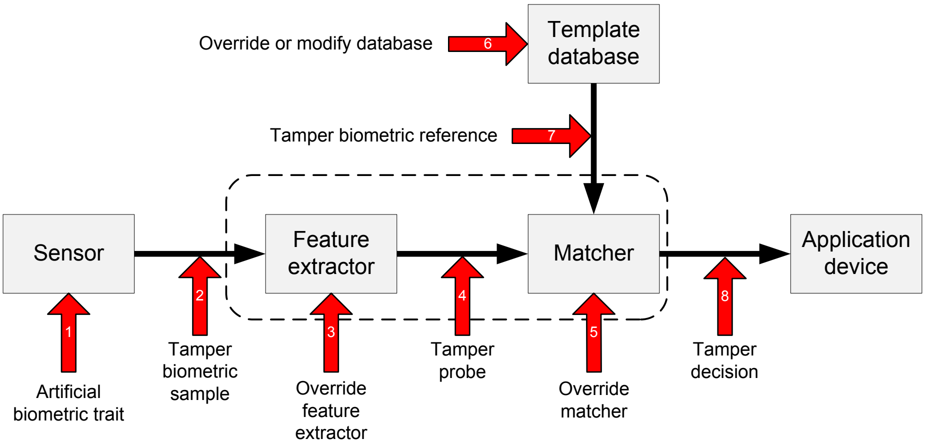

Fig 1. Possible attack points in a generic biometric system, inspired by Ratha *et al.* (2001).

2.1.1 Presentation attacks

Despite the recent huge efforts, the biometric community has not yet managed to agree on the best terminology for vulnerability related concepts, such as presentation attacks, but the standardization is nevertheless a very active process at the moment (Galbally *et al.* 2014). The motivation behind the attacks, i.e. false negative or false positive identification, and the origin of the presented biometric trait, i.e. artificial or human-based, are important factors of presentation attacks when investigating appropriate solutions to counter the different ways for circumventing the biometric systems. For instance, the focus of this thesis is on detecting artificial face biometric traits that are used for obtaining targeted false positive identification. Thus, one intuitive way is to divide presentation attacks into three subcategories: impersonation, obfuscation and spoofing (Jain *et al.* 2011).

Impersonation refers to the scenario where an imposter tries to intrude the system simply by posing as himself (zero-effort attack), i.e. the attacker attempts to gain positive identification by exploiting the false acceptance rate (FAR) of the biometric system. In targeted impersonation, the attacker tries to mimic the behaviour of a specific enrolled identity (e.g. gait, signature or voice). In contrast, an impostor can obfuscate his or her genuine biometric trait in order to deliberately attempt to evade recognition (or detection) by the biometric system as the attacker wants to hide his or her true identity due to watch list applications, e.g. video surveillance or border control. A spoofing attack occurs when an attacker tries to fool a biometric system into recognizing an illegitimate user as a targeted person using any counterfeit trait that is not originated from a live person, i.e. artificial biometric traits (also referred to as artefacts) and dismembered body parts.

Without understating the issue of obfuscation, e.g. in video surveillance and border control, spoofing is probably the most malicious type of attack as it poses a serious threat to the security and privacy of the people enrolled to systems using biometric-enabled authentication. It has been shown recently that some (if not all) conventional biometric techniques, such as fingerprint or face recognition, are vulnerable to spoofing attacks. Currently, spoofing attacks are one of the main problems for companies willing to market identity management solutions based on biometric technologies. The existing biometric authentication systems are beginning to be mature enough for higher-security consumer-level applications but the lack of public confidence due to vulnerabilities to spoofing attacks prevents biometric authentication from making its breakthrough. Thus, there is an urgent need to find efficient and reliable solutions for detecting and rejecting synthetic artefacts and to bring the technology into practical use. As a consequence, in recent years, there have been several initiatives focusing on the problem of anti-spoofing, like the EU FP7 project entitled Trusted Biometrics under Spoofing Attacks[4] (TABULA RASA) within also the work presented in this thesis was conducted.

Contrary to zero-effort attacks, spoofing cannot be countered by improving the accuracy of the biometric system (in terms of FAR). Furthermore, where indirect attacks require advanced programming skills, direct attacks do not require any special abilities because the end-users have naturally the access to the acquisition device. As every user can be considered as a potential attacker, the presentation of biometric trait is the most susceptible point for an attack (Nixon *et al.* 2008). Among the different vulnerabilities, spoofing is purely a biometric vulnerability because, e.g. unlike passwords, our biometric traits, e.g. face, iris, fingerprints and ever DNA, are widely available and some of them are easy to sample, which is probably the most crucial drawback of biometric authentication (Galbally *et al.* 2014). Producing an artificial biometric trait is rather straightforward because after a quick browsing on the Internet, it is not very hard to find step-by-step tutorials on how to make low-costs synthetic artefacts, including face masks or artificial fingerprints. Since sensor (or user interface) level attacks are performed outside the control of the biometric systems manufacturer, no digital protection mechanisms can be deployed to prevent them. Thus, specialized countermeasures are needed for dealing with the forged biometric samples.

[4] `https://www.tabularasa-euproject.org/`

2.1.2 Spoof detection schemes

The task of automatically differentiating whether the presented biometric trait originates from a living legitimate subject or from some other source is referred to with many different terms as anti-spoofing, spoof detection and presentation attack detection (Marcel *et al.* 2014). Also liveness detection is quite often used as a synonym for spoof detection in some fields but, in general, it can be used to refer to a more limited problem of sensing vitality signs, like eye blinking or heartbeat. In this thesis, this term is treated as a subcategory for anti-spoofing methods.

Spoofing attack detection can be performed before the actual biometric data is captured or when processing the acquired sample using three different approaches: 1) use only the information acquired for identification purposes, 2) further process the data or acquire additional information over time to find clues of a possible spoofing attack, or 3) use additional sensors and software to find out a representation that is more suitable for capturing inherent differences between genuine subjects and fake ones than the original biometric data (Nixon *et al.* 2008). The anti-spoofing research has mainly concentrated on further processing and collecting the biometric data, or using additional instruments because based on the acquired biometric sample it is really hard to tell if the presented biometric trait is valid or not.

Like in the case of attack terminology, there exists no unified taxonomy for the different spoof detection approaches. The aforementioned techniques can be categorized in several ways, e.g. based on the working mechanisms into methods utilizing the intrinsic properties of the biometric samples, liveness cues or contextual information (Marcel *et al.* 2014), or based on the biometric system module in which they are integrated into sensor-level (hardware-based), feature-level (software-based) or score-level techniques (Galbally *et al.* 2014). In this thesis, however, the score-level techniques are not considered as a separate technique but as a part of the whole biometric system design which is a research topic of its own that is not related only to anti-spoofing and has not been explored much yet. Thus, a three-part categorization is followed dividing the individual spoof detection schemes into hardware-based, software-based and multi-modal techniques.

Hardware-based methods introduce some custom sensor into the biometric system that is designed particularly for capturing the differences between a valid living biometric trait and others, including, e.g. heartbeat, blood pressure, sweating, thermogram and multi-spectral reflection properties of the skin or eye. The measured characteristics can be divided into three groups: 1) intrinsic properties, e.g. capacitance and reflectance at specific wavelengths, 2) involuntary signals because of the nervous system, e.g. pulse, perspiration, EEG and EKG, or 3) voluntary or involuntary responses to external stimuli (challenge-response based liveness detection), e.g. pupil dilation due to illumination changes (Galbally *et al.* 2014).

Software-based (data-driven) anti-spoofing techniques are exploiting only the same data that is used for the actual biometric purposes or additional data capture with the standard acquisition device, thus they can be basically integrated into the feature extractor module of the biometric system. The data-driven methods can work either on single samples (static) or a sequence of samples acquired over time (dynamic) (Galbally *et al.* 2014). Dynamic approaches are probably more robust in spoof detection because temporal information might provide useful cues for anti-spoofing which might not be present during that particular moment when a single snapshot of a biometric trait is captured. On the other hand, the use of dynamic techniques is not always possible to deploy and, ideally, static methods would be preferable because they are faster and less intrusive than their dynamic counterparts. Software-based countermeasures might be used also for detecting indirect attacks besides spoofing because they are not operating directly on the biometric trait itself but on the acquired biometric sample. Thus, software-based approaches could reveal also when the input sensor has been circumvented, i.e. when replay-attack or synthetic biometric sample is injected to the communication channel between the acquisition device and the feature extractor, which makes the use of data-driven methods very appealing.

One of the reasons why research on anti-spoofing used to receive less attention was that it was assumed that a combination of different biometrics increases system robustness to spoofing attacks as intuitively it is difficult to present multiple fake biometric traits simultaneously. However, this hypothesis suggests that all individual modalities need to be fooled in order to circumvent multi-modal biometric system. Unfortunately, it has been shown that multi-biometrics itself is not inherently robust to spoofing attacks because successfully spoofing one of the unimodal subsystems might be enough to break naively tuned fusion strategies (Rodrigues *et al.* 2009).

The problem is that multi-modal systems have usually combined complementary unimodal techniques at score level in order to improve recognition performance without evaluating the robustness of different fusion rules to spoofing attacks. Thus, more research effort in developing more robust fusion strategies is needed. Multi-biometrics can be also explicitly used in spoof detection. For instance, many multi-biometric anti-spoofing techniques have been proposed by combining face and voice modalities (talking-face verification) that exploit the correlation between observed lip movement and speech (Chetty & Wagner 2004, Bredin & Chollet 2008) or challenge-response based approaches for verifying if a preassigned pin code can be recognized in both modalities (Kollreider *et al.* 2007). In the ideal case, iris and face recognition could be performed from a single image if the resolution of the acquired image is high enough, which might enable new directions for multi-biometric anti-spoofing.

Anti-spoofing techniques, like any other consumer-level product, have some desirable properties. Thus, the developed spoof detection schemes should be non-invasive, user friendly, fast, low-cost and robust without degrading the overall recognition performance in terms of false rejection rate (FRR). The aforementioned different categories of anti-spoofing techniques have their own pros and cons. Hardware-based methods are the most robust ones since the dedicated sensors are used for spoof detection. On the other hand, the tailored sensors are usually expensive and not compact, thus not (yet) available in mobile devices, for instance, which prevents their wide deployment. Furthermore, many sensor based liveness detection measurements, e.g. EEG, are too intrusive. In general, software-based techniques, on the other hand, can be seen non-invasive, user-friendly (if challenge-response approach not utilized) and low-cost but their main drawback is their robustness because only the standard biometric sensor data is utilized. Multi-biometrics provides the best recognition performance but its potential robustness to spoofing is not yet very clear.

These conclusions should be, however, considered as very broad indication because individual techniques have their own advantages and limitations within each category. It is also worth highlighting that some spoof detection schemes are more effective in detecting other types of presentation attacks in addition to synthetic artefacts, such as concealed, mutilated or dismembered traits. Thus, a proper combination of complementary software and hardware-based techniques could be utilized in order to overcome the drawbacks of individual countermeasures and to increase level of protection to various types of attacks. While nowadays every mobile phone and laptop are equipped with a microphone and a camera, other sensors, such as 3D or multispectral imaging, are emerging in mobile devices[5,6], which opens up new possibilities for both hardware-based anti-spoofing and multi-biometrics. Thus, it is important to investigate novel approaches in all three categories because one can never know what

[5] https://www.google.com/atap/projecttango/

[6] http://structure.io/

new possibilities the future devices provide in addition to increase in computational resources.

2.2 2D Face biometrics under spoofing attacks

Because of its natural and non-intrusive interaction, identity verification and recognition using facial information is among the most active and challenging areas in computer vision research. Although wide range of viewpoints, ageing of subjects and complex outdoor lighting conditions are still major research challenges, the recognition algorithms are beginning to be mature enough for consumer-level applications. A good example of this is the latest advances in face verification using deep convolutional neural networks as the obtained error rates are closely approaching human-level performance on very challenging datasets (Taigman *et al.* 2014). Discussion on different face recognition algorithms falls out of the scope of this thesis. For further details on the advances in the field of face detection and recognition, interested readers are referred to surveys by Jain *et al.* (2008) and Li & Jain (2011), for instances.

Among different biometric traits, face has probably (one of) the highest potential in biometric-enabled applications because almost every laptop and mobile phone is equipped with a front-facing camera. According to the International Biometric Group (IBG), face is the second most utilized modality after fingerprints. For instance, it is accepted in many official large-scale identifications documents, e.g. International Civil Aviation Organization (ICAO) compliant biometric passport and other national ID cards. Furthermore, face authentication based unlocking features have been introduced on laptops and mobile phones, such as Face Unlock on Android devices. Some companies are promoting payment systems relying solely on face modality in supervised environments and even on mobile devices.

In this section, the scope of biometric anti-spoofing is narrowed down to face modality. First, the problem of face spoofing is discussed in detail. Then, an extensive literature review is given with particular focus on software-based countermeasures. In the end, the publicly available spoofing datasets are introduced because they have been an important landmark in the recent face anti-spoofing research and been indispensable tools for the research community, including the present author.

2.2.1 Face spoofing

Face authentication systems, in particular, are known to respond weakly to spoofing attacks for a long time. Duc & Minh (2009) demonstrated at Black Hat 2009 conference, the world's premier technical security conference, how to easily spoof and bypass commercial face authentication systems embedded in laptops (Lenovo Veriface III, Asus SmartLogon V1.0.0005 and Toshiba Face Recognition 2.0.2.32) - each set to its highest security level by presenting mere photographs of the legitimate user and thereby gaining access to the laptops. This single example among others (Chingovska *et al.* 2012, Erdogmus & Marcel 2013b, Li *et al.* 2014, Wen *et al.* 2015) have revealed the vulnerabilities of 2D face authentication systems to spoofing attacks and have been a final wake-up call for the research community to put more weight on anti-spoofing.

Falsifying face biometric data is straightforward compared with other biometric modalities, e.g. fingerprint and iris, because no special skills or tutorial are required for creating high-quality synthetic artefacts of the targeted face. For instance, Biggio *et al.* (2012) suggested that the corresponding fake score distributions are very similar to those of the genuine users, i.e. the worst-case scenario, in the case of face spoofing but the same does not hold for latex based fake fingerprints. Thus, producing a perfect replica of the targeted face is easy compared with the recapturing process of a fingerprint.

First of all, acquiring a high-quality sample of the targeted face is extremely easy. Hiding your face in public is extremely difficult if one does not prefer to wear some sort of disguise all the time, thus our facial information can be captured even from long distance. Furthermore, a great deal of pre-labeled multimedia content, i.e. photographs and videos, is openly available in the Internet due to the increasingly popular social networking services (Facebook, Flickr, Youtube, Instagram, etc.) and their automatic face tagging features. For instance, Li *et al.* (2014) inspected the threat of the online social networks based facial disclosure against the latest version of six commercial face authentication systems (Face Unlock, Facelock Pro, Visidon, Veriface, Luxand Blink and FastAccess). While on average only 39% of the images published on social networks could be successfully used for spoofing, the relatively small number of vulnerable images was enough to fool face authentication software of 77% of 74 users.

In films, spoofing attacks are usually performed using very sophisticated elastic masks or surgically altered faces while a mere photograph of the targeted person might be enough for fooling a face authentication system in reality. Plastic surgery (Singh *et al.* 2010) and make-up (Dantcheva *et al.* 2012) pose problems for face recognition algorithms and degrades their performance in terms of FRR, which makes them very useful in deliberately obscuring one's true identity. While plastic surgery is gradually becoming more affordable, it is less likely that surgically altered faces are used for spoofing purposes considering the costs and potential gained benefit. In a live demonstration during the International Conference on Biometric (ICB 2013), a female intruder with a specific make-up succeeded in targeted impersonation against a face authentication system[7]. However, this kind of attack could probably be prevented by increasing the accuracy of the recognition algorithm in terms of FAR.

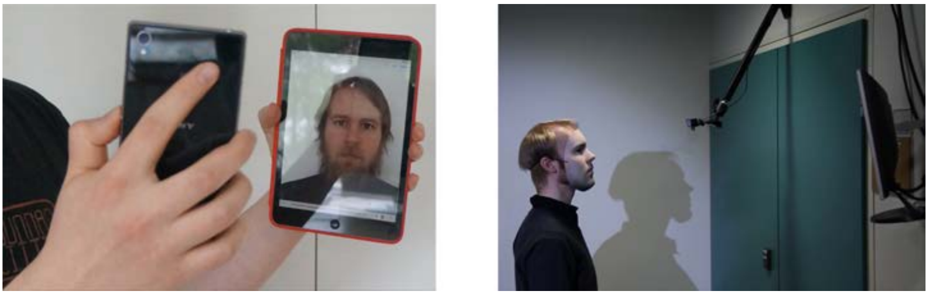

Fig 2. Illustration of spoofing attacks executed with a hand-held photo and a wearable 3D mask in mobile authentication and access control scenario.

Compared with the previous techniques, it is much easier to download or capture a photograph or video of the targeted person and present the acquired biometric sample using a print or display device (also referred to as spoofing or display medium) to the input sensor (see, Figure 2). Moreover, the costs of producing nearly photo-realistic 3D masks have become reasonable and one can buy wearable masks manufactured from just two photographs of a person's face (frontal and profile) in online stores[8] (see, Figure 2). Therefore, most of the prior studies in the anti-spoofing literature propose countermeasures to two categories of synthetic artefacts: 2D surfaces, such as prints and video screens, or 3D shapes, such as masks. In order to understand the problem of face spoofing better and the rationale behind the approaches for countering the attacks, it is useful to take a closer look at the different types of fake faces and also how they are presented to the camera.

A photo attack is performed by presenting a photograph of the targeted identity in front of the camera of a face recognition system. The compromised biometric sample can

[7] `http://www.biometrics-center.ch/testing/tabula-rasa-spoofing-challenge-2013`

[8] `http://www.thatsmyface.com/`

be displayed by printing the photograph on paper or by using video screen, e.g. mobile phone or tablet. Recent work by Duc & Minh (2009) and Li *et al.* (2014) indicates that many available commercial systems are vulnerable to this kind of attack at their highest security level. Thus, countermeasures are needed also for this kind of primitive attack. Since photographs describe only the appearance of the face, motion or liveness analysis based methods are in principle very effective for detecting plain photo-attacks. However, the attacker might try to fool this kind of countermeasures by simulating facial and head movements by translating, rotating or warping (bending inward and outward) a face print (see, Figure 3), which poses new challenges to motion based methods (Pan *et al.* 2008). In addition, the attacker can cut the eyes and mouth out when the imposter can use the face print as a photographic mask and exhibit facial motion like eye blinking and mouth movements behind the mask or by using two photographs (Kollreider *et al.* 2008) in order to fool challenge response based methods based on eye blink detection, for instance.

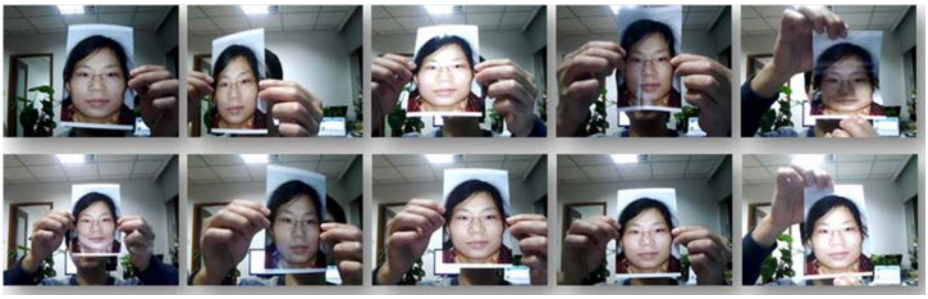

Fig 3. Attack samples from the NUAA Photo Imposter Database by Tan *et al.* (2010). Vivid photo attacks are simulated by translating, rotating and bending (warping) a photograph.

Video (replay) attacks are bit more advanced than photo attacks because also dynamics (liveness) of the targeted face are copied. Video attacks are performed using display devices, such as mobile phones and tablets. Although video attacks are potentially a much more serious problem compared to photo attacks, high-quality videos of a targeted person are much more difficult to acquire compared to frontal face photographs that people share openly on social networks. However, state-of-the-art commercial animation software[9], tailored for modelling avatars, are very impressive nowadays. Thus, a frontal face image can be animated to exhibit realistic liveness characteristics and even 3D head motion. This kind of attacks have been created within

[9] `http://www.reallusion.com/crazytalk/`

the context of talking-face verification in the form of face (Verdet & Hennebert 2008) and audio-visual (both face and voice) (Karam *et al.* 2009) impostures. Naturally, the animated faces suffer from quality and unnatural motion artefacts but, in principle, they can react to randomly prompted challenges unlike replayed videos. Primitive facial animation of eye blinking and head rotation can be performed also by applying basic image editing on one or more digital photos and flash the original and modified images on a display device[10]. Fortunately, this causes abrupt changes in the observed motion.

A photo-realistic 3D mask of the genuine face would be the ultimate spoofing artefact because the complete 3D shape is represented. Thus, depth cues cannot be exploited for anti-spoofing while they work fine for detecting previously presented 2D surfaces. Manufacturing 3D masks is becoming more affordable due to 3D printing devices. As a consequence, also mask attacks are getting increasing interest in face anti-spoofing literature. With printed rigid 3D masks, it is hard to exhibit non-rigid facial movements, thus only cut out eyes and mouth regions can be utilized, like in the case of photographic masks (see, Figure 4). Moreover, the facial appearance quality (amount of details) is still quite rough, as can be seen in Figure 4.

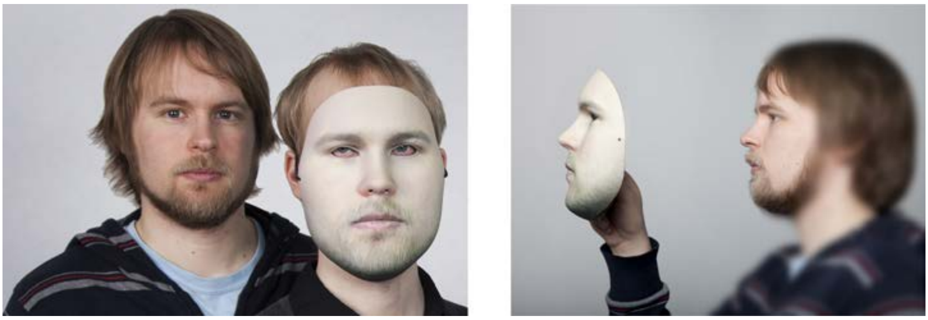

Fig 4. A wearable 3D mask which has been ordered from http://www.thatsmyface.com/.

In addition to the aforementioned properties, there are several other important factors that need to be considered while designing anti-spoofing techniques. The environmental conditions (e.g. illumination and background scene) and the used camera setup (e.g. quality and support) set some operational requirements for the countermeasures. For example, a fixed camera is used in access control scenario, whereas a hand-held camera is used in mobile applications, which might cause problems for methods utilizing motion

[10] `http://www.androidpolice.com/2012/08/03/android-jelly-beans-face-unlock-liveness-check-circumvented-with-simple-photo-editing/`

information (see, Figure 2). Furthermore, the forged biometric trait and its presentation using display medium defines cues that could be exploited for spoof detection, including the quality (resolution) of the display medium or the sample of the targeted face, used support (e.g. fixed or hand-held), and trimming of the fake face (e.g. whether the frame of the display device is visible). For instance, two example spoofing scenarios are presented in Figure 2. In the first one, the attacker uses his or her hands to support the video display showing the targeted face, thus a rectangular frame can probably be observed around the face. In the second attack, the mask is mounted on the attackers face, thus no clear boundary can be seen and the attacker is able to exhibit eye blinking through the holes in the eyes. Thus, the second attack is much more sophisticated and is therefore much more challenging to detect.

Hardware-based solutions using 3D or multi-spectral imaging provide efficient means for detecting face spoofs because they offer additional useful information on the surface reflectance properties or depth of the observed face. For instance, a low-cost depth sensor, i.e. Microsoft Kinect, can be utilized for differentiating a real face from a planar surface, e.g. video display or photograph, in a quite straightforward manner (Erdogmus & Marcel 2013a). Skin reflectance measurements at two specific wavelengths can be used to distinguish a genuine face from artificial materials used in 3D masks and 2D surfaces because human skin has extremely low reflectance in the upper-band of near-infrared (NIR) spectrum which is a universal property among human race (Pavlidis & Symosek 2000, Zhang *et al.* 2011). Also thermal information could be used for detecting prints and video screens. Adding and subtracting skin tissue using redistributed fat or creating or removing scars with silicone are typical operations of plastic surgery. Furthermore, surgical operations usually cause alteration in blood vessel flow that can be seen as cold spots in the thermal domain. These kinds of physiological changes can be detected in the thermal infrared (IR) region (Pavlidis & Symosek 2000).

On the other hand, depth sensors are powerless under 3D mask attacks if depth cue is the only utilized countermeasure. It is a known fact that thermal radiation can pass through materials, which causes problems when thermal IR information is used against wearable mask attacks (Zhang *et al.* 2011). Also the existing NIR based techniques have difficulties in capturing the reflectance disparities between genuine faces and 3D masks due to the 3D shape and variety of artificial materials (Zhang *et al.* 2011). Furthermore, the use of NIR imaging is restricted to indoor use only since the sunlight causes severe perturbation. The dedicated imaging solutions are indeed effective in detecting various kinds of artificial faces if they are coupled in the same system (Pavlidis & Symosek

2000). Unfortunately, the problem with hardware-based techniques is that, in general, they are either quite intrusive, expensive or impractical because unconventional imaging devices (e.g. active lighting) are required.

As a consequence, it would be rather appealing to utilize anti-spoofing techniques using the conventional cameras included in the existing face authentication systems. However, already photo attacks have shown the problem to be very difficult. Even though print attacks itself are very primitive attacks, it is very hard to derive robust and reliable software-based approaches for detecting them in various acquisition conditions, not to mention variations in the attacks, e.g. photographic masks. Thus, a variety of software-based countermeasures have been proposed. The following section presents a comprehensive review of the most successful data-driven anti-spoofing techniques proposed within 2D face modality.

2.2.2 Software-based face anti-spoofing

The following overview on software-based spoof detection schemes is largely based on the literature review presented in Paper I. Like mentioned in Section 2.1.2, feature level approaches can be categorized into static and dynamic techniques based on whether temporal information or features are utilized. On the other hand, the dynamic methods in the related literature are mainly based on analysing the motion or liveness and while the static methods on analysing the facial appearance or quality based cues. Therefore, the taxonomy in Paper I is based on the inspected visual cues: liveness, motion, facial appearance and context. The main emphasis of this thesis is on non-intrusive techniques but also challenge-response approaches are introduced because random user interaction demand provides an important liveness cue in addition to visual ones.

The LBP based method proposed in Paper II has inspired several other researchers to extend the work on facial texture analysis based anti-spoofing. For clarity, this section introduces only those texture based methods that have been proposed before the publication of Paper II. Section 3.4 and Chapter 5 provide more detailed discussion on how the facial appearance based anti-spoofing has been further developed.

Liveness detection

Typical countermeasure to face spoofing is liveness detection that aims at detecting physiological signs of life, such as eye blinking, facial expression changes and mouth movements. For instance, Pan *et al.* (2008) exploited the observation that humans blink once every 2-4 seconds and proposed an eye blink-based anti-spoofing method which uses Conditional Random Field (CRF) framework to model and detect eye blinking. The authors provided also a publicly available data set that contains short video clips of eye blinks and vivid spoofing attacks using photographs. Obviously, such technique can only be considered with photographs while it can be easily fooled with video replay or photographic mask attacks.

In order to provide more evidence of liveness, Eulerian motion magnification (Wu *et al.* 2012) has been applied for to enhancing subtle changes in the face region that may not be otherwise observed without a closer inspection (Chingovska *et al.* 2013b, Bharadwaj *et al.* 2013). Within the context of the 2nd competition on counter measures to 2D facial spoofing attacks (Chingovska *et al.* 2013b), one team presented a technique for magnifying the small color and motion changes that appear on the face due to the natural human blood flow, thus the algorithm amplifies a set of frequencies within the range of human pulse. Bharadwaj *et al.* (2013) used Eulerian motion magnification as a preprocessing stage for exaggerating macro and micro facial expressions in the input video. Moreover, inspired by the use of optical flow in micro expression detection, a histogram of oriented optical flow (HOOF) features (Chaudhry *et al.* 2009) was considered for describing observed facial motion patterns. Very impressive results were reported on the Print-Attack and Replay-Attack Databases but the algorithm needs to be improved in order to increase its performance in more challenging and adversarial acquisition conditions.

Motion analysis

In addition to facial motion used in liveness detection, also other motion cues can be exploited for face anti-spoofing. For example, it can be assumed that the movement of planar objects, e.g. video displays and photographs, differs significantly from real human faces which are complex non-rigid 3D objects. Kollreider *et al.* (2009) presented an optical-flow based method to capture and track the subtle relative movements between different facial parts, assuming that facial parts in real faces move differently than on photographs. The method was able to achieve an equal error rate (EER) of 0.5% on a private data set consisting of real client accesses from XM2VTS database and hard-copy attacks from the corresponding live samples. The same authors proposed a method for fusing scores from multiple experts to combine the results of 3D face motion analysis and liveness detection, e.g. eye-blink detection and mouth movement analysis (Kollreider *et al.* 2008). However, the experiments were conducted on short image sequences which were not made publicly available and no specific error rates were reported.

In another work, Bao *et al.* (2009) also used optical flow for motion estimation for detecting attacks produced with planar media such as prints or screens. The movement of planar objects is categorized as translation, rotation, normal or swing and the eight quantities extracted from the cropped face are used to express the amount of these movements. The eight values are then given to an ad- hoc equation that outputs the probability of a spoofing attack. Experiments on a private database showed a 6% false-alarm against about 14% false-acceptance.

If a face spoof is not tightly cropped around the targeted face or it has an incorporated background scene, scenic fake face (see, Figure 7), it should be possible to observe high correlation between the overall motion of the face and the background regions for stationary face recognition systems. Anjos & Marcel (2011) proposed a straightforward motion-based anti-spoofing technique to measure the overall motion correlation between the face and the background regions through simple frame differences. The motion correlation analysis based technique showed to be efficient for measuring synchronized shaking of hand-held attacks and static nature of fixed photo attacks in the publicly available Print-Attack Database. However, a drawback is that it can get confused between a fixed support photo attack and a motionless person while being recognized. Hence, the algorithm was improved by utilizing optical flow instead of plane averaged frame differences in order to get more accurate motion description, e.g. direction (Anjos *et al.* 2013). The upgraded algorithm outperformed the previous motion based techniques (Kollreider *et al.* 2009, Bao *et al.* 2009, Anjos & Marcel 2011) on the publicly available Photo-Attack Database. The authors have also made the source codes of the proposed algorithm and the three reference systems publicly available. Similar approach measuring the face and background motion correlation was also proposed within the IJCB 2011 competition on counter measures to 2D facial spoofing attacks (Chakka *et al.* 2011) by Yan *et al.* (2012).

Facial appearance analysis

The main problem of motion analysis and liveness detection based anti-spoofing techniques is that the verification process takes some time or the user needs to be very cooperative. Even though motion is an important visual cue, vitality and non-rigid motion detectors relying only on spontaneous facial movements are powerless under video replay attacks and the lack of motion may lead to a high number of authentication failures if user cooperation demand is not deployed. Assuming that the inherent disparities between genuine faces and artificial material can be observed in single images (or a sequence of images), another category of anti-spoofing techniques is based on the analysis of static (and dynamic) facial appearance properties, such as reflectance, shading, texture and quality. Intuitively, the main advantage of single image based spoof detection schemes is that they treat video playback attacks as if they were photo attacks, since individual video frames are considered (Bai *et al.* 2010).

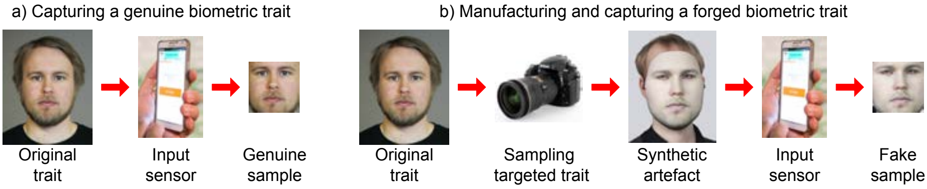

Fig 5. The process of capturing a genuine biometric sample and a fake one, inspired by Gao *et al.* (2010).

Usually, the key idea behind facial appearance analysis based spoof detection is that an image of a fake face is actually an image of a face which passes through two different camera systems and a printing system or a display device (see, Figure 5), thus it can be referred to in fact as a recaptured image. As a consequence, the observed fake face image is likely to have lower image quality compared to a genuine one captured in the same conditions due to lack of high frequency information. Furthermore, the recaptured images may suffer from other quality issues, such as printing or video encoding artefacts. In the literature, the facial appearance analysis based methods are usually referred to as texture analysis based techniques because the aforementioned properties can be considered as variations in the facial texture information. Also in this thesis, facial texture and quality are treated as synonyms in general.

Li *et al.* (2004) described a method for detecting print-attack face spoofing. The method is based on the analysis of 2D Fourier spectrum, assuming that photographs are usually smaller in size and they would contain fewer high frequency components compared to real faces. Such an approach may work well for down-sampled photos but is likely to fail for higher-quality images. The database used in the experiments is unfortunately not publicly available.

In a more recent work, Tan *et al.* (2010) considered the Lambertian reflectance to discriminate between the 2D images of face prints and 3D live faces. The method extracts latent reflectance features using either a variational retinex-based method or a much simpler difference of Gaussians (DoG) based approach. The features are then fed to different types of classifiers. The idea behind DoG filters is that their bandpass behaviour is able to exclude the low frequency information and the very high frequencies (noise), and to preserve only the relevant frequency information for spoof detection. The authors reported promising results of area under curve (AUC) from 0.69 to 0.95 on the publicly available NUAA Photograph Imposter Database composed of real-accesses and attacks to fifteen subjects using both photo-quality and laser-quality prints of different sizes.

Zhang *et al.* (2012) modified the method of Tan *et al.* (2010) by introducing multiple DoG filters because they suggested that there is no prior knowledge which frequency component is the most discriminative. Thus, they used the concatenated filtered images and a support vector machine (SVM) classifier to conduct more experiments on a new publicly available CASIA Face Anti-Spoofing Database containing more versatile set of spoofing attacks, including cut-photo and video replay attacks with three different imaging qualities. They noticed that the DoG filter based method is able to detect well the similar sharp edges in the eye regions of cut-photo attacks, which make the spoofing attack samples less variational. On the other hand, the performance was very low under video replay attacks and when high imaging quality was used. Thus, they claim that it may not always be a good idea to pursue high quality in imaging.

Alternatively, it is likely that real faces and fake ones present different texture patterns because of facial texture quality degradation due to recapturing process and disparities in surface and reflectance properties. Bai *et al.* (2010) used micro-textures extracted from the specularity component of a recaptured image and a linear SVM classifier to detect printed photo based spoofing attacks. The authors report a performance of 2.2% in terms FLR against 13% in terms of FFR but also the dataset for this experiment was not made public. The used features try to estimate how smooth the surface of the client face is, i.e. smoother face texture is more likely to come from a print attack because the reflection from a natural face tends to be more diffuse. The major drawback of this method is that it requires high resolution input images in order to discriminate the fine micro-texture of the used display medium. In the work by Pinto *et al.* (2012), the dynamic disturbances (texture) of display devices were exploited for detecting video replay attacks. More specifically, visual rhythms were computed from the Fourier spectrum of the extracted video noise signatures and the resulting textural information was compressed with gray level co-occurrence matrices (GLCM).

Also features explicitly describing the inherent image quality differences between genuine faces and fake ones have been presented in the literature. Gao *et al.* (2010) proposed a general physical model for describing the recapturing process and used a set of physical features and contextual background information from single images to discriminate the recaptured images from real scenes. This is a more general concept of recognizing images of natural scenes and the recaptured natural-scene images, thus it could be used for more general object recognition and scene understanding (e.g. in robotics) in addition to anti-spoofing. The physics based features consist of the spatial distribution of specularity that is related to the surface geometry, the image gradient that captures nonlinearity of the recaptured image rendering process, and color, contrast and blurriness properties that describe the quality of the reproduction. The proposed approach outperformed a wavelet-based method (Farid & Lyu 2003) and a publicly available data set was also released for evaluating methods for recaptured image detection on mobile devices.

Galbally & Marcel (2014) introduced fourteen image quality assessment based features that showed comparable performance to texture analysis based algorithms. However, as expected, the method was robust in detecting fake faces presented on mobile phone displays, whereas high-definition face spoofs caused problems. In addition, the spoof detection performance of the proposed feature set was highly dependent on the used imaging quality, i.e. the method performed well on high quality input images, whereas the results degraded dramatically at lower acquisition qualities.

Contextual information

In addition to motion (Anjos & Marcel 2011, Yan *et al.* 2012), the provided view provides also actual contextual cues for anti-spoofing. In order to make eye-blink based liveness detection more robust to video replay attacks, Pan *et al.* (2011) included scene context matching for checking if the background scene of the stationary face recognition system suddenly changes. Some carefully chosen fiducial points outside the face region are used to describe the expected background scene. The scores of the eye blink and scene context detector components are then fused together and the new setup obtained 0.5% false acceptance against 0% false rejection on a new private data set.

Fusion of countermeasures

Intuitively, a combination of several complementary countermeasures increases robustness to various types of face spoofs. Therefore, it was not unexpected that fusion of several methods analysing the motion and facial appearance has been a common trend in the recently organized two competitions on software-based anti-spoofing (Chakka *et al.* 2011, Chingovska *et al.* 2013b). In the IJCB 2011 competition on counter measures to 2D facial spoofing attacks (Chakka *et al.* 2011), facial texture analysis based methods were dominating because the photo attacks in the competition dataset suffered from severe print quality defects. However, since the attack scenarios in the 2nd competition on counter measures to 2D facial spoofing attacks (Chingovska *et al.* 2013b) were more diverse and challenging, all the best-performing methods were utilizing some sort of combination of both motion and texture analysis.

Nevertheless, the fusion of different anti-spoofing techniques, like motion and texture, has not been investigated more closely besides the algorithms (Schwartz *et al.* 2011, Tronci *et al.* 2011, Yan *et al.* 2012) proposed within the context of the IJCB 2011 competition on counter measures to 2D facial spoofing attacks (Chakka *et al.* 2011). Tronci *et al.* (2011) and Schwartz *et al.* (2011) were able to obtain impressive performance using motion and texture information but at the cost of high complexity. Tronci *et al.* (2011) utilized several visual features, e.g. eye blinking and multiple low-level features, and SVM for detecting print-attacks, whereas Schwartz *et al.* (2011) accumulated temporal information from videos by concatenating appearance descriptions of individual frames, which results in very high-dimensional feature vectors. Conversely, Yan *et al.* (2012) wanted to achieve better generalization capabilities and proposed novel liveness clues with clear semantic definitions instead of just extracting off-the-shelf features and training a "black box" classifier. However, the algorithm utilized mainly two uncorrelated motion cues, non-rigid motion and face and background consistency analysis, while the only spatial cue, banding analysis, was discarded unless uniform background was observed, since both face and background regions were used for image quality assessment.

User-interaction based methods

Liveness and motion analysis based spoof detection is rather difficult to perform by observing only spontaneous facial motion during short video sequences but the amount of distinctive motion can be increased with user cooperation demand. More importantly, user collaboration itself can be used for revealing spoofing attacks because we humans tend to be interactive, whereas a photo or video replay attack cannot respond to randomly specified action requirements. In particular, a face authentication system prompts a specific action request to the user (challenge), such as a facial expression (Kollreider *et al.* 2007, Ng & Chia 2012), mouth movement (Chetty & Wagner 2004, Kollreider *et al.* 2007) or head rotation (Frischholz & Werner 2003, De Marsico *et al.* 2012, Wang *et al.* 2013), and then analyses the user activity in order to check whether the required action was actually performed (response).

While spontaneous non-rigid facial motion is likely to occur during a relatively short authentication process, usually, only small head pose changes can be observed. Therefore, the challenge-response approach is particularly useful when head pose (Frischholz & Werner 2003) or 3D structure estimation (De Marsico *et al.* 2012, Wang *et al.* 2013) is utilized for spoof detection. Frischholz & Werner (2003) proposed to avoid photo and video replay attacks by giving the user a random head pose as a challenge and tracking the head pose in real time using 3D model and suitable facial feature points. De Marsico *et al.* (2012) reduced the system complexity by measuring the three-dimensionality using projective invariants. Instead of tracking the exact 3D head pose changes, the user can move more freely as long as minimum continuous motion requirement is met, thus making the authentication process more comfortable. Furthermore, the movement challenge at random intervals and expected response time is assumed to be sufficient to avoid video replay attacks. Wang *et al.* (2013) presented an approach for measuring three- dimensionality of the face without continuous motion requirement by recovering sparse 3D facial structure from two or more images (or video) captured from different viewpoints. Ng & Chia (2012) performed liveness detection by prompting the user random sequence of facial expressions. Assuming that the consecutive frames in videos of valid users contain smooth and gradual changes, the presence of tampered or stitched video or image sequences is detected by observing image properties for abrupt changes. More specifically, SIFT flow energy (Liu *et al.* 2011) between consecutive frames is computed because the videos with sudden changes typically result in high SIFT flow energy.

The drawback of the challenge-response approach is that it requires user cooperation, thus making the authentication process a time-consuming and unpleasant experience. Another advantage of non-intrusive techniques is that from challenge-response based countermeasures it is rather easy to deduce which liveness cues need to be fooled.

For instance, the request for uttering words suggests that analysis of synchronized lip movement and lip reading is utilized, whereas rotating head in a certain direction reveals that the 3D geometry of the head is measured. For non-intrusive approaches, it is usually not known which countermeasures are used, thus the system might be harder to deceive (Pan *et al.* 2011).

2.2.3 Databases

The evaluation of biometric systems under spoofing attacks can be conducted either at algorithm or system level. In the first case, the performance of the anti-spoofing modules is evaluated independently from the performance of the rest of the system, e.g. the actual recognition stage. The latter option considers the performance of the whole biometric system as a whole. The advantage of system based evaluations is that it provides better insight what the overall robustness of the system to spoofing attacks is and how the proposed anti-spoofing technique affects the system accuracy (in terms of FRR).

Currently, there exists no clear consensus on the best protocols and metrics for assessing the robustness of a biometric system under spoofing attacks. However, one generalized evaluation protocol is based on two scenarios, licid and spoofing (Galbally *et al.* 2014). The first scenario measures the traditional recognition accuracy, i.e. only genuine accesses and zero-effort imposter access attempts are considered when the performance is reported, usually in terms of FRR and FAR. In the latter scenario, the imposter attacks are substituted with spoofing attacks when the evaluation metrics when the used evaluation metrics would be FRR (like in the licit scenario) and spoofing false acceptance rate (SFAR). With these measures, the accuracy and the vulnerability to spoofing of a particular biometric system can be determined.

In this thesis, the proposed anti-spoofing techniques are assessed using algorithm based evaluation, thus the used evaluation metrics are referred to as false fake rate (FFR) and false living rate (FLR). This is mainly due to the lack of specific enrolment data in the publicly available datasets, which makes it impossible to evaluate the performance of recognition algorithms under spoofing attacks. From the existing benchmark datasets, only Replay-Attack Database (Chingovska *et al.* 2012) provides means for evaluating the joint operation of face recognition and anti-spoofing algorithms (Chingovska *et al.* 2013a).

Constructing a large scale spoofing related database poses additional challenges compared to acquisition of standard biometric data. Where it is feasible to capture the biometric data of a statistically meaningful number of identities with variations in user demographics (e.g. age, gender and race) and environmental conditions, manufacturing a huge amount of artificial biometric traits and then simulating various types of attack scenarios (e.g. use-cases) for the different subjects is extremely time-consuming and expensive. It is relatively cheap for an attacker to exploit a known vulnerability of a face authentication system (a "golden fake"), such as a realistic 3D mask, Unfortunately, the same does not hold for researchers who should ensure the robustness of the developed anti-spoofing mechanisms on a diverse dataset with a meaningful number of subjects and several mask materials, let alone various ways of performing the attacks.

Precisely for this reason, the early works in face anti-spoofing have been utilizing mainly small proprietary datasets. This is especially true in the case of hardware-based approaches because capturing new sensor-specific data is always required. Therefore, sensor based techniques have been usually evaluated just to demonstrate a proof of concept or have not been experimentally validated at all in the worst case (Pavlidis & Symosek 2000). As a consequence, it is extremely hard to directly compare sensor based approaches with other related biometric solutions.

An advantage of software-based countermeasures is that they can be assessed on common benchmark datasets or, even better, if any new data is collected, it can be distributed to other researchers. Although the number of publicly available datasets is still quite scarce, new anti-spoofing databases appear gradually due to the increasing interest in anti-spoofing by the research community (Tan *et al.* 2010, Chingovska *et al.* 2012, Zhang *et al.* 2012, Erdogmus & Marcel 2013b, Wen *et al.* 2015) and international competitions (Chakka *et al.* 2011, Chingovska *et al.* 2013b). The benchmark datasets have been indispensable tools for the researchers by giving them a chance to concentrate on investigating the problem of anti-spoofing, which has had a significant impact on the amount of papers on data-driven countermeasures during the recent years.

Also the directions for face anti-spoofing explored in this thesis are largely based on three publicly available databases, the NUAA Photo Imposter Database (Tan *et al.* 2010), the Replay-Attack Database (Chingovska *et al.* 2012) and the CASIA Face Anti-Spoofing Database (Zhang *et al.* 2012) which are introduced in this section. It is worth mentioning that two valuable datasets have been released after the experiments reported in this thesis have been conducted: the 3D Mask Attack Database[11] (3DMAD) by Erdogmus & Marcel (2013b) which is the first public dataset to include 3D masks in

[11] `http://www.idiap.ch/dataset/replayattack`

2D face anti-spoofing, and the MSU Mobile Face Spoofing Database[12] (MFSD) by Wen *et al.* (2015) that facilitates spoof detection research on mobile phone applications.

The NUAA Photo Imposter Database

The NUAA Photograph Imposter Database[13] (PID) by Tan *et al.* (2010) can be considered as the first publicly available spoofing database because the evaluation is based on binary classification task of differentiating genuine faces from fake ones with a predefined protocol. The dataset contains images of both real client accesses and photo attacks using both photo-quality and laser-quality prints that were collected in three sessions at intervals of about two weeks. During each session, the environmental and illumination conditions were different. Examples of cropped facial images from the database can be seen in Fig. 6. The client accesses and spoofing attacks were recorded using a generic webcam with resolution of 640×480 pixels and altogether there are about 500 images (20fps) for each subject's recording. When capturing the data, the main idea was to make the live subjects look like a static as much as possible by minimizing the movements and the eye-blinking, i.e. resembling a photograph. In contrast, five different vivid photo-attacks were simulated using 2D facial prints with varying motions. The high-quality photos of the targeted person were printed on photographic paper of two sizes 6.8cm × 10.2cm (small) and 8.9cm × 12.7cm (big) using a traditional development method, or on a 70g white A4 paper using a conventional Hewlet-Packard color printer. Unfortunately, the printing option is not included in the meta data of the database.

The dataset is composed of images of fifteen subjects (in three sessions for most of the subjects) that are decomposed into two separate sets for training and test purposes. The training set consists of images from the first two sessions only. The test set consists of the images from the remaining third session. The training set contains altogether 1,743 face images of nine real clients (889 and 854 from the first and the second sessions, respectively) and 1,748 imposter images of the same nine clients (855 and 893 images from the first and the second sessions, respectively). The test set is constructed from 3,362 client samples and 5,761 imposter images taken during the third session. Only three clients who took part in the first two sessions attended the third session.

[12] `http://www.cse.msu.edu/rgroups/biometrics/Publications/Databases/MSUMobileFaceSpoofing/index.htm`

[13] `http://parnec.nuaa.edu.cn/xtan/NUAAImposterDB_download.html`

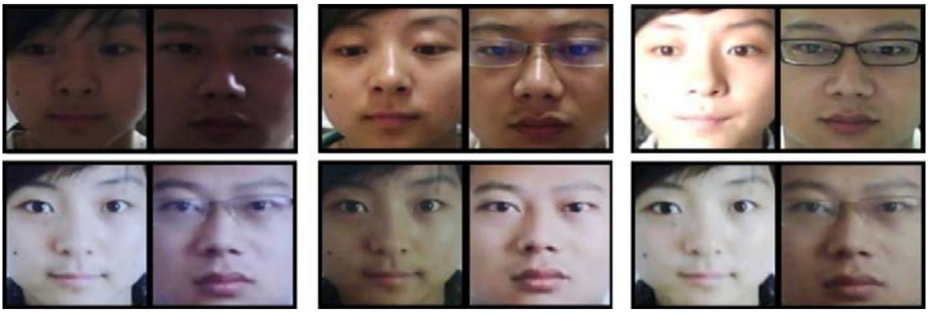

Fig 6. Cropped face samples from the NUAA PID by Tan *et al.* (2010). In each group of samples (from left to right) are from session 1, session 2 and session 3, respectively. Upper rows are from a live human and the lower row from a corresponding photo.

Furthermore, six new clients and their photographs are introduced in the test set to further increase the level of difficulty. There is no specific development set provided in the database, thus cross-validation or fixed validation set has to be used for tuning the algorithms. The database contains also the data needed for face normalization and the geometrically normalized face images of 64 × 64 pixels which were used in the experiments by Tan *et al.* (2010), thus making it easier to compare the results between different spoof detection techniques.

The Replay Attack Database

The Replay-Attack Database[14] by Chingovska *et al.* (2012) and its subsets (the Print-Attack Database (Anjos & Marcel 2011) and the Photo-Attack Database (Anjos *et al.* 2013)) consist of short video recordings (roughly ten seconds) of both real accesses and corresponding attack attempts. The studied attack scenarios in the dataset can be categorized based on display media, A4 sized hard copy (print), iPhone 3GS (mobile) and iPad with a resolution of 1024×768 (highdef) and two attack types, photo and video. Also two illumination conditions are introduced: controlled with uniform background scene and fluorescent lamp illumination and adverse with non-uniform background scene and day-light illumination. The videos clips were captured using an Apple 13-inch MacBook laptop and its embedded webcam with a relatively low-quality resolution of 320 × 240 pixels (QVGA) at 25 fps.

[14] `http://www.idiap.ch/dataset/replayattack`

When recording the real client accesses of fifteen seconds, the subjects were asked to look at the laptop camera as during normal authentication process. Unlike in traditional laptop authentication scenario, the laptop was placed on top of a small stand in order to capture frontal-pose faces. For creating the attacks, two photographs and two video clips were taken of each person in each of the two illumination and background settings used for recording the real accesses. The first photograph/video clip was recorded using iPhone 3GS (3.1 megapixel camera) and the second using a high-resolution 12.1 megapixel Canon PowerShot SX200 IS camera. To maximize the attack quality, the subjects were asked to look up-front like in the case real access attempts. As will be seen in Section 4.2.1, the frame of a spoofing medium might be easy to detect, thus the fake faces are executed in a way that the border of the display media is not visible in the final video clips of spoofing attacks. Furthermore, each spoofing attack video clip of ten seconds was recorded with two different support modes, hand-held and fixed-support. Figure 7 shows examples of the genuine and attack samples in the different conditions explored by the Replay Attack Database.

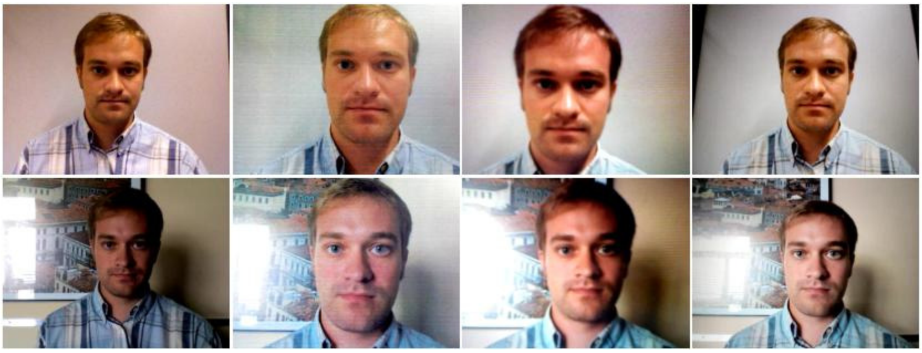

Fig 7. Example of a valid client and the corresponding scenic attacks in different scenarios and with different lighting conditions, extracted from the Replay-Attack Database by Chingovska *et al.* (2012). On the top row, attacks in the controlled scenario. At the bottom, attacks with samples from the adverse scenario. Columns from left to right show examples of real accesses, hard-print, photo and video attacks.

In total, the Replay-Attack Database contains 50 different identities and 1,300 video clips of which 300 correspond to real-accesses (three trials in two different conditions for each of client. The first trial for each subject is dedicated solely for evaluating face verification systems, i.e. not used for evaluating anti-spoofing performance. The remaining 200 real-accesses and 1,000 attack video clips are divided into training, development and test sets (360, 360 and 480 videos, respectively) for evaluating the binary spoof detection classifiers. The subject-disjoint subsets were randomly selected, i.e. identities that are on one of the subsets do not appear in any other set. Thus, the anti-spoofing models are not trained for detecting person-specific appearance or facial dynamics. In order to enable system-level evaluation, i.e. the joint operation of recognition and anti-spoofing algorithm, the identities between the verification protocol and anti-spoofing protocols match. The dataset provides eighteen protocols for evaluating the effectiveness of the anti-spoofing methods under different conditions, including support, fake face type and quality.

The training set is used for training the countermeasure, whereas the development set operates as a separate validation set for estimating a threshold value to be used on the test set. The database protocol defines the EER as a decision threshold. The actual test set is used only to report results. As a performance measure, the protocol suggests reporting the HTER on the test data. The dataset provides also automatically annotated face bounding boxes for convenience.

The CASIA Face Anti-Spoofing Database

The CASIA Face Anti-Spoofing Database[15] (FASD) by Zhang *et al.* (2012) introduces some significant improvements to previous databases because it provides more variations in the collected data. The authors indicated that imaging quality of different cameras is an important factor that may influence the robustness of anti-spoofing techniques, especially methods analysing the facial texture. Thus, the database contains data from 50 real clients and the corresponding forged samples collected using three different devices with varying quality, old webcam (low quality) with resolution of 480×640, new webcam (normal quality) both with resolution of 640×480 and a Sony NEX-5 digital system camera with a resolution of 1920×1080 (high-quality). However, in order to save memory and computational burden, the original 1920×1080 resolution videos have been cropped into patches of 1280×720 pixels which contain only the face region, thus maximizing the appearance quality of the target faces. Example images of a genuine face at the different imaging qualities can be seen in Figure 8.

Both real client accesses and the corresponding attack attempts are captured in natural office scenes. The subjects are required to exhibit eye blinking during data capture as the authors argue that motion is crucial cue for face spoof detection, thus it

[15] `http://www.cbsr.ia.ac.cn/english/FaceAntiSpoofDatabases.asp`

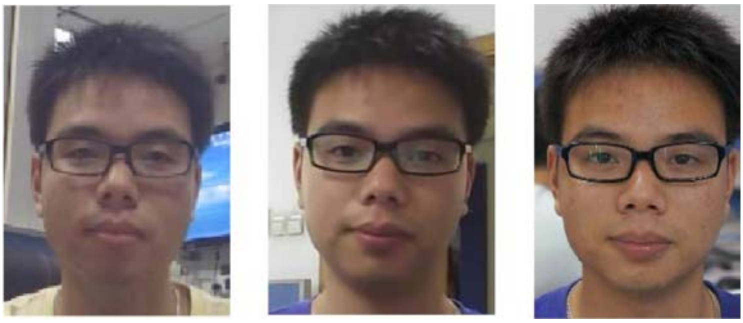

Fig 8. Samples showing the different imaging qualities (low, normal and high, respectively) extracted from the CASIA FASD by Zhang *et al.* (2012).

is important to provide motion like in challenge-response based systems. The attack scenarios in the dataset are based on three types of fake faces which include warped photo, cut photo (photographic mask) and video attacks. The high-quality samples of the targeted faces were generated from the videos captured with the Sony NEX-5. The facial prints were printed on copper paper in order to achieve better quality spoofs compared to conventional A4 printing paper and to avoid printing artefacts that are very obvious in the Replay-Attack Database (Chingovska *et al.* 2012). The warped photo attacks were performed like in the work by (Pan *et al.* 2008), Kollreider *et al.* (2008) and Tan *et al.* (2010). The eye regions were cut off in order to create photographic masks and eye blinking was simulated either by the attacker or by sliding another piece of paper behind the resulting cut photo (see, Figure 9). The video attacks were executed using iPad with a screen resolution of 1024×768, thus the original high resolution of 1280×720 is downsized compared to the photo attacks (see, Figure 9).

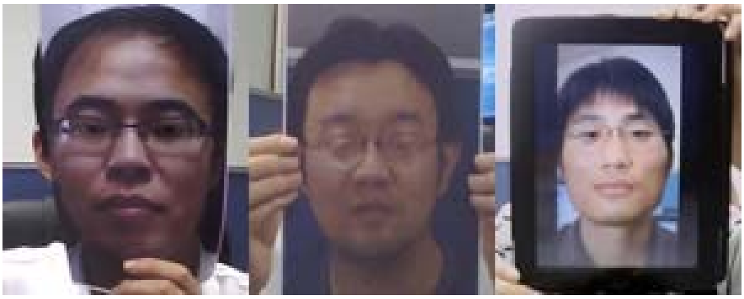

Fig 9. Cut-photo attacks (imposter blinking behind the mask and intact photo moving behind the cut one) and a video attack, extracted from the CASIA FASD (Zhang *et al.* 2012).

Altogether the database consists of 600 video clips and the identities are divided into subject-disjoint subsets for training and testing (240 and 360, respectively). Since the main purpose of the database is to investigate the possible effects of different fake face types and imaging qualities, the test protocol consists of seven scenarios in which particular train and test samples are to be used. The quality test considers the three imaging qualities separately, low (1), normal (2) and high quality (3), and evaluates the overall spoof detection performance under a variety of attacks at the given imaging quality. Similarly, the fake face test assesses how robust the anti-spoofing measure is to specific fake face attacks, warped photo (4), cut photo (5) and video attacks (6), regardless of the imaging quality. In the overall test (7), all data is used to give a more general evaluation. Examples of the different scenarios in the database can be seen in Figure 10. The results of each scenario are reported as Detection-Error Trade-off (DET) curves and EER. Results of a baseline system are also provided along the database. Inspired by the work by Tan *et al.* (2010), the baseline system considers the high frequency information in the facial region using multiple DoG features and SVM classifier.

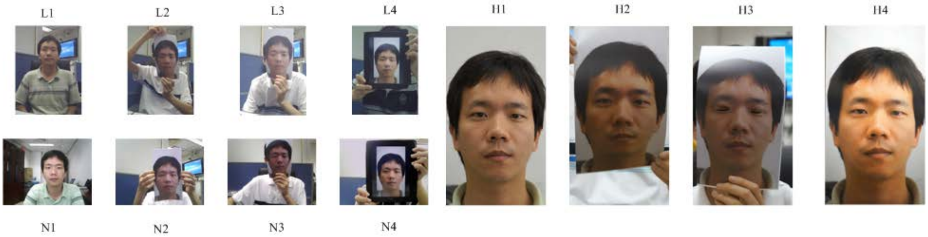

Fig 10. Samples showing a genuine face (1) the corresponding three types of attacks, including warped photo (2), cut photo (3) and video (4) attacks, across the different imaging qualities, low (L), normal (N) and high (H), present in the CASIA FASD by Zhang *et al.* (2012).

2.3 Discussion

This section gave an extensive overview on the vulnerabilities of biometric systems and anti-spoofing strategies with particular focus on face modality and software-based countermeasures. It is no secret that biometric systems without dedicated countermeasures are vulnerable to spoofing attacks because they try to maximize the discriminability between identities without regards to whether the presented trait originates from a living legitimate client or not. This is especially true for face based identity verification because falsifying face biometric data is very straightforward compared to other modalities, e.g. fingerprint or iris. For instance, one can download a high quality facial image of the targeted person from social networks and present the compromised biometric data using a photograph or a display device which are not very expensive to acquire considering the costs and potential benefit.

The performance of the actual biometric recognition algorithms is beginning to be mature enough for consumer-level applications, e.g. e-commerce or online banking, but the vulnerability to spoofing is one of the main problems for companies willing to market identity management solutions based on biometric technologies. Therefore, the research community has had a significant increase in anti-spoofing which can be attested by the growing number of articles, publicly available datasets and the various competitions that started to appear in major biometric forums. Still, there exists no clear consensus on the terminology related to spoofing, let alone presentation attacks in general, evaluation protocols or best anti-spoofing practices.

Intuitively, hardware-based techniques, e.g. NIR, thermal and 3D imaging, provide protection against several types of fake face attacks from 2D surfaces, e.g. prints and video displays, to 3D masks. However, constructing a large scale spoofing related dataset is even more expensive and time-consuming than the collection of conventional biometric data because also a statistically significant amount of forged biometric traits, like 3D masks, needs to be manufactured, not to mention that there are numerous ways of presenting the synthetic artefacts to the input sensor. Thus, custom imaging solutions have been usually evaluated on small proprietary datasets if they have been validated at all, which makes their performance comparison impossible. Another problem with sensor based approaches is that they can be seen as expensive and impractical due to the unconventional imaging devices that are not yet widely available in consumer-level products.

In the meanwhile, it is rather appealing to use software-based countermeasures that are operating on data acquired with the cameras embedded in the existing face authentication systems. Another advantage of software-based countermeasures is that, in most cases, they can be assessed using the same data. Few publicly available databases covering a wide range of attack scenarios have been released and the number is gradually increasing. This has also had a huge impact on the software-based anti-spoofing research and many data-driven anti-spoofing techniques have been proposed during the recent years. The early works in software-based face anti-spoofing have been studying motion based visual cues because they are very intuitive from liveness detection point of view. Also facial appearance analysis based approaches have been proposed in the literature because there are some inherent disparities between genuine faces and fake ones due to differences in reflection, shading and pigment properties. The prior work has been an important kick-off but the research on software-based anti-spoofing is still in its infancy. Thus, many questions need to be answered in order to develop robust and reliable software-based solutions for increasing the security of 2D face biometrics under spoofing attacks. This thesis complements the prior works by exploring new approaches for non-intrusive data-driven anti-spoofing.

The major drawback of early works in texture based face anti-spoofing is that high resolution input images are required in order to discriminate the fine micro-texture of the used display medium or the methods can detect only down-sampled fake faces. In Chapter 3, this issue is addressed by exploring the structure of facial micro-textures and gradient structures using well-known low-level feature descriptors on conventional webcam-quality images. Furthermore, the micro-texture analysis based spoof detection is extended to the spatiotemporal domain in order to exploit the facial dynamics information in addition to appearance.

Another limitation with the prior work is that the proposed countermeasures have been developed for performing anti-spoofing in a generic manner, i.e. treating all sorts of attack scenarios equally. Since it is reasonable to assume that no single superior technique is able to perform robustly under all known, let alone unseen, attack scenarios, in Chapter 4, the problem of face anti-spoofing is broken down into attack-specific sub-problems that are easier to solve using a proper combination of countermeasures. Two attack-specific countermeasures are proposed for complementary attack scenarios. Furthermore, the fusion of countermeasures is also investigated more closely because it has not been explicitly studied besides the international competitions.

3 Facial texture analysis based anti-spoofing

As mentioned in Section 2.2.2, it can be assumed that the inherent disparities between genuine faces and fake ones can be captured by describing the static and dynamic facial properties, such as reflectance, shading, texture or quality. Since these properties can be considered also as variations in the facial texture information, methods analysing them are usually referred to as texture based techniques. This chapter summarizes the main ideas and findings presented in Papers II-V that propose to approach the problem of face anti-spoofing from the texture analysis point of view.

3.1 Static texture based approach

Face images captured from printed photos may visually look very similar to the images captured from genuine ones. Consequently, all these images would be largely overlapping in the original input space. Therefore, a suitable feature space is needed for separating the two classes (genuine against fake face images). The main issue is how to derive such a feature space.

A closer look at the differences between real faces and face prints reveals that human faces and photographs reflect light in different ways because a human face is a complex non rigid 3D object, whereas a photograph can be seen as a planar rigid object. The surface properties of real faces and prints, e.g. pigments, are also different. These intrinsic properties may cause different (specular) reflections and differences in shading. In addition, face prints often contain printing artefacts, such as jitter and banding (Eid *et al.* 2011), that can be observed well, especially on uniform or smooth areas of the face like cheeks. Skin has also a very particular texture with, for example, pores, whereas fake faces have seldom such a level of detail. Spoofing attacks when executed with face prints tend to engender some overall image blur because, for example, of a low-resolution printing device or original image of the target person, or rapid motion caused by simulated vivid photo-attacks performed like in the NUAA PID (Tan *et al.* 2010). Furthermore, glossy surface of a photograph may cause problems for auto-focusing of cameras. Example images of possible cues used for face print detection are presented in Figure 11.

Inspired by the observations above, Papers II and III study whether the characteristics between a real face and a face print could be captured by describing the texture and

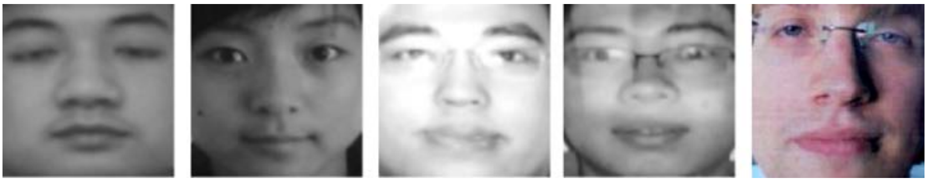

Fig 11. Examples of face print properties that could be used for spoof detection, for example, overall image blur, low contrast, characteristic specular reflections and printing artefacts. Paper III © IET.

gradient structure of the face images with a set of well-known low-level features. The facial texture representation in Paper II was based solely on LBP (Ojala *et al.* 2002) features. The work was extended in Paper III by including Gabor wavelet features (Manjunath & Ma 1996) into the texture description and combining local shape information to the face representation using HOG features (Dalal & Triggs 2005).

The LBP texture analysis operator, introduced by Ojala *et al.* (2002), is defined as a grey-scale invariant texture measure, derived from a general definition of texture in a local neighbourhood. It is a powerful means of texture description and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic grey-scale changes. Uniform LBP patterns were chosen in Paper II and Paper III in order to keep the feature histograms more compact histogram (the interested reader is referred to see the work by Ojala *et al.* (2002) for more details).

The basic idea behind Gabor based texture features is to extract features at multiple scales and orientation using Gabor wavelet decomposition. The Gabor filters can be seen as orientation and scale tunable edge and line detectors whose statistics can be used for encoding the texture information in a region (Manjunath & Ma 1996). For classification purposes, a feature vector is constructed using the mean and standard deviation of the magnitude of the transform coefficients at different scales and orientations (Manjunath & Ma 1996).

The local shape characteristics are introduced to the face representation using HOG descriptor (Dalal & Triggs 2005) that has been widely used for object detection. HOG feature captures the edge or gradient structures of the facial image by decomposing the given image into square cells and computing a histogram of the oriented gradients over each cell. HOG representation is invariant to local geometric transformations if translations or rotations are much smaller than the local spatial or orientation bin size.

The pipeline for performing the face anti-spoofing from single images is presented in Figure 12. First, the face is detected, cropped and normalised into an M×M pixel grayscale image. Since some important visual cues, like printing artefacts, shading and characteristic specular reflections, are observed better locally, the facial images are spatially partitioned into several sub-regions, each of which corresponds to a local patch of the face image. Then, the descriptions are extracted from each block and the resulting feature vectors are concatenated into an enhanced feature vector. Such a representation has shown to be very adequate for face recognition (Ahonen *et al.* 2006) and face detection (Hadid *et al.* 2004). In order to get an overall description of the face image, also a holistic description can be included in the spatial information by computing the features over the whole face image, like it is done in the case of the LBP based representation.

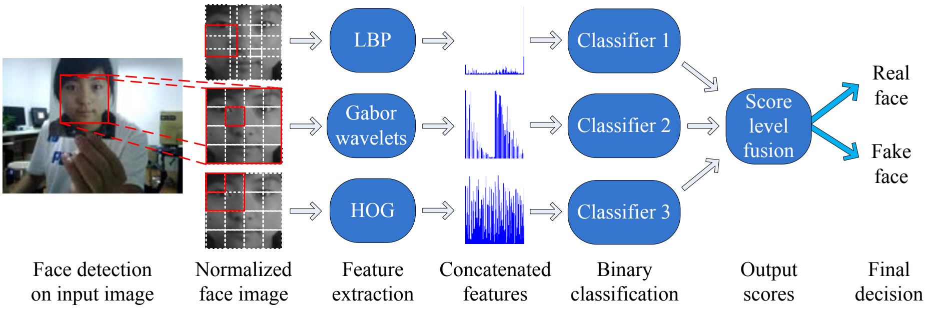

Fig 12. Block diagram illustrating face spoof detection from a single image.

Each low-level descriptor produces its own face representation which is then fed to a classifier. The output score value describes whether there is a live person or a fake one in front of the camera. In Paper II and Paper III, an SVM based classification schemes were used. In Paper III, the outputs of complementary face representations were fused at score level. Since some of the feature spaces are more discriminative than others, the final score value was obtained by computing the weighted average of the individual z-scores.

3.2 Dynamic texture based approach

Motion is indeed a very powerful cue for liveness detection because it can be assumed that a live human being will exhibit some involuntary facial movements, including eye blinking, mouth movements and facial expression changes, in front of the camera. However, these vitality cues can be easily deceived using video replay attacks, for instance, and the lack of motion during short video sequences may lead to high FFR.

There exists also another group of motion patterns useful for face anti-spoofing that do not originate from genuine faces. The movement of the display medium may cause several distinctive dynamic patterns that do not describe genuine faces. As shown in Figure 13, the use of (planar) spoofing medium might cause sudden characteristic reflections when a photograph is warped or because of a glossy surface of the display medium. It can be also seen in Figure 13 that warped photo attacks may cause also distorted facial motion patterns. Furthermore, it is likely that hand-held attacks engender synchronized shaking of the face and spoofing medium, which can be observed as excessive unnatural motion in the view if the distance between the display medium and the camera is relatively short. Like printed photographs, also replayed videos suffer from dynamic quality defects, either due to the video itself (compression artefacts) or the captured display device, e.g. moiring and flickering.

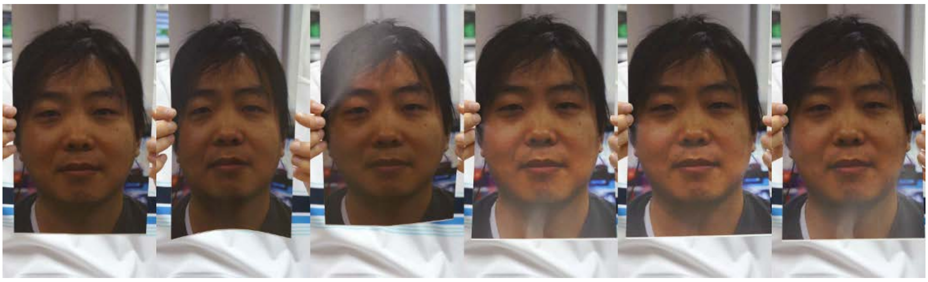

Fig 13. Example sequence of a warped photo attack from the CASIA FASD (Zhang *et al.* 2012). This describes the characteristic reflections (flickering) of a planar spoofing medium and the distorted motion patterns. Paper V © Springer.

All aforementioned temporal visual cues can be seen as temporal variations in texture patterns (dynamic texture). Some of them can be detected from single face images using static texture analysis but they might not be present during that particular moment of capture (see, Figure 13). However, if the temporal dimension is considered, it is more likely that these visual cues are observed. Where traditional motion based countermeasures are limited for exploiting specific liveness and shape cues, dynamic texture analysis provides means for describing both static texture patterns and various motion patterns that are characteristic to either a genuine human face or a fake one.

Since multiple visual cues from static and dynamic artefacts to liveness are utilized at the same time, the approach should be more robust to various attack types, including video replays and animated fake faces.

The micro-texture analysis based spoof detection was extended into the spatiotemporal domain in Paper IV and Paper V by exploring the dynamic texture content of the facial region. A compact face liveness description was proposed that combines both facial appearance and motion based visual cues using the spatiotemporal (dynamic texture) extensions of the LBP approach. More specifically, local binary patterns from three orthogonal planes (LBP-TOP) were considered which have shown to be very effective in dynamic texture recognition, face and facial expression recognition, lip-reading, and activity and gait recognition (Pietikäinen *et al.* 2011).

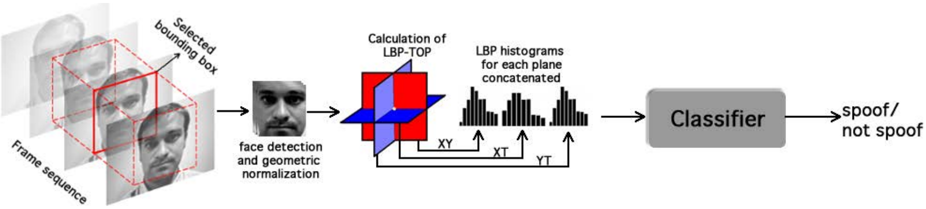

Fig 14. Block diagram of a generic LBP-TOP based countermeasure. Paper V © Springer.

Zhao & Pietikäinen (2007) extended the LBP operator to a spatiotemporal representation for dynamic texture analysis. The dynamic LBP description from three orthogonal planes of a space time volume is formed by applying LBP operator on the individual planes and concatenating the resulting histograms (see, Figure 14). In this manner, also the vertical and horizontal motion patterns (of xt and yt planes, respectively) can be encoded in addition to appearance (xy plane).

Figure 14 illustrates how the LBP-TOP description based face anti-spoofing is performed from short video sequences (space time volumes) in general. First, face detection is applied on each video frame and the resulting face bounding boxes are geometrically normalized. In order to reduce the noise of face detection, the detected faces are not cropped to form a space time volume. Instead, the LBP-TOP description of each frame is extracted from a volume that is determined by the face bounding box in the current frame (see, Figure 14). The histograms of each frame are accumulated over the whole input video sequence to form the final dynamic texture description of the facial region. The final feature vector is then fed to a binary classifier.

3.3 Experiments

This section summarizes the main results and key findings of Papers II-V. For more detailed results, please refer to the original papers.

3.3.1 Print attack detection using static texture

In Paper II, the performance of the LBP based print attack detection was evaluated on the NUAA PID. The combination of local and holistic texture descriptions yields total classification accuracy of 98.0%, FLR of 0.6% and FFR of 4.4%.

A closer look at the misclassified samples revealed that mainly over-exposed and very blurry images of client faces were labeled as face prints, whereas the successful attacks were mainly face prints with better recapture quality (see, Figure 15). In the NUAA PID, the main idea was to make the live subjects look like a static as much as possible by minimizing the movements and the eye-blinking. In contrast, the photo-attacks were simulated with varying vivid motions in order to fool motion (liveness) based countermeasures. Therefore, the facial texture quality of the fake faces is much worse compared to the live ones and also strong characteristic specular reflections are commonly observed, which explains the low FLR of the proposed method.

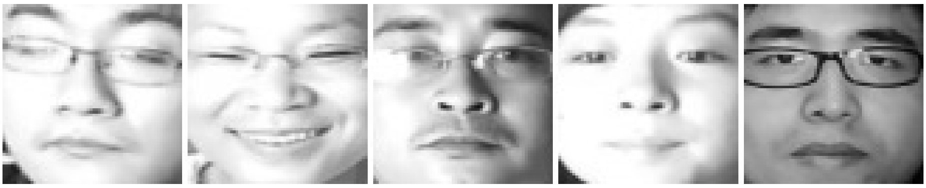

Fig 15. Examples of misclassified images. The first four images from the left represent rejected genuine subjects and one on the right is a photograph. Paper II © IEEE.

The proposed LBP based face representation is indeed adequate for measuring the facial texture quality and determining whether other inherent differences, like differences in shading and reflections, are observed. On the other hand, the main drawback of the method is that the quality of genuine face images is assumed to be reasonable. For instance, in the NUAA PID, the acquisition conditions of some clients are very challenging most of time during the recording session, which explains rather high FFR of the proposed approach (see, Figure 15).

Table 1. A summary of results from the original papers in the print-attack detection experiments.

Paper II. Overall performance (in %) of the proposed LBP based face description on the NUAA Photograph Imposter Database.

Method	FFR	FLR	Classification rate
LBP	**4.4**	**0.6**	**98.0**

Paper III. Performance comparison between the different feature combinations on the NUAA Photograph Imposter Database.

Method	EER (%)
LBP	**2.8**
Gabor + HOG	**2.4**
LBP + Gabor	**2.0**
LBP + HOG	**1.5**
LBP + Gabor + HOG	**1.1**

Paper III. Performance comparison between the method approach and the teams who participated in the IJCB 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks (Chakka *et al.* 2011).

	Devel	Test		
Method	EER	FLR	FFR	HTER
AMILAB	**0.00**	**0.00**	1.25	0.63
CASIA	1.67	**0.00**	**0.00**	**0.00**
IDIAP	**0.00**	**0.00**	**0.00**	**0.00**
SIANI	1.67	**0.00**	21.25	10.63
UNICAMP	1.67	1.25	**0.00**	0.63
UOULU	**0.00**	**0.00**	**0.00**	**0.00**
LBP + Gabor + HOG	1.67	**0.00**	**0.00**	**0.00**

In Paper III, the experimental analysis on NUAA PID was extended and the performance of different feature combinations was evaluated. As seen in Table 1, LBP based face representation alone is already very discriminative but more robust performance is obtained when it is fused with the Gabor and HOG feature based descriptions. Table 1 shows a performance comparison between the proposed approach, consisting of LBP, Gabor and HOG, features, and the teams who participated in the IJCB 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks (Chakka *et al.* 2011). Almost all methods seem to work extremely well on the data set, including the proposed approach which is able to obtain perfect results on the test set and only two videos of the development set were incorrectly classified. It is worth mentioning that the proposed countermeasure considered only the face region where the best-performing algorithms utilized also the background region.

3.3.2 Generic anti-spoofing using static and dynamic texture

In Paper IV, the effectiveness of the LBP-TOP based countermeasure was assessed on the CASIA FASD. When deriving the proposed face liveness description, scaling was avoided during geometric face normalization in order to keep all valuable information about the facial texture quality. The face bounding box was expanded to cover roughly also the contour and forehead regions of the face based on the detected eye locations. On the test set, the first two seconds from the beginning of each video sequence are used for determining whether a genuine face or a fake one is observed.

Table 2. Summary of results from the original papers in the dynamic texture experiments on the CASIA FASD.

Paper IV. EER (in %) comparison between the DoG baseline method, regular LBP and LBP-TOP descriptions extracted from expanded face bounding boxes.

Scenario	Low	Normal	High	Warped	Cut	Video	Overall
Zhang *et al.* (2012)	13	13	26	16	6	24	17
LBP	4	10	**0**	4	4	**1**	4
LBP-TOP	**3**	**3**	1	**1**	**2**	**1**	**2**

Paper V. EER (in %) comparison between the DoG baseline method, LBP^{u2}, LBP-TOP^{u2} and regular LBP-TOP descriptions extracted from 64×64 pixel images.

Scenario	Low	Normal	High	Warped	Cut	Video	Overall
Zhang *et al.* (2012)	13	13	26	16	**6**	24	17
LBP^{u2}	11	17	**13**	13	16	16	16
LBP-TOP^{u2}	**10**	**12**	**13**	**6**	12	**10**	**10**
LBP-TOP	**7**	**10**	**7**	**2**	**9**	**9**	**10**

The main results of Paper IV are summarized in Table 2. In general, the spatiotemporal face representation (LBP-TOP) is more robust when compared to two static methods, LBP and the DoG baseline, thus confirming the benefits of encoding and exploiting not only the facial appearance but also the facial dynamics information. However, the facial appearance description (LBP) works perfectly at the highest imaging quality because the skin texture of genuine faces looks strikingly sharper compared to the fake ones. On the other hand, the imaging quality tests show that the use of facial dynamics enhances the spoof detection results at lower imaging qualities. Furthermore, the fake face test indicates that adding temporal planes to the face description improves the robustness to different types of spoofing attacks across the different imaging qualities. Thus, the measurement of facial texture quality seems to provide sufficient means for face anti-spoofing if the imaging quality is good enough to capture the fine details, whereas motion patterns are more important visual cues at lower imaging qualities. The limited iPad screen resolution (Zhang *et al.* 2012) and the occasionally visible video screen frame within the face bounding boxes explain partially the less challenging nature of the video replay attacks.

The texture analysis based face anti-spoofing was extended into the spatiotemporal domain coincidentally in Paper IV and by de Freitas Pereira *et al.* (2013a). Even though LBP-TOP based dynamic texture description was considered in both works, very dissimilar strategies were introduced for exploring the benefits of temporal dimension. de Freitas Pereira *et al.* (2013a) extracted the LBP-TOP based face liveness description from short time windows using the dense sampling of multiresolution approach, whereas an average of LBP-TOP features over longer temporal windows was used in Paper IV.

Paper V consolidated the work in Paper IV and by de Freitas Pereira *et al.* (2013a) and provides an in-depth analysis on the use of dynamic texture for face liveness description. In Paper V, unified experimental setup and evaluation methodology was applied for assessing the effectiveness of the different temporal processing strategies on Replay-Attack Database and CASIA FASD. In order to isolate the effect of face normalization, i.e. to be consistent with previous studies (Chingovska *et al.* 2012, Tan *et al.* 2010), the analysed facial region was tightly cropped and normalized into 64×64 pixels. Furthermore, uniform LBP pattern encoding (LBP^{u2}) was applied when extracting the LBP-TOP descriptions instead of regular mapping used in Paper IV.

In some attack scenarios, the characteristic motion patterns are hard to observe within very short time windows. The LBP-TOP based face description can be accumulated over longer periods of time, either by averaging the features within a time window like in Paper IV or by classifying each sub-volume and then averaging the scores within the current window (de Freitas Pereira *et al.* 2013a). Figure 16 illustrates how the used temporal window size affects the performance when the facial appearance and dynamics information are accumulated over time using the two temporal processing strategies. The results indicate that when the amount of temporal information increases, the better discrimination between real faces and fake ones is obtained. Furthermore, by

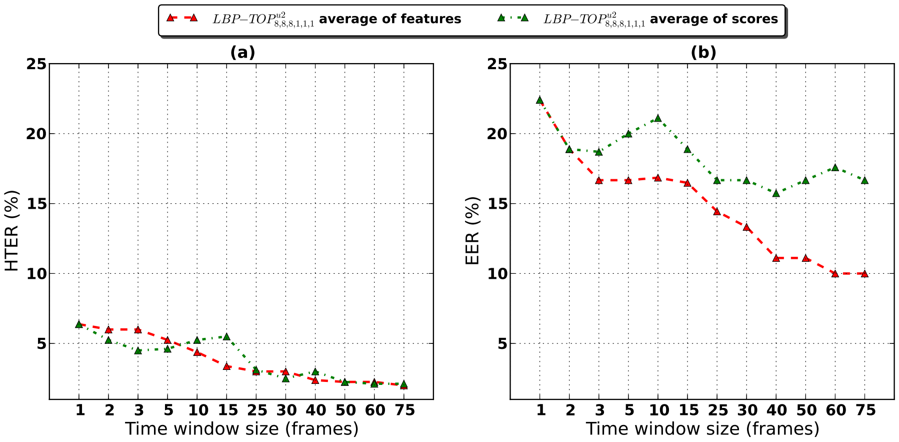

Fig 16. Evaluation of the LBP-TOP based face description with two temporal processing and different time window sizes (a) Replay-Attack Database (HTER %) (b) CASIA FASD (EER %). Paper V © Springer.

averaging features, more stable and robust spoof detection performance is achieved on both databases. It is also worth mentioning that the performance on the Replay-Attack dataset is consistently worse compared to the CASIA FASD. These observations can be explained by taking a closer look at the differences in the databases and their spoofing attack scenarios.

The Replay-Attack Database consists of scenic attacks in which the background scene is incorporated in the face spoof and the resulting scenic fake face is placed very near to the sensor. In such cases, the description of facial appearance leads to rather good performance because the proximity between the spoofing medium and the camera causes the recaptured face image to be defocused, also revealing other facial texture quality issues. Furthermore, the attacks in Replay-Attack Database are performed using two types of support conditions, fixed and hand-held. Naturally, the LBP-TOP based face representation can easily detect fixed photo and print attacks since there is no variation in the facial texture over time. On the other hand, the hand-held attacks introduce synchronized shaking of the face and spoofing medium. This can be observed as excessive relative motion in the view, again, due to the proximity between the display medium and the sensor. Since the aforementioned visual cues can be captured within relatively short temporal windows, the use of longer temporal windows does not provide too much gain in performance or difference between the two temporal processing strategies.

In contrast, the CASIA FASD consists of close-up fake faces that describe only the facial area. The distance between the camera and the display medium is much farther compared to the attacks on Replay-Attack Database. Furthermore, the display medium does not usually move much during the attack attempt. Therefore, the overall global movement of a fake face is much closer to the motion of a genuine face. Due to the lack of distinctive shaking of the display medium, the CASIA FASD can be considered to be more challenging from the dynamic texture point of view. However, when longer video sequences are explored, we are more likely to observe other specific dynamic events, such as different facial motion patterns, e.g. eye blinking, lip movements and facial expression changes or sudden characteristic reflections of planar spoofing media which can be used for differentiating real faces from fake ones.

It can be seen in Table 2 that the results on the CASIA FASD in Paper IV are consistently better compared to ones in Paper V. The main reason for the performance gap is that the experiments were carried out using different face normalization techniques. In Paper V, the LBP-TOP description was extracted from down-sampled face images which reduces the effectiveness of static texture analysis, whereas the importance of motion cue stands out. For instance, warped photo attacks can be detected very well because of their characteristic specular reflections (flickering) and excessive and distorted motion. In contrary, the cut-photo and the video replay attacks are more challenging because their motion patterns are similar to the ones of genuine faces, i.e. eye-blinking and mouth movements can be observed. Furthermore, valuable details about the facial texture quality are lost when the face images are down-sampled to 64×64 pixels. Intuitively, the static texture information is more distinctive when face images with higher quality (e.g. resolution) are analysed. In Paper IV, this can be especially observed at the highest imaging quality.

The additional experiments, i.e. left out from Paper V, in Table 2 depict that the use of uniform patterns has a negative impact on the performance compared to regular LBP encoding, especially at high quality and photo-attack scenarios. Thus, the non-uniform patterns may contain some useful information on the dynamic texture that is lost due to the uniform encoding. Also the occasionally visible scenic cues in Paper IV explain partially the results. Nevertheless, a common finding is that the spatiotemporal face representation (LBP-TOP) is more robust than its static counterpart (LBP) in general.

3.4 Further work on texture based face anti-spoofing

After the publication of Paper II and Paper III, texture based face anti-spoofing has also been widely adopted by other researchers. For instance, in the work by Erdogmus & Marcel (2013b), the LBP based face description presented in Paper II showed promising performance, also in 3D mask attack detection, especially when block-based LBP descriptions were utilized on color or depth images instead of plain holistic description. Like in Paper IV, Yang *et al.* (2013) expanded the face bounding box and investigated more closely the benefits of extracting features over specific face components including eyes, nose, mouth, forehead, cheeks and contour regions instead of rigid block division used in Paper II and Paper III. The authors computed densely several low-level features, e.g. LBP and HOG, over the different components and formed a high-level face representation by pooling codes with weights derived from Fisher criterion. Both the use of expanded face bounding box and the proposed component dependent feature description had a positive impact on the performance because the inherent different appearances and shapes among different regions of the real 3D face are considered explicitly in the low-level feature description. It is also worth mentioning that some kind of LBP based face representation was used as a component in all best-performing algorithms in the latest competition on counter measures to 2D facial spoofing attacks (Chingovska *et al.* 2013b).

3.5 Discussion

In this chapter, the problem of non-intrusive face anti-spoofing was approached from the texture analysis point of view. It was investigated whether well-known low-level feature descriptors can be used for describing the inherent disparities between genuine faces and fake ones. In the ideal case, face spoof detection could be performed from a single face image, thus the static texture based method was proposed for analysing the appearance. Since also motion cue is very important for anti-spoofing, the approach was extended to the spatiotemporal domain and dynamic texture analysis was applied for exploiting both appearance and motion information within relatively short video sequences.

The extensive experimental analysis on the latest benchmark datasets depicted that static texture analysis provides means for face anti-spoofing in general if the input image quality (e.g. resolution) is high enough. More importantly, the results suggest that the cropped face images should not be down-sampled too much during preprocessing stage in order to keep all valuable information about the facial texture (quality) which is a crucial visual cue in spoof detection. If the performance of a new feature description is compared to the related literature, it might not be a good idea to use just small down-sampled face images to be consistent with previous studies, like (Tan *et al.* 2010, Chingovska *et al.* 2012) and Papers II, III and V, but to recompute the results using different face scales and bounding boxes. A good example of the latter is the work by Yang *et al.* (2013) and Wen *et al.* (2015) in which the effect of face normalization has been mitigated and studied more closely.

While lower imaging quality might be enough for detecting the most crude attack attempts, such as small mobile phone displays and prints with strong artefacts, and differences in shading, other fine details cannot be distinguished any more using static texture information. Also high false rejection rate might be an issue if acquisition quality is not good enough. The use of facial motion patterns showed its importance when face anti-spoofing is performed at lower imaging qualities or if the quality of the input face images is otherwise poor. The main drawback with dynamic approach is that its robustness improves when longer temporal windows are analysed because people tend to start moving more in front of the camera as time passes, thus reducing the false rejection rate as the amount of involuntary movement increases. However, face authentication process duration of two or three seconds should not be an issue in real-life applications, especially if basically no user interaction is required.

4 Attack-specific face anti-spoofing

As seen in previous chapters, many directions for non-intrusive anti-spoofing have been already explored and impressive results have been reported on individual databases. However, the proposed countermeasures have been designed in a generic way. In other words, the algorithms try to detect all sorts of attack types equally well by exploiting their common visual cues that are present in the development datasets. The varying nature of spoofing attacks and acquisition conditions in open environments makes it impossible to predict how single anti-spoofing techniques, e.g. facial texture analysis, can generalize the problem in real-world applications. Moreover, we cannot foresee all possible attack scenarios and cover them in databases because the imagination of the human mind always finds out new tricks ("golden fakes") to fool existing systems.

It is reasonable to assume that no single superior technique is able to detect all known, let alone unseen, spoofing attacks because every countermeasure most likely has its own golden fake that can be exploited by an attacker. Therefore, the problem of spoofing attacks should be broken down into attack-specific sub-problems that are solvable with a proper combination of countermeasures. Intuitively, an anti-spoofing solution consisting of several complementary spoof detectors probably performs more robustly also under unseen attack scenarios. Thus, it is important to find out which countermeasures are complementary and how the different techniques should be combined. This would provide valuable insight into how to construct a flexible anti-spoofing framework consisting of a network of attack-specific spoof detection modules. In this manner, the discovered vulnerabilities could be patched in no time when new countermeasures appear. This chapter summarizes the main ideas and findings presented in Papers VI and VII that propose to approach the problem of face anti-spoofing from the attack-specific point of view.

4.1 Complementary countermeasures

To follow the principle of attack-specific spoof detection and to find out well-generalizing countermeasures to particular attack scenarios, we first need to define clearly the spoofing and use-case scenarios that we are dealing with. Face spoofing attacks can be categorized in several ways. For instance, an obvious way is based on the used display medium type, such as photograph, video screen or mask. Paper VI and Paper VII approach the problem of face anti-spoofing by categorizing the 2D spoofing attacks (executed with a photograph or display device), into two groups based on whether the spoofing medium is visible in the view or not. In both attack scenarios, common, and more importantly, their own distinctive visual cues can be exploited in spoof detection schemes.

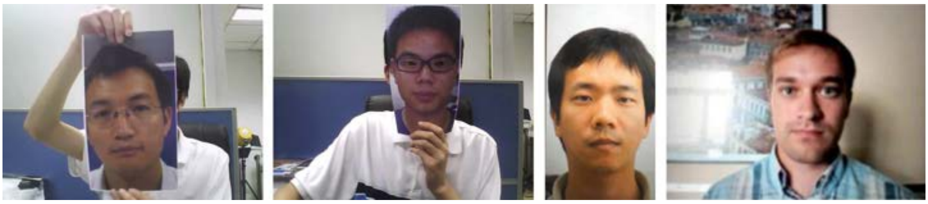

Fig 17. Examples illustrating the relative position of fake face to upper body, i.e. a simple photo attack and a photographic mask, and different ways of concealing the used display medium outside the provided view, i.e. a close-up face spoof placed very near to the sensor and a scenic attack.

In the scenario, in which the spoofing medium is visible, the attacks are usually performed using a close-up fake face that describes mainly the facial area which is presented to the sensor. The facial region is maximized on the display medium in order to reproduce the targeted face with the best possible quality. The main weakness with the tightly cropped face spoofs is that the boundaries of the spoofing medium, e.g. video screen frame or photograph edges, or the attacker's hands are usually visible during the attack, thus they can be detected in the scene. Moreover, if the fake face is not well aligned with the upper half of the torso of the imposter, a natural upper-body profile cannot be observed (see, Figure 17).

As illustrated in Figure 17, there are two ways for concealing the used display medium outside the view, either by placing a close-up face spoof very near to the sensor or by incorporating background scene in the face spoof (scenic attack). In the first case, the proximity between the fake face and the camera might cause the recaptured face image to be defocused and expose also other facial texture quality issues if the imaging quality is good enough for capturing the differences in fine details of surface properties between a human face and spoofing medium, as seen in Paper V and in the work by Bai *et al.* (2010). In the case of scenic attacks, the incorporated scene occupies a lot of the limited display area, thus the resolution and facial texture quality of the presented face is likely to be low compared to genuine faces. Furthermore, for stationary camera based systems, it should be possible to observe high correlation between the overall motion of the face and the background regions (Anjos & Marcel 2011, Yan *et al.* 2012) or to check if the background scene suddenly changes if the operating environment is fixed (Pan *et al.* 2011).

The following sections summarize how the aforementioned visual cues were exploited for face anti-spoofing in Paper VI and Paper VII. Paper VI introduces an approach for detecting the presence of the display medium in the view by analysing the available contextual information, whereas Paper VII concentrates on detecting scenic face attacks using a combination of motion correlation and texture based methods, and studying how different complementary countermeasures could be coupled.

4.1.1 Detecting the presence of spoofing medium

Scene information has been neglected in the previous works, i.e. the existing countermeasures are mainly based on different facial texture and motion analysis techniques. The background information has been exploited only for measuring the overall motion correlation between face and background (Anjos & Marcel 2011, Yan *et al.* 2012) or for checking if the background scene of a stationary face recognition system suddenly changes due to (scenic) face (Pan *et al.* 2011).

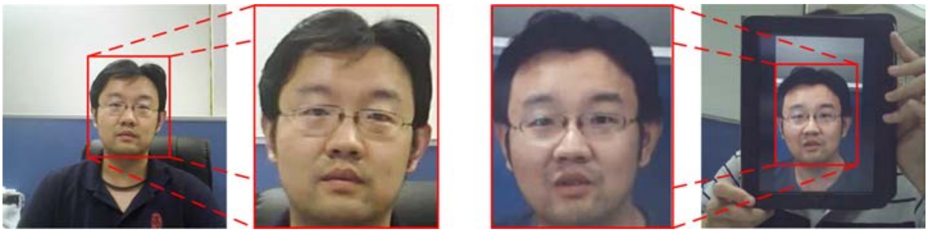

Fig 18. Example highlighting the importance of context information.

Face images captured from face spoofs may visually look very similar to the images captured from live faces, thus face spoof detection is rather difficult to perform based on a single face image or a relatively short video sequence only. Depending on the imaging and fake face quality, even for us humans, it is almost impossible to tell the difference between a genuine face and a fake one without any scene information or any unnatural motion or facial texture patterns. However, we can immediately notice if there is something suspicious going on in the view, e.g. if someone is holding a video display or a photograph in front of the camera (see, Figure 18).

Inspired by how we humans can perform reliable spoof detection only based on the available scene and context information, Paper VI presents a countermeasure that tries to mimic human behaviour for determining whether a spoofing medium is present in the observed scene. The proposed algorithm is based on analysing the surrounding region of the observed face with a HOG descriptor based detector that determines whether proper alignment of the face and the upper half of the torso or presence of the display medium is detected.

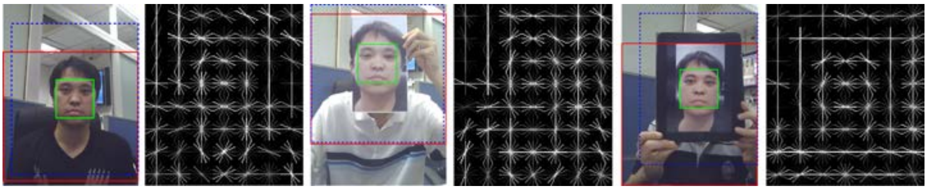

Fig 19. Example images from the CASIA FASD highlighting the differences between a genuine face and two spoofing medium types (photograph and video screen) and the corresponding HOG descriptors that are extracted from spoofing medium detection window (blue dashed lines). The red and green bounding boxes represent the detected upper-body and face location, respectively. Paper VI © IEEE.

The boundaries of the used spoofing medium, e.g. video screen frame or photograph edges, can be usually easily observed as very distinctive discontinuities around the face. Thus, the target face does not blend in the background scene as well as a genuine face and looks also a bit disconnected of the upper-body. Figure 19 presents example images of genuine face and two close-up fake faces, a photo and a video attack, and the corresponding HOG descriptors computed around the detected face. As can be seen, the local shape features can capture well the continuous edges of the used display medium that form a closed frame around the face, especially in the case of a video attack. The proposed method for spoofing medium detection computes the HOG descriptors from a bounding box that contains roughly the upper-body area and is symmetrically expanded also above the detected face. Once the features are computed, a linear SVM classifier is used for determining whether the presence of spoofing medium or other abnormalities, such as the attacker's hands or misaligned upper-body, is detected in the provided view.

4.1.2 Scenic attack description

Since there is no universal anti-spoofing technique for detecting all kinds of attacks, the system design, e.g. the fusion of countermeasures, is a very important research topic of its own. Unfortunately, it has not been explicitly studied much apart from the methods (Schwartz *et al.* 2011, Tronci *et al.* 2011, Yan *et al.* 2012) proposed within the context of the IJCB 2011 competition on counter measures to 2D facial spoofing attacks (Chakka *et al.* 2011) because the research focus has been on developing generic spoof detection schemes. In Paper VII, this very issue was addressed by investigating more closely the fusion of complementary anti-spoofing techniques for detecting various scenic attacks. In addition to the static texture analysis based approach introduced in the previous chapter, a face and background motion correlation measure based countermeasure (Anjos & Marcel 2011) was selected for the study.

The motion correlation analysis based technique is efficient for measuring excessive synchronized shaking of hand-held attacks within the scene and no movement of fixed support photo attacks. However, a drawback is that it can get confused between a motionless person while being recognized and a fixed support photo-attack (Anjos & Marcel 2011). Moreover, the method was originally proposed for detecting photo-attacks, while the assumption of decorrelated movement between face and background is true also in the case of video replay-attacks. On the other hand, the performance of LBP based countermeasures is not dependent on the spoofing attack scenario if disparities in the facial texture properties exist. More importantly, the two countermeasures exploit independent visual cues, motion and texture, thus, intuitively, they should be able to provide complementary information about the nature of the observed access attempt.

System design is an important factor when transferring anti-spoofing software into practice because the environmental conditions and possible spoofing attack scenarios are unpredictable in unsupervised real world applications. It can be assumed that the generalization ability and stability of the individual countermeasures could be improved by reducing the complexity of individual countermeasures. Thus, we also considered utilizing linear discriminant analysis (LDA) instead of the complex classifiers (multilayer perceptron (MLP) and SVM) used in the original methods to avoid over-fitting on the development dataset and possibly increasing robustness in real-world applications. Moreover, only a single holistic LBP description is computed over the facial region.

The block diagram of the proposed fusion strategy is illustrated in Figure 20. In order to combine the motion and micro-texture analysis based techniques, the video sequences are divided into overlapping windows of N frames with an overlap of N-1 frames and each observation generates a score for each countermeasure. The fusion of the two visual cues is performed at score level using logistic linear regression (LLR). The proposed anti-spoofing framework was implemented using the free signal processing and machine learning toolbox Bob (Anjos *et al.* 2012). The source code of the fusion algorithm as well as the individual countermeasures are available as add-on packages to this framework. After installation, it is possible to reproduce all experiments reported in Paper VII.

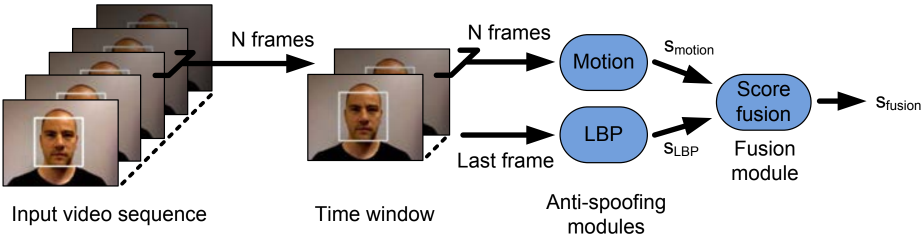

Fig 20. Block diagram of the used fusion strategy. Paper VII © IEEE.

4.2 Experiments

This section summarizes the main results and key findings of the attack-specific countermeasures proposed in Paper VI and Paper VII. For more detailed results, please refer to the original papers.

4.2.1 Display medium detection

To assess of the effectiveness of the display spoofing medium detection scheme proposed in Paper VI, a set of experiments was performed on the CASIA FASD. In the high imaging quality scenario of the dataset, the scenic cues are usually missing in the view, i.e. the upper-body of the subject and the used display medium are not visible. The reason for this is that the original 1920×1080 resolution videos have been cropped into patches of 1280×720 pixels in order to save memory and computational burden. Therefore, the resulting videos contain only the face region because the main purpose is to capture the appearance of the target faces with maximum precision. However, as seen in the previous chapter, the measurement of facial texture quality seems to provide sufficient means to reveal whether degradation due to recapturing process is observed as the imaging quality is good enough to capture the differences in fine details of surface properties between a human face and spoofing medium. On the other hand, the performance or generalization capabilities of texture based methods can be questioned at lower imaging qualities. Therefore, only the access attempts at the lowest two imaging qualities were considered for the experiments.

Table 3. EER (%) comparison on CASIA FASD (*scenarios in the official test protocol).

Scenario	Low*	Medium*	Photo	Video	Photo & video	Inter-test
DoG (Zhang *et al.* 2012)	13.3	13.3	-	-	-	-
LBP (Paper IV)	4.4	10.0	-	-	-	-
LBP-TOP (Paper IV)	3.3	3.3	-	-	-	-
Spoofing medium detector	**2.2**	**1.1**	**3.3**	**0.0**	**3.3**	**6.8**

The results of the experiments are presented in Table 3. The first evaluation tested how well the context based anti-spoofing algorithms perform under different fake face types across the two lower imaging qualities. The spoofing medium detector is able to capture the nature of both attack types well, thus confirming the very benefits of exploiting contextual cues. As shown in Figure19, the HOG features can easily describe the black screen frames of the video displays that cause strong discontinuities around the face, leading to perfect detection performance.

In order to get comparable results with previous studies, the proposed spoofing medium detector was assessed by following also two official test protocols. The performance comparison for the two quality tests, low and normal, can be seen in Table 3. The results indicate that the proposed display medium detector is able to improve the state of the art under the two official test protocols, especially at the normal imaging quality as nearly perfect detection performance is achieved. Thus, the scene information is indeed a very important visual cue for face spoof detection and should be considered as one building block when constructing anti-spoofing solutions.

In real-world applications, face anti-spoofing techniques must be working well under varying attack scenarios in different acquisition conditions. Therefore, inter-database test was also conducted in order to see whether the display medium detector is able to generalize its good performance beyond CASIA FASD. The proposed countermeasures were trained and tuned solely using the photo-attacks of the CASIA FASD and the resulting models were then tested on the NUAA PID.

The results of the inter-database experiment are shown in Table 3. The display medium detector managed the cross-database testing well as the intra-test performance decreases only from 3.3% to 6.8% in terms of EER. The main reason for the performance drop of the display medium detector is the more challenging acquisition conditions of the NUAA dataset, i.e. the target faces are more vivid and the background scene is not static (other people moving around) that cause two problems. First, the face localization data was not as accurate and stable, thus also the predicted window locations for detecting the presence of spoofing medium were noisier. Secondly, the complex and vivid background scenes introduce new artefacts inside the spoofing medium detection window. Although the results of the cross-database evaluation were promising, there is still room for improving the appearance description and segmentation of the spoofing medium in order to solve the aforementioned problems.

4.2.2 Scenic attack detection

The experiments in Paper VII were conducted on the Replay-Attack Database. The purpose of the experimental analysis is to first determine if the two countermeasures have fusion potential and then see what the actual fusion performance is. More importantly, it is also studied how the reduced complexity of the individual methods affects the performance of the anti-spoofing framework.

Table 4. Overall performance (HTER in %) for time windows. Complex classifiers means that MLP is used for motion correlation and SVM for LBP based method.

	Devel		Test	
Method	Complex	LDA	Complex	LDA
Motion	11.13	15.16	11.2	16.05
LBP	14.72	19.08	15.06	17.12
Mutual errors	**2.25**	**2.27**	**1.37**	**1.76**
LLR	**4.57**	**5.48**	**5.11**	**5.47**

Since the complementarity of different countermeasures could be determined before blindly trying fusion, the total errors for each individual anti-spoofing technique were considered and their mutual mistakes were determined. The number of samples that both countermeasures fail to detect can be seen as an upper bound for their combined spoof detection accuracy. As we can see in Table 4, the percentage of common errors is extremely low compared to the moderate accuracy of the individual methods. These observations indicate that the motion and LBP based countermeasures are indeed complementary, thus the fusion of these anti-spoofing approaches should improve robustness to scenic attacks. More importantly, only a minor increase from 1.37% to 1.76% in the number of mutual errors is observed when LDA is used on individual countermeasures to reduce complexity. Even though the performance of the individual countermeasures degrades substantially, the fusion potential is unaffected by the proposed simplification.

The overall accuracy for the fusion of motion correlation and texture based countermeasures is shown in Table 4. The results support the complementary hypotheses as significant performance enhancement is obtained when both techniques are used together. For example, the HTER of motion correlation based approach can be improved from 11.20% to 5.1% by utilizing also the LBP based face description. Moreover, the simplification of classification schemes reduces the performance of the individual methods, whereas the fusion performance remains nearly the same. This observation is also consistent with the mutual error analysis, thus suggesting that the complementarity of countermeasures is somewhat independent of the complexity of the individual classification techniques.

It is also worth mentioning that the HTER evolution of access attempt based evaluation in Paper VII depicted another benefit that supports the idea of using simpler classifications schemes. The simplified anti-spoofing framework actually works significantly better when spoofing decision is made within 100 frames. Furthermore, the HTER of combination of LDA-based classifiers drops much faster and saturates within 75 frames (three seconds), whereas it takes more than 100 frames (four seconds) when the outputs of more complex classification schemes are combined.

4.3 Discussion

There exists no generic spoof detection scheme that is able to perform robustly in all known, let alone unseen, attack scenarios. Therefore, it is important to define the use case scenarios in which it is feasible to develop well-generalizing countermeasures. In this chapter, the problem of 2D face anti-spoofing was approached by dividing the attack scenarios into two categories based on whether the boundaries of the used display medium are visible in the view or not, and investigating countermeasures to each scenario. Since it is reasonable to assume that a framework of several complementary spoof detector modules is likely to perform more robustly under unseen spoofing attacks, an open-source anti-spoofing framework was introduced and the fusion of countermeasures was explored more closely under scenic attacks.

The experimental analysis showed that it is rather straightforward to detect whether someone is presenting a fake face on a display medium to the camera in the provided view if contextual cues, e.g. the boundaries of the spoofing medium, can be exploited. The proposed HOG feature based upper-body representation could easily describe the discontinuities around the detected face, especially the black screen frames of the video screen based attacks. Moreover, the inter-test evaluation depicted that the proposed approach had promising generalization capabilities for detecting print-attacks.

Although contextual information is indeed a powerful visual cue for face spoof detection, the proposed spoofing medium detector can be circumvented by concealing the boundaries of the display medium outside the provided view either by placing fake face close to the camera or by incorporating a background scene into the face spoof. The first scenario can be tackled by using texture or image quality based anti-spoofing techniques because the recaptured face is likely to be blurry and defocused due to the limited size of the display medium and the proximity of the camera. In the case of scenic attacks, a combination of face and background motion correlation measurement and facial texture analysis showed excellent results at low imaging quality.

The fusion study depicted that the unsatisfying performance of the individual techniques can be significantly improved if the countermeasures are exploiting complementary visual cues. More importantly, the results indicate that the complementarity of anti-spoofing techniques is somewhat independent of the complexity of the individual classification techniques. Since it is very expensive to create large training and tuning datasets consisting of various fake faces in different acquisition conditions, the development datasets for practical applications are likely to be rather small. The lack of variation in the development data increases the chance of over-fitting, especially with powerful "black-box" features and complex classification schemes. Since the generalization ability of very complex classification schemes can be questioned and the gain in fusion performance on benchmark databases is very small, the use of simple and computationally efficient classifiers should be indeed considered when constructing real-world anti-spoofing solutions. Similar findings have been also reported in related studies (Chingovska *et al.* 2012, Erdogmus & Marcel 2013b).

5 Discussion and conclusions

This thesis gave an overview on the state-of-the-art approaches for software-based 2D face anti-spoofing and proposed new methods that explored the problem from generic and attack-specific viewpoints. In this chapter, the final discussion and conclusions are presented with particular focus on future research challenges.

It was shown that the static and dynamic characteristic differences between genuine faces and fake ones, such as shading, specular reflections, quality and motion patterns, can be exploited for generic face anti-spoofing using off-the-shelf feature descriptors, such as LBP. The proposed approaches have been successfully applied in two competitions on countermeasures to 2D facial spoofing attacks (Chakka *et al.* 2011, Chingovska *et al.* 2013b) and some kind of LBP based face representation was used as a component in all best-performing algorithms in the latest competition (Chingovska *et al.* 2013b). It is also worth mentioning that the texture analysis based face anti-spoofing has been widely adopted in other works as well, in 3D mask detection (Erdogmus & Marcel 2013b, Kose & Dugelay 2014), for instance.

The performance of the proposed texture based countermeasures is very encouraging when following the intra-test protocols of the publicly available databases, i.e. when the operating conditions are known. Since mobile applications are one of the most probable use cases for face biometrics, robustness to different acquisition conditions and spoofing scenarios is an important property when transferring the developed countermeasures into practice. Although the current publicly available face spoofing databases are beginning to cover a variety of spoofing attacks from high-quality photo attacks to video replay attacks, and even 3D masks, the generalization capabilities of the texture based methods are not yet clear due to the lack of variation between training and test sets, e.g. illumination, sensor quality and user demographics, let alone unknown attack scenarios.

The initial inter-database tests (de Freitas Pereira *et al.* 2013b, Wang *et al.* 2013) have suggested that the performance of texture based techniques degrades dramatically when the face models learned from one dataset are tested on another dataset. However, this kind of experiments cannot be regarded as the final word on the generalization capabilities of texture based approaches. While there is not much variation in the collected data within a dataset, the conditions between databases are completely different. Due to the limited amount of representative training and tuning data, inevitable bias will be seen in the inter-tests. The differences between conditions could be compensated by combining multiple datasets for training and tuning the models (de Freitas Pereira *et al.* 2013b, Yang *et al.* 2014) but, then again, the algorithms are no more dealing with unknown scenarios.

The research community has just begun to focus on the problem of spoofing attacks and the current publicly available databases have been a very important kick-off for finding out best practices for spoof detection. The impressive results on the existing benchmark datasets indicate that more challenging configurations are needed before the research on non-intrusive face anti-spoofing can reach the next level. In the future, more work should be carried out for designing and collecting new databases with more representative and diverse development set but still with unseen scenarios in the test set simulating the unknown attacks that will be faced in real operational conditions. Also enrollment data should be included in the spoofing databases when the joint-operation of spoof detection and recognition stages could be investigated. This is a very important research topic that has not been studied apart from the work by Chingovska *et al.* (2013a). In addition to variations in the collected data, well-defined test protocols with clear training, development and test sets are needed for unbiased comparison between various approaches across different databases.

One limitation with the existing software-based anti-spoofing approaches has been that the mainly well-known low-level feature descriptors developed within other computer vision problems have been applied for face anti-spoofing. Since they might not be optimal for explicitly describing the intrinsic differences between genuine faces and fake ones, problem-specific feature learning or design have been proposed for improving the generalization capabilities (inter-test performance) of methods utilizing the facial appearance. The initial studies using deep convolutional neural network (CNN) have resulted in the state-of-the-art intra-test performance but the inter-test results have still been unsatisfactory (Yang *et al.* 2014). However, the current publicly available datasets may not provide enough training data to exploit CNN to their full potential, thus also application-specific learning needs to be further explored when more comprehensive databases are available.

In a very recent work, Wen *et al.* (2015) argued that commonly used features, e.g. LBP, might be too person-specific or contain too much redundant information for face anti-spoofing because they are capable of capturing the facial details, i.e. differentiating individuals for face recognition purposes. Hence, they proposed to extract features that do not try to capture the facial details but the characteristic differences between genuine faces and fake ones, including characteristic reflection and quality properties, e.g. blur and color diversity. The experimental validation showed promising generalization capabilities compared to LBP based methods but only short distance spoofing attacks were considered and the features could generalize to cameras with similar quality but not to cameras with distinctively different quality. However, their argument on features describing facial details suggests that person-specific anti-spoofing models (Chingovska & Anjos 2015, Yang *et al.* 2015) might improve the generalization of texture based approaches, for instance.

On the other hand, due to the varying nature of spoofing attacks in open environments, we cannot foresee every possible attack scenario and cover them in databases because the imagination of the human mind always finds out new tricks ("golden fakes") to fool existing systems. Thus, it is reasonable to assume that no single superior technique is able to detect all known, let alone unseen, spoofing attacks. In this thesis, the problem of face spoofing was approached by dividing the attack scenarios into two complementary attack-specific categories based on whether the boundaries of the used display medium are visible in the provided view. It was shown that the boundaries of the spoofing medium, such as video screen frame, are rather straightforward to detect, whereas a combination of face and background motion correlation and facial texture analysis was effective when the used display medium is concealed from the input camera. Furthermore, the experiments with the developed fusion framework demonstrated the benefits of combining complementary countermeasures using independent visual cues. Thus, it is indeed important to define clearly the attack and use-case scenarios the biometric system is dealing with in order to understand the problem better and to overcome the limitations of generic anti-spoofing algorithms.

Even though the proposed two-part attack-specific categorization is indeed very useful for face spoof detection, there is still room for future work in order to solve the two attack scenarios more robustly and to improve the system-level performance. For instance, the main drawback of the spoofing medium detector is that the accuracy of the used face detector sets the upper limit to the performance of the spoofing medium detector. More sophisticated segmentation and scene motion analysis could be applied to mitigate the face detection noise and fixed bounding box limitation. New anti-spoofing modules could be included in the fusion framework by developing novel countermeasures and by implementing techniques proposed in related works. Also an ensemble of separate detectors for face prints, video displays and masks should be considered (Wen *et al.* 2015). Furthermore, training and updating an anti-spoofing framework is an open research topic that needs to be addressed in the future. Intuitively, an anti-spoofing solution consisting of several complementary spoof detectors probably performs more robustly, also under unseen attack scenarios. Thus, it would be interesting to see how the use of complementary countermeasures with different fusion strategies affects the generalization performance.

In its present state, the fusion framework provides only score level fusion techniques but due to the flexible architecture, it is simple to add new fusion rules or fusion strategies to couple several complementary countermeasures at multiple stages. A good example of this was the algorithm used in the 2nd Competition on Counter Measures to 2D Face Spoofing Attacks (Chingovska *et al.* 2013b) where LBP and GLCM based texture representations were fused before combining the outputs of LBP-TOP and motion correlation based countermeasures. In the same competition, however, the best fusion performance was achieved when the countermeasures were combined at feature level. Conversely, the results by Kose & Dugelay (2014) suggested that fusion at score level performed better than feature level fusion. Still, also feature-level fusion should be added into the developed open-source framework and investigated more closely. The main drawback with feature-level fusion is, however, that the models need to be retrained every time when new feature representation is included to the anti-spoofing framework.

Besides focusing on software-based countermeasures to face spoofing attacks, multi-modal and hardware-based solutions should not be forgotten in future research, either. While nowadays every mobile phone and laptop are equipped with a microphone and camera, other sensors, such as 3D and NIR imaging, are emerging in mobile devices which opens up new possibilities for face anti-spoofing. Furthermore, already the existing (mobile) devices provide means for novel spoof detection schemes. For instance, Smith *et al.* (2015) proposed to analyse dynamic reflections from the observed person's face caused by varying illumination due to a sequence of images (digital watermarks) presented on the used display device, e.g. a tablet or a laptop, for replay-attack detection. It would be interesting to see whether similar digital watermarks could be coupled for performing both replay-attack and spoof detection simultaneously.

While the methods proposed in this thesis are not able to solve the problem, many potential visual cues and approaches for face anti-spoofing were nevertheless explored. It is also worth mentioning that the security of the 2D face authentication systems can already be improved by utilizing existing countermeasures. A good example of this was the eye blink detection based liveness check that was introduced to the Face Unlock feature on Android phones. Also the proposed spoofing medium detector would be practical in detecting crude attack attempts in which the border of the used display media is visible in the view. Although these kinds of security updates are not capable of dealing with all sorts of attack types, they still manage to boost the robustness of the biometric system and pose new challenges to the attackers.

6 Summary

This thesis introduced the work conducted in face anti-spoofing with particular focus on software-based techniques. First, the problem of spoofing was introduced and the state of the art was reviewed. Then, different approaches for software-based face spoof detection were explored and new methods were proposed. The main contributions of this thesis are the texture based countermeasures and the proposed attack-specific spoof detection schemes using contextual information and fusion of complementary techniques. The proposed algorithms operate either on a single image or a short video sequence, thus practically real-time response can be achieved.

The key idea of the developed facial texture representations is to capture the inherent disparities in quality, light reflection, shading and motion between genuine faces and fake ones. The proposed methods were evaluated on the latest benchmark datasets and promising spoof detection performance was obtained under various types of print and video replay attacks. The texture based approaches have been widely adopted by other researchers. For instance, some kind of facial texture description was utilized as a component in all best-performing algorithms in the latest competition on countermeasures to 2D facial spoofing attacks. Even though the texture based face representation was originally developed for detecting print attacks, it has later shown to be useful in 3D mask attack detection as well.

It is reasonable to assume that there exists no single superior anti-spoofing technique because every countermeasure most likely has its own vulnerability ("a golden fake") that can be exploited by an attacker. In order to find out well-generalizing spoof detection schemes, the problem of anti-spoofing was broken into two complementary attack-specific sub-problems based on whether the used spoofing medium can be detected in the camera view. The experimental analysis showed that it is rather straightforward to detect the boundaries of the display medium, e.g. a video screen frame, by describing the discontinuities in the gradient structures around the detected face, whereas a combination of face and background motion correlation and facial texture analysis showed to effective for spoof detection when the display medium is concealed outside the provided view. Furthermore, an open-source anti-spoofing fusion framework was introduced and it was utilized to study the effects on combining different anti-spoofing modules.

Also the current state of face anti-spoofing was discussed. The existing software-based countermeasures in the literature have shown very encouraging results on individual databases but may still lack generalization to varying nature of spoofing attacks that are encountered in real-world applications. While the current publicly available datasets are still useful for developing new countermeasures, more representative and challenging databases are yet needed before the software-based face anti-spoofing can reach the next level. It is also worth mentioning that multi-modal and hardware-based solutions should not be forgotten in future research, either.

References

Ahonen T, Hadid A & Pietikäinen M (2006) Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12): 2037–2041.

Anjos A, Chakka MM & Marcel S (2013) Motion-based counter-measures to photo attacks in face recognition. IET Biometrics 3(3): 147–158.

Anjos A, El Shafey L, Wallace R, Günther M, McCool C & Marcel S (2012) Bob: a free signal processing and machine learning toolbox for researchers. Proc. Proceedings of the ACM Multimedia Conference.

Anjos A & Marcel S (2011) Counter-measures to photo attacks in face recognition: a public database and a baseline. Proc. Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB).

Bai J, Ng TT, Gao X & Shi YQ (2010) Is physics-based liveness detection truly possible with a single image? Proc. IEEE International Symposium on Circuits and Systems (ISCAS), 3425–3428.

Bao W, Li H, Li N & Jiang W (2009) A liveness detection method for face recognition based on optical flow field. Proc. 2009 International Conference on Image Analysis and Signal Processing, IEEE, 233–236.

Bharadwaj S, Dhamecha TI, Vatsa M & Richa S (2013) Computationally efficient face spoofing detection with motion magnification. Proc. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Biometrics.

Biggio B, Akhtar Z, Fumera G, Marcialis GL & Roli F (2012) Security evaluation of biometric authentication systems under real spoofing attacks. IET Biometrics 1(1): 11–24.

Bredin H & Chollet G (2008) Making talking-face authentication robust to deliberate imposture. Proc. ICASSP 2008, IEEE International Conference on Acoustics, Speech, and Signal Processing.

Buhan I & Hartel P (2005) The state of the art in abuse of biometrics. Technical Report CTIT-05-41, Center for Telematics and Information Technology, University of Twente.

Chakka MM, Anjos A, Marcel S, Tronci R, Muntoni D, Fadda G, Pili M, Sirena N, Murgia G, Ristori M, Roli F, Yan J, Yi D, Lei Z, Zhang Z, Li SZ, Schwartz WR, Rocha A, Pedrini H, Lorenzo-Navarro J, Castrillón-Santana M, Määttä J, Hadid A & Pietikäinen M (2011) Competition on counter measures to 2-d facial spoofing attacks. Proc. Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB).

Chaudhry R, Ravich A, Hager G & Vidal R (2009) Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1932–1939.

Chetty G & Wagner M (2004) Liveness verification in audio-video speaker authentication. Proc. in Proc. 10th Australian International Conference on Speech Science and Technology, 358–363.

Chingovska I & Anjos A (2015) On the use of client identity information for face anti-spoofing. IEEE Transactions on Information Forensics and Security .

Chingovska I, Anjos A & Marcel S (2012) On the effectiveness of local binary patterns in face anti-spoofing. Proc. IEEE International Conference of the Biometrics Special Interest Group.

Chingovska I, Anjos A & Marcel S (2013a) Anti-spoofing in action: joint operation with a verification system. Proc. Proceedings of IEEE Conference on Computer Vision and Pattern

Recognition, Workshop on Biometrics.

Chingovska I, Yang J, Lei Z, Yi D, Li SZ, Kähm O, Glaser C, Damer N, Kuijper A, Nouak A, Komulainen J, Pereira T, Gupta S, Khandelwal S, Bansal S, Rai A, Krishna T, Goyal D, Waris MA, Zhang H, Ahmad I, Kiranyaz S, Gabbouj M, Tronci R, Pili M, Sirena N, Roli F, Galbally J, Fierrez J, Pinto A, Pedrini H, Schwartz WS, Rocha A, Anjos A & Marcel1 S (2013b) The 2nd competition on counter measures to 2d face spoofing attacks. Proc. IAPR International Conference on Biometrics, ICB.

Dalal N & Triggs B (2005) Histograms of oriented gradients for human detection. Proc. International Conference on Computer Vision & Pattern Recognition, 2: 886–893.

Dantcheva A, Chen C & Ross A (2012) Can facial cosmetics affect the matching accuracy of face recognition systems? Proc. IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 391–398.

de Freitas Pereira T, Anjos A, De Martino J & Marcel S (2013a) Lbp-top based countermeasure against face spoofing attacks. In: Computer Vision - ACCV 2012 Workshops, volume 7728 of *Lecture Notes in Computer Science*, 121–132.

de Freitas Pereira T, Anjos A, De Martino JM & Marcel S (2013b) Can face anti-spoofing countermeasures work in a real world scenario? Proc. International Conference on Biometrics.

De Marsico M, Nappi M, Riccio D & Dugelay JL (2012) Moving face spoofing detection via 3D projective invariants. Proc. ICB 2012, 5th IAPR International Conference on Biometrics (ICB).

Duc NM & Minh BQ (2009) Your face is not your password. Proc. Black Hat Conference.

Eid AH, Ahmed MN, Cooper BE & Rippetoe EE (2011) Characterization of electrophotographic print artifacts: Banding, jitter, and ghosting. IEEE Transactions on Image Processing 20: 1313–1326.

Erdogmus N & Marcel S (2013a) Spoofing attacks to 2d face recognition systems with 3d masks. Proc. IEEE International Conference of the Biometrics Special Interest Group.

Erdogmus N & Marcel S (2013b) Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. Proc. Biometrics: Theory, Applications and Systems Conference (BTAS).

Farid H & Lyu S (2003) Higher-order wavelet statistics and their application to digital forensics. Proc. in IEEE Workshop on Statistical Analysis in Computer Vision.

Frischholz RW & Dieckmann U (2000) Bioid: A multimodal biometric identification system. Computer 33(2): 64–68.

Frischholz RW & Werner A (2003) Avoiding replay-attacks in a face recognition system using head-pose estimation. Proc. Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures.

Galbally J & Marcel S (2014) Face anti-spoofing based on general image quality assessment. Proc. Proc. IAPR/IEEE Int. Conf. on Pattern Recognition, ICPR, 1173–1178.

Galbally J, Marcel S & Fiérrez J (2014) Biometric antispoofing methods: A survey in face recognition. IEEE Access 2: 1530–1552.

Gao X, Ng TT, Qiu B & Chang SF (2010) Single-view recaptured image detection based on physics-based features. Proc. IEEE International Conference on Multimedia & Expo (ICME), 1469–1474.

Hadid A, Pietikäinen M & Ahonen T (2004) A discriminative feature space for detecting and recognizing faces. Proc. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 797–804.

Jain A, Ross A & Nandakumar K (2011) Security of biometric systems. In: Introduction to Biometrics, 259–306. Springer-Verlag.

Jain AK, Flynn P & Ross AA (eds) (2008) Handbook of Biometrics. Springer-Verlag.

Karam W, Bredin H, Greige H, Chollet G & Mokbel C (2009) Talking-face identity verification, audiovisual forgery, and robustness issues. EURASIP J. Adv. Signal Process 2009: 4:1–15.

Kollreider K, Fronthaler H & Bigun J (2008) Verifying liveness by multiple experts in face biometrics. Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops : CVPR 2008, 1200–1205.

Kollreider K, Fronthaler H & Bigun J (2009) Non-intrusive liveness detection by face images. Image and Vision Computing 27: 233–244.

Kollreider K, Fronthaler H, Faraj MI & Bigun J (2007) Real-time face detection and motion analysis with application in liveness assessment. Trans. Info. For. Sec. 2(3): 548–558.

Kose N & Dugelay JL (2014) Mask spoofing in face recognition and countermeasures. Image and Vision Computing 32(10): 779–789.

Li J, Wang Y, Tan T & Jain AK (2004) Live face detection based on the analysis of fourier spectra. Proc. In Biometric Technology for Human Identification, 296–303.

Li S & Jain AK (eds) (2011) Handbook of face recognition. Springer-Verlag.

Li Y, Xu K, Yan Q, Li Y & Deng RH (2014) Understanding osn-based facial disclosure against face authentication systems. Proc. Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security, ACM, 413–424.

Liu C, Yuen J & Torralba A (2011) Sift flow: Dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. 33(5): 978–994.

Manjunath BS & Ma WY (1996) Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell. 18(8): 837–842.

Marcel S, Nixon M & Li S (eds) (2014) Handbook of Biometric Anti-Spoofing. Springer-Verlag.

Ng ES & Chia AYS (2012) Face verification using temporal affective cues. Proc. International Conference on Pattern Recognition (ICPR), 1249–1252.

Nixon K, Aimale V & Rowe R (2008) Spoof detection schemes. In: Handbook of Biometrics, 403–423. Springer-Verlag.

Ojala T, Pietikäinen M & Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7): 971–987.

Pan G, Sun L, Wu Z & Wang Y (2011) Monocular camera-based face liveness detection by combining eyeblink and scene context. Telecommunication Systems 47(3-4): 215–225.

Pan G, Wu Z & Sun L (2008) Liveness detection for face recognition. In: Recent Advances in Face Recognition, 109–124. In-Teh.

Pavlidis I & Symosek P (2000) The imaging issue in an automatic face/disguise detection system. Proc. Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (CVBVS), 15–24.

Pietikäinen M, Hadid A, Zhao G & Ahonen T (2011) Computer Vision Using Local Binary Patterns. Springer.

Pinto AdS, Pedrini H, Schwartz WR & Rocha A (2012) Video-based face spoofing detection through visual rhythm analysis. Proc. Conference on Graphics, Patterns and Images (Sibgrapi).

Ratha NK, Connell JH & Bolle RM (2001) An analysis of minutiae matching strength. Proc. Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), Springer-Verlag, 223–228.

Roberts C (2007) Biometric attack vectors and defences. Computers & Security 26(1): 14–25.

Rodrigues RN, Ling LL & Govindaraju V (2009) Robustness of multimodal biometric fusion methods against spoof attacks. Journal of Visual Language and Computing 20(3): 169–179.

Schwartz WR, Rocha A & Pedrini H (2011) Face Spoofing Detection through Partial Least Squares and Low-Level Descriptors. Proc. Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB).

Singh R, Vatsa M, Bhatt HS, Bharadwaj S, Noore A & Nooreyezdan SS (2010) Plastic surgery: a new dimension to face recognition. IEEE Transactions on Information Forensics and Security 5(3): 441–448.

Smith DF, Wiliem A & Lovell BC (2015) Face recognition on consumer devices: Reflections on replay attacks. IEEE Transactions on Information Forensics and Security 10(4): 236–245.

Taigman Y, Yang M, Ranzato M & Wolf L (2014) Deepface: Closing the gap to human-level performance in face verification. Proc. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Tan X, Li Y, Liu J & Jiang L (2010) Face liveness detection from a single image with sparse low rank bilinear discriminative model. Proc. Proceedings of the 11th European conference on Computer vision: Part VI, 504–517.

Tronci R, Muntoni D, Fadda G, Pili M, Sirena N, Murgia G, Ristori M & Roli F (2011) Fusion of multiple clues for photo-attack detection in face recognition systems. Proc. Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB).

Verdet F & Hennebert J (2008) Impostures of Talking Face Systems Using Automatic Face Animation. Proc. IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS).

Wang T, Yang J, Lei Z, Liao S & Li SZ (2013) Face liveness detection using 3d structure recovered from a single camera. Proc. IAPR International Conference on Biometrics, ICB.

Wen D, Han H & Jain A (2015) Face spoof detection with image distortion analysis. Transactions on Information Forensics and Security 10(4): 746–761.

Wu HY, Rubinstein M, Shih E, Guttag J, Durand F & Freeman WT (2012) Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics. (SIGGRAPH) 31(4).

Yan J, Zhang Z, Lei Z, Yi D & Li SZ (2012) Face liveness detection by exploring multiple scenic clues. Proc. 12th International Conference on Control, Automation, Robotics and Vision, (ICARCV), 188–193.

Yang J, Lei Z & Li SZ (2014) Learn convolutional neural network for face anti-spoofing. CoRR abs/1408.5601.

Yang J, Lei Z, Liao S & Li SZ (2013) Face liveness detection with component dependent descriptor. Proc. IAPR International Conference on Biometrics, ICB.

Yang J, Yi D & Li SZ (2015) Person-specific face anti-spoofing with subject domain adaptation. IEEE Transactions on Information Forensics and Security .

Zhang Z, Yan J, Liu S, Lei Z, Yi D & Li SZ (2012) A face antispoofing database with diverse attacks. Proc. 5th IAPR International Conference on Biometrics (ICB), 26–31.

Zhang Z, Yi D, Lei Z & Li SZ (2011) Face liveness detection by learning multispectral reflectance distributions. Proc. International Conference on Face and Gesture, 436–441.

Zhao G & Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6): 915–928.

Original publications

I Anjos A, Komulainen J, Marcel S, Hadid A & Pietikäinen M (2014) Face anti-spoofing: Visual approach. In: S. Marcel, M.S. Nixon & S.Z. Li, (eds) Handbook of Biometric Anti-Spoofing, Springer Verlag, 65-82.

II Määttä J, Hadid A & Pietikäinen M (2011) Face spoofing detection from single images using micro-texture analysis. Proc. International Joint Conference on Biometrics (IJCB).

III Määttä J, Hadid A & Pietikäinen M (2012) Face spoofing detection from single images using texture and local shape analysis. IET Biometrics, 1(1):3-10.

IV Komulainen J, Hadid A & Pietikäinen M (2013) Face spoofing detection using dynamic texture. In: ACCV 2012 Workshops, Part I (LBP 2012), Lecture Notes in Computer Science, 7728:146-157.

V de Freitas Pereira T, Komulainen J, Anjos A, De Martino JM, Hadid A, Pietikäinen M & Marcel S (2014) Face liveness detection using dynamic texture. EURASIP Journal on Image and Video Processing, 2014:2.

VI Komulainen J, Hadid A & Pietikäinen M (2013) Context based face anti-spoofing. Proc. the IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS).

VII Komulainen J, Anjos A, Hadid A, Marcel S & Pietikäinen M (2013) Complementary countermeasures for detecting scenic face spoofing attacks. Proc. IAPR International Conference on Biometrics (ICB).

Reprinted with permission from Springer (I, IV, V), IEEE (II, VI, VII) and IET (III).

Original publications are not included in the electronic version of the dissertation.

521. Putaala, Jussi (2015) Reliability and prognostic monitoring methods of electronics interconnections in advanced SMD applications

522. Pirilä, Minna (2015) Adsorption and photocatalysis in water treatment : active, abundant and inexpensive materials and methods

523. Alves, Hirley (2015) On the performance analysis of full-duplex networks

524. Siirtola, Pekka (2015) Recognizing human activities based on wearable inertial measurements : methods and applications

525. Lu, Pen-Shun (2015) Decoding and lossy forwarding based multiple access relaying

526. Suopajärvi, Terhi (2015) Functionalized nanocelluloses in wastewater treatment applications

527. Pekuri, Aki (2015) The role of business models in construction business management

528. Mantere, Matti (2015) Network security monitoring and anomaly detection in industrial control system networks

529. Piri, Esa (2015) Improving heterogeneous wireless networking with cross-layer information services

530. Leppänen, Kimmo (2015) Sample preparation method and synchronized thermography to characterize uniformity of conductive thin films

531. Pouke, Matti (2015) Augmented virtuality : transforming real human activity into virtual environments

532. Leinonen, Mikko (2015) Finite element method and equivalent circuit based design of piezoelectric actuators and energy harvester dynamics

533. Leppäjärvi, Tiina (2015) Pervaporation of alcohol/water mixtures using ultra-thin zeolite membranes : membrane performance and modeling

534. Lin, Jhih-Fong (2015) Multi-dimensional carbonaceous composites for electrode applications

535. Goncalves, Jorge (2015) Situated crowdsourcing : feasibility, performance and behaviours

536. Herrera Castro, Daniel (2015) From images to point clouds : practical considerations for three-dimensional computer vision

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

UNIVERSITY of OULU
OULUN YLIOPISTO

ISBN 978-952-62-0872-5 (Paperback)
ISBN 978-952-62-0873-2 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)