WARSAW UNIVERSITY OF TECHNOLOGY

INFORMATION AND COMMUNICATION TECHNOLOGY

ENGINEERING AND TECHNOLOGY

Ph.D. Thesis

Countermeasure algorithms against subterfuge in mobile biometric systems

Jalil Nourmohammadi Khiarak, M.Sc.

Supervisor:

Professor **Andrzej Pacut**, Ph.D., D.Sc. (deceased)

Supervisor:

Professor **Włodzimierz Kasprzak**, Ph.D., D.Sc.

WARSAW 2025

STATEMENT

The Author acknowledges contributions from co-author of the papers relating to the topic of this doctoral dissertation: Professor Andrzej Pacut. Precise descriptions of the contributions from each of the co-authors are included in the publication list in Appendix B.

AKNOWLEDGMENTS

To all those who have been deprived of education in their MOTHER TONGUE.

Abstract

Biometric authentication, facilitated by advanced mobile phones and technologies such as ear recognition, has had a profound impact on modern societies. Ear recognition, enabled by artificial intelligence, the Internet of Things (IoT), and computer vision, has emerged as a prevalent biometric authentication technique in mobile devices. With the widespread use of biometric applications in smartphones, laptops, airport services, and banking services, ensuring their security becomes imperative. The abundance of sensitive information stored on these devices underscores the need to safeguard data and identity. Consequently, it is critical to study methods for detecting presentation attacks and liveness as countermeasures against spoofing attacks. New techniques must address evolving acquisition scenarios and increased noise in collected biometric data.

This thesis focuses on investigating countermeasure methods against presentation attacks in ear recognition, specifically ear-touch and ear photo attacks, through the development of a PAD system and its analysis. Two key biometric challenges are addressed: ear-touch and ear photo PAD.

Ear biometrics, as a sub-branch of computer vision, have gained significant attention for their reliable and effective person recognition capabilities. However, ear recognition systems (ear photo and ear touch) are vulnerable to presentation or spoofing attacks, where attackers attempt to conceal their true identity and deceive the biometric system. To protect these systems from attacks, recent advancements in Ear PAD or ear anti-spoofing methods have been developed. This thesis explores and compares various novel approaches for detecting ear presentation attacks, utilizing both traditional features and deep learning techniques. Temporal data, capturing dynamic properties of presentation attacks, is examined through a range of characteristics. The application of deep neural architectures and handcrafted features, along with potential extensions in PAD, is investigated. Additionally, deep learning methods employing different pre-trained neural architectures are explored to enhance detection performance. The visualization of internal representations of networks is studied to improve method performance.

The proposed techniques are evaluated and analyzed using publicly available databases and data obtained during the course of this doctoral research. Experimental results confirm the effectiveness of temporal information and the advantages of deep learning

methods in PAD. The explanations employed for enhancing reliability and designing automatic PAD neural architectures demonstrate significant potential for future advancements.

This research contributes to the field by addressing the security challenges of ear recognition systems against presentation attacks. The findings underscore the importance of temporal information and deep learning techniques in achieving robust presentation attack detection. The study also highlights the prospects of explicable deep learning techniques and the automation of PAD system design for future developments.

Keywords: biometrics, ear-touch, ear recognition, presentation attack detection

Streszczenie

Biometryczna autentykacja, wspierana przez zaawansowane telefony komórkowe oraz technologie takie jak rozpoznawanie ucha, wywarła znaczący wpływ na współczesne społeczeństwa. Rozpoznawanie ucha, umożliwione dzięki sztucznej inteligencji, Internetowi Rzeczy (IoT) oraz wizji komputerowej, stało się powszechną techniką biometrycznej autentykacji w urządzeniach mobilnych. Wraz z szerokim wykorzystaniem aplikacji biometrycznych w smartfonach, laptopach, usługach lotniskowych i bankowych, zapewnienie ich bezpieczeństwa staje się kluczowe. Obfitość wrażliwych informacji przechowywanych na tych urządzeniach podkreśla potrzebę ochrony danych i tożsamości. W związku z tym konieczne jest badanie metod wykrywania ataków prezentacyjnych oraz technik weryfikacji żywotności jako środków przeciwdziałania atakom fałszywym. Nowe techniki muszą uwzględniać zmieniające się scenariusze akwizycji oraz rosnący poziom zakłóceń w zebranych danych biometrycznych.

Niniejsza praca doktorska koncentruje się na badaniu metod przeciwdziałania atakom prezentacyjnym w rozpoznawaniu ucha, w szczególności w kontekście ataków bazujących na dotyku ucha oraz zdjęciach ucha, poprzez opracowanie systemu PAD i jego analizę. Poruszono dwa kluczowe wyzwania biometryczne: PAD dla dotyku ucha oraz PAD dla zdjęć ucha.

Biometria ucha, jako poddziedzina wizji komputerowej, zyskała dużą uwagę ze względu na swoje niezawodne i skuteczne możliwości rozpoznawania osób. Jednak systemy rozpoznawania ucha (zarówno zdjęć, jak i dotyku) są podatne na ataki prezentacyjne lub fałszywe, w których atakujący próbują ukryć swoją prawdziwą tożsamość i oszukać system biometryczny. Aby chronić te systemy przed atakami, rozwijane są nowe metody PAD, czyli przeciwdziałania oszustwom w rozpoznawaniu ucha. W tej pracy zbadano i porównano różne nowatorskie podejścia do wykrywania ataków prezentacyjnych na ucho, wykorzystując zarówno tradycyjne cechy, jak i techniki głębokiego uczenia. Dane czasowe, odzwierciedlające dynamiczne właściwości ataków prezentacyjnych, zostały przeanalizowane pod kątem różnych charakterystyk. Zbadano zastosowanie głębokich architektur neuronowych oraz cech ręcznie definiowanych, a także potencjalne rozszerzenia w zakresie PAD. Ponadto zbadano metody głębokiego uczenia wykorzystujące różne wstępnie wytrenowane

architektury sieci neuronowych w celu poprawy wydajności wykrywania. Analizowano także wizualizację wewnętrznych reprezentacji sieci w celu poprawy skuteczności metod.

Proponowane techniki zostały ocenione i przeanalizowane przy użyciu publicznie dostępnych baz danych oraz danych uzyskanych w trakcie badań doktorskich. Wyniki eksperymentalne potwierdzają skuteczność informacji czasowej oraz przewagę metod głębokiego uczenia w PAD. Wyjaśnienia wykorzystane do zwiększenia niezawodności oraz projektowania automatycznych architektur PAD wykazują znaczący potencjał w przyszłych badaniach.

Praca ta wnosi istotny wkład w dziedzinę, rozwiązując wyzwania bezpieczeństwa systemów rozpoznawania ucha wobec ataków prezentacyjnych. Uzyskane wyniki podkreślają znaczenie informacji czasowej oraz technik głębokiego uczenia w osiąganiu skutecznego wykrywania ataków prezentacyjnych. Badanie wskazuje także perspektywy wyjaśnialnych technik głębokiego uczenia oraz automatyzacji projektowania systemów PAD w przyszłych badaniach.

Słowa kluczowe: biometria, dotyk uszu, rozpoznawanie uszu, wykrywanie ataków prezentacji

اؤزت

بیومتریک کیملیک دؤغرولاماسی، یاپای ذکاء، اشیالارین اینترنتی (IOT) و کامپیوتر گؤرمه تکنولوژیلاری ایله تأمین ائدیلن قؤلاق تانیما تکنولوژوسو واسیطه‌سیله معاصیر جمعیتلره اهمیتلی ائتکی گؤسترمیشدیر. قؤلاق تانیما، موبایل جیهازلاردا گئنیش تطبیق اؤلونان بیومتریک کیملیک دؤغرولاماسی اؤصوللاریندان بیری کیمی اؤرتایا چیخمیشدیر. بیومتریک اؤلانیشلاری اسمارت فونلاردا، نوتبوک لاردا، هاوالیمانی خیدمتلرینده و بانکچیلیق خیدمتلرینده گئنیش یایقینلاشماسی ایله بیرلیکته بو سیستملرین گؤونلیکلری‌نین ساغلانماسی واجیبدیر. بو جیهازلاردا ساخلانیلان حساس معلوماتلارین بؤللوغو وئریلرین و کیملیگین قؤرونماسی‌نین واجیبلیگینی وؤرغولاییر. بؤنا گؤرا دا سونوم سالدیری‌نین آلقیلانماسی و جانلیلیق یؤخلانیشلاری‌نین اؤصوللارینی اؤیرنمک، خؤصوصاً ده قوندارما هؤجوملارینا قارشی عکس تدبیرلر کیمی تحقیق ائتمک واجیبدیر. یئنی تکنیکلر گلیشن ائدینیم سناریولارینی و تؤپلانان بیومتریک وئریلردکی آرتان گؤرولتویو اله آلمالیدیر.

بو تئز قؤلاق تانیما سیستملرینده سونوم سالدیرینا قارشی عکس تدبیرلرین تحقیقینه یؤنلیب. خصوصیله قؤلاق تؤخونوشو و قؤلاق شکیلی هؤجوملارینا قارشی PAD (Presentation Attack Detection) سیستمی‌نین حاضیرلانماسی و تحلیلی آپاریلمیشدیر. ایکی اساس بیومتریک پروبلئم حل اؤلونور: قؤلاق تؤخونوشو و قؤلاق شکیلی سؤنوم سالدیری‌سی آلقیلاما.

کامپیوتر گؤرمه‌سی‌نین بیر آلت ساحه‌سی کیمی قؤلاق بیومتریکی اینسانلاری دقیق و ائتکیلی شکیلده تانیما قابیلیتینه گؤره دیقت مرکزینده‌دیر. بؤنونلا بئله، قؤلاق تانیما سیستملری (قؤلاق شکیلی و قؤلاق تؤخونوشو) سؤنوم و یا قوندارما هؤجوملارینا قارشی حساسدیر، بؤرادا سالدیری ائدنلر اؤز گرچک کیملیگلرینی گیزلتمه‌یه و بیومتریک سیستمی آلداتماغا چالیشیر. بو سیستملری سالدیریجیلاردان قؤروماق اؤچون قؤلاق PAD و یا قؤلاق آلادماقا قارشی قوراق مئتودلاریندا سؤن اینکیشافلار تطبیق ائدیلمیشدیر. بو تئز سؤنوم سالدیری‌سی آلقیلاما اؤچون مؤختلیف یئنی یاناشمالاری آراشدیریر و مقاییسه ائدیر، هم اننوی خؤصوصیتلردن، هم ده درین اؤیرنمه تکنولوژیلاریندان ایستیفاده ائدیر. سؤنوم سالدیریلاری‌نین دینامیک خؤصوصیتلرینی عکس ائتدیرن مؤقتی وئریلر مؤختلیف کاراکترلر اساسیندا تحقیق اؤلونور. درین اؤیرنمه معماریلاری‌نین و ال ایله حاضیرلانمیش خؤصوصیتلرین تطبیقی ایله یاناشی PAD ساحه‌سینده‌کی گئنیشلنمه پرسپئکتیولری ده آراشدیریلمیشدیر. مختلف اؤلچه‌دن اؤیردیلمیش نئیرون شبکه معماریلاری‌نیدان ایستیفاده ائده‌رک درین اؤیرنمه اؤصوللاری آشکارلاما پرفورمانسینی آرتیرماق اؤچون اؤیره‌نیلمیشدیر. شبکه‌لرین داخیلی تمثیلاتلاری‌نین ویزوواللاشدیریلماسی مئتدلارین پرفورمانسینی آرتیرماق اؤچون تحقیق ائدیلمیشدیر.

تکلیف اؤلونان مئتودلار هم ایجتیمایته آچیق معلومات بازالاری، هم ده بو دؤکتورلوق تحقیقاتی چرچیوه‌سینده الده ائدیلمیش معلوماتلار اساسیندا قیمتلندیریلمیش و آنالیز ائدیلمیشدیر. تجربه اولونموش نتیجه‌لر مؤقتی وئریلرین ائتکیلیگینی و PAD ساحه‌سینده درین اؤیرنمه مئتودلاری‌نین

اۆستونلوکلرینی تصدیق ائدیر. آوتوماتیک PAD نئرون معماریلاری‌نین دیزاینینی و اعتیبارلیلیغی آرتیرماق اۆچون ایستیفاده اولونان یاناشمالار گله‌جک اینکیشافلار اۆچون بۆیوک پتانسیل گوستریر.

بو تحقیقات قۆلاق تانیما سیستملری‌نین سۆنوم سالدیریلارینا قارشی تهلوکه‌سیزلیک پروبلئملرینی حل ائتمکله ساحه‌یه تۆحفه وئریر. تاپینتیلار، سۆنوم سالدیری‌سی آلقیلاماسیندا موقتی وئریلرین و درین اۆیرنمه اۆصوللاری‌نین اهمیتینی وۆرغولاییر. تحقیقات، بیرده ایضاح ائدیله بیلن درین اۆیرنمه تکنولوژیلاری‌نین و PAD سیستملری‌نین آوتوماتلاشدیریلمیش دیزاینی‌نین گله‌جک اینکیشاف پرسپکتیولرینی گۆستریر.

آچار سۆزجوکلر: بیومتریک، قۆلاق تۆخونوشو، قۆلاق تانیما، سۆنوم سالدیری‌سی آلقیلاما

Contents of Thesis

List of Abbreviations

PAD Presentation Attack Detection

APCER Attack Presentation Classification Error Rate

IAPMR Impostor Attack Presentation Match Rate

NPCER Normal Presentation Classification Error Rate

BPCER Bona-Fide Presentation Classification Error Rate

IQA Image Quality Assessment

T-SNE t-distributed stochastic neighbor embedding

KLT Kanade-Lucas-Tomasi

EER Equal Error Rate

FTE Failure To Enroll

FMR False Match Rate

FNMR False Non-Match Rate

LFM Light Field base Methods

AMI Mathematical Analysis of Images

UBEAR University of Beira EAR

ReLU Rectified Linear Unit

XOR Exclusive OR

HD Hamming Distance

VIS VISible light

NIR Near InfraRed

SDK Software Development Kit

CDF Cumulative Distribution Function

ROC Receiver Operating Characteristic

AUC Area Under Curve

IoU Intersection over Union

EVM Eulerian Video Magnification

NN Neural Network

CNN Convolutional Neural Network

DCNN Deep Convolutional Neural Network

FCEDN Fully Convolutional Encoder-Decoder Network

PCA Principal component analysis

LDA Linear Discriminant Analysis

HOG Histogram of Oriented Gradients

FFT Fourier Transform

LBP Local Binary Patterns

1. Introduction

The use of biometric technologies has become increasingly common in enhancing security and user convenience across various platforms, particularly in mobile devices. Among the various of biometric modalities, ear photo and ear-touch have emerged as promising identifiers due to their unique and consistent characteristics. This thesis explores the potential of these ear-based biometrics, addressing not only their implementation but also the challenges associated with presentation attack detection (PAD) in mobile contexts. By investigating these aspects, the research aims to advance the field of ear PAD and contribute to the development of more robust and secure biometric systems. This introductory chapter provides an overview of the key topics covered, including the uniqueness of ear photo and ear-touch, the critical importance of PAD, the new challenges faced in ear PAD research, and outlines the thesis statements, scope, and structure of the subsequent chapters.

1.1. Ear photo and ear-touch, as a biometric characteristic

Ear-based biometrics, including ear photo and ear-touch, are increasingly recognized for their unique characteristics, independent of the broader facial recognition field [1, 2]. These specific modalities, which capture the distinct contours and textures of the ear, are unique to each individual and have found practical applications in mobile devices. The use of ear photo and ear-touch as biometric identifiers offers a reliable and convenient solution for authentication, leveraging the ear's unique anatomical features [3-8].

Ear photo recognition has been used in large-scale applications due to its uniqueness and also does not depend on the cooperativeness of subjects. Examples of ear photo recognition are the Helix from Descartes Biometrics [9]. According to the most recent Unconstrained Ear Recognition Challenge 2019 organizer’s report, which evaluates the current state-of-the-art on ear biometrics, the best performing approach is 0.966 based on the area under curve (AUC) values [10].

Recent studies even experimentally validated existing assumptions that certain features of the ear are distinct for identical twins [11]. This fact has significant implications for security-related applications and puts ear photos matching for advanced biometric modalities, such as the iris. Ear photos can also serve as supplements for other biometric modalities in automatic recognition systems and provide identity cues when other information is unreliable or even unavailable. In surveillance applications, for example, where face recognition technology may struggle with profile faces, the ear can serve as a source of information on the identity of people in the surveillance footage. The importance and potential value of ear recognition technology for multimodal biometric systems are also evidenced by large scale of studies on this topic, e.g. [12, 13]. Today, ear recognition represents an active research area, for which new techniques are developed on a regular basis and several datasets needed for training and testing of the technology are publicly available, e.g., [8], [9].

An ear-touch impression is a biometric to use for smartphone security. Ear-touch provides users with the ability to adjust their security level to meet their unique requirements. There is an obvious trade-off between device security and convenience. If device security is set too high users may be prompted to submit more than one scan to unlock their device. Conversely, if device security is set too low users run the risk of having their smartphone unlocked by an unauthorized user. The convenience and ease of use of ear-touch biometrics for mobile device security are provided – an individual user simply lifts their device to their ear and presses their ear to the touch screen to authenticate and unlock their mobile device. To our best knowledge, ear-touch biometrics on mobile devices has been introduced for first time in this thesis

1.2. PAD for ear photo and ear-touch on mobile devices

Over the past few years, smartphones and tablets have become a primary means of accessing information in both business and personal contexts. These devices are now the preferred choice for accessing services such as social networks, email, e-commerce, and banking, surpassing traditional computers and becoming indispensable tools in daily life. Mobility and ubiquity work are powerful tools for increasing efficiency and productivity in business (and also in personal life). However, without the proper usage, companies and users may be exposed to security risks and threats. Security in the access to information is one of the most important issues to consider in mobility scenarios.

Passwords have been the usual mechanism for user authentication for many years. However, there are many usability and security concerns that compromise their effectiveness. People use

simple passwords, they reuse them on different accounts and services, passwords can be shared and cracked, etc. The number of different accounts and passwords we deal with these days contributes in making harder the proper usage and maintenance. As a result, we often see news and reports that alert of stolen accounts and passwords. This problem becomes critical in mobile devices, since they can be easily lost or stolen. Nevertheless, mobile devices can also become part of the solution, providing increased levels of security due to their new authentication options and capabilities.

The use of biometrics brings a more secure and convenient authentication method than traditional passwords. The use of biometrics for mobile access control has been established as the most significant development in the biometrics world over the last years [14].

There are so many biological characteristics that can be considered as a biometric system which are divided into two categories; Physiological and behavioral [15]. Iris, Face, Voice, Ear, hand geometry, fingerprint, Palm, vein, retina, and etc. are physiological and keystroke, gait, signature, and etc. are behavioral biometrics, yet there is a question which one of the modalities are possible for mobiles? The modalities with no need of special sensors. If we answer the question, we remove some of the modalities and we should focus on possible ones in practice. In addition to the mentioned biometrics, emerging modalities such as ear-touch, ear photo, and PAD for ear are gaining attention. Ear-touch involves using the unique shape and texture of an individual's ear to authenticate identity, similar to how fingerprints are used. Ear photo, on the other hand, captures the unique contours and patterns of the outer ear using a camera, allowing for non-intrusive biometric verification. Presentation attack detection for ear refers to techniques designed to detect and prevent fraudulent attempts, such as using masks or printed images, to spoof ear-based recognition systems.

In this thesis, we explored the feasibility of conducting biometric research using common mobile devices. For instance, only a limited number of companies can work on fingerprint recognition due to restricted access to sensor outputs on mobile devices. Therefore, we focused exclusively on ear and ear-touch PAD on mobile platforms.

The biometrics of the ears (ear-touch, ear photo) have both advantages and disadvantages compared to other physical attributes. The small surface and the relatively simple structure have a controversial effect [16]. In a positive way, these features provide faster processing compared to face detection and make detection easier compared to fingerprints [17]. On the other side, like other biometrics, current ear biometric recognition systems are vulnerable to

attacks [18-20]. A spoofing attack occurs at sensor level and every impostor can pretend as someone else by altering data, thus, obtaining an illegitimate access. As a first modality, this research analyzes different aspects of ear-touch and ear photo recognition and presentation attacks, such as state-of-the-art presentation attacks detection algorithms, various kinds of artifacts, the accessibility of public databases, and summary of standardization in this area. Due to a lack of anti-spoofing databases, that would support this thesis, ear fake databases have been built using different mobile phones. Releasing the first database in ear PAD can open new ways for investigating on ear biometrics systems more confidently to use future research on mobile smartphones.

1.3. New challenges ear PAD research

Even though ear biometric recognition system has been consequently establishing its position as a secure, reliable, and fast means of biometric authentication over more than 30 years, the last decade has brought the attention of the biometrics community to several new challenges.

Ear-touch recognition: In an increasingly mobile society, where access to banking and business applications over a personal smartphone or tablet is commonplace, personal identification and authentication have emerged as a global imperative. Biometrics is a rapidly evolving technology found not only in forensic science but also in a broad range of consumer applications. An ear-touch analysis is utilized for identification purposes in forensic applications similar to a fingerprint. By combining the most natural of all phone gestures – lifting your phone to your ear – with the unique geometry of your ear, could be created a robust and reliable authentication security solution. When the ear-touch recognition system is running, the operator simply raises the device to their ear in a natural way, and data is captured at the precise moment the ear taps the screen. This biometric characteristic is being introduced in this thesis.

Ear Photo Presentation Attack Detection Challenges

Ear recognition is emerging as a reliable biometric modality within the domain of image-based biometrics for human identification [5]. It can complement facial profile-based recognition and integrate with other biometric modalities such as palm print and gait for enhanced accuracy and robustness [21]. However, like other biometric systems, ear recognition faces significant challenges, particularly in the domain of PAD.

While PAD solutions have been extensively explored for modalities like face [22], fingerprint [23], and iris recognition [24], research on ear PAD remains limited. This gap highlights the need for focused efforts to address vulnerabilities in ear biometric systems.

Ear photo presentation attack detection(PAD) challenges: Ear recognition is emerging as a reliable biometric modality in image-based biometrics for human identification [5]. It can also be used as a complement to facial profile-based recognition and fused with other biometric modalities, notably palm print and gait [21]. As for other biometric modalities, there are challenges for ear recognition systems. Despite the fact that PAD solutions have been widely used for face [22], fingerprint [23], or iris recognition [24], so far there is a few research activities addressing ear PAD. Recent advancements in imaging sensor technologies present new opportunities for both ear PAD, yet the lack of large, diverse datasets has been a barrier to progress in this field.

To address this challenge, we have developed and recorded a substantial dataset specifically designed for ear PAD research. This dataset not only provides the diversity required to analyze presentation attacks comprehensively but also aims to bridge the gap caused by the scarcity of publicly available resources. Through this effort, we hope to advance the development of robust ear PAD solutions and establish a foundation for further research in this critical area of biometric security.

Ear-touch recognition challenges: While ear-touch recognition presents an innovative approach to biometric authentication, it faces several challenges that need to be addressed for it to be widely adopted. The primary challenge lies in the variability of the ear's contact with the device. Factors such as the angle of contact, pressure applied, and environmental conditions (e.g., moisture or dirt on the screen) can significantly affect the consistency and accuracy of the biometric data captured. Moreover, the diverse range of ear shapes and sizes across different individuals adds complexity to the development of universally effective recognition algorithms. Another challenge is ensuring that the ear-touch recognition system can distinguish between genuine and fraudulent attempts, such as using a photograph or a 3D model of an ear. Additionally, user acceptance and comfort are critical, as frequent and precise alignment may not always be practical or user-friendly. Thus, improving the robustness of the technology while maintaining ease of use is essential for its success in mobile applications.

Ear-touch PAD challenges: The deployment of ear-touch biometrics in mobile devices introduces new vulnerabilities to presentation attacks, where an attacker might attempt to spoof

the system using fake ears, such as molds or high-quality images. Unlike more traditional biometric systems like fingerprint or face recognition, there is currently limited research and development in ear PAD, which presents a significant gap. The unique nature of ear-touch recognition requires specialized PAD techniques that can effectively differentiate between a live ear and a presentation attack instrument (PAI). Challenges include detecting subtle differences in texture and temperature that may be used to distinguish between a real ear and a fake one. Additionally, the variability in the quality of sensors across different mobile devices can affect the effectiveness of PAD measures. Therefore, developing robust, device-agnostic PAD techniques that can accurately detect a wide range of presentation attacks is a critical area of research. This thesis aims to explore these challenges and propose solutions to enhance the security and reliability of ear-touch biometric systems.

1.4. Thesis statements

In this thesis, we attempted to investigate the various ear PAD techniques that are available on mobile devices. We discover that combining single biometric variables improves PAD's protection against related attacks on mobile devices.

As previously said, we have classified PAD approaches into models inside faces, which include ear-touch, and ear photo biometrics. The following statement will be proven as part of the approach to improve within-face biometrics (ear photo, and ear-touch) PAD presented in this doctorate dissertation:

S1. We introduced and investigated on an ear-photo presentation attacks detection and recognition system on mobile device.

The idea is illustrated by collecting data and building an ear-photo PAD using deep neural networks as a classifier along with a verification algorithm.

Chapter 3 explores ear-photo PAD and a verification system to show the first statement (S1). Various types of attacks and ear recognition have been developed. To address the PAD issue, we may develop powerful PAD algorithms that increase attack detection performance. Our classifier into real and fake classes, with near-76 percent accuracy, is still unable to distinguish specific presentation attack of ear-photo.

The next statement related to the ear-touch characteristic this statement is used to prove the third statement (S3) which is related to multimodal recognition system:

S2. Ear-touch characteristic is an effective biometric, which is based on mobile devices with multiple touchscreens.

To show this statement, we investigate the newly proposed biometric feature in the recognition system in Chapter 4.

An analytical approach is then proposed to detect ear-touch with high accuracy which has achieved more than 97% in the experiments.

The third statement is related to the combined effects of ear-touch and ear-photo for recognition system on the performance evaluation results:

S3. Ear photo and ear-touch multimodal biometrics improve ear recognition system accuracy in comparison with its single model.

To show third statement, we include an inquiry into the merging of multimodal ear-touches and ear photos in Chapter 5. We suggested a Siamese neural network-based classifier with an equivalent error rate of 0.105 for a recognition system.

S4. "Ear photo and ear-touch multimodal biometrics improve ear PAD accuracy concerning for to a single model."

To show this statement, we include an inquiry into the merging of multimodal ear-touches and ear photos PAD in Chapter 6. We suggested a transformer-based classifier with an equivalent error rate of 0.105 for a PAD system.

1.5. Scope of this Thesis

This doctoral dissertation focuses on the four problems introduced above, namely on introducing a new biometric characteristic ear-touch recognition, proposing an ear photo and ear-touch PAD method and dataset collection on mobile devices for ear biometric.

1.6. Thesis layout

This thesis is structured as follows. Chapter 1 provides an introduction to PAD systems on mobile devices, highlighting recent trends and addressing the emerging challenges in the field.

In Chapter 2, we offer a comprehensive exploration of the WUT-Ear V1.0 database. We begin with an overview that outlines the scope of the database and the rationale behind its creation. The chapter then delves into detailed descriptions of the four key datasets within the database, each focusing on different aspects of ear biometrics and PAD: the ear real photo dataset, ear fake photo dataset, ear real touch dataset, and ear fake touch dataset. For each dataset, we cover the data collection methodologies, preprocessing techniques, and specific applications. We also address the challenges encountered during data integration, especially when combining data from multiple sources and formats. Additionally, we discuss the ethical considerations and data privacy measures implemented to ensure the responsible handling of sensitive biometric data. The chapter concludes with a summary that highlights the key points discussed and the significance of the WUT-Ear V1.0 database for advancing research in ear biometrics and PAD.

Chapter 3 delves into a detailed study that focuses on quantifying the effectiveness of ear authentication on mobile devices. Specifically, it investigates the performance of presentation attack detection and verification algorithms based on a newly developed dataset. The chapter describes the careful collection and characterization of a suitable database comprising diverse ear photos. Through a series of experiments, the extent and underlying reasons for performance degradation in existing ear recognition methods when exposed to challenging samples are thoroughly examined. The findings are summarized, and appropriate conclusions are drawn based on the results obtained.

In Chapter 4, the emphasis shifts to ear-touch based mobile user authentication. A dedicated database of ear-touch patterns is meticulously designed and assembled to support the research objectives. This chapter explores the utilization of ear-touch as a modality for mobile user authentication and access control. A proposed method for capturing ear-touch data is presented, along with a solution to address the challenge of missing data points. The performance of the proposed method is evaluated, demonstrating its effectiveness in achieving high performance in terms of equal error rate.

In Chapter 5, the thesis introduces the initial step towards achieving fusion ear-touch and ear-photo recognition in the two aforementioned scenarios. The chapter proposes a novel approach

that combines ear-photos and ear-touches, integrating this newly designed recognition method into the existing ear recognition pipeline. The experimental results demonstrate a high level of performance in terms of ear recognition accuracy, particularly when dealing with challenging fusion scenarios involving ear-photos and ear-touches.

Chapter 6 focuses on the development of a novel fusion ear-touch and ear-photo presentation attack detection technique. This innovative method aims to further enhance the accuracy of ear recognition, specifically when confronted with samples obtained from both genuine and fake ears. The chapter presents the details of this approach, highlighting its effectiveness in improving the overall performance of presentation attack detection in the context of ear biometrics.

Finally, **Chapter 7** serves as a summary of the entire project, providing a comprehensive overview of the key findings, contributions, and implications of the research conducted throughout the thesis.

Additionally, the thesis includes several appendices. **Appendix A** contains a compilation of biometrics-related vocabulary, error metrics, and testing protocols utilized throughout the doctoral dissertation. **Appendix B** presents some used peso code in Chapter 4. **Appendix C** provides a comprehensive list of the author's publications, including statements outlining their specific contributions to each publication and lists the author's active conference participations. Lastly, **Appendix D** provides a compilation of grants and projects in which the author has participated, along with details regarding their roles within each project.

These appendices offer supplementary information and resources that further enhance the understanding and contextualization of the research presented in the thesis.

2. Datasets and Experimental Data

2.1. Introduction

The absence of a PAD dataset in ear biometric research has motivated us to gather a comprehensive database specifically tailored for mobile devices. Recognizing the increasing importance of robust and reliable biometric systems, we have compiled the Multimodal Mobile Biometric Dataset, a substantial database that focuses on ear verification and, more crucially, ear PAD. This database, known as Warsaw University of Technology Ear Version 1.0 (WUT-Ear V1.0), encompasses a wide range of ear images and data collected from diverse individuals using various mobile devices. The dataset not only includes authentic ear images but also features a comprehensive collection of presentation attack instruments (PAIs), such as printed images and 3D models, to facilitate the development and testing of PAD systems. By providing this rich resource, we aim to advance the research in ear biometrics and PAD, enabling the development of more secure and reliable biometric authentication methods in mobile applications.

This chapter offers a detailed examination of the WUT-Ear V1.0 database, beginning with an overview that explains its purpose, scope, and the reasons for its development. We then delve into detailed descriptions of the four key datasets that comprise the database, each focusing on different aspects of ear biometrics and PAD, including:

- Ear real photo dataset.
- Ear fake photo dataset.
- Ear real touch dataset.
- Ear fake touch dataset.

These sections cover the data collection methodologies, preprocessing techniques, and specific applications of each dataset. We also address the challenges faced during data integration, particularly when combining data from multiple sources and formats. Additionally, we discuss the ethical considerations and data privacy measures taken to

ensure the responsible handling of sensitive biometric data. The chapter concludes with a summary, highlighting the key points discussed and the significance of the WUT-Ear V1.0 database for advancing research in ear biometrics and PAD.

2.2. Overview of the Database

WUT-Ear V1.0 database is composed of ear touches and photos which were collected from students around the world. All participants were located in Poland. The multi-biometric data has been gathered in wild conditions. It was primarily designed for testing algorithms against presentation attacks, specifically on mobile phones [19]. Let's see Table 2-1 in detail, which provides a comprehensive overview of the WUT-Ear V1.0 database. This table delineates the total number of real and fake photos, as well as real and fake ear-touches, alongside the distribution of female and male participants for each category. The dataset is specifically designed for research in ear verification and PAD, offering a balanced representation of genders and a substantial number of samples across all categories.

Table 2-1: Detailed Composition of the WUT-Ear V1.0 Database

Dataset / Params	Real Photos	Fake Photos	Real Ear-Touches	Fake Ear-Touches
Total number of crops	8000	8662	960	520
Total number of female participants	28	28	20	20
Total number of male participants	109	109	72	52

Real Photos and Fake Photos: The dataset includes a substantial number of both real and fake ear photos, with the real photos amounting to 8,000 and fake photos to 8,662. Each gender is represented in this subset, with 28 female participants and 109 male participants contributing to the dataset.

Real Ear-Touches and Fake Ear-Touches: This subset is focused on the touch-based aspect of ear biometrics. It contains 960 real ear-touch samples and 520 fake ear-touch samples. The distribution of participants shows a higher representation of females in real ear-touches (20 females) compared to males (72 males), whereas the fake ear-touches have 20 female and 52 male participants.

This detailed dataset can be utilized in various research applications, including but not limited to:

- Developing and testing ear verification algorithms.

- Enhancing the robustness of PAD techniques.
- Conducting gender-based analysis of biometric systems.
- Evaluating the performance of multimodal biometric systems combining ear photos and ear-touches.

By providing a balanced and comprehensive database, WUT-Ear V1.0 stands as a valuable resource for researchers and developers working towards the next generation of biometric security solutions.

2.3. Dataset 1: Ear Real Photo Dataset

The Ear Real Photo Dataset (ERP dataset) was meticulously compiled to support the development and testing of ear recognition systems, focusing on capturing the natural variability found in real-world conditions. This dataset includes contributions from 137 unique users, representing a wide range of age groups to ensure inclusivity and generalizability. For each individual, more than 70 images were taken, capturing the ear from six distinct angles. This multi-angle approach is essential for creating a comprehensive dataset that can handle variations in ear orientation, a common challenge in practical applications. Figure 2-1 provides an example of the raw ear images from our database, highlighting the dataset's richness and the detailed features captured at various angles.

Figure 2-2 presents an example of preprocessed ear images from the same databases. In this step, we meticulously cropped the images to isolate the ear region, ensuring that extraneous parts of the image are removed. Additionally, any images that were deemed to be of poor quality or unsuitable for analysis were discarded. This preprocessing step is crucial as it enhances the quality of the dataset, ensuring that the subsequent analysis and algorithm development are based on clean and relevant data.

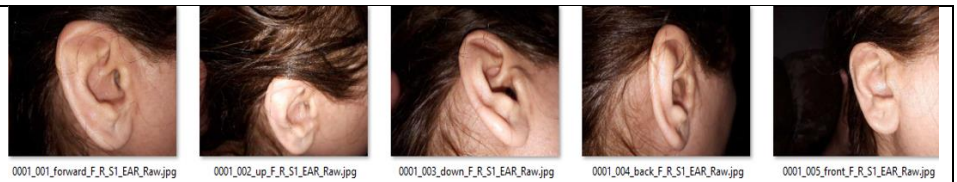
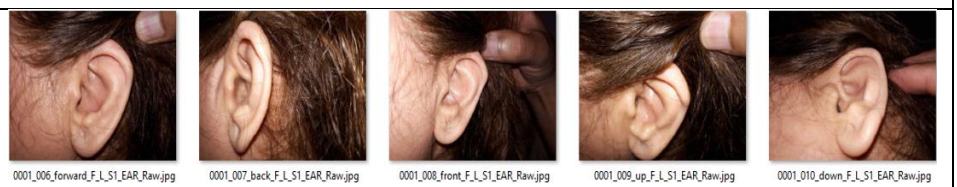
Right Ear	
Left Ear	

Figure 2-1: An example of raw ear photos from ear photo databases

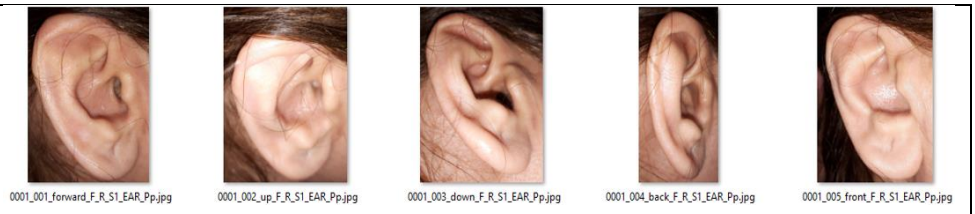
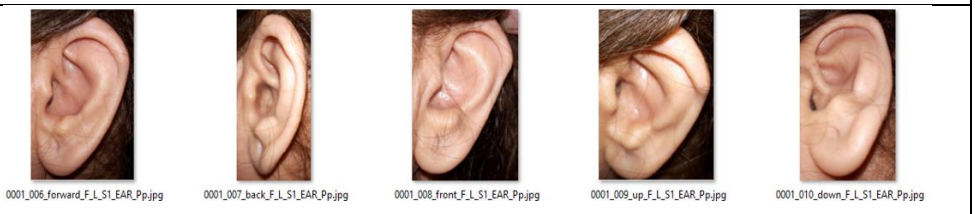
Right Ear	
Left Ear	

Figure 2-2: An example of preprocessed ear photos from ear photo databases

The Table 2-2 shows that the ear photo database includes 8,000 samples from 137 subjects, with each subject contributing 60 ear images for training and testing ear authentication algorithms.

Table 2-2: Database distribution in preprocessed ear photo databases

Condition		Gender (%)	Total number of ear images	Number of Subjects
Male				
Left	Up (10%) Down (9%) Front (8%) Back (10%) Forward (10%)	Male (51%)	3482	109
Right	Up (10%) Down (10%) Front (9%) Back (9%) Forward (11%)	Male (49%)	3373	
Female				
Left	Up (8%) Down (8%) Front (10%) Back (9%) Forward (11%)	Female (46%)	864	28
Right	Up (8%) Down (8%) Front (8%) Back (9%) Forward (11%)	Female (54%)	984	

Table 2-3 shows some basic descriptive statistics for a dataset of real ear photos. The mean value represents the average of all measurements in the dataset, calculated as 63.5 units. The standard deviation, 14.34 units, reflects how much the data points vary around the mean. The minimum and maximum values, 20 and 99 units respectively, show the overall range of the measurements. The first quartile (25%) value of 57 indicates that 25% of the measurements in the dataset fall below 57 units. The median (50%) value of 64 indicates that half of the

measurements in the dataset fall below 64 units. The third quartile (75%) value of 73 indicates that 75% of the measurements in the dataset fall below 73 units. Overall, this Table provides basic information about the distribution of measurements in the real ear photo dataset, including measures of central tendency and variability.

Table 2-3: Description of the real ear photo dataset

Measures	Values
Total	137
Avg.	63.5
std	14.34
min	20
25%	57
50%	64
75%	73
max	99

Figure 2-3 shows the frequency of subjects in a dataset based on the number of images per subject. The x-axis represents the number of images per subject, and the y-axis represents the frequency or number of subjects that have that number of images. The bars in the chart represent the number of subjects in the dataset that have 8000 number of images.

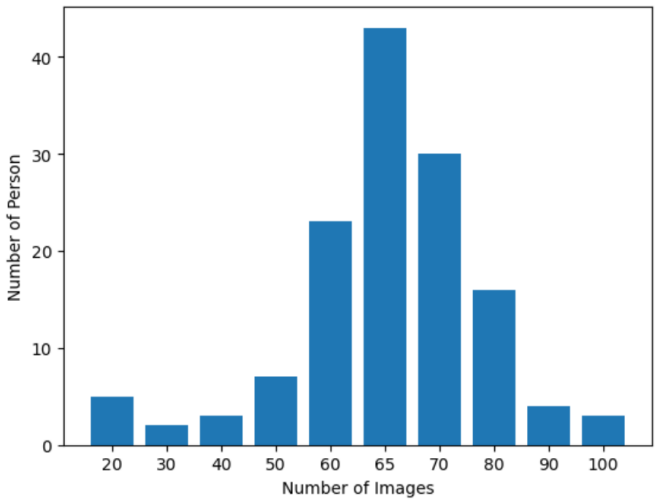


Figure 2-3: Frequency of subjects per number of images.

Furthermore, the chart shows a well-distributed number of images per subject, which means that the number of subjects is roughly the same across different numbers of data samples. The

dataset contains a relatively distribution of photos across subjects, which can be useful for analyzing trends and patterns in the data.

2.3.1. Data collection procedure

The Samsung Galaxy A7 is used to take images and touch information. To take images flash is used. The distance between a subject's ear and mobile phone is about 15 cm. Almost 35 images from 5 different positions are taken in one session from each side therefore totally there are almost 70 images however some subjects have more than 70. It should be noted that more than 10 subjects have earrings. The positions of camera are: up, down, front, forward, and back.

In addition, all images were preprocessed and numbered with a property name. The dimensions of the preprocessed images remained the same as their raw counterparts. On average, the raw images have a resolution of 4608×3456 pixels, while the preprocessed images have 1992×3120 pixels. All images in the dataset are stored in JPEG format.

2.3.2. Data visualization

WUT-Ear V1.0 is a large database therefore, since we are not going to use all extracted features from the images in the visualization, we take a subset of the images which the images are selected randomly. Hence, to visualize we just used 10 subjects overall 100 images from genuine images. We have used t-SNE to visualize images using pre-processed images as input and feature-extracted by a model as shown in Figure 2-4 and Figure 2-5 using a feature-extractor model on the images as input can help to reduce the dimensionality of the data, while retaining the most relevant information about the images. By reducing the dimensionality of the data, t-SNE can more effectively capture the underlying structure of the data and identify meaningful clusters.

Our dataset, which includes real ear photos, can be a valuable resource for researchers and practitioners in ear verification field. The real photos can help to address some of the limitations of existing datasets, such as limited variability and a lack of representative samples. The large number of images can help to reduce the risk of over fitting and increase the generalizability of ear photo verification systems.

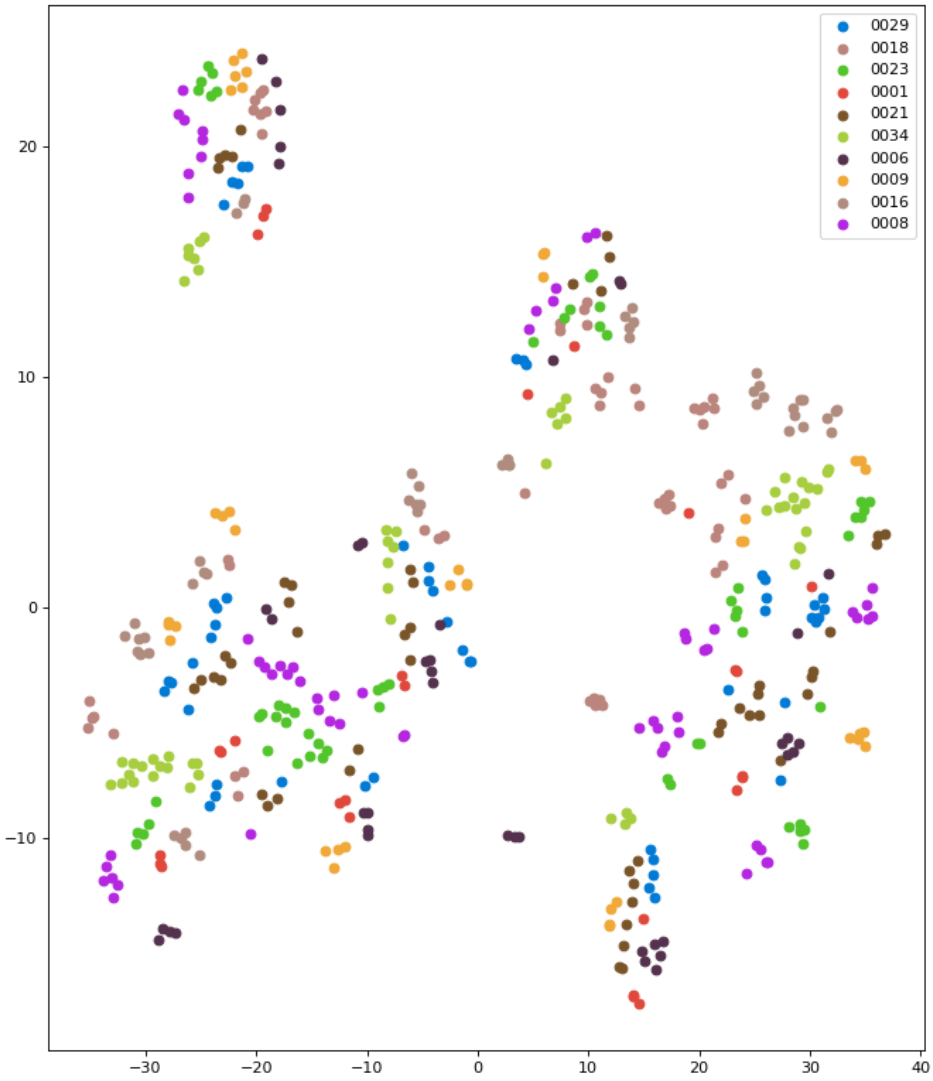


Figure 2-4: Visualizing the dataset using tSNE for the 10 classes of the dataset, dimensions are reduced using pre-processed images without any feature extractor model. Each point on the plot shows an image.

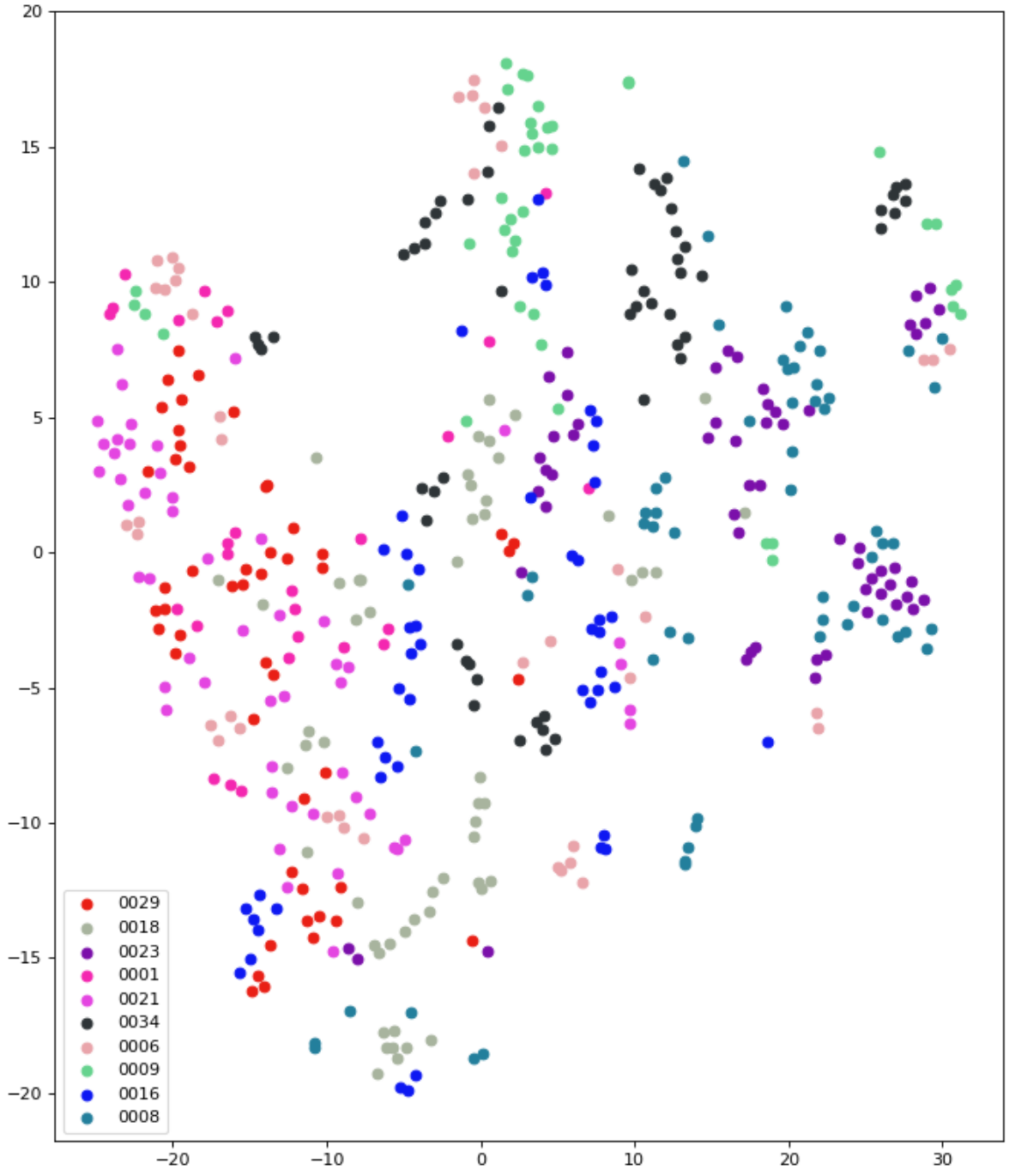


Figure 2-5: Visualizing the dataset using tSNE for the 10 classes of the dataset, dimensions are reduced using feature vectors which are extracted using a feature extractor model. Each point on the plot shows an image of each class.

2.4. Dataset 2: Ear Photo PAD dataset

To address the growing need for robust biometric systems capable of resisting presentation attacks, the Ear Photo PAD Dataset (EPPAD dataset) was developed. This dataset includes a variety of ear fake photos created under different scenarios to simulate potential spoofing attempts. **Error! Reference source not found.** illustrates examples of spoof and real ear p

hotos from a single subject within our dataset, showcasing the variability and complexity of the data.

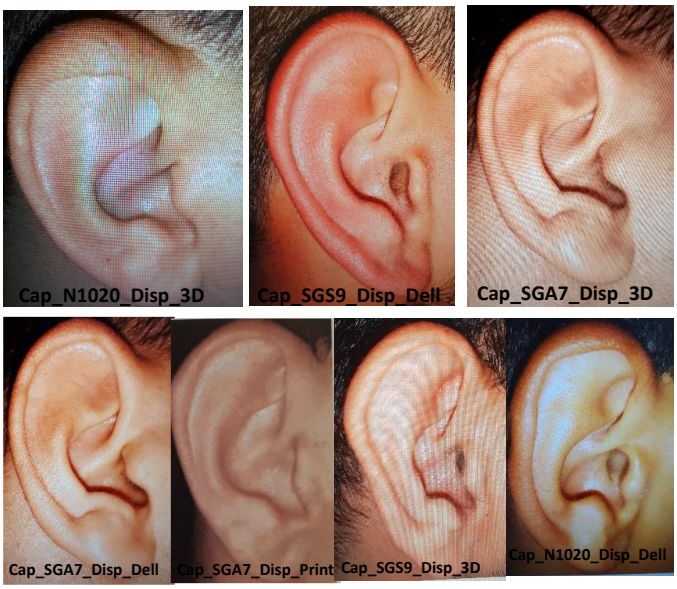


Figure 2-6: Ear fake images samples, the written code bottom of each images shows information about what the recognition devices and display screen are used to take photos. For instance, Cap_N1020_Disp_3D means, we used Nokia lumia 1020 as a recognition system to capture photo and SAMSUNG C27JG50QQUX monitor as a display or representor.

The dataset incorporates three distinct types of attacks, each executed using different display sensors to perform the attacks. Detailed information about these attacks is provided in Table 2-4, which includes the type of attack, the display screen used, the recognition devices involved, the number of fake ear photos, and the total number of photos in the dataset.

The first row of the table describes a display attack using a Dell UltraSharp 32 Ultra HD 4K Monitor. Photos were captured using three different devices: Samsung Galaxy A7, Samsung Galaxy S9, and Nokia Lumia 1020. The dataset comprises 2134 photos captured by the Samsung Galaxy A7, 2827 images by the Samsung Galaxy S9, and 101 images by the Nokia Lumia 1020.

The second row of the table also shows a display attack, but using a different monitor (SAMSUNG C27JG50QQUX) and with different recognition devices. The dataset contains a total of 2339 ear photos captured by the Samsung Galaxy A7, 2026 images captured by the Samsung Galaxy S9, and 1369 images captured by the Nokia Lumia 1020.

The third row of Table 2-4 represents a photo attack, created using a Brother MFC-9340CDW multifunction printer and photos taken with a Samsung Galaxy A7. This section of the dataset contains 189 photos in total.

Table 2-4: Details of synthetic images database

Type of attack	Representor (display screen)	Recognition devices	Number of Fake Ears	Total Number of Images
Display attack	Dell UltraSharp 32 Ultra HD 4K Monitor	Samsung Galaxy A7	2134	5062
		Samsung Galaxy S9	2827	
		Nokia lumia 1020	101	
Display attack	SAMSUNG C27JG50QQUX monitor	Samsung Galaxy A7	16	3411
		Samsung Galaxy S9	2026	
		Nokia lumia 1020	1369	
Photo attack	Brother MFC-9340CDW - multifunction printer	Samsung Galaxy A7	189	189

Table 2-5 shows that the dataset includes a total of 7 different types of attacks, and the number of photos varies across these different attacks. The mean number of photos in the dataset is 1237, with a standard deviation of 1144.

Table 2-5: Description of the ear PAD photo dataset

Measures	**Value**
Number of attacks	7
Avg. number of images	1237
std	1144
min	16
25%	144
50%	1369
75%	2080
max	2827

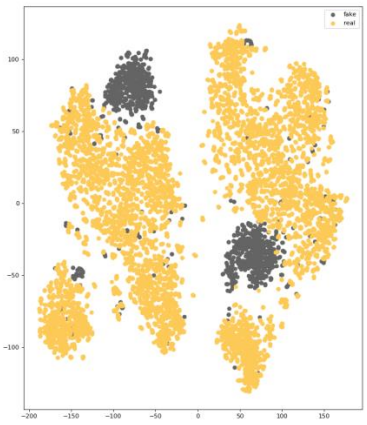
This indicates that the number of images can vary widely across the different attacks, with some attacks having a relatively small number of images and others having a much larger number of images. The minimum number of images in the dataset is 16, indicating that some attacks may have very few images available for analysis. The 25th percentile of the dataset is

144, meaning that at least 25% of the attacks have 144 or fewer images. The median, or 50th percentile, of the dataset is 1369, indicating that half of the attacks have 1369 or fewer images, and the other half have more than 1369 images. The 75th percentile of the dataset is 2080, indicating that at least 75% of the attacks have 2080 or fewer images. The maximum number of photos per attack in the dataset is 2827, suggesting that some attacks have a very large number of photos available for analysis.

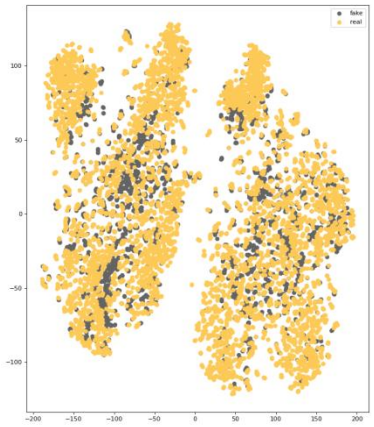
2.4.1. Importance of PAD photos

The sub-figures in Figure 2-7 can be used to demonstrate the importance and potential usefulness of the dataset for ear PAD research. The sub-figure (a) shows the "Fake Capture N1020 mobile device Display 3D Samsung monitor" attack scenario is relatively easy to classify compared to the other sub-figures, which does not require a perfect classification deep learning model. Based on this observation we could conclude that:

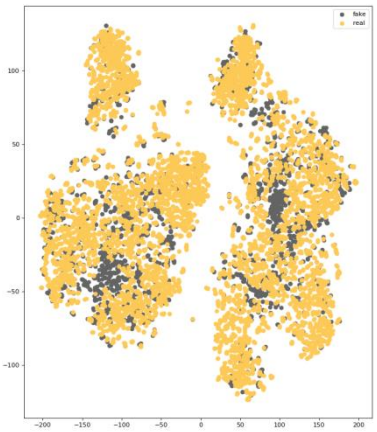
1. Our ear PAD dataset includes a diverse range of attack scenarios: By including multiple attack scenarios in our dataset, we can ensure that our ear PAD model is able to detect attacks across a variety of different devices and methods of capture.
2. Our ear PAD dataset includes both easy and difficult-to-classify attack scenarios: By including both easy and difficult-to-classify attack scenarios, we can ensure that our ear PAD model is robust and effective across a variety of different scenarios.
3. Our ear PAD dataset is well-suited for training and evaluating deep learning models: By using t-SNE to visualize the distribution of images in our dataset, we can demonstrate that our dataset is well-structured and can be used for training and evaluating deep learning models for ear PAD.
4. Our ear PAD dataset is representative of real-world scenarios: By including images captured from real-world devices and displays, our dataset is representative of the types of scenarios that might be encountered in the real world, making it a valuable resource for ear PAD research and development.



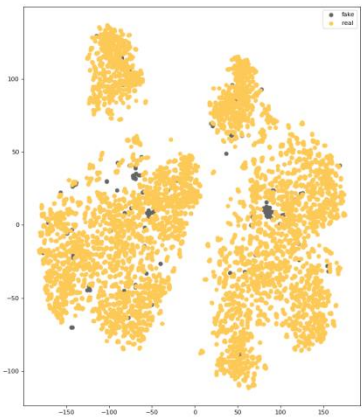
(a) The fake photos are captured by N1020 mobile device and are displayed by 3D Samsung monitor



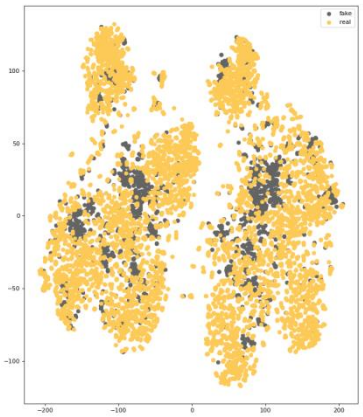
(b) The fake photos are captured by N1020 mobile device and are displayed by Dell monitor



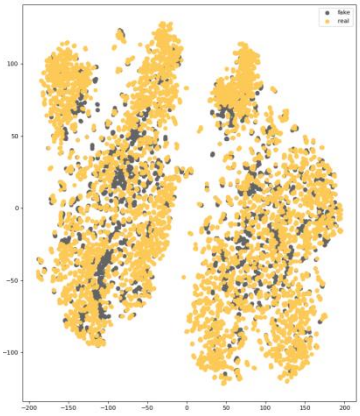
(c) The fake photos are captured by Samsung A7 mobile device and are displayed by Dell monitor



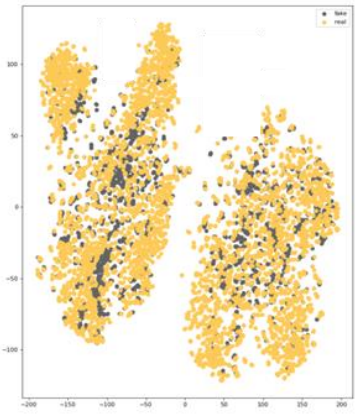
(d) The fake photos are captured by Samsung A7 mobile device and are displayed by Printed using colorful printer



(e) The fake photos are captured by Samsung S9 mobile device and are displayed by 3D Samsung monitor



(f) The fake photos are captured by Samsung S9 mobile device and are displayed by Dell monitor



(b) The fake photos are captured by Samsung A7 mobile device and are displayed by 3D Samsung monitor

Figure 2-7: Sample visualization in two-dimensional space by t-SNE. Samples with multi dimensions can be mapped to two-dimensional space to visualize the original dataset.

2.4.2. Presentations attack instruments

In the realm of biometric security, the ability to detect and mitigate such attacks is crucial to maintaining the integrity and reliability of the system. This section explores two primary types of presentation attacks: display attacks and photo attacks, detailing the methods and tools used to create these spoofing attempts.

Display attack: Display attack is considered one of the most important attacks in presentation attack detection. In these scenarios, attackers try to unlock the system by showing the fake image instead of having a real enrolled person in the recognition system via a monitor or a smartphone. To make fake images, we used a **Dell UltraSharp 32 Ultra HD 4K Monitor** and **SAMSUNG C27JG50QQUX monitor** which had good quality and are suitable for making fake images. Then a Samsung Galaxy A7, Samsung Galaxy S9, and Nokia lumia 1020 smartphone were used for taking photos of the displayed images.

Photo attack: First, we print the genuine photos on A4 glossy paper. For doing this, we used a **Brother MFC-9340CDW - multifunction printer - color Specs** printer which had good quality and is suitable for making fake images. Then a Samsung Galaxy A7 smartphone was used for taking photos of the printed images. The average distances for printed photos are considered as genuine images (~15cm). The sizes of printed photos and the genuine ear photo are the same. The dataset has two advantages: a) all of the images have been captured by the mobile device and it can be used on mobile phone applications, and b) the printed photos are generated with a good quality printer and the size of the fake ear is the same size as the genuine ear.

2.5. Dataset 3: Ear real touch dataset

Touchscreens are one of the sensors used to capture data, in particular biometric data. They have been used in many real-world applications, such as Bodyprint on the Yahoo smartphone. As a matter of fact, the outer ear (shown in Figure 2-8) is composed of different parts including the tragus, antitragus, helix, root of helix, crus of helix, antihelix, lower crus of antihelix, lobule anterior notch, navicular fossa, crus of antihelix, anti-helical fold, lobule, scapha and concha. However, only the first eight parts are touched or captured by a touchscreen, an example of which is shown in Figure 2-9 (b).

Figure 2-9 (a) shows how to capture the data. As shown in Figure 2-9 (b), the subject holds a smartphone to their ear, and then a mobile application captures all possible touch points

touched on the touchscreen. The touched points are extracted and considered as biometric features to authenticate an individual. To collect touchscreen ear biometric data, a mobile application was created and developed in the Android Studio mobile application environment. The application was able to take touchscreen data from several points simultaneously. First, the user completed some information about them in the application. Then the touches for the left and right ears of each subject were captured separately.

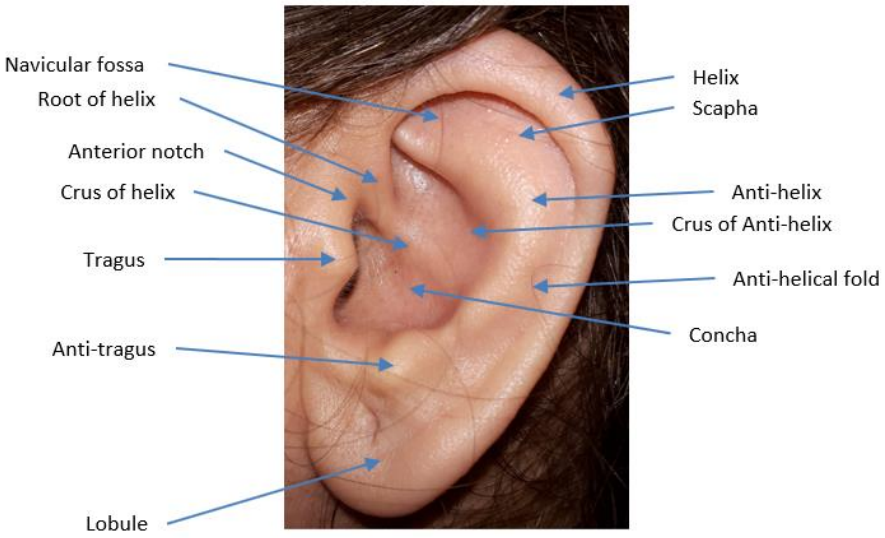


Figure 2-8: Structure of a human ear; eight parts of an ear can be touched or captured by a mobile touchscreen

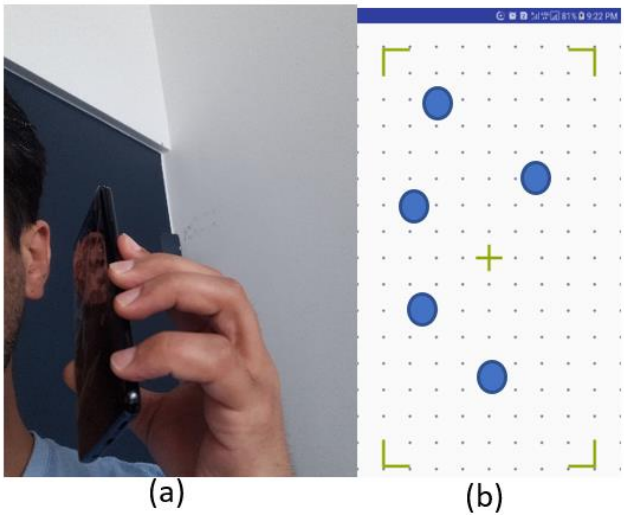


Figure 2-9: Setup of capturing data using a touchscreen on a smartphone – a) shows how to press the smartphone to the ear and b) illustrates a sample of the touched area on a smartphone. The blue circle shapes are touch points; hence, in this sample five touch points are touched by the smartphone

2.5.1. Dataset collection methodology

The ear-touch data associated with almost half of the participants in whole database collection were collected. More details about the dataset are presented in Table 2-6. There were almost 20 ear-touches per subject. At each session, seven touches from each ear were taken, among which there were some unsuccessful acquisitions of data, considered as an unsuccessful attempt. To determine which were unsuccessful presentations, it was decided that sample presentations with less than four touch points would not be considered. Therefore, presentations giving four or more touch points and additional information were considered as a successful presentation. This rule was applied to ensure the presence of enough touch points to carry out the calculations.

According to hardware limitation of touchscreens on the mobile devices based on Android operating system, we could not acquire more than one touch point from each part (mentioned in Figure 2-8) of an ear. According to Table 2-6, there were 92 subjects – 72 men and 20 women. If two ears (left and right ear) were taken for each subject, we would have a total subject of 92. Despite the higher number of male subjects, we had a greater number of data acquisitions from female subjects, as there were more unsuccessful attempts using the males during data collection.

Table 2-6: Dataset distribution in ear-touches of view at WUT-Ear databases. Each ear from a person is considered as a subject

Sex	Single/ Multi	Total number of ear-touches	Number of Subjects
Male	Multi Ear-touch	585	42
Female	Multi Ear-touch	343	18
Male	Single Ear-touch	30	30
Female	Single Ear-touch	2	2
Total		960	92

In Figure 2-10, the distribution of the number of ear-touches is shown. The vertical and horizontal axes show the number of subjects and the number of ear-touches respectively. According to our data collection experience, we were able to record ear-touches with just one ear touch, but the graph shows ear-touches with at least two ear-touches. Therefore, in the graph, the minimum and the maximum touch points were two and 30 respectively.

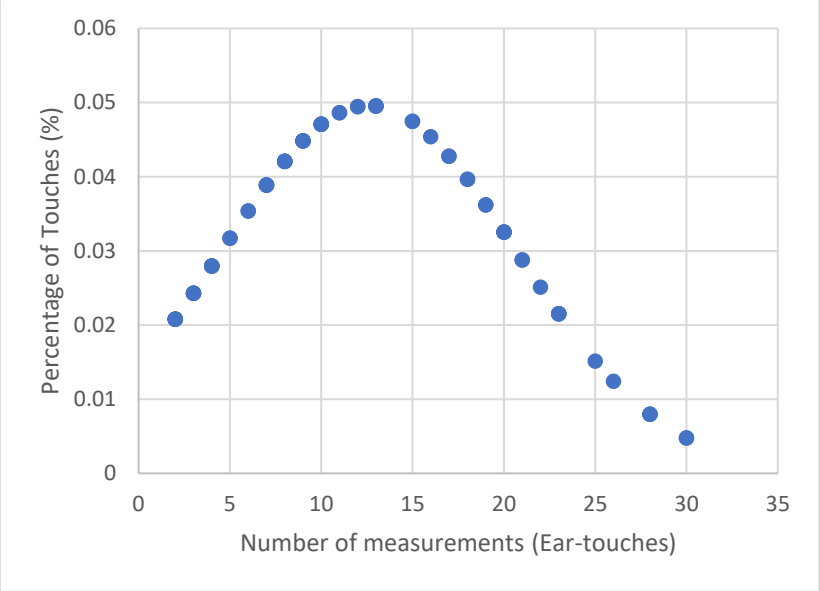


Figure 2-10: The distribution of ear-touches from 72 subjects. The number of measurements and the frequencies are shown in the horizontal and vertical axis respectively.

In Figure 2-11, a distribution of the maximum number of touch points is shown. The vertical and horizontal axes show the number of subjects and the maximum number of touch points respectively. According to our data collection experience, we could record ear-touches with fewer touch points (less than four touch points); hence, we captured ear-touches with at least four touch points. Therefore, in the graph, the minimum and the maximum touch points were four and eight respectively.

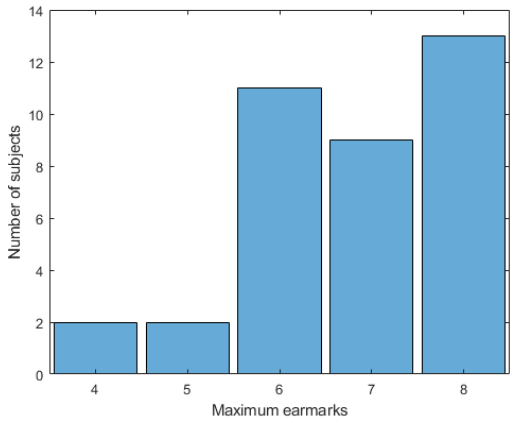
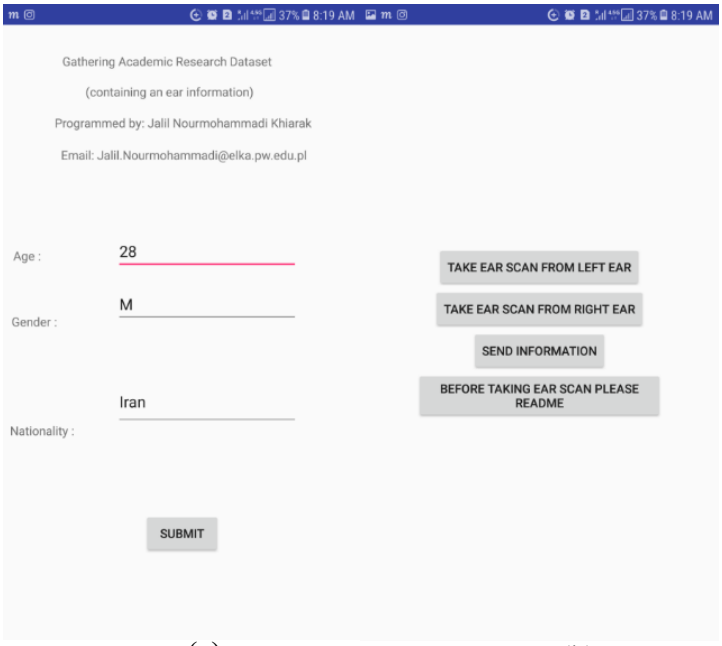


Figure 2-11: Maximum number of touch points in all subjects

2.5.2. Collection of measurements, procedures and problems (missing data)

In this section, the data collection experiments are explained. To collect touchscreen ear biometric data, we created a mobile application, as shown in Figure 2-12. It is developed in the Android Studio mobile application environment. The application is able to take touchscreen

data from several points simultaneously. First of all, the application asks the user for some information about them (Figure 2-12(a)). Then, as shown in Figure 2-12 (b), ear-touches are captured for the left and right ear of each subject separately.



(a) (b)

Figure 2-12: Ear touch data collection mobile application: a) The first page of the developed application, personal information should be filled; this is done to analysis data in terms of gender, age, and nationality (or race), b) the second page of the developed application, this page has three parts; take ear scan from left ear – this is used to capture the ear-touch from the person's left ear; Take ear scan from right ear – this is used to capture the ear-touch from the person's right ear. Send information is used to send the collected data to the server. The last button is used to explain the consent form and all about the research purposes.

Like a fingerprint recognition system, the ear touch needs to be enrolled several times. At each recording session we took seven touches from each ear, though some data acquisition attempts proved to be unsuccessful. We considered these as unsuccessful attempts. In total, we recorded 1427 ear-touch presentations. Of these, 467 had fewer than four touch points and were excluded as unsuccessful presentations. Presentations with four or more touch points were considered successful, resulting in 960 successful ear-touch presentations.

2.6. Dataset 4: Ear touch PAD dataset

Figure 2-13(a) shows how to capture the data. As shown in Figure 2-13(b), the subject holds a smartphone to their hand, and then a mobile application captures all possible touch points touched on the touchscreen. The touched points are extracted and considered as fake biometric features to authenticate an individual.

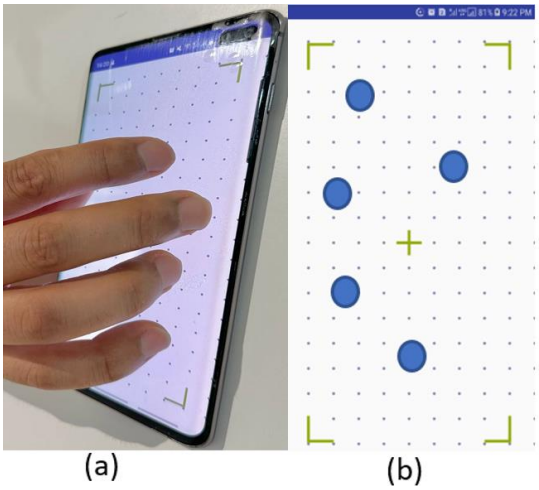


Figure 2-13: Setup of making fake data using a touchscreen on a smartphone – a) shows how to press hand on the smartphone and b) illustrates a sample of the touched area on a smartphone. The blue circle shapes are touch points; hence, in this sample five touch points are touched by the smartphone

2.6.1. Dataset collection methodology

To acquire fake ear-touches, a Samsung Galaxy S10+ and Galaxy A21 were used. To make fake data four hands (Two a man hands and two a woman hands) were used. To make more than five touch points both hands used to capture touch information. To take similar but fake ear-touch all subjected looked and tried to imitated that subject.

Fake ear-touch data were created for 72 subjects (52 male and 20 female), with an average of 7 ear-touches per subject, resulting in a total of 504 ear-touches. At each session, seven touches from each ear were taken, among which there were some unsuccessful acquisitions of data, considered as an unsuccessful attempt. To determine which unsuccessful presentations were, it was decided that sample presentations with less than four touch points would not be considered. Therefore, presentations giving four or more touch points and additional information were considered as a successful presentation. This rule was applied to ensure the presence of enough touch points to carry out the calculations.

2.6.2. Pre-processing of the data

Some participants did not participate in the ear-touch data collection. To address this, we organized the dataset based on available ear-touch samples to facilitate the creation of fake ear-touch data. Specifically, we generated a new dataset combining ear-touch and ear-photo data, and subsequently created synthetic samples for these subjects.

During data acquisition from mobile devices, additional metadata unrelated to the touch points was recorded. This extraneous information was filtered out to ensure a clean dataset of touch

points. The processed data was then prepared for use in training the feature extractor, ensuring consistency and accuracy in the dataset.

2.7. Ethical Considerations and Data Privacy

In conducting research involving human participants, following ethical guidelines is essential to ensure the protection of participants' rights and privacy. The following section outlines the ethical considerations and procedures that were followed in the collection and handling of data for this study. This information was submitted as part of the "Application to Warsaw University of Technology for Research with Human Participation" and approved by the relevant ethical review board.

Participants for this study included university students, family, and friends of the applicant. This diverse group ensured a wide range of biometric data for comprehensive analysis. The selection criteria were designed to include individuals who could provide varied biometric samples under different conditions.

Participants were involved in the following activities:

- Making ear photographs and ear touch images with smartphones for biometric purposes.
- Multi-modal presentation attack detection based on gaze information (Collected data is not used in this thesis).
- Pupil dynamics for iris liveness detection using the visible light spectrum on mobile devices (Collected data is not used in this thesis).

These activities were designed to gather diverse biometric data to test and validate the novel PAD algorithms developed during the study.

There was no significant risk involved for participants in this study. All procedures were designed to ensure participant safety and comfort. The activities conducted were non-invasive, and participants were fully informed about the study's procedures and goals. The study did not involve the participation of minors (below 16 years) and followed strict guidelines to ensure compliance with ethical standards. Data were collected in Iran (home country of the participant) and Poland. This cross-cultural data collection enhanced the robustness of the research findings by incorporating biometric data from diverse demographic backgrounds. No payments were planned for participants in this study. Participation was entirely voluntary, and participants contributed to advancing scientific knowledge.

A detailed consent form was provided to each participant, outlining the study's purpose, procedures, risks, and benefits. Participants had the right to withdraw their data at any time and for any reason without penalty. This ensured that participants were fully aware of their rights and the nature of the study. Data were stored securely on a disk safeguarded by the researcher. Each usage of the data was registered to ensure accountability. There were no plans to share data outside the European Union. Personal data such as names and addresses were stored separately from biometric data to prevent identity disclosure. These details were used only for possible data removal at the participant's request. No individual experimental data were shared. Biometric data were aggregated statistically, ensuring no risk of divulging individual information. Data will be stored indefinitely, allowing for ongoing and future research while ensuring data integrity and security. Findings from this study will be disseminated through diploma dissertations, reports, conference presentations, and journal publications. This ensures that the research contributes to the broader scientific community and supports advancements in mobile biometric security.

This ethical framework ensured that the research was conducted responsibly, respecting participants' rights and maintaining the integrity of the data collected. By adhering to these ethical guidelines, the study was able to gather valuable biometric data while ensuring the privacy and safety of all participants. This information was crucial for developing and validating new algorithms for mobile biometric systems, contributing to the advancement of secure biometric technologies.

2.8. Challenges

The WUT-Ear V1.0 database, while comprehensive and innovative, presents several challenges that must be addressed to enhance its utility for ear biometric and PAD research. One primary challenge is the variability in data quality due to the "in the wild" conditions under which the data was collected. These uncontrolled environments can introduce inconsistencies in lighting, background, and positioning, potentially affecting the reliability of the data for training and testing algorithms. Another challenge lies in the inherent complexity of distinguishing between real and fake ear samples. For instance, the fake ear photos and ear-touches were generated using various methods and materials, which may not cover all possible types of presentation attacks. This could limit the dataset's representativeness and the robustness of algorithms developed using it.

Additionally, the gender distribution in the dataset, particularly in the real ear-photo samples, where there is a higher representation of male participants, may introduce bias. Such an imbalance can skew the performance of biometric systems, potentially leading to less accurate results for underrepresented groups. Ensuring an equitable representation across different demographics is crucial for the generalizability of biometric systems.

Moreover, the technical limitations in capturing touch-based data for ear-touches pose a significant challenge. Variations in the pressure applied during ear-touch and the angle of contact can affect the consistency of the biometric data collected. This variability necessitates advanced preprocessing and normalization techniques to ensure that the data is consistent and reliable.

Finally, ethical considerations and data privacy concerns also present challenges. As the dataset includes sensitive biometric data, strict measures must be in place to protect the privacy and security of participants' information. Ensuring compliance with relevant data protection regulations, such as GDPR, is essential to maintaining the ethical integrity of the research.

Addressing these challenges is crucial for maximizing the dataset's value and ensuring that it supports the development of robust, fair, and effective biometric systems.

2.9. Conclusions

In conclusion, the WUT-Ear V1.0 database provides a comprehensive and diverse collection of ear biometric data, encompassing both real and fake samples of ear photos and ear-touches. This dataset is instrumental in advancing research in ear verification and PAD, particularly within the context of mobile devices. The inclusion of both real and fake samples, along with a balanced representation of male and female participants, ensures a robust foundation for developing and testing biometric algorithms.

However, several challenges must be addressed to fully utilize the potential of this database. The variability in data quality due to uncontrolled collection environments poses a challenge for consistent algorithm training. The potential biases arising from gender imbalances in certain subsets, as well as the technical difficulties in capturing tactile data, highlight the need for careful consideration in data preprocessing and analysis.

Additionally, the dataset underscores the importance of addressing ethical considerations and maintaining data privacy, especially given the sensitive nature of biometric information.

Implementing strict measures to safeguard participants' privacy and ensure compliance with data protection regulations is critical for ethical research practices.

Overall, the WUT-Ear V1.0 database is a valuable resource for the biometric community, offering significant opportunities for innovation in ear biometric recognition and PAD. As the field progresses, addressing the challenges identified will be key to improving the effectiveness and fairness of biometric systems developed using this database.

3. Ear Authentication on Mobile Devices: An Investigation into Presentation Attack Detection and Recognition Algorithm based on a New Dataset

3.1. Introduction

Ear authentication is an emerging field of biometrics that utilizes the unique physical characteristics of an individual's ear for identity verification [1, 25-27]. Compared to other biometric modalities, ear authentication offers several advantages, including non-intrusiveness, contactless nature, and high recognition accuracy. With the increasing use of mobile devices for authentication, there is a growing interest in developing ear authentication systems for mobile platforms.

However, ear authentication systems on mobile devices can be vulnerable to presentation attacks, where an attacker can present a fake or altered ear photo to the system [18-20]. Presentation attack detection and verification algorithms are crucial for ensuring the security and reliability of ear authentication systems on mobile devices. Additionally, mobile platforms often have resource constraints such as limited processing power, storage, and memory, making it challenging to implement robust and efficient presentation attack detection and verification algorithms.

In this chapter, we investigate presentation attack detection and verification algorithms for ear authentication on mobile devices using a new dataset. The dataset contains a diverse set of ear photos with various presentation attack types, including print attacks and replay attacks. We evaluate the performance of existing presentation attack detection and verification algorithms on the new dataset and compare the results with previous studies.

Our investigation has several contributions. Firstly, we provide a benchmark for evaluating the performance of ear authentication systems on mobile devices against various presentation attack types. Secondly, our study provides insights into the effectiveness and efficiency of

existing presentation attack detection and verification algorithms on mobile platforms. Thirdly, our study highlights the importance of developing robust and efficient ear authentication systems for mobile devices to ensure their practicality and usability.

The remainder of this chapter is structured as follows. Section 3.2 surveys prior work on presentation-attack detection and verification for ear-based authentication on mobile devices. Section 3.3 details the new dataset, covering data collection and preprocessing. Section 3.4 outlines our methodology for ear presentation-attack detection and verification. Section 3.5 reports the experimental results and contrasts them with earlier studies. Finally, Section 3.6 offers concluding remarks and directions for future work.

3.2. Literature review

PAD is crucial to ensure the security of ear authentication systems against spoofing attacks. Several techniques have been proposed to detect presentation attacks, including the use of machine learning algorithms to differentiate between genuine and fake ear images. In 2018, a small anti-spoofing dataset from the existing dataset called AMI and their own dataset has been made for ear PAD [19]. They have used Image Quality Assessment for feature extraction and in following on an SVM classification was applied to the features to distinguish between real and fake input. Three kinds of attacks have been designed for ear PAD namely; video attack, display attack, and photo attack. Half Total Error Rate (HTER) for their own dataset and AMI dataset were 22.4 and 0 respectively.

In [28], a no-reference(such as Distortion Specific Measures, Training Based Measures and Natural Scene Statistics Measures are implemented to distinguish fake and real ear images) and full-reference(such as such as Error Sensitivity Measures, Correlation-Based Measures, Pixel Difference Measures, Edge-Based Measures, Gradient-Based Measures, Spectral Distance Measures, Information Theoretic Measures and Structural Similarity Measures) IQA methods are used for detecting print attacks. They used 21 full-reference and 4 no-references IQA methods and the outputs are classified by a K-Nearest Neighbor classifier. To test the proposed method, they have used AMI and University of Beira EAR (UBEAR) as a real image and they have made 200 fake images using a printer. The HTER in the best condition is respectively 15.5% and 8.5% for the UBEAR and AMI databases.

Also, In [20], both full-reference and no-reference IQA methods are utilized to detect attacks on ear biometrics with more experiments on AMI and UBEAR datasets. They proposed a

three-level fusion method including; a) feature extraction using both full-reference and no-reference IQA methods, b) score normalization to move them into the same scale and c) decision level. Printed images on paper are used as attack purposes which are made using AMI and UBEAR datasets. The HTER in the best condition is respectively 1% and 10% for the AMI [9] and UBEAR databases [29]. Moreover, they have shown that CNN-based method doesn't achieve high accuracy in comparison with their proposed method on UBEAR database in which HTER for their methods and CNN-based method 14.5% and 10.0% has been reported respectively. As a matter of fact, CNN-based methods need more dataset to get better results.

In [18] a dataset is collected for ear PAD, the Lenslet Light Field Ear Artefact Database (LLFEADB), including light field and 2D ear artifact images. The dataset has two subsets; Baseline Set (BS) is formed 268 ear images from 67 users, Lytro ILLUM lenslet light field camera for image capturing, and the images are resized to 192*128 pixels and Extended High-Resolution Set (EHRS) is formed 60 images from 14 users from IST-EURECOM LLFEDB bona fide images and the images are resized to 1152*768 pixels. Four attacks are considered; Laptop Attack (MacBook Pro 13'' display device), Tablet Attack (iPad Air2), Mobile Attack 1, 2(iPhone 6S and Sony Xperia z2). Conventional 2D Methods (C2DM) and Light Field base Methods (LFM) (such as Light Field Angular Local Binary Patterns (LFALBP), exploiting ray and edge difference features, and light field histogram of gradients (LFHoG) descriptor) are used as PAD algorithms. Attack Presentation Classification Error Rate (APCER) for LFM was better than other methods in the paper and has achieved 0.0% for all existence attacks in the dataset. Based on 64-bit Intel PC with a 3.40 GHz processor and 16 GB RAM, MATLAB R2015b, running time was 217ms per image.

Verification algorithms, on the other hand, aim to verify the identity of the user based on their ear features. The most commonly used approach for ear verification is feature-based recognition, which involves extracting distinctive features from ear images, such as the shape, texture, and contour. A number of studies have investigated the effectiveness of different feature extraction methods for ear verification, such as the Log-Gabor filter-based approach proposed by Heng Liu [30], which achieved an accuracy of 93.3% on a dataset of 480 ear images with eight large view variation ear images. However, a key challenge in ear authentication is the lack of publicly available datasets for evaluating the performance of different algorithms. In conclusion, the investigation into presentation attack detection and verification algorithms for ear authentication on mobile devices has shown promising results.

3.3. Ear photo recognition methodology

To have unique features of an individual's ear for recognition purposes there is a methodology for ear photo recognition:

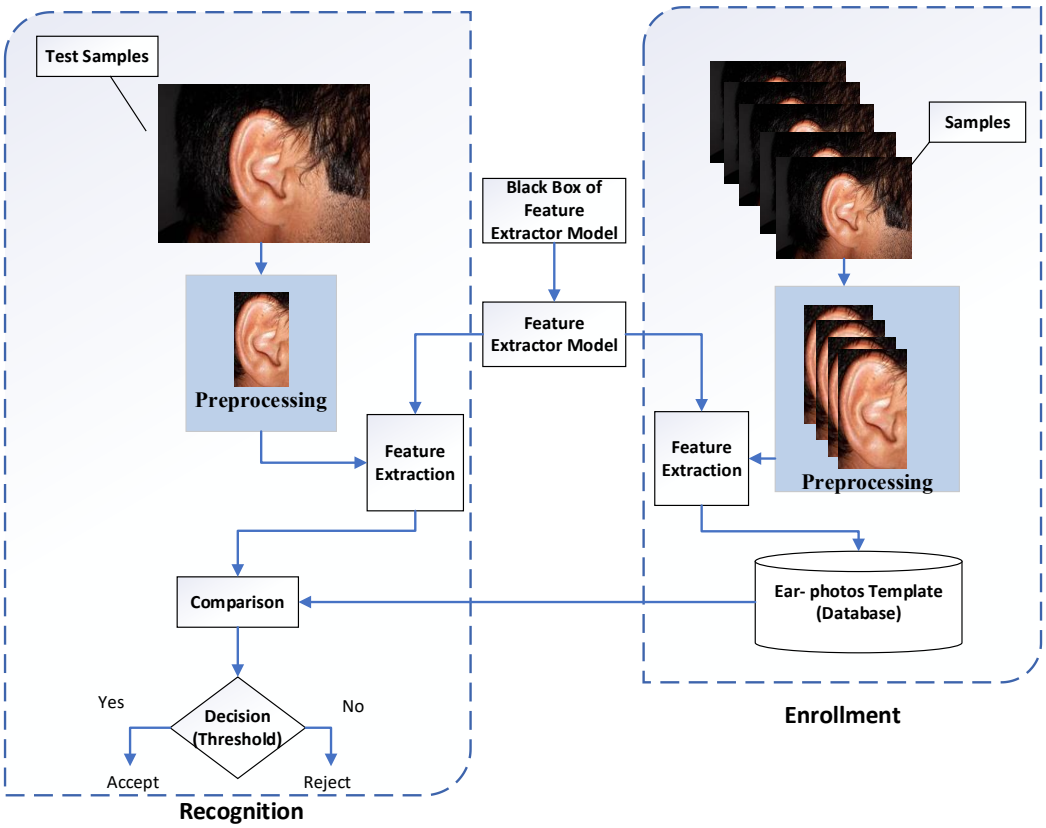


Figure 3-1: Ear photo recognition methodology

The flowchart in Figure 3-1 describes an ear photo recognition methodology for biometric authentication. Here is a brief explanation of each step in the process:

Input data (samples): This refers to the ear photos that are captured from a user using a mobile device.

Test samples: This refers to the ear photos that are given to recognition system to test the system.

Preprocessing: This step involves manually crop of images.

Black Box of Feature extractor model: In this step, distinctive features are extracted from the preprocessed ear photos using a Siamese network. These features are then used to create a template that represents the unique characteristics of the user's ear.

Feature extractor model: In this step, distinctive features are extracted from the preprocessed ear photos using a Siamese network. These features are then used to create a template that represents the unique characteristics of the user's ear.

Enrollment database: This is a database that stores the templates created from the feature extraction step. The enrollment process involves creating a template for each user and adding it to the database for future use in verification.

Comparison: During the verification process, the feature extraction step is repeated on a new ear photo, and a new template is created. The matching step involves comparing this new template to the templates in the enrollment database to determine if there is a match.

Decision: Finally, the decision step involves making a decision about whether the user is authenticated or not based on the outcome of the matching step. If there is a match between the new template and a template in the enrollment database, the user is authenticated, and access is granted. Otherwise, the user is not authenticated, and access is denied.

3.3.1. Feature extractor models

This section discusses the methodologies employed to extract features from ear photos, which are crucial for accurate recognition [1, 17, 27]. The methodology incorporates two main approaches: deep learning-based feature extraction using a Siamese network, and a classical image processing technique known as Local Binary Patterns (LBP).

Siamese Network with Pre-trained Backbones: A Siamese network learns a similarity function between pairs of inputs by passing each input through twin subnetworks that share parameters [31]. Each branch produces a feature embedding, and a distance metric over these embeddings quantifies likeness between the inputs. For ear-image recognition, this setup encourages features that discriminate among individuals' ear shapes. In our approach, the Siamese branches use pre-trained convolutional backbones for feature extraction (see Fig. 3.2). We evaluate two such backbones: MobileNetV2 [32] and VGG16 [33].

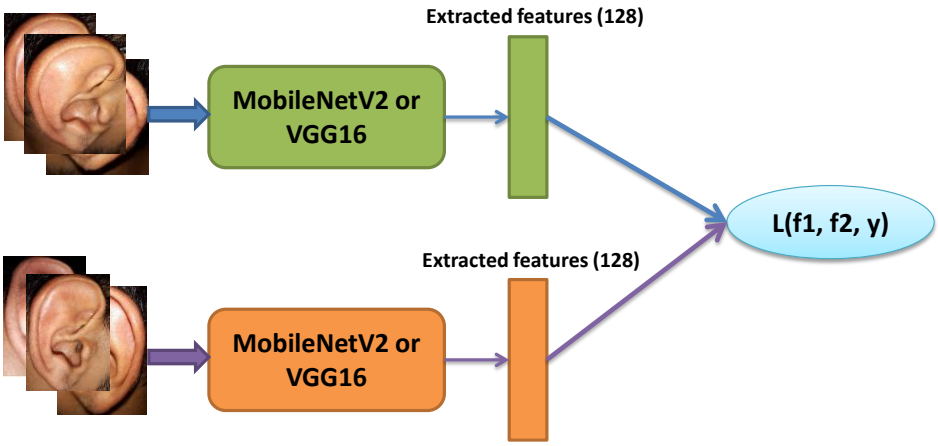


Figure 3-2: A level architecture of our used model with pre-trained deep neural network

Siamese network with a MobileNetV2 backbone: MobileNetV2 is a lightweight CNN tailored for mobile and edge deployment. By using depthwise–separable convolutions with inverted residuals and linear bottlenecks, it drastically reduces parameters and compute, making it suitable for real-time use. Pre-trained on ImageNet [34], it provides strong generic visual features that we then fine-tune on our ear-image dataset to specialize the representation. Within the Siamese framework, each branch uses MobileNetV2 to encode an input ear image into a compact, discriminative embedding that summarizes the salient structural characteristics of the ear.

Training a Siamese Network with MobileNetV2 Backbone: To train a Siamese network for ear photo recognition, we integrate MobileNetV2's feature extraction capabilities with the Siamese framework. The process includes input preparation, network architecture setup, and training.

Input Layer: Each subnetwork receives an ear image as input.

Feature Extraction: MobileNetV2 processes the image and produces a feature vector representing its learned features.

Output Layer: The outputs of both subnetworks are compared using a distance metric (cosine similarity) to measure similarity between the two input images.

Pair Formation:

Positive pair: two ear images from the same individual.

Negative pair: two ear images from different individuals.

Preprocessing:

Resizing – both images are resized to 224×224pixels to match MobileNetV2 input requirements.

Normalization – pixel values are scaled appropriately (typically between 0 and 1).

Augmentation – transformations such as rotation, scaling, and flipping are applied to increase model robustness.

Feature Vector Extraction: Each image passes through MobileNetV2 to generate a high-dimensional feature vector.

Shared Weights: The subnetworks share parameters, ensuring identical processing and comparable feature representations.

Similarity Computation: After obtaining feature vectors $f_1$and f_2, the distance between them is computed to assess image similarity.

Cosine similarity: As an alternative metric, we compute the cosine of the angle between two embeddings to quantify their likeness (Eq. 1). The corresponding cosine distance is [10].

$$Cosine\ Distance = 1 - cos(\theta) = 1 - \frac{f_1 \cdot f_2}{\| f_1 \| \| f_2 \|} \quad (1)$$

Where $f_1 \cdot f_2$ denotes the dot product and $\| f_1 \|$ and $\| f_2 \|$ are the magnitudes of the feature vectors f_1 and f_2, respectively.

Contrastive Loss: A standard objective for Siamese models encourages embeddings of positive pairs to be close and those of negative pairs to be separated by at least a margin. Using label $y = 0$ for positive pairs and $y = 1$ for negative pairs, the loss is.

$$L(y, D) = (1 - y)\frac{1}{2}D^2 + y\frac{1}{2}max(0, m - D)^2 \quad (2)\ [35]$$

Here, y is 0 for positive pairs and 1 for negative pairs, D is the distance between the feature vectors, and m is a margin parameter that defines the minimum distance for negative pairs.

Data collection: We assemble a labeled ear-image dataset with identity annotations. From these labels, we form image pairs: multiple **positive** pairs per identity (same person) and **negative** pairs drawn across identities (different people).

Training procedure: Each iteration samples a mini-batch of positive and negative pairs. Both images in a pair are passed through the Siamese branches (MobileNetV2 encoders) to obtain embeddings; a distance (e.g., cosine-based) is computed for each pair, and Eq. (2) is minimized via backpropagation.

The contrastive loss is computed based on the distances and the labels (positive/negative). The network weights are updated using backpropagation and an optimization algorithm such as Adam or SGD. The training process continues for several epochs until the model converges, with the loss decreasing and the accuracy improving. A separate validation set is used to monitor the performance of the Siamese network during training. After training, the network is evaluated on a test set using metrics such as False Match Rate (FMR) and False Non-Match Rate (FNMR) to measure its effectiveness in distinguishing between different individuals based on ear photos. Since MobileNetV2 is pre-trained on ImageNet, we use transfer learning to adapt it to the ear recognition task. Initially, the network may be fine-tuned with a lower learning rate to retain useful features from ImageNet and then gradually trained on the ear dataset. Initially, some layers of MobileNetV2 may be frozen (not trained) to preserve the learned features from the pre-trained model, especially in the early stages of training.

Siamese network based on VGG16:

VGG16 is a deeper network with 16 layers, known for its simplicity and effectiveness in capturing high-level features through its sequential convolutional layers. Like MobileNetV2, VGG16 is pre-trained on ImageNet and then fine-tuned on ear photos. Its architecture allows it to capture more detailed features, which can be particularly useful for distinguishing between subtle variations in ear shapes. VGG16 extracts a feature vector from the input ear image, representing the key aspects of the ear's appearance.

To train a Siamese network using VGG16 for ear photo recognition, we follow a systematic approach that closely resembles the process used for MobileNetV2, but with certain key differences due to the specific characteristics of the VGG16 architecture.

The Siamese network architecture involves two identical subnetworks that share weights. Each subnetwork in this case is based on the VGG16 architecture. The input to the Siamese network consists of a pair of ear images, where each pair can be either a positive pair (images of ears from the same person) or a negative pair (images of ears from different people). These pairs

are processed by the two subnetworks to produce feature vectors, which are then compared using a distance metric to measure the similarity between the two images.

For input preparation, the ear images are resized to match the input size required by VGG16, typically 224x224 pixels. The pixel values of the images are normalized, and data augmentation techniques such as rotation, scaling, and flipping may be applied to improve the model's robustness. Each image in the input pair is passed through the VGG16 subnetwork, which outputs a high-dimensional feature vector. Because the two subnetworks share weights, the feature vectors produced by them are directly comparable.

To measure the similarity between the two feature vectors, cosine similarity is used. The contrastive loss function is employed to train the network. This loss function encourages the network to reduce the distance between feature vectors for positive pairs and increase it for negative pairs, helping the model learn to distinguish between different individuals based on their ear structures.

The training process involves preparing a dataset of ear images, generating pairs of images (both positive and negative), and then passing these pairs through the Siamese network. In each training iteration, a batch of image pairs is formed, and each pair is processed by the VGG16 subnetworks. The contrastive loss is computed based on the distances between the output feature vectors and their corresponding labels (positive or negative), and the network weights are updated using backpropagation and an optimization algorithm such as Adam. This process is repeated for multiple epochs until the network converges, meaning the loss decreases and the accuracy improves.

Since VGG16 is pre-trained on ImageNet, transfer learning is employed to adapt it to the ear recognition task. Initially, some layers of VGG16 may be frozen to preserve the features learned during pre-training, especially in the early stages of training. As training progresses, these layers may be unfrozen to allow the network to fine-tune its feature extraction capabilities specifically for the ear recognition task.

Local binary patterns (LBP): We use LBP as our second feature extractor model. A hand-crafted based algorithm (Local binary patterns (LBP)) and a CNN based algorithm (VGG-16) has been introduced as a baseline algorithm for feature extracting in the ear recognition in the Unconstrained Ear Recognition Challenge 2019 [10]. In addition, these two algorithms are well-known and have been used extensively in biometric recognition (Face [36], ear [37], iris

[38], etc.). Therefore, in our experiments, we use LBP and pre-trained VGG-16 and MobileNetV2 models [32] with Siamese network. We used MobileNetV2 as a baseline because it has a better result than VGG-16 on the UERC 2019 dataset [10].

In our study, LBP is used to extract local texture features from ear images. The ear structure has unique texture patterns that can be captured effectively using LBP. These patterns are invariant to monotonic illumination changes, making LBP robust to varying lighting conditions. The ear image is divided into small regions (often called blocks). For each image block, the Local Binary Pattern (LBP) is formed by comparing the center pixel with its neighborhood: a neighbor is assigned 1 if its intensity is greater than or equal to the center, and 0 otherwise. The resulting binary sequence is interpreted as a decimal value—the block's LBP code. Aggregating the LBP codes across all blocks yields a histogram that captures the image's texture characteristics. This histogram is then used as the feature vector, summarizing local texture patterns of the ear. The LBP histograms of different ear images are compared using a similarity measure, cosine distance. This comparison determines the similarity between ear images, aiding in the verification process.

3.3.2. Ear photo recognition analysis results

Dataset: The Ear Real Photo Dataset (ERP dataset) has been collected from 137 unique participants, covering a diverse range of age groups to enhance the dataset's generalizability. For each participant, over 70 images were captured, with the ear photographed from six distinct angles: up, down, front, back, left, and right. This multi-angle capture approach is crucial for accommodating the natural variability in ear orientation, a common challenge in practical recognition scenarios. The dataset is categorized by gender and ear orientation, with specific distributions as follows:

Male Subjects: Comprising 109 participants, with a total of 6,855 images (3,482 for the left ear and 3,373 for the right ear).

Female Subjects: Comprising 28 participants, with a total of 1,848 images (864 for the left ear and 984 for the right ear).

Each gender group includes images captured from various angles, ensuring comprehensive coverage of possible ear orientations. This dataset serves as a foundational resource for testing and developing ear recognition algorithms.

In a recognition scenario, our fine-tuned DNN achieves an equal error rate of 0.12, in test set. Figure 3-3 shows false non-match error rate (FNMR) and false match error rate (FMR) for test set. Comparing FNMR/FMR curves shows differences in the used methods as a result, input size of network effect on a biometric recognition system. Based on the results in the figure, it appears that the VGG-16 feature extractor performed the best, with an EER of 0.12. The MobileNetV2 feature extractor also performed well, with an EER of 0.22. The LBP feature extractor had the highest EER of 0.42, indicating that it was the least accurate of the three methods evaluated.

It's important to note that these results are specific to the ear images database used in our study, and may not generalize to other datasets or applications. Additionally, other factors such as the size of the dataset, the quality of the images, and the specific implementation details of each method can also influence performance.

In conclusion, the results presented in Figure 3-3 suggest that using VGG-16 or MobileNetV2 feature extractors may be more effective than LBP for ear recognition tasks on this particular dataset.

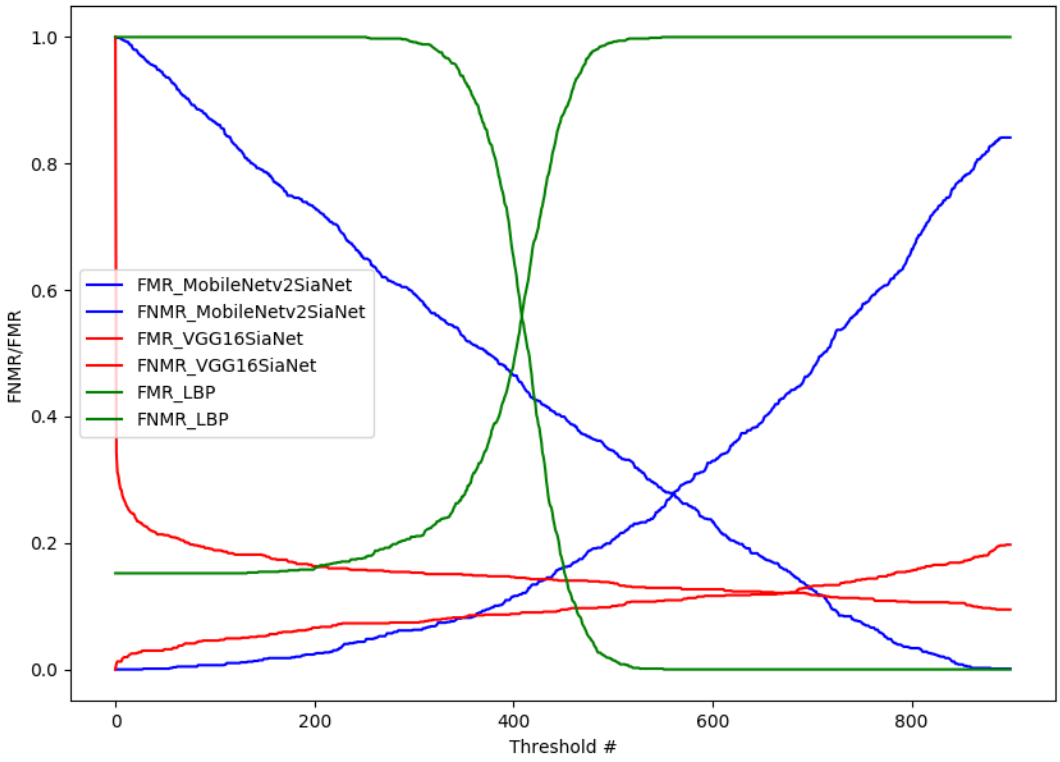


Figure 3-3: FNMR and FMR for the ear images database based on MobileNetV2 and VGG-16 feature extractors and LBP

Figure 3-4 shows the Detection Error Tradeoff (DET) curves for the ear images database based on MobileNetV2 and VGG-16 feature extractors, using a Siamese neural network. The DET curve is a plot of False Rejection Rate (FRR) against False Acceptance Rate (FAR). According to the results in the figure, the VGG-16 feature extractor achieved better performance than the MobileNetV2 feature extractor based on the DET curve. This is indicated by the fact that the DET curve for VGG-16 is lower than that of MobileNetV2 at all points, indicating a lower combined FAR and FRR.

It's important to note that these results are specific to the ear photo recognition database used in the study and the Siamese neural network architecture used. Other neural network architectures or datasets may yield different results. Additionally, as with any evaluation metric, the choice of DET curve as a measure of performance may depend on the specific requirements and goals of the application.

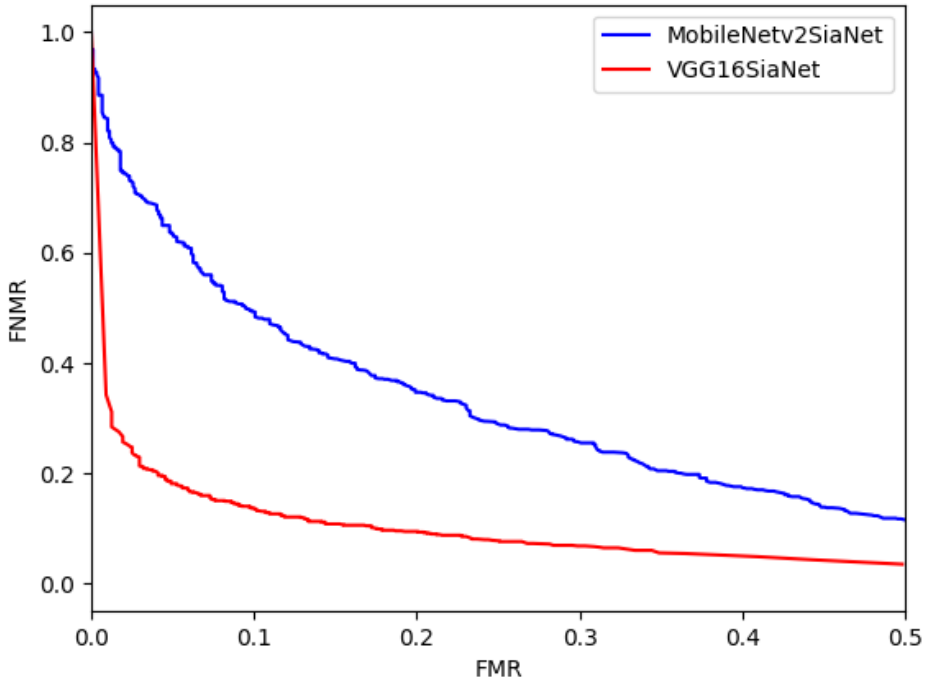


Figure 3-4: DET curve for the ear images database based on MobileNetV2 and VGG-16 feature extractor based on Siamese neural network

3.4. Ear photo PAD methodology

The goal of an ear presentation attack detection methodology (it is shown in Figure 3-5) is to develop an algorithm that can accurately detect and prevent presentation attacks, thereby improving the security of ear biometric authentication systems.

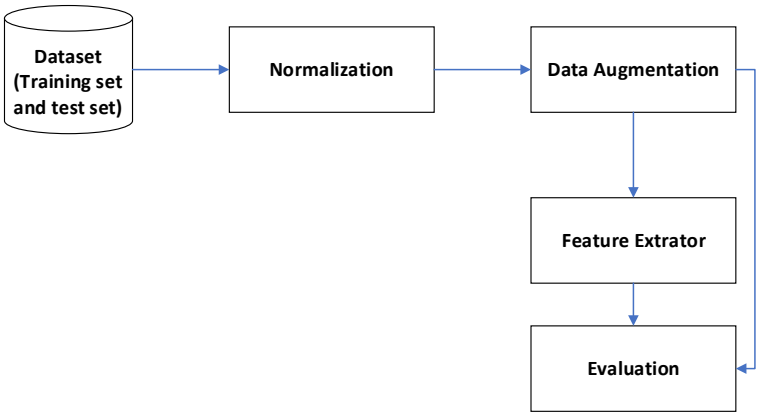


Figure 3-5: The flowchart of ear photo PAD using a deep neural network structure.

Dataset: The input data consists of a collection of ear images, including both genuine and fake (spoofed) samples. The dataset is divided into 60% for training, 20% for validation, and 20% for testing.

Normalization: In our work we use normalization with the following parameters:

- mean: Sequence of means for each channel. (R=0.4915, G= 0.4823, B=0.4468)
- std: Sequence of standard deviations for each channel. (R=0.2470, G=0.2435, B=0.2616)

Feature extraction: In this step, features are extracted from the preprocessed ear images to differentiate between genuine and fake photos.

Training data: A dataset of genuine and fake ear photos is needed to train the algorithm. This dataset is diverse and representative of the types of presentation attacks.

Testing and evaluation: The next step is to test and evaluate the performance of the algorithm on a separate dataset of genuine and fake ear images. The algorithm should be evaluated using standard metrics, such as the Attack Presentation Classification Error Rate (APCER), Bona-Fide Presentation Classification Error Rate (BPCER).

3.4.1. Feature extractor model

The proposed ear photo PAD network (PADNet) as shown in Figure 3-6 follows the MobileNetV2 [32] architecture except for the added top layers. Our PADNet model adopts MobileNetV2—pre-trained on ImageNet (1,000 classes)—as its backbone. We replace the original classification head by removing the final fully connected layer and attaching a task-

specific head for ear presentation-attack detection, then fine-tune the entire network on our collected dataset.

Backbone details: MobileNetV2 is a 2D convolutional network that accepts RGB inputs of $224 \times 224 \times 3$by default. It is built from inverted residual blocks [32] with linear bottlenecks and depthwise–separable convolutions, which substantially reduce computation while preserving representational power. In our implementation, layers use batch normalization and ReLU/ReLU6 activations; dropout can be applied in the new head to mitigate overfitting. A global average pooling layer aggregates the final feature maps, and a dense layer consumes this pooled representation to produce the prediction. For binary PAD, we employ a sigmoid output as the decision function [39].

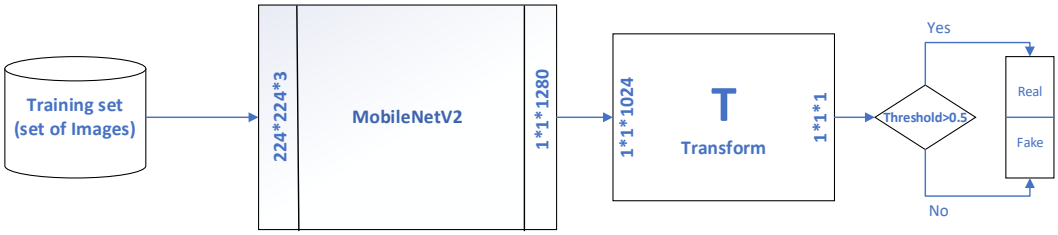


Figure 3-6: PADNet-1 architecture. The top layers of MobileNetV2 are changed to execute the binary classification for the presentation attack detection task. The layers highlighted in orange represent the added top layers.

We implement two fine-tuned variants of MobileNetV2 and refer to these modified models as PADNet; note that PADNet is not a new architecture but a MobileNetV2 backbone with a task-specific head. After the base network, we add fully connected layers to tailor the generic features to ear PAD. The head begins with two large dense layers (1024 units each) to provide sufficient capacity for modeling fine-grained patterns, followed by a reduced layer (512 units) that compresses the representation while preserving salient information. This progressive narrowing helps regularize the model and can improve generalization, particularly when the number of target classes is small. Overall, the added dense layers specialize the pretrained features of MobileNetV2 to our dataset and task.

In PADNet-1 we froze 26 layers and added three dense layers which are shown in Table 3-1. Then flattening layer, two dense layers with 1024, one dense layer with 512 and final output layer is with two classes with sigmoid activation.

Table 3-1: Transfer learning parameters on PADNet-1

Type/Stride	Layer Statues or size of layer	Activation Function
1-26 Layers	Frozen	-
27th Layer	Trainable	RELU
Dense 1	1024	RELU
Dense 2	1024	RELU
Dense 3	512	RELU
Output	1	Sigmoid

We split our database into two classes; real and fake photos. As another experiment, we changed the frozen layers and parameters to get better results in the PADNet and called PADNet-2. The details of PADNet-2 are shown in Table 3-2.

Table 3-2: Transfer learning parameters on PADnet-2

Type/Stride	Layer Statues or size of layer	Activation Function
1-16 Layers	Frozen	-
17-27 Layers	Trainable	RELU
Dense 1	1024	RELU
Dense 2	1024	RELU
Dense 3	512	RELU
Output	1	Sigmoid

3.4.2. Ear photo PAD analysis

In this section, anti-spoofing methods have been evaluated for ear biometric. Five different attack scenarios have been made. They were supplied to a DNN to train a classifier for distinguishing between fake and real ear biometric. With the proposed algorithm the accuracy of classification has been determined. All fake images have been considered as a class called fake class, however, in the test part, they are evaluated separately.

Evaluation metrics: In the ISO/IEC PAD framework, the verification terms False Match Rate (FMR) and False Non-Match Rate (FNMR) are replaced by Attack Presentation Classification Error Rate (APCER) and Bona fide Presentation Classification Error Rate (BPCER), respectively. Impostor Attack Presentation Match Rate (IAPMR) denotes the rate at which an impostor presentation attack is (incorrectly) accepted—conceptually inherited from FMR in

the spoof-detection setting. In consequence of testing PAD systems, APCER and BPCER are defined as;

$$APCER = \frac{1}{N_{PAIS}} \sum_{i=1}^{N_{PAIS}} (1 - RES_i) \quad (25)[41]$$

Where, N_{PAIS} is a quantity of attack presentations and RES_i shows number of success presentation attack it means:

$$\begin{cases} \text{if} \quad RES_i = 1 \quad \text{attack presentation} \\ \text{if} \quad RES_i = 0 \quad \text{bona fide presentation} \end{cases}$$

And for BPCER;

$$BPCER = \frac{\sum_{i=1}^{N_{BF}} (1 - RES_i)}{N_{BF}} \quad (26)[41]$$

N_{BF} is the number of Bona fide presentation. In addition, half total error rate (HTER), which are utilized for evaluating biometrics systems [42]. Since the deep neural network outputs are probabilities, the threshold for classifying the output is set at $\tau = 0.5$, assuming a symmetrical distribution of the probabilities around this midpoint. This assumption of symmetry implies that the model's predictions are equally likely to be above or below the threshold when classes are balanced, making 0.5 a natural choice for distinguishing between class 0 and class 1.

In this chapter, all presentation attack detection standards and metrics follow ISO/IEC 30107-3 standards. In PADNet-1 we use Adam optimizer, input size 224*224*3 pixels, batch size 32, and 50 epoch iterations. Furthermore, in PADNet-2 we use a slightly different style of parameterization. We froze first 16 layers, input size 224*224*3 pixels, batch size 64, and Stochastic Gradient Descent (SGD) (with learning rate 0.0001 and momentum 0.9). The loss values for PADNet-1 and PADNet-2 are 0.0040 and 8.7142e-04 respectively.

The BPCER (Bona Fide Presentation Classification Error Rate) for both models on bona fide presentations is also 1%. This BPCER can be attributed to the limited variability in the training dataset, which might not fully reflect the diversity of real-world conditions. In controlled training environments, where the data may not capture the full range of potential variations and challenges present in actual usage, models can achieve near-perfect performance. This can occur even when the dataset is limited in size or diversity, leading to high accuracy on the training set. However, it should be noted that such results may not necessarily translate to equally high performance in real-world scenarios.

Table 3-3 displays the results of an investigation into the performance of two different PAD models, PADNet-1 and PADNet-2, on various attacks using different display and recognition devices. The APCER (attack presentation classification error rate) is reported for each combination of display and recognition device, abbreviation, detection methods, and model.

Display device: This column lists the name of the display device used in the experiment, such as Dell Ultra Sharp 32 Ultra HD 4K Monitor or SAMSUNG C27JG50QQUX monitor.

Recognition device: This column lists the name of the recognition device used in the experiment, such as Samsung Galaxy A7, Samsung Galaxy S9, or Nokia Lumia 1020.

Abbreviation: This column provides an abbreviation for the combination of display and recognition devices, such as Dell-GA7, S3D-GS9, or Print-GA7.

Detection methods: This column specifies which detection method was used in the experiment, such as PADNet-1 or PADNet-2.

APCER (%): This column lists the APCER, which is the percentage of presentation attacks that were incorrectly classified as genuine by the PAD model. A lower value PAD indicates higher security vulnerability, as they indicate that the model is more accurate at detecting presentation attacks.

Table 3-3: The APCER for both models PADNet-1 and PADNet-2 on various attacks

Display device	Recognition device	Abbreviation of Display and Recognition device	APCER (%)	
			PADNet-1	PADNet-2
Dell Ultra Sharp Monitor	Samsung Galaxy A7	Dell-GA7	23.26	24.16
	Samsung Galaxy S9	Dell-GS9	17.78	15.11
	Nokia Lumia 1020	Dell-NL1020	0.61	0.79
SAMSUNG monitor	Samsung Galaxy A7	S3D-GA7	**0.17**	0.34
	Samsung Galaxy S9	S3D-GS9	0.52	8.33
	Nokia Lumia 1020	S3D-NL1020	1.8	1.8
Brother printer	Samsung Galaxy A7	Print-GA7	2.43	3.03

The results in Table 3-3 provide insights into the performance of two different PAD models, PADNet-1 and PADNet-2, on various attacks using different display and recognition devices.

The APCER values for both PADNet-1 and PADNet-2 are generally lower for the Nokia Lumia 1020 recognition device than for the Samsung Galaxy A7 or Samsung Galaxy S9. This suggests that the Nokia Lumia 1020 may be easier to detect presentation attacks on than the Samsung devices. The Dell Ultra Sharp 32 Ultra HD 4K Monitor appears to be more challenging to detect presentation attacks on than the SAMSUNG C27JG50QQUX monitor. The APCER values for the SAMSUNG monitor are consistently lower than those for the Dell monitor.

In general, PADNet-1 appears to perform slightly better than PADNet-2, as indicated by the lower APCER values for most of the device combinations. However, there are some cases where PADNet-2 performs better, such as with the Samsung Galaxy S9 on the Dell monitor. The APCER values for some of the device combinations are quite high, indicating that the PAD models may not be effective at detecting certain types of presentation attacks. For example, the APCER values for the Nokia Lumia 1020 on the S3D-GA7 combination are ~24 % for both PADNet-1 and PADNet-2, suggesting that these models are not effective at detecting 3D-printed presentation attacks on this device. The APCER values are generally higher for the Print-GA7 combination than for the other combinations, indicating that the printer-based presentation attacks may be more difficult to detect than attacks using other display devices.

These results suggest that the performance of presentation attack detection models can vary depending on the specific combination of display and recognition devices used, as well as the type of attack being attempted. It is important to thoroughly evaluate the performance of these models under a wide range of conditions to ensure their effectiveness in detecting presentation attacks in real-world scenarios.

Table 3-4 shows the Half Total Error Rate (HTER) results of the two proposed PADNet models (PADNet-1 and PADNet-2) on different display and recognition devices for various presentation attacks. HTER is a widely used metric for evaluating the performance of presentation attack detection systems, which is defined as the average of the false acceptance rate (FAR) and false rejection rate (FRR).

In Table 3-4, each row represents a combination of display and recognition devices used in the experiment, with a corresponding abbreviation. The "Detection Methods" column shows which PADNet model was used for the experiment, either PADNet-1 or PADNet-2. The "HTER (%)" column displays the HTER values obtained by each model on the different attacks.

Both models achieved low HTER values, indicating their effectiveness in detecting presentation attacks. The results vary depending on the type of attack and the combination of display and recognition devices. For example, for the Dell Ultra Sharp 32 Ultra HD 4K Monitor and Samsung Galaxy A7 combination, both PADNet-1 and PADNet-2 achieved HTER values around 12%, which is relatively high compared to other combinations. However, for the Nokia Lumia 1020 and Dell-NL1020 combination, both models achieved very low HTER values of 0.30% and 0.39%, respectively. The SAMSUNG C27JG50QQUX monitor and Samsung Galaxy A7 combination achieved the lowest HTER value of 0.08% for PADNet-1 and 0.17% for PADNet-2, indicating their high performance in detecting presentation attacks.

In addition, we have not used Print-SGA7 attacks in the PADNet-1 and PADNet-2 training stage. Despite the fact that Print-SGA7 was not used in the training, the result of presentation attack detection was 2.43% which is acceptable in comparison with other attack presentations results in this chapter.

Table 3-4: The HTER for both models PADNet-1 and PADNet-2 on various attacks

Display device	Recognition device	Abbreviation of Display and Recognition device	HTER (%)	
			PADNet-1	PADNet-2
Dell Ultra Sharp Monitor	Samsung Galaxy A7	Dell-GA7	11.63	12.08
	Samsung Galaxy S9	Dell-GS9	8.84	12.555
	Nokia Lumia 1020	Dell-NL1020	0.30	0.39
SAMSUNG monitor	Samsung Galaxy A7	S3D-GA7	0.08	0.17
	Samsung Galaxy S9	S3D-GS9	2.76	4.11
	Nokia Lumia 1020	S3D-NL1020	0.9	0.9
Brother printer	Samsung Galaxy A7	Print-GA7	1.21	1.51

3.5. Conclusions

In this chapter, ear recognition and ear PAD were explored by considering mobile devices application and biometric multimodality. The proposed system uses a smartphone's camera to capture ear shape. Subsequently, to make fake or unreal data for attack purpose, different kinds of material has been used. Different presentation attacks are made for ear PAD. According to the achieved APCER values for different presentation attacks, using different devices with variable quality achieves different results. Fine-tuned deep neural networks achieved very good results (in the best condition 1%) on just some of the attacks.

4. Ear-touch Based Mobile User Authentication

4.1. Introduction

As mobile devices continue to grow, the need for reliable and secure authentication and access control becomes increasingly important. Biometrics such as fingerprints, iris patterns, faces, voices, and ears have been explored as possible solutions. However, incorporating additional sensors can increase the cost of mobile devices, making them less accessible to users.

Ear-touch refers to a method where the distinct touch patterns and contours of the ear are captured by several touches on a smartphone screen in our case maximum 10 touch points, focusing on specific points rather than a detailed image. In contrast, ear-print, often implemented in modified Android kernels, captures touch data at a fixed resolution, such as 12x27 pixels, to create a detailed representation of the ear's contours through multiple touch points. While earprints are typically captured using specialized hardware or modifications to mobile devices, a similar biometric feature known as ear-touch can be captured easily using a smartphone's touchscreen. Ear-touch has lower image quality compared to earprints but can be captured using a normal smartphone with multi-touch ability, making it a cost-effective and accessible biometric measure.

An ear-touch is new biometric feature being introduced in this research. To capture an ear-touch, a smartphone’s multi-touch screen is used. As a result, ear-touches could be used for authentication on mobile devices in the same way as fingerprints, facial images, iris prints etc. Ear-touches have not been used yet as a biometric measure in any research, though similar research, called earprint recognition, has been carried out on mobile devices. In this research, we propose a method for an ear-touch recognition system on mobile devices. With ear-touches we have other problem called ‘missing points’. It occurs due to the physical features of ears and the way that people pressing smartphones to their ears. Each captured ear-touch could have either the same number of touched points or a different number of touched points. In this research, we also take the missing points into account.

Earprints have been extensively used in forensic research, as they provide unique identifying information for individuals [43-45]. Recent studies have demonstrated that the shape and geometry of an ear can be used to identify individuals with a high degree of accuracy, even between twins [46].

This chapter explores the use of ear-touch for mobile user authentication and access control. We present a method for capturing ear-touch data and address the challenge of missing data points. Our method achieves high performance in terms of equal error rate.

The rest of this chapter is structured as follows: Section 4.2 reviews related work; Section 4.3 presents the methodology and results; Section 4.4 provides a discussion; and Section 4.5 concludes with a summary of key findings.

4.2. Literature solutions

Touchscreens have been used to capture biometric data already, but by making changes inside the smartphones. In [47], the authors used touchscreen to capture fingers, fists, ears, and palms through their system called Bodyprint. The capacitive touchscreen sensor was used to capture biometrics from 12 users. The touchscreen had an input resolution of ~6 dpi and the sensor was able to capture an image with 27*15 pixels and 8-bit size. They used SURF descriptors for feature extraction. For matching features, the L2 distance was used, based on extracted 12 key frames by SURF descriptors. The performance of the method was tested on 12 participants in 12-fold cross validation. Overall, they had 864 samples. Despite the fact that the false rejection rate was 26.8% for all biometrics (fingers, fists, ears, phalanges and palms) in this evaluation. For just ears, the false rejection rate was reported as 7.8%.

In [48], they manipulated touchscreen of a smartphone to capture all touch points on the capacitive screen. The resolution of the touchscreen was 6 dpi or 27*15 pixels. The database used contained 1520 images, collected from 37 subjects, within which 40 images were taken from each subject. Recall and precision were reported as performance evaluation metrics and were 0.5960 and 0.8761, respectively.

In [49], the use of earprints in the forensic field, the stability and variability of ears for earprints are reviewed. They showed the substantial features in earprints that can be used in forensic identification.

In [50], a FearID earprint identification system was proposed, utilizing a dataset of 7364 earprints collected from 1229 participants. For feature extraction, the system employed three

methods. The first was the weighted width comparison method, which identifies connected structures representing the imprints and weights the corresponding intensities to extract local intensity signals. The second was vector template matching (VTM), a method based on the anatomical annotation of earprints. Each print is annotated to generate a template of labeled points representing landmarks and minutiae, categorized into different classes, and comparisons are made by assessing the similarity between templates. The third method was the angular comparison method, which compares signals by tracking the angle of the medial axes of connected structures relative to the x-axis. Logistic regression was used for classification, with the data split into training and testing sets. The system reported an Equal Error Rate (EER) of 9.3% on the test set.

In [51], The authors proposed a hybrid method combining global and local features for earprint feature extraction. To identify global features, binary images were used, and comparisons were made between the model and the query earprint. For local features, the method employed several techniques. The scale-space extrema detection technique involved calculating the Difference of Gaussian(DoG) function to identify regions invariant to scale and rotation. The keypoint localization method evaluated local maxima and minima of the DoG by comparing each point to its 16 neighbors, selecting candidate points larger or smaller than all their neighbors, and rejecting those with the lowest contrast after detailed fitting for location, scale, and curvature ratios. The orientation assignment method represented each keypoint by 16 orientations derived from local image gradient directions. The proposed method was tested on the FearID database and achieved an Equal Error Rate (EER) of 1.87%.

To recognize the extracted ear-touch via touchscreen on mobile devices, it is necessary to know how to matching two-dimensional finite set points as a result. There are various algorithms that have been used to match two-dimensional finite set points.

In [52], to find exact patterns of points to match two sets, they found the centre of each set and then calculated polar coordinates based on the centroids. In another paper [53], the authors used the one-to-one matching method, although this method needs exactly the same number of points for two set points.

The studies reviewed suggest that ear-touch based mobile user authentication has potential for practical application in mobile device security.

4.3. Ear touch methodology

The ear touch verification system is shown in Figure 4-1. The input – a number of touch points with (x, y) coordinates, is presented. Ear touches information would be features for the next step in our ear-touch verification procedure. The similarity function is calculated based on F_p (presented input) and F_t(the stored featured of the ear-touch in the server or database). These scores are compared and if Th (threshold) is smaller than $S_{p\&t}$ then it would be accepted.

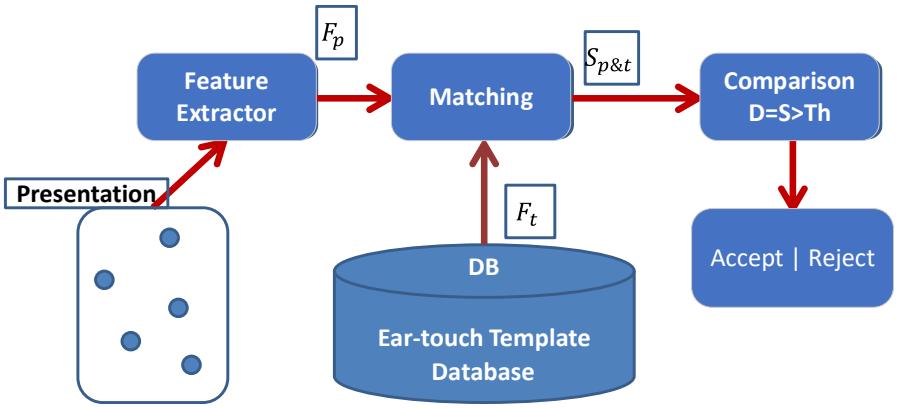


Figure 4-1: Ear-touch recognition procedure: presentation – the captured ear touches from ears; F_p – extracted features after cleaning the data; F_t – the stored ear-touch or ear-touches in the database; $S_{p\&t}$ – similarity values after matching the system.

Missing data problem:

In ear-touch data we are faced with the problem of missing points. This means that the touches do not have a fixed number of points in each presentation. Normally, during the data acquisition, the distances between points are not close.

Figure 4-2 shows an example of missing points in various touches. It should be noted that, in order to find the best matching algorithm, we must consider the problem of missing points, because even if we solve all translation and revolution problems, the missing point's problem can still lead to significant errors in the matching.

- **First problem:** when looking at Ear-touch 1 and Ear-print 2 in Figure 2-6, they have the same number of points, they are from the same ear, but because those points are in a different region, our matching algorithm does not recognise it as the same ear.
- **Second problem:** If we consider Ear- touch 1 and Ear- touch 3, there are a different number of points. So, in the Ear- touch 3 there are points that do not appear in Ear-touch 1.

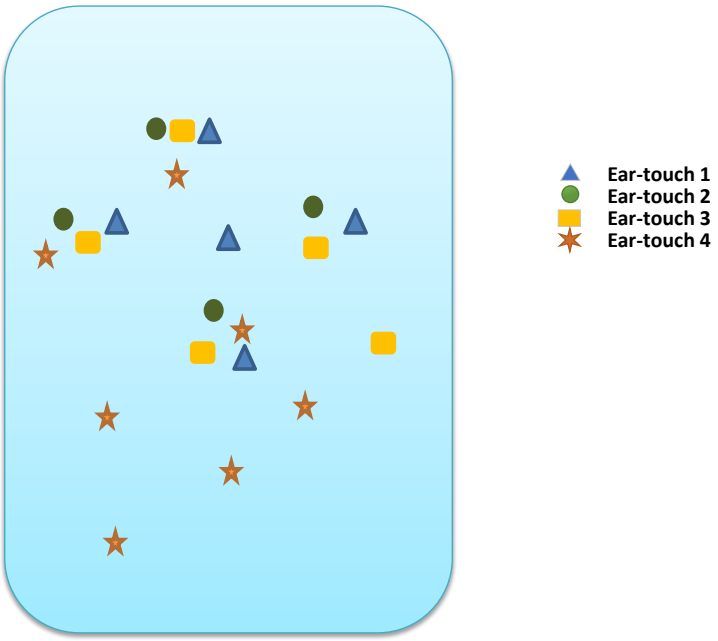


Figure 4-2: Ear-touches that have a different number of touch points and are in a different region in some cases. These samples are shown in 2D coordinates.

Alignment of the ear-touches: The solution for alignment of the ear-touches has two major parts: "matching" between given ear-touches and a given template, used for authentication, and "template creation/extraction" based on a set of related ear-touches (e.g. ear-touches of some known subject), which can be used as a reference for authentication purposes. These are two basic tasks in our proposed method and there are several challenges to performing these tasks.

The first challenge is in connection with the "missing points". In each experiment, the touchscreen will return up to eight touch points, based on the ear contact position with respect to the device coordinate system. There is no guarantee that all the points will be measured in each experiment, and each experiment may contain some missing points.

The second challenge concerns "permutations". For each ear-touch, the sensor returns a "set" of points. They have no consistent or meaningful anatomical order. So, when matching between two sets, the algorithm should pair points one by one. In fact, it considers different permutations between these two sets to find the most meaningful pairing.

The third challenge is to do with "rotations and translations". The touchscreen measures the touch points with respect to its own coordinate system, not fixed coordinate based on the subject's ear. So, even when comparing two ear-touches of the same ear, the algorithm should consider random rotations and translations that may have occurred during the different data acquisition phase.

To resolve these challenges, we used an optimization-based approach that tries to find the best matches by minimizing some relevant loss functions. To explain this approach, we first consider a simplified scenario, in which there are no missing points, before moving on to the more challenging scenarios, which are more suitable for real world applications.

4.3.1. Biometric problem: A matching scenario without missing points

In an ideal world, with no missing points and permutations, each ear-touch would be represented by a sequence of touch points of fixed length. In this scenario, comparing two ear-touches is as simple as measuring the distance between all corresponding 2D points, if a fixed coordinate system exists. Zero distance means all the corresponding points match each other, meaning that the two ear-touches match, whereas any non-zero distance mean they do not match. Such a distance measure can be defined by Eq.1, in which N is the number of touch points for each ear-touch, $\|.\|$ represents the Euclidean norm, $T = (t_1, t_2, \dots, t_N)$ is the template or reference ear-touch, $X = (x_1, x_2, \dots, x_N)$ is the given ear-touch we want to authenticate, and t_i and x_i represent the i-th touch points of the template and the given ear-touch respectively.

$$D_0(T, X) = \sum_{i=1}^{N} \|t_i - x_i\|^2 \tag{1}$$

In a more sophisticated scenario, when rotations, translations, and permutations are present, we have to consider all possible modes of these transformations. So, Eq.1 is modified into Eq.2, containing an optimization problem. In Eq.2, $\mathcal{R}$ is the set of 2D rotation matrices, $\mathcal{L}$ is the set of all possible translations presented by 2D-vectors, and $\mathcal{P}$ is the set of all possible permutations between N points. It is assumed that each permutation is presented in$(\pi_1, \pi_2, \dots, \pi_N)$.

$$D_1(T, X) = \min_{(R,l,P)\in\mathcal{R}\times\mathcal{L}\times\mathcal{P}} \sum_{i=1}^{N} \left\|t_i - (Rx_{\pi_i} + l)\right\|^2 \tag{2}[54]$$

In fact, D_1 is measuring the distance between T and some transformed version of X using the D_0 distance we defined in Eq.1. We call this transformed version as the best matching form of X w.r.t. to T, which will be used to create a template later. This best matching form or "best match" can be represented by Eq.3, while its parameters defined in Eq.4.

$$b_i = R^* x^*_{\pi_i} + l^* \qquad i = 1,2, \dots, N$$
$$Best(T, X) \stackrel{\text{def}}{=} (b_1, b_2, \dots, b_n) \tag{3}$$

$$(R^*, l^*, P^*) = \underset{(R,l,P)\in\mathcal{R}\times\mathcal{L}\times\mathcal{P}}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| t_i - (Rx_{\pi_i} + l) \right\|^2 \quad (4)$$

Pacut [54] resolved the optimization problem of Eq.4 based on the Procrustes problem and the Kabsch-Umeyama algorithm. His solution is straight-forward and efficient, without requiring common iterative schemes of general optimization algorithms.

However, his proposed method does not consider permutations and thus it has to be extended in order to solve the simplified scenario is detailed in the pseudo code provided in Appendix B, Section B.1. It returns the mismatch based on Eq.2 and the best match based on Eq.3, without considering any missing points. For aligning ear-touch samples without missing points it operates under the constraint that each ear-touch must have between 4 and 10 touch points. This constraint ensures a manageable number of permutations, as exceeding 10 touch points would result in a computationally prohibitive number of permutations, making the alignment process infeasible.

The algorithm begins by initializing the mismatch to infinity and setting the best match B to X. For each possible permutation P of the touch points in X, the algorithm permutes according to P and saves the result as S. Using Pacut's method, it then solves for (R^*, l^*) that minimizes the sum of squared differences between the template touch points t_i and the transformed touch points in S, specifically $Rx_{\pi_i} + l_i$ where R is a rotation matrix, l is a translation vector, and π_i denotes the permutation of indices.

For each permutation, the algorithm computes the auxiliary best match B_{aux}, which consists of the transformed points $R^* x^*_{\pi_i} + l_i$. It then calculates min_{aux}, the distance $D_0(T, B_{aux})$, representing the mismatch for the current permutation. If this mismatch min_{aux} is smaller than the current best mismatch, the algorithm updates the mismatch to min_{aux} and set B to B_{aux}. This process continues until all permutations have been evaluated. The algorithm then returns the smallest mismatch found and the corresponding best match B. The constraint of having a maximum of 10 touch points ensures that the number of permutations remains computationally feasible, as the factorial growth of permutations would otherwise lead to excessive computational demands.

4.3.2. Template creation for no missing points

Creating a reference template from a group of known ear-touches was another part of our problem. So, we need an algorithm to perform this task as well. In the absence of missing points, every known ear-touch has the potential to be used as a template. However, in practice it is preferable to construct or extract the template from a bunch of related or known ear-touches. This template can be found by the Eq.5, where $\mathcal{E}$ is the set of all possible templates or ear-touches, T is the unknown template we are looking for, $Y = (Y^1, Y^2, \dots, Y^M)$ is the set of input ear-touches and Q is some mismatch measure, which can be devised as Eq.2.

$$T = \underset{T \in \mathcal{E}}{\operatorname{argmin}} \frac{1}{M} \sum_{j=1}^{M} Q(T,\ Y^j) \qquad (5)[54]$$

Pacut also solved the optimization problem of Eq.5, although it does not consider permutations [54]. It can simply be extended as Pseudo-code 2.

The main idea of this method is the notion of "Best Match" introduced earlier. It is based on the properties of 2-Norm or Euclidean Norm which already used in Eq.1 and Eq.2. To create a template from a series of ear-touches that minimizes the average mismatch according to a specified equation (Eq. 5). The algorithm, is detailed in the pseudo code provided in Appendix B, Section B.2, for template creation begins with a series of M related ear-touches, denoted as $Y = (Y^1, Y^2, \dots, Y^M)$, as its input. The goal of the algorithm is to produce a template that minimizes the average mismatch to the ear-touch samples according to a specified equation (Eq.5). Initially, the iteration counter k is set to 0, and the initial value of L, denoted as L_0 is set to infinity. The template T is initialized with the first ear-touch sample Y^1, so $T_0 = Y^1$.

The algorithm then enters a while loop, which continues to execute as long as either it is the first iteration ($k = 0$) or the difference between the previous and the current L values ($L_{k-1} - L_k$) is greater than a specified tolerance (tol). Within the loop, the iteration counter k is incremented by 1. For each ear-touch sample Y^j where $j = 1,2, \dots, M$, the algorithm computes the best match B_k^j to the current template T_{k-1} according to Pseudo code 1 (Appendix B, Section B.1). Next, the algorithm updates the template T. For each element i in the template, the new template value $t_{i,k}$ is calculated as the average of the corresponding best match values $b_{i,k}^j$ over all samples Y^j. Specifically, $t_{i,k}$ is the sum of $b_{i,k}^j$ from $j = 1$ to M, divided by M. After updating the template, the algorithm calculates the new L value L_k, which represents the

average distance D_0 between the current template T_k and each of the best matches b_k^j. The average is taken over all M samples. The while loop continues to iterate until the difference between successive L values is less than the specified tolerance. Once the loop exits, the algorithm returns the final template T_k, which is the template that minimizes the average mismatch to the series of ear-touches.

Numeric example:

Consider a scenario where a user is trying to authenticate using ear-touch on a mobile device. Let's say that the set of points on the touchscreen represents the user's unique ear-touches as shown in Figure 4-3.

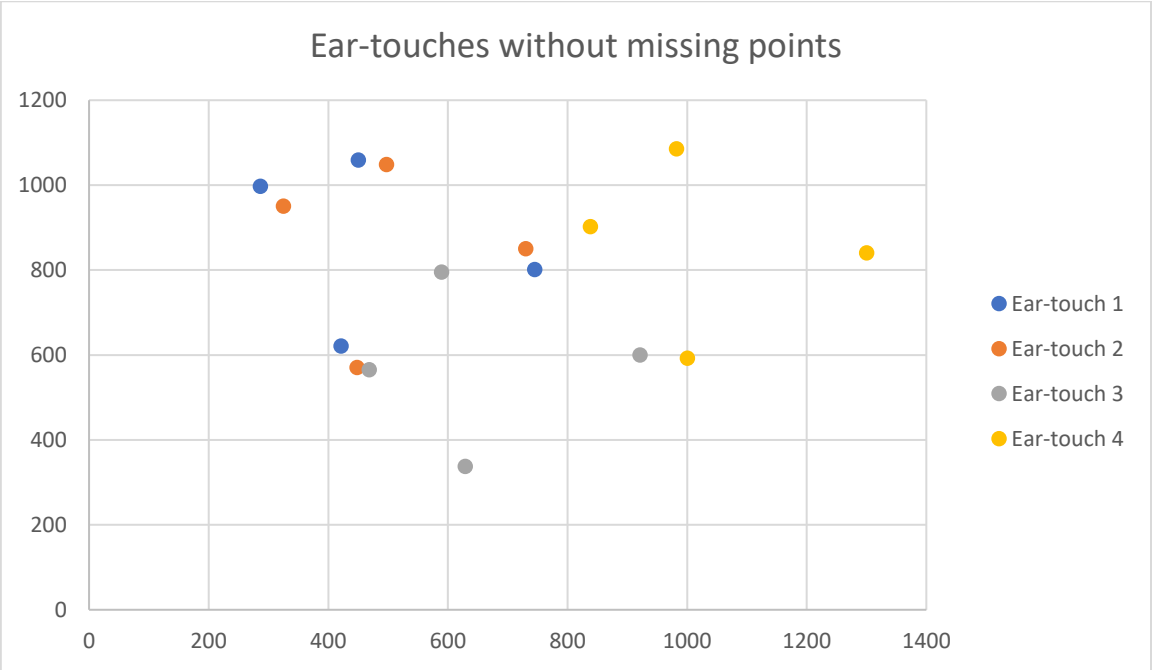


Figure 4-3: Four ear-touches without missing points; the samples are taken from a participants

Based on these samples let's make a template using three of the ear-touches for analyzing pseudo code 1(Appendix B, Section B.1). So, based on pseudo code 2 (Appendix B, Section B.2) the template creation would be as shown in Figure 4-4.

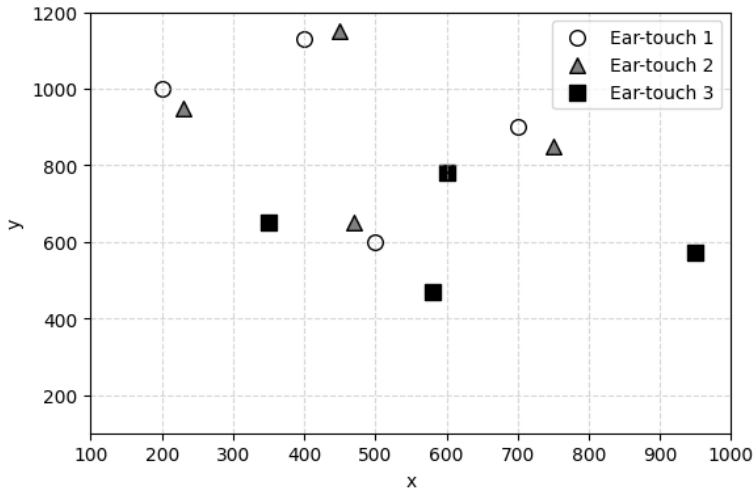


Figure 4-4: Three ear-touches without missing points are chosen to create template

Ear-touch 1 is considered as $T_0 = ear - touch\ 1$ the rest are applied to find the best match. We move T_0 to the origin then we align the first ear-touch to x-axis. So, the result would be shown in Figure 4-5:

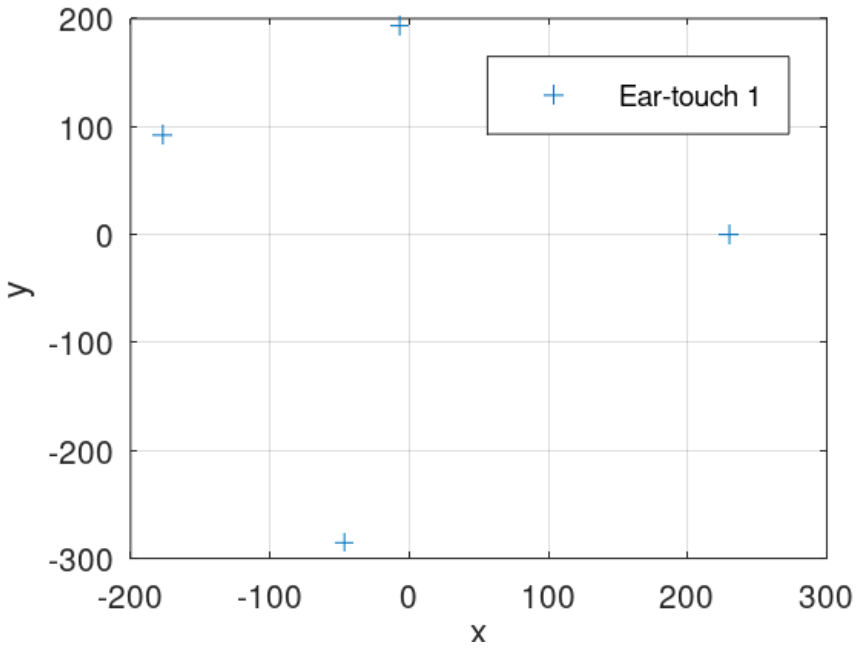


Figure 4-5: Ear-touch 1 is moved to origin and considered as T0 to find the best match

The created template would be shown in Figure 4-6:

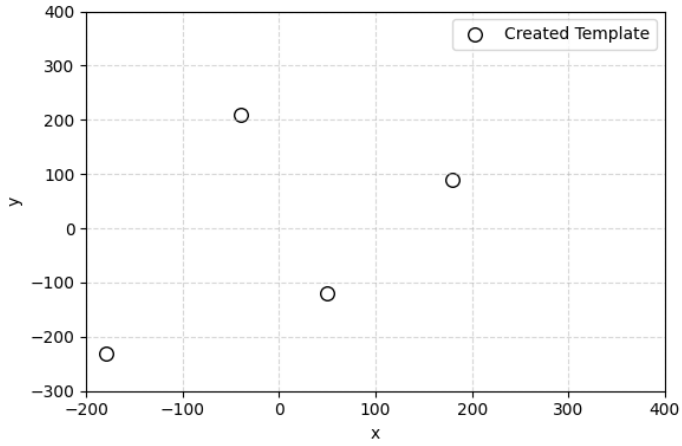


Figure 4-6: Created template using Pseudo code 2

Applying Pseudo code 1 on ear-touch 4 and getting result for that would have the results on Figure 4-7. After calculation of distance between template and test we get 25070. This result is achieved for the identical ear-touch.

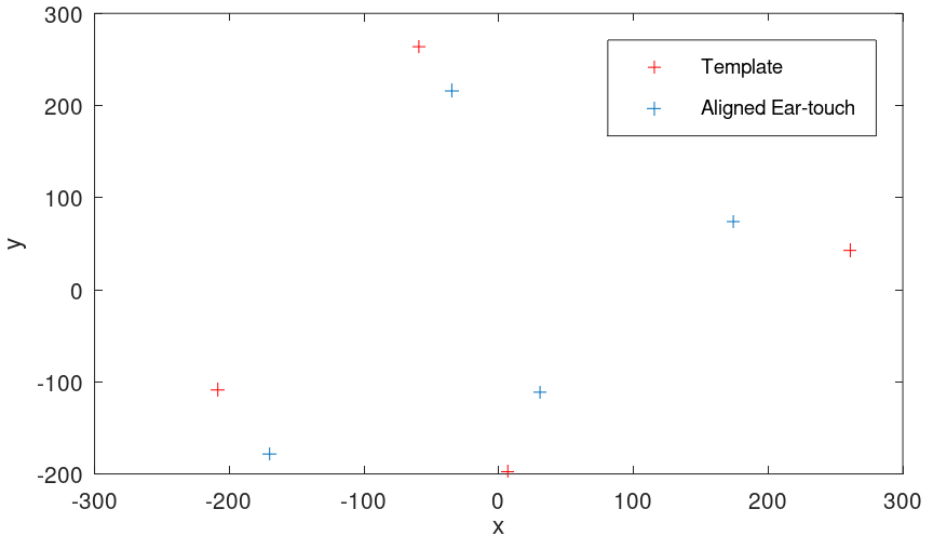


Figure 4-7: Ear-touch 4 is being considered as test sample and result after aligning

Let's consider a dissimilar ear-touch to test how to algorithm works. Figure (3-13) shows an example of ear-touch which is not identical with template that we have for this scenario.

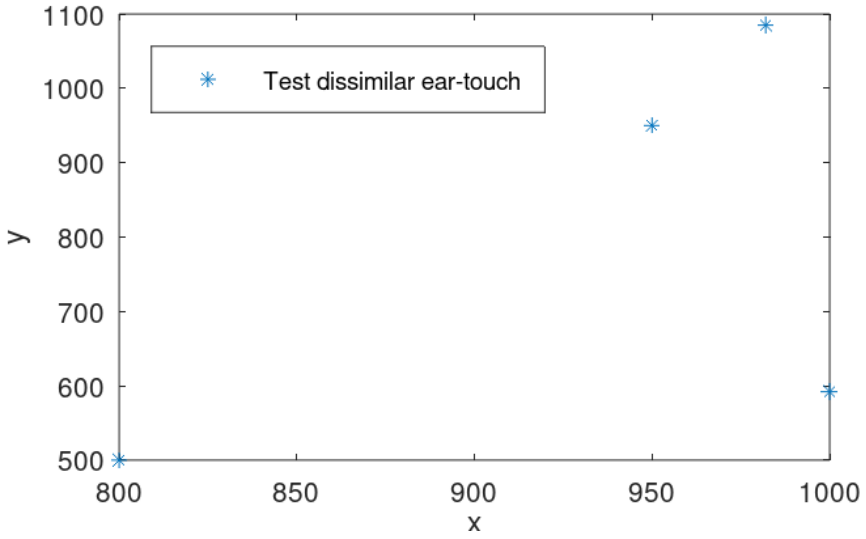


Figure 4-8: Test dissimilar ear-touch for testing if the distance is far from created template or not

If we consider the shown ear-touch in Figure (3-14), we will have distance equal 42489. We see that when we have different ear-touch that template we will have worse results.

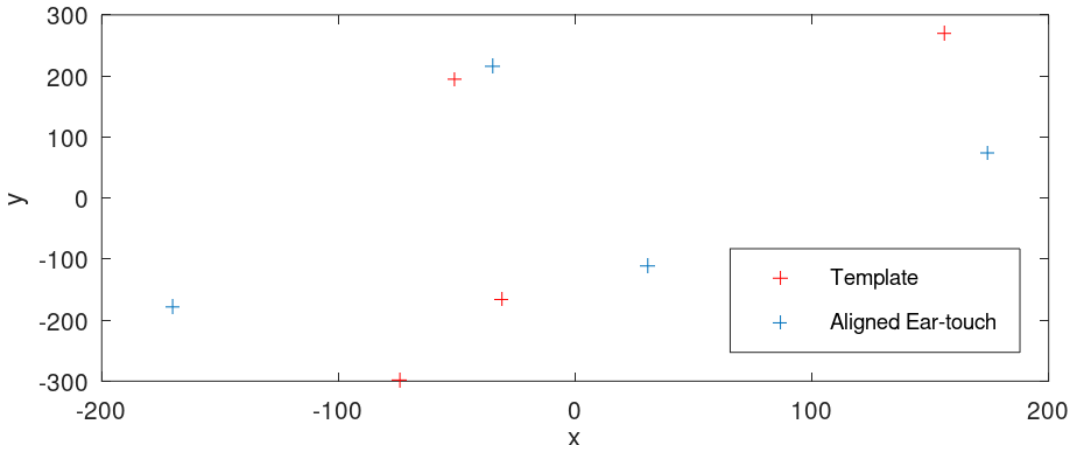


Figure 4-9: Aligned ear-touch and template; the aligned ear-touch is dissimilar ear-touch

4.3.3. Creating template in presence of missing points

In pseudo-code 2 (Appendix B, Section B.2), the template is constructed by averaging related ear touches on all Best Matches on the input ear-touches. Those Best Matches are dependent on the template itself, thus an iterative scheme is employed to reach the fixed point of equation, which is the optima of Eq.5. Now to create a template in presence of missing points we will do as described in Appendix B, Section B.3.

The algorithm begins by sorting the inputs Y and E based on the number of existing touch points in descending order. This sorted data is then stored in the original variables. The initial values for the iteration counter k, the loss function L_0, and the template T_0 are set. Specifically, k is initialized to 0, L_0 is set to infinity, and the initial template T_0 is assigned the value of the first incomplete ear-touch $Y_1{}'$. The initial existence sequence E_0^T is set to $E_1{}'$.

The algorithm then enters a while loop, which continues to execute as long as it is the first iteration ($k = 0$) or the difference between the previous and the current loss function values ($L_{k-1} - L_k$) is greater than a specified tolerance (tol). Within the loop, the iteration counter k is incremented by 1. For each incomplete ear-touch $Y_j{}'$ and its corresponding existence sequence $E_j{}'$, the algorithm calculates the best match B_k^j and the updated existence sequence E_k^j with respect to the current template T_{k-1} and the initial existence sequence E_0^T .

$$n_i = \sum_{j=1}^{M} e_{i,k}^j \quad t_i^k = \begin{cases} \frac{1}{n_i} \sum_{j=1}^{M} b_{i,k}^j & n_i > 0 \\ 0 & n_i = 0 \end{cases} \quad i = 1, 2, \dots, N \qquad (6)$$

Next, the algorithm updates the template T. For each element i in the template, it calculates the sum of the corresponding existence sequence values $e_{i,k}^j$ over all M samples. The new template value t_i^k is then determined by averaging the corresponding best match values $b_{i,k}^j$ for all samples where the existence sequence value n_i is greater than 0. If n_i is 0, the template value t_i^k is set to 0. This results in a new template $T_k = (t_{1,k}, t_{2,k}, \dots, t_{N,k})$.

After updating the template, the algorithm calculates the new loss function value L_k, which is the average distance D_0 between the current template T_k and each of the best matches B_k^j over all M samples. The while loop continues to iterate until the difference between successive loss function values is less than the specified tolerance.

$$L_k = \frac{1}{M} \sum_{j=1}^{M} D_0(T_k, B_k^j) \qquad (7)$$

Once the loop exits, the algorithm returns the final template T_k and the initial existence sequence E_0^T. This approach ensures that the template is constructed in a way that accounts for

the incomplete nature of the input ear-touches, providing a robust solution for biometric template creation.

4.3.4. Biometric problem: Matching in the presence of missing points

If the template (T') is incomplete, some touch points of X' may lose their equivalents in T'. So, it is important to know which ear touches are common in both ear-touches. If total number of touch points is denoted by N, T' and X' have $N_{T'}$ and $N_{X'}$ touch points respectively, and $N_{X'} < N_{T'}$ then number of common touch points (N_c) obeys the inequality: $N_{T'} + N_{X'} - N < N_c < N_{X'}$. So, our algorithm should search all these possible modes of this interval, which is described by Pseudo-code in Appendix B, Section B.5.

In practice, both template (T) and imposter (X) are variable length sequences of touch points, so the one-to-one correspondence between their touch points may not apply. This means that the previously mentioned algorithm for matching requires some modifications to overcome this challenge. First, a simpler scenario is assumed, in which the template is assumed to be complete, and then it is extended to the case where the template itself is an incomplete touch points.

If the template is complete, then there must be an injection from a set of imposter ear touches to the set of template ear touches. We represent this injective function by a permutation. If we assume the imposter (X') has $N_{X'}$ touch points and the template (T) has N_T touch points, then each injection can be shown by $N_{X'}$-permutation of N, i.e. as a n-tuple $(\pi_1, \pi_2 \ldots, \pi_{N_{X'}})$, where π_is are distinct and $1 \leq \pi_i \leq N_{X'},\ i = 1,2, \ldots, N_{X'}$. The prime symbol (‘) on X’, is just added to emphasize its incompleteness. If we denoted the set of all such permutations by $\mathcal{P}(N, N_{X'})$, then D_1 can be reformulated as D_2 in Eq.6 to consider missing points in its calculations. The optimization problem can also be solved by some extension of Pseudo-code 1, which is described in pseudo-code 3, allowing the Best Match to be incomplete and returning some additional output, the index sequence E to distinguish between existing and non-existing ear touches. The definition of E is presented in Eq.7 and the incomplete $Best\ Match\ (B, E)$ satisfies the equality in Eq.8.

$$D_2(T, X') = \frac{1}{N_{X'}} \min_{(R,l,P) \in \mathcal{R} \times \mathcal{L} \times \mathcal{P}(N, N_{X'})} \sum_{i=1}^{N_{X'}} \left\| t_{\pi_i} - (Rx_i + l) \right\|^2 \tag{8}$$

$$E \stackrel{\text{def}}{=} (e_1, e_2, \ldots, e_N) \tag{9}$$

$$e_i = \begin{cases} 1 & if\ i-th\ earmark\ exists \\ 0 & otherwise \end{cases} \qquad i = 1, 2, \ldots, N$$

$$D_2(T, X') = \sum_{i=1}^{N} e_i \|t_i - b_i\|^2 \tag{10}$$

Using Pacut's method, the algorithm finds the optimal rotation and translation parameters $(\bar{R}, \bar{l})$ that minimize the sum of squared differences between S and the transformed incomplete ear-touch X' [54]. Specifically, it minimizes $\sum_{i=1}^{N_{X'}} \|s_i - (Rx_i + l)\|^2$.

The virtually complete best match of X' with respect to S is then stored as $C = (c_1, c_2, \ldots, c_{N_{X'}})$, where $c_i = \bar{R}x_i + \bar{l}$. The mismatch between S and C is calculated using D_0 (S, C). If this mismatch is less than the current minimum mismatch, min is updated to this new value, and the corresponding best match B and indicator vector E are updated as well. B is set such that $b_i = c_i$ for indices in the permutation and zero otherwise, while E is an indicator vector marking the positions in the permutation.

The algorithm continues iterating through all permutations, updating min, B, and E whenever a lower mismatch is found. Finally, it returns the minimum mismatch, the best match B, and the indicator vector E.

However, a constraint is imposed on the number of touch points N_X'. If N_X' exceeds 10, the computational complexity becomes prohibitive due to the factorial growth in the number of permutations $P(N, N_X')$. This exponential increase in permutations would make the algorithm computationally infeasible, as evaluating each permutation involves significant processing time. Therefore, the algorithm is designed to handle cases where N_X' is less than or equal to 10, ensuring practical and efficient computation.

Numeric example:

Consider a scenario where we have missing points and would like to authenticate a user. Let's say that the set of points on the touchscreen represents the user's unique ear-touches as shown in Figure 4-10.

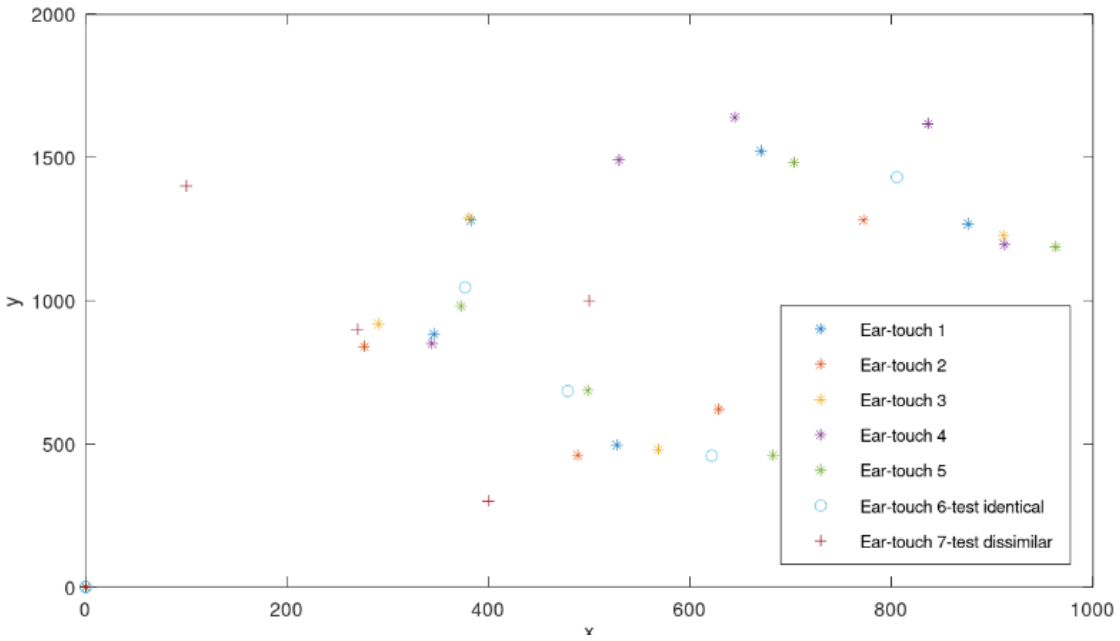


*Figure 4-10: Seven ear-touches with missing points; Six samples (showed by * and o) are taken from a participant and a sample (showed by +) from another one.*

Figure 4-10 shows template creation samples, identical sample and dissimilar sample which will be used in our analysis. Five ear-touches are used to create template in presence of missing points. Based on these samples (shown in Figure 4-11) let's make a template using five of the ear-touches for analyzing pseudo code 3 and 4. So, based on pseudo code 5 the template creation would be as follows:

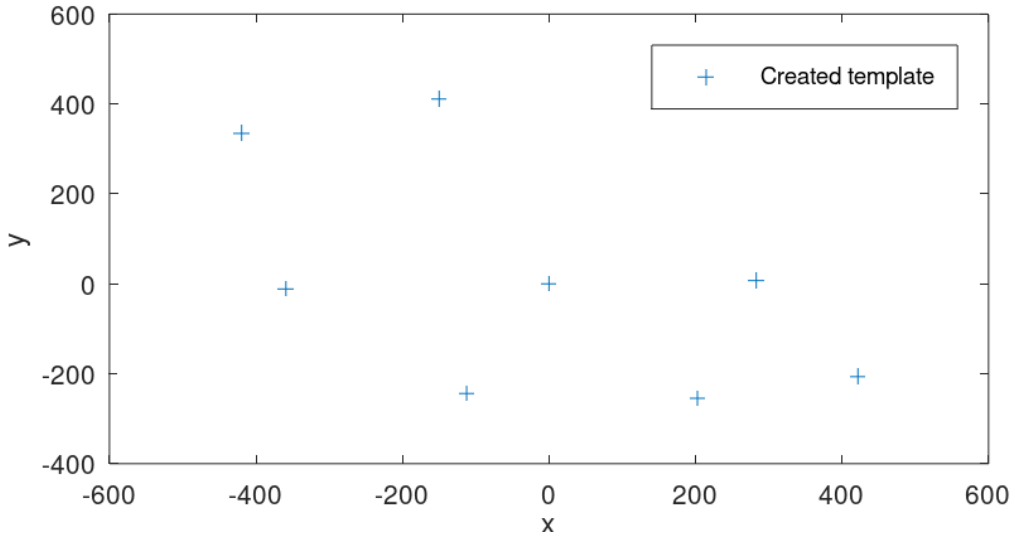


Figure 4-11: Created template using Pseudo code 5

Applying Pseudo code 4 on ear-touch 6-test identical and getting result for that would have the results on Figure 4-12. After calculation of distance between template and test we get 1849. This result is achieved for the identical ear-touch.

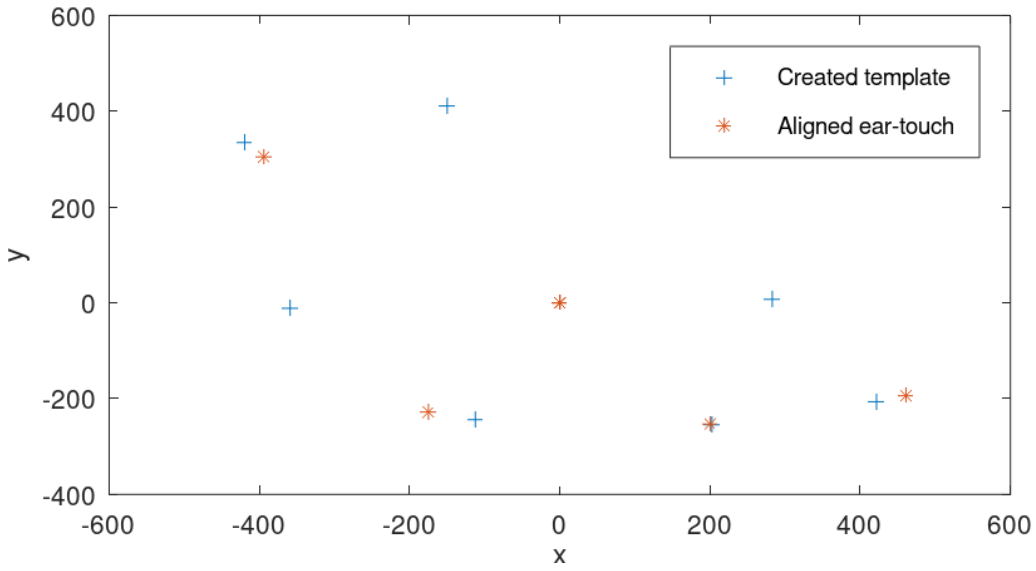


Figure 4-12: Aligned ear-touch and template; the aligned ear-touch is identical ear-touch

Let's consider a dissimilar ear-touch to test how to algorithm works. Figure (3-18) shows an example of ear-touch which is not identical with template that we have for this scenario (called "Ear-touch 7 dissimilar").

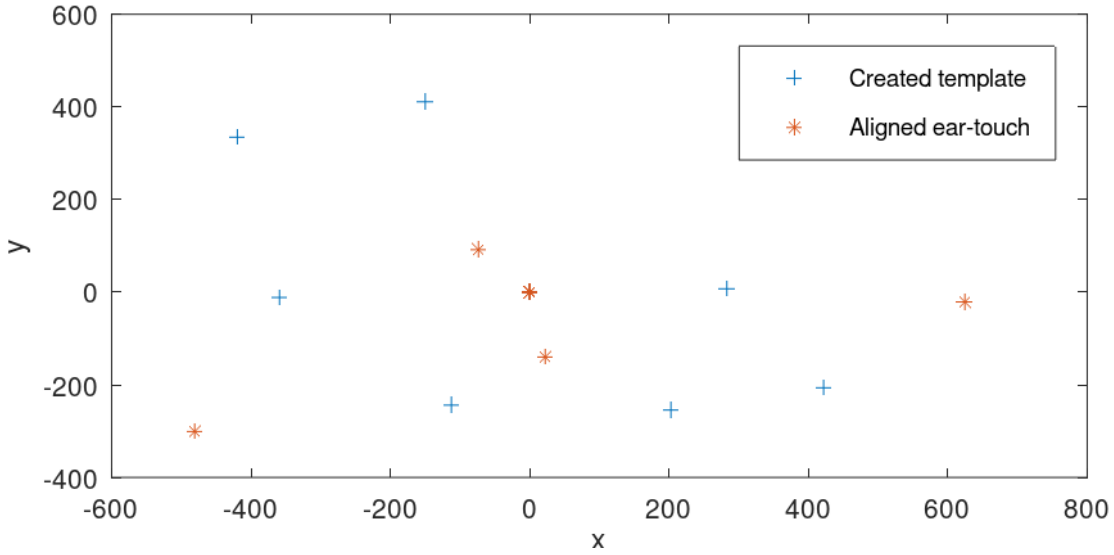


Figure 4-13: Aligned ear-touch and template; the aligned ear-touch is dissimilar ear-touch

If we consider the shown ear-touch in Figure 4-13, we will have distance equal 10970. We see that when we have different ear-touch that template we will have worse results.

4.4. Experimental results

The experimental scenario involved the limited ear-touch database. Each user considered in the enrolment process was chosen randomly. Then the extracted features for the test user were calculated. We carried out the tests on our own ear-touch dataset. In the verification systems, the problem of missing points, because of the physical properties of ear, is marginal, since it is always possible to acquire a proper ear-touch. Pressing a bit longer and keeping the touch screen on ear takes only a few seconds and usually users cooperate with the process. Hence, in the experiment, we concentrated on the images of the ear-touches with over four touch points, and ignored images with less than four.

Experiments were carried out to calculate the performance gain of using touch point coordinates in a matching system. For each subject, the number of imposters and genuine matcher were almost 36315 ((270*269)/2) and 1110 (30*37), respectively. It should be mentioned that we did not consider the symmetric similarity of the same subject, or the similarity between the same ear-touches. The average times it took to create a template (feature extraction from the minimum four touch points and the maximum eight touch points) and matching were 0.22s and 0.003s, respectively. We used a PC with 8 GB RAM and a 2.6 GHz core i3 CPU. Octave was used to implement all the programs currently.

The query images were captured for the subjects in a similar condition. The features of ear touches were computed and the ear recognition decision was taken based on the calculated features.

In the coming subsections, we denote the results for the proposed methods. The False Rejection Ratio (FRR) and the False Acceptance Ratio (FAR) parameters were calculated, thanks to which the Equal Error Ratio (EER) was computed. The False Match Rate (FMR) is the rate at which a biometric process mismatches biometric signals from two distinct individuals as coming from the same individual.

4.4.1. Evaluate the recognition system without missing points

In our dataset we have some touches with no missing points. We have 17 users which have ear-touch with no missing point. In total, we have 72 ear-touches with no missing points. It means these 17 users had at least three ear-touches which had the same number of touch points and they were located in almost the same places. In this section we evaluate how our proposed method works on this part of dataset. Figure 4-14 indicates FMR and FNMR for data with no

missing points. We used single enrollment and multi enrollment scenario. For evaluation without missing data, we have used just 3 enrollment images because there is no possibility to have more enrollment images. We can see that the proposed method achieved EER 0.037 when there is single enrollment image whereas it achieved EER 0.032 when there are multi enrollment images. Figure 4-15 depicts the Detection error trade-off (DET) curves for the proposed method on the ear-touches database. The DET curve has been shown to find a trade-off between FNMR and FMR.

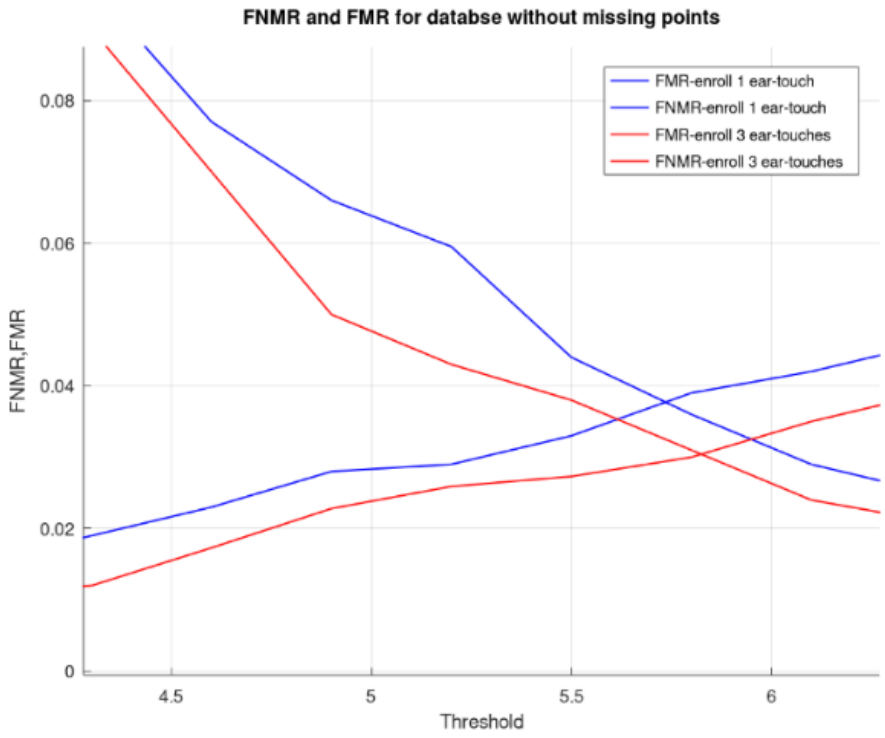


Figure 4-14: False Match Rate and False Non-Match Rate for ear-touches dataset with no missing data

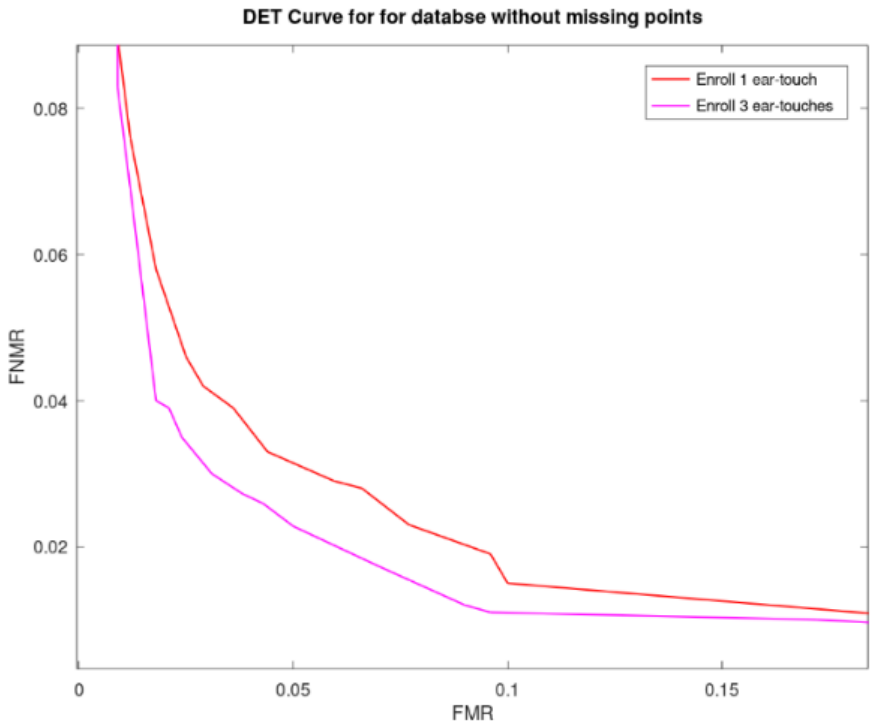


Figure 4-15: DET curve for ear-touches dataset with no missing data

4.4.2. Evaluate the system in the presence of missing points

In this section we used all dataset to evaluate the system performance. We have 92 users which have ear-touch with no missing point. In total, we have 960 ear-touches include missing points. In this section we evaluate how our proposed method works on whole dataset. The recognition outcome (RMR vs. FNMR) for the proposed method is shown in Figure 4-16. Figure 4-17 depicts the Detection error trade-off (DET) curves for the proposed method on the ear-touches database. The experiments showed that the proposed method could improve the results by about 0.17 (from 0.27 to 0.10) when eight ear-touches are used for enrolment. This result was probably achieved because in the single ear-touch, we assumed that all possible points in an ear-touch would appeared in the ear-touch with the maximum points; whereas, in the method with eight ear-touches for enrolment, we considered possible points and the computations were done based on all touch points.

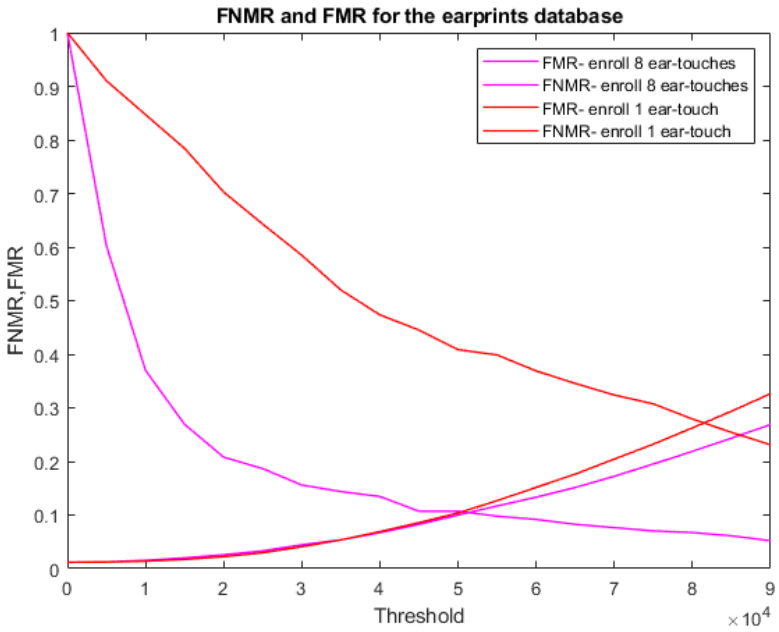


Figure 4-16: False Match Rate and False Non-Match Rate for ear-touches data using the proposed method

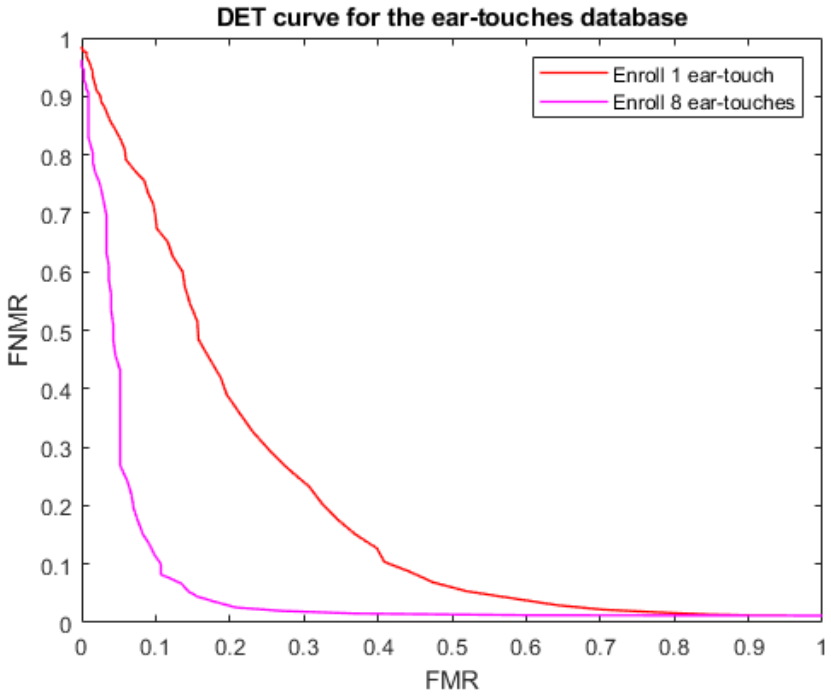


Figure 4-17: DET curve for ear-touches data using the proposed method

4.4.3. Different sample numbers for template creation

We evaluated the proposed method on our dataset based on various numbers of ear-touches for template creation because we wanted to explain how the number of ear-touches could affect the results. Table 3-2 depicts the comparison of performance with our proposed method in terms of using various numbers of ear-touches for enrolment in the ear-touch recognition

system. There are four subsets of the enrolment model. For instance, the "Enrol 1 ear-touch" means that one and the rest of ear-touches in a subject are used for training (template creation) and testing respectively, and so on. It was observed that the more ear-touches were used for enrolment, the better the performance that was achieved. Consequently, Figure 4-18 indicates the comparison of FMR and FNMR with our proposed method in terms of using various numbers for enrolment.

Table 4-1: Performance result comparison with various types of the enrollment model on the ear-touch recognition system

Enrolment model	EER
1 ear-touches for enrollment	0.178
2 ear-touches for enrollment	0.093
4 ear-touches for enrollment	0.05
6 ear-touches for enrollment	0.04
8 ear-touches for enrollment	0.037

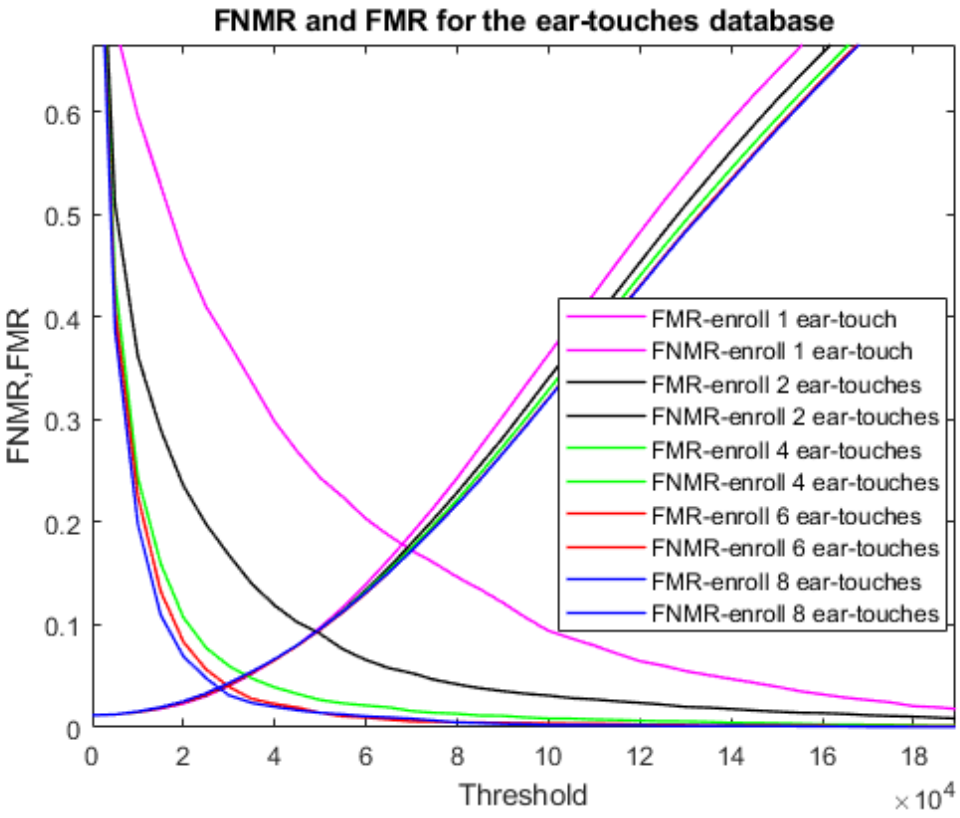


Figure 4-18: FMR and FNMR for ear-touches data with various types of the enrollment model on the ear-touch recognition system

Subsequently, Figure 4-19 depicts the DET curves for the proposed method on ear-touches database. The experiments showed that mean cross validation results is about EER =0.04. As a result, the mean EER was 0.04 for all folds, acceptable for a recognition system.

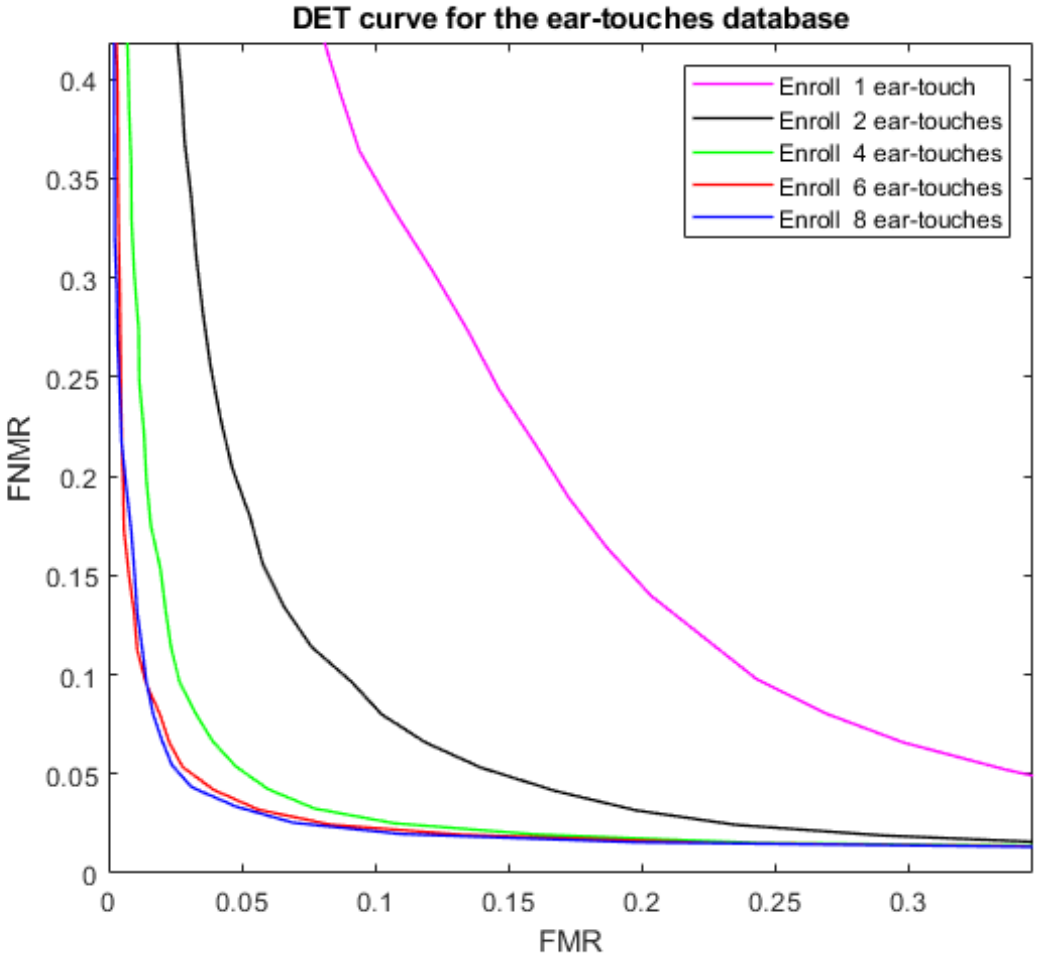


Figure 4-19: DET curve for ear-touches data based on the enrollment with different number of ear-touches

As a whole, our experiment indicated remarkable improvement in performance, since we tested different number of ear-touch as enrollment presentation using the proposed method. It was shown that the method provided precise and additional information and might be used for authentication and control access on smartphones. The results of this research strongly recommend that considering missing points is crucial and must be considered as part of features.

4.5. Discussion

The goal of this research study was to introduce a novel touch base biometric characteristic on mobile devices. This section comprise a discussion of main findings as related to the written

works on ear-touch recognition method in mobile devices. The section sums up with a discussion of limitation of the research, areas for future research, and a short summery. Then include discussion and future study possibilities to assistant answer the research questions:

Question 1: What motivates ear-touch person authentication to be useful on mobile devices?

Question 2: How the used method for alignment of the ear-touches could show performance of the biometric characteristic.

To answer the first question, let's consider the existing ear-touch biometric characteristic. A proposed method in [16] shows a similar biometric system but with different type of data acquisition. They scan all ears with specific screen which is not possible on normal smartphone. It means they changed kernel of the mobile devices. Therefore, to make available ear-touch biometric system we used touch screen in normal mode. According to the result, we could use the acquired data as biometric characteristic.

The proposed method in [55] for alignment of the ear-touch achieve equal error rate 0.04. We might could consider it as a biometric system even for identification but it needs to evaluate on large database.

This biometric system has advantage in terms of acquiring data can be done easily. It could be installed on all multi touch screen mobile devices. However, the number of features as a biometric characteristic system might be less.

4.6. Conclusion

In this work, we have introduced a biometric characteristic for mobile devices that have multi touch screen. The proposed method can identify persons based on their ear-touches, and as a result, person could authenticate themselves easily. The authentication system utilizes the multi touch screen of mobile devices as a sensor to capture biometric features. We have collected a database contain 92 subjects and 960 images (ear-touches). To extract and match the ear-touches, we used a method proposed in [55]. Based on the method, we achieved EER 0.04. This research can be a starting point to work with a new biometric characteristic acquired with multi touchscreen on most of mobile devices.

5. Fusion ear-touch and ear-photo recognition

5.1. Introduction

Unimodal biometric systems often encounter difficulties such as noise, limited degree of freedom, non-universality of biometric traits, and high error rates. To address these challenges, multimodal biometrics has emerged as a promising approach in the field of biometrics [56]. With the advancement of technology, it has become possible to combine multiple biometric modalities to create more accurate and reliable recognition systems.

One such combination that shows great potential is the fusion of ear-touch and ear-photo recognition. Due to the unique physiological structure and location of the ear, the combination of these two biometric modalities can fully utilize their complementary relationship and overcome the limitations of unimodal biometric systems. Moreover, this approach does not require the subject's cooperation, making it non-intrusive and more user-friendly.

Information fusion in multimodal biometric systems can occur at three levels: feature level, matching score level, and decision level [57]. Among these, decision level fusion plays a crucial role as it consolidates individual modality decisions into a final output, ensuring robustness and reliability in the system. Although feature level fusion has the potential to retain more detailed classification information, decision level fusion remains vital for integrating diverse biometric modalities effectively.

In this chapter, we propose a fusion method for ear recognition systems that uses ear-photo and ear-touch biometric modalities. Our approach utilizes a Siamese neural network with pre-trained MobileNetV2 and VGG-16 as feature extraction. Additionally, we use an analytic method, consisting of translation, rotation, and permutation, to extract ear-touch features. Through our proposed method, we aim to demonstrate the potential of ear-touch and ear-photo recognition fusion for improving the performance of biometric recognition systems.

5.2. Literature Review

Nowadays, mobile devices have more capabilities to obtain information from the environment around us for biometric purposes, such as accelerometers, cameras, gyroscopes, etc. However, weaknesses in the computing process force developers to design suitable algorithms to adapt to these situations. In [58], a multimodal biometric has been proposed for person authentication on mobile smartphones, where ear and arm physical and behavioral biometrics have been considered as a multi-biometric. Their proposed method has three subsystems: Arm-Gesture-Acquisition, Ear-Acquisition, and a fusion decision system. Fourier transform (FFT) and local binary patterns (LBP) have been used for extracting arm gesture and ear shape features, respectively. They have database of 300 ear images, 600 accelerometer recordings (300 standing + 300 sitting), and 600 gyroscope recordings (300 standing + 300 sitting). The ROC and EER for accelerometer data in the best performance was 0.93 and 0.13, respectively.

In [59], the authors explore the fusion of ear and palmprint biometric modalities at the feature-level. Ear and palmprint patterns are highly distinctive and stable, providing a wealth of information for discriminating individuals. The authors employ various local texture descriptors, including local binary patterns, weber local descriptor, and binarised statistical image features, to extract discriminant features for accurate human identification. Through extensive experimental analysis using the IIT Delhi-2 ear and IIT Delhi palmprint databases, the authors demonstrate that the proposed multimodal biometric system outperforms single-modal biometrics, achieving a recognition rate of 100%.

Fusion ear-touch and ear-photo recognition is a promising approach for improving the accuracy of ear recognition systems. The combination of these two modalities can provide complementary information that can enhance the recognition performance.

5.3. Fusion recognition methodology

In this section, we introduce the fusion recognition methodology employed in our ear recognition system, which integrates both ear photos and ear-touch data for enhanced accuracy and reliability. The fusion model leverages the unique strengths of each biometric modality, combining the detailed visual features captured in ear photos with the touch patterns obtained from ear-touch interactions. By enrolling input from both sources, our system aims to create a robust and comprehensive template that improves the overall recognition performance. The

proposed method is shown in Figure 5-1. The following subsections will detail the process of data acquisition, feature extraction, and the specific fusion techniques utilized to merge the information from ear photos and ear-touch data.

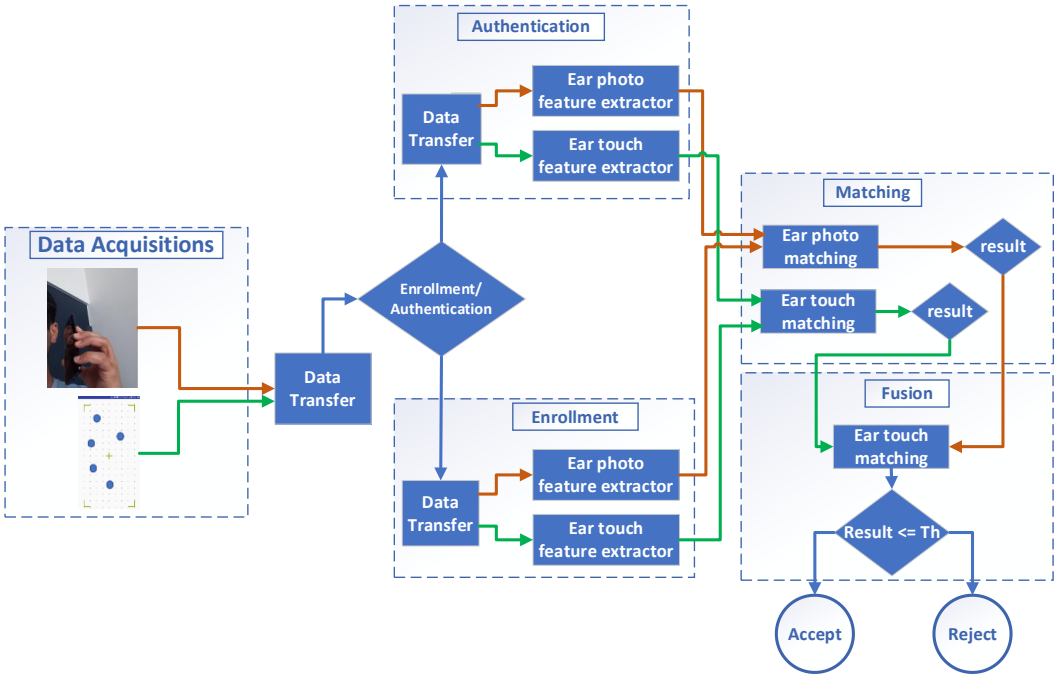


Figure 5-1: The proposed method for combining ear photos and ear-touch to make an integrated system

Acquisition: This first stage is essential, supplying the raw inputs used in both enrollment and authentication.

Data transfer: Following capture, a transfer module routes the signals into two parallel streams—ear photo data and ear-touch data.

Mode selection: enrollment vs. authentication.

- **Enrollment:** Features are extracted from the captured samples and stored for later matching.
- **Authentication:** Features are extracted from the probe samples and compared against the stored enrollment representations.

Enrollment:

- **Ear Photo Feature Extraction:** The Siamese network, as detailed in Chapter 3 section 3-3-1, is employed to extract features from the ear photo. This deep learning model is adept at generating distinctive feature vectors that capture the unique characteristics of each ear image.

- **Ear Touch Feature Extraction:** The analytical method proposed in Chapter 4 section 4-3 is used to extract features from the ear touch data. This method analyzes the geometric properties of the set points to produce a representative feature vector.
- The extracted features from both modalities are then stored as part of the enrollment data, ready for future comparisons.

Authentication phase:

- **Ear photo feature extraction:** The same Siamese network used at enrollment encodes the test ear images to produce embeddings.
- **Ear-touch feature extraction:** The analytical pipeline is reapplied to the test ear-touch samples to obtain their feature representations.

Matching:

- **Ear photo matching:** The test embedding is compared with stored enrollment embeddings using a cosine-distance metric.
- **Ear-touch matching:** Likewise, the ear-touch feature vector from the probe sample is matched against the enrolled templates.

Fusion: The results from the matching phase are combined to produce a final authentication decision. Given the higher importance assigned to ear photos, the fusion process prioritizes this modality:

- **Weighted Fusion:** A weighted sum approach is used to combine the match scores from the ear photo and ear touch data. If S_{photo} is the score from the ear photo and S_{touch} is the score from the ear touch, the fused score S_{fused} is calculated as:

$$S_{fused} = w_{photo} * S_{photo} + w_{touch} * S_{touch}$$

Where w_{photo} is greater than w_{touch} to reflect the higher importance of the ear photo. To have a correct scale between S_{touch} and S_{touch} the S_{touch} is scaled to between 0-1. In the experiment we considered $w_{photo} = 0.7$ and $w_{touch} = 0.3$.

- **Threshold-Based Decision:** The fused score is compared against a predefined threshold. If S_{fused} exceeds the threshold, the authentication is accepted; otherwise, it is rejected. In the experiment we considered threshold=0.2.

The final output of the system indicates whether the authentication attempt is accepted or rejected based on the fused score and the threshold comparison.

The proposed fusion-based ear recognition model effectively integrates ear photo and ear touch data, leveraging the strengths of both modalities to achieve high accuracy and robustness. The process encompasses data acquisition, data transfer, feature extraction, matching, and fusion, ensuring a comprehensive approach to biometric authentication. By prioritizing the more reliable ear photo data in the fusion process, the model provides a robust solution for secure and reliable ear biometric recognition.

5.4. Fusion dataset

To evaluate the performance of the proposed fusion ear-touch and ear-photo recognition system, a dataset consisting of ear-touch and ear-photo images was created. The dataset was collected from 92 volunteers (72 males and 20 females) aged between 18 and 55 years old. The volunteers were asked to provide both ear-touch and ear-photo samples.

The dataset consists of 960 ear-touches and 5132 ear-photos, with 55 ear photos per individual for each modality. The dataset is divided into two parts: a training set and a testing set. The training set consists of 750 ear-touches and 4200 ear-photos, while the testing set consists of the remaining 210 ear-touches and 1132 ear-photos. It should be noted that we don't use all subject because lack of ear-touch samples for some of them. As a result, there are 92 subjects which have both ear-touch and ear photos.

Table 5-1: Dataset of ear-touches and ear-photo. Each ear from a person is considered as a subject

Sex	**Single/ Multi**	**Number of ear-touches**	**Number of ear-photos**	**Number of Subjects**
Male	Multi Ear-touch	595	2352	42
Female	Multi Ear-touch	345	1710	18
Male	Single Ear-touch	30	954	30
Female	Single Ear-touch	2	116	2
Total		960	5132	92

The dataset is further divided into two categories: single ear-touch and multi ear-touch. For multi ear-touch samples, each subject contributed multiple ear-touch samples, while for single ear-touch samples, each subject contributed only one sample. Among the multi ear-touch

samples, males contributed 595 samples, while females contributed 345 samples. On the other hand, among the single ear-touch samples, males contributed 18 samples, while females contributed only 2 samples. The table also provides information on the total number of ear-touch and ear-photo samples used in the dataset.

5.5. Experiments Results

In this chapter, we present a comprehensive analysis of our fusion-based ear recognition model, which integrates ear-touch and ear-photo data to enhance the accuracy and reliability of biometric authentication. To provide a clear understanding of the individual and combined contributions of each modality, we structure the experimental section to first present the results of ear-photo and ear-touch recognition separately, followed by the results of the fusion-based approach.

5.5.1. Ear photo recognition results

The utilized model in this research is a Siamese neural network with pre-trained deep neural network. We use MobileNetV2 and VGG16 as two pre-trained models for Siamese network.

Local binary patterns (LBP)) and a CNN based algorithm (VGG-16) has been introduced as a baseline algorithm for feature extracting in the ear recognition in the Unconstrained Ear Recognition Challenge 2019 [10]. These two algorithms are well-known and have been used extensively in biometric recognition (Face [36], ear [37], iris [38], etc.). Therefore, in our experiments, we use LBP and pre-trained VGG-16 and MobileNetV2 models [32] with Siamese network. We used MobileNetV2 as a baseline because it has a better result than VGG-16 on the UERC 2019 dataset. We use a well-known feature matching method in computer vision namely; cosine distance.

Figure 5-2 presents the False Match Rate (FMR) and False Non-Match Rate (FNMR) curves for three different biometric recognition methods: LBP, pre-trained MobileNetV2 with Siamese network (MobNetV2-SiaNet), and pre-trained VGG-16 with Siamese network (VGG16-SiaNet). The curves are plotted across varying threshold values to illustrate the performance of each method in distinguishing between genuine and imposter matches.

For the LBP method, the FMR and FNMR curves intersect at an EER of 0.59, indicating that at this point, the probability of a false match is equal to the probability of a false non-match. Similarly, the MobNetV2-SiaNet method achieves a lower EER of 0.31, and the VGG16-

SiaNet method further reduces the EER to 0.17. These EER points are marked on the plot with red dots and annotated with their respective threshold values.

The comparative analysis shows that as we move from LBP to more advanced models like MobNetV2-SiaNet and VGG16-SiaNet, the EER decreases, demonstrating improved recognition performance. This improvement is visually evident from the reduced area between the FMR and FNMR curves at the EER points for MobNetV2-SiaNet and VGG16-SiaNet compared to LBP.

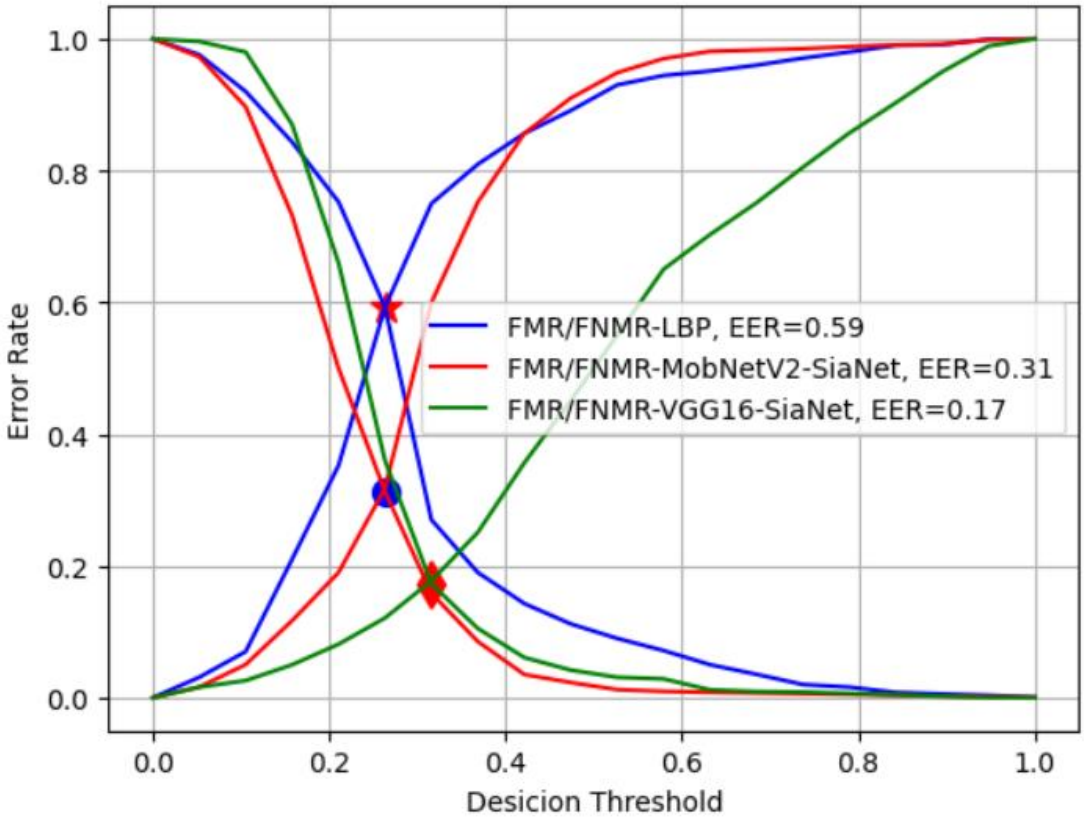


Figure 5-2: FNMR and FMR for the ear images database based on MobileNetV2 and VGG-16 feature extractors and LBP

Figure 5.3 illustrates the Detection Error Tradeoff (DET) curves for the ear photo test set, comparing three different feature extraction methods: Local Binary Patterns (LBP), MobileNetV2 based on a Siamese neural network, and VGG-16 based on a Siamese neural network.

The LBP method exhibits a relatively high Equal Error Rate (EER) of 0.59, indicating that at the point where FMR equals FNMR, both types of errors occur with a significant frequency.

The MobileNetV2 model based on a Siamese neural network demonstrates a lower EER of 0.31. This indicates better overall performance compared to LBP, with a more favorable trade-off between FMR and FNMR.

The VGG-16 model based on a Siamese neural network achieves the lowest EER of 0.17. The DET curve for VGG-16 shows the lowest FMR and FNMR values, demonstrating its ability to minimize both false matches and false non-matches more effectively than the other two methods. The VGG-16 DET curve demonstrates the best performance, with the lowest EER and the most favorable balance between FMR and FNMR. The curve rises slowly, indicating that the model effectively reduces both false matches and false non-matches across a wide range of thresholds.

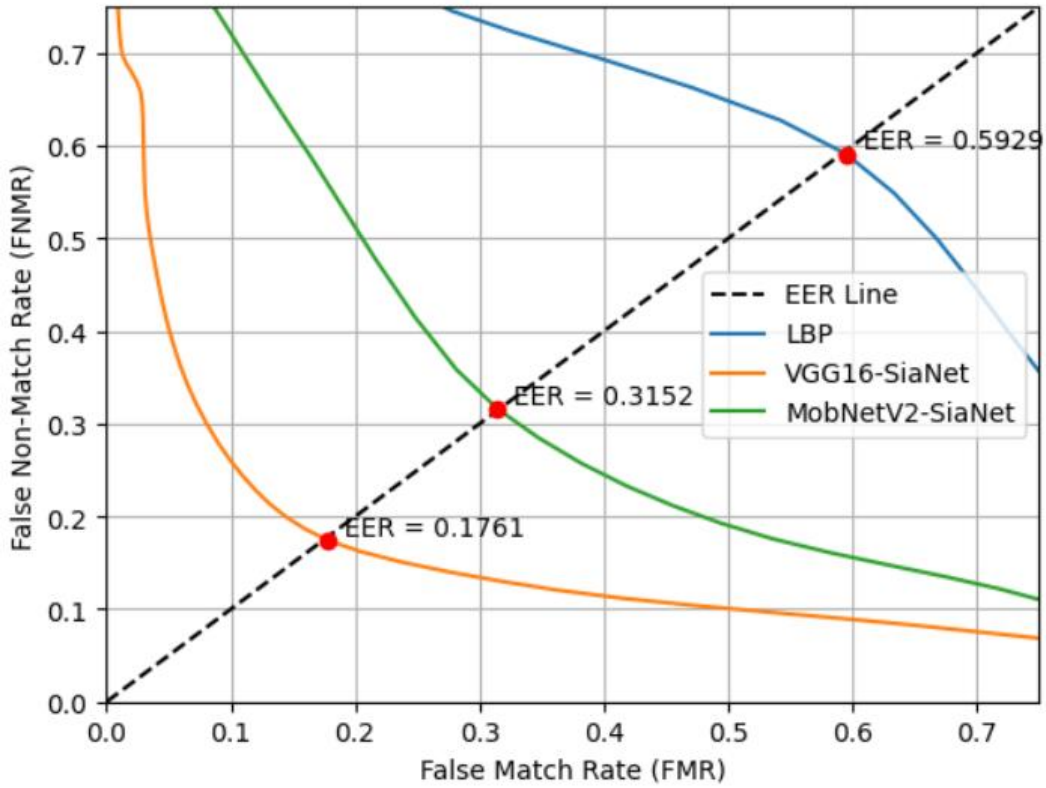


Figure 5-3: DET curve for the ear photo test set for MobileNetV2 and VGG-16 feature extractor based on Siamese neural network and LBP.

5.5.2. Ear touch recognition results

We used an optimization-based approach to address the challenges in ear touch recognition system in Chapter 4 section 4.3. This method, inspired by Pacut [54], aims to find the best matches by minimizing relevant loss functions and involves permutation, rotation, and translation. The approach effectively resolves the optimization problem by leveraging the

Procrustes problem and the Kabsch-Umeyama algorithm. Notably, Pacut's solution is straightforward and efficient, avoiding the need for common iterative schemes typical of general optimization algorithms. This section includes performance metrics such as FMR, FNMR, and EER, obtained using the analytical method described in Chapter 4 section 4.3.

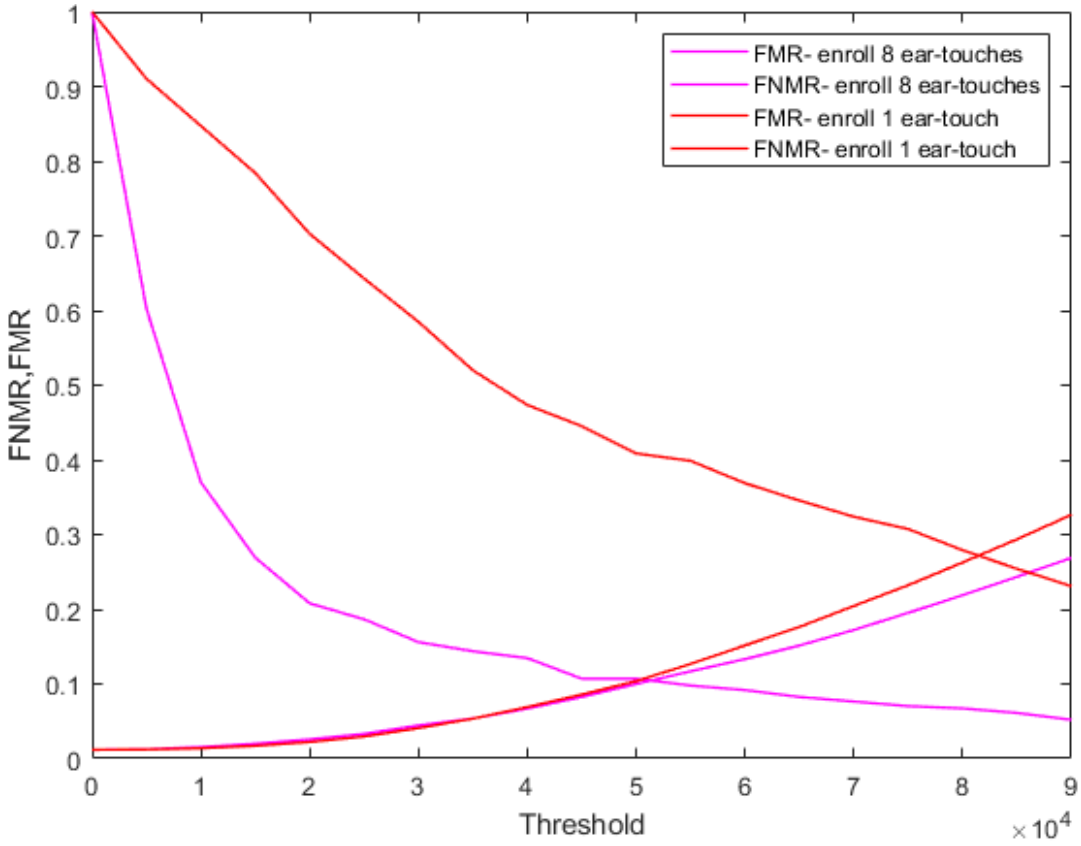


Figure 5-4: FNMR and FMR for the ear-touch database based on a proposed method at Chapter 4 section 4.3.

5.5.3. Fusion recognition results

We have shown the evaluation of the limited data to merge ear photos and ear-touch. Figure 5-5 and Figure 5-6 together provide a comprehensive comparison of the performance of two biometric recognition models: the VGG16-SiaNet fusion model and the VGG16-SiaNet model based on ear photos.

Figure 5-5 presents the False Non-Match Rate (FNMR) and False Match Rate (FMR) curves for both the VGG16-SiaNet fusion model and the VGG16-SiaNet model using only ear photos. The FNMR and FMR curves indicate that while the VGG16-SiaNet model using only ear photos is capable of distinguishing between genuine and impostor samples, it does so with moderate accuracy, as evidenced by an Equal Error Rate (EER) of 0.17. In contrast, the VGG16-SiaNet fusion model, which integrates both ear photo and ear-touch data, achieves a

lower EER of 0.106. The FNMR and FMR curves for the fusion model demonstrate improved performance, with a more favorable balance between the two error rates across various thresholds.

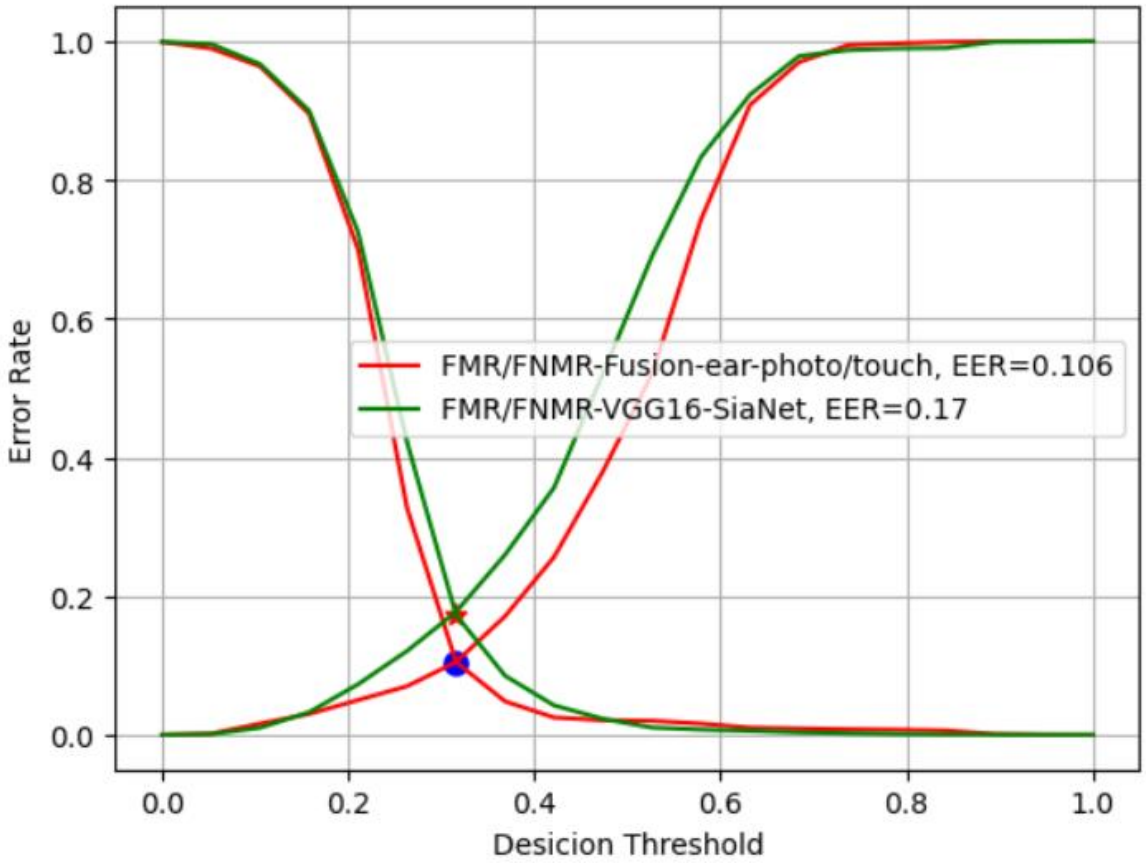


Figure 5-5: FNMR and FMR curve for the VGG16-SiaNet fusion model and VGG16-SiaNet on just ear photos on test set

Figure 5-6 complements the insights from Figure 5-5 by presenting the Detection Error Tradeoff (DET) curves for the same two models. The DET curve for the VGG16-SiaNet model using only ear photos shows a typical trade-off between the two error rates, with a higher curve indicating that the model has more difficulty maintaining low levels of both FMR and FNMR simultaneously. This aligns with the higher EER observed in Figure 5-5, confirming that the model struggles more with balancing these errors. The DET curve for the fusion model lies consistently below the curve for the model using only ear photos, demonstrating that the fusion approach achieves better performance across all thresholds.

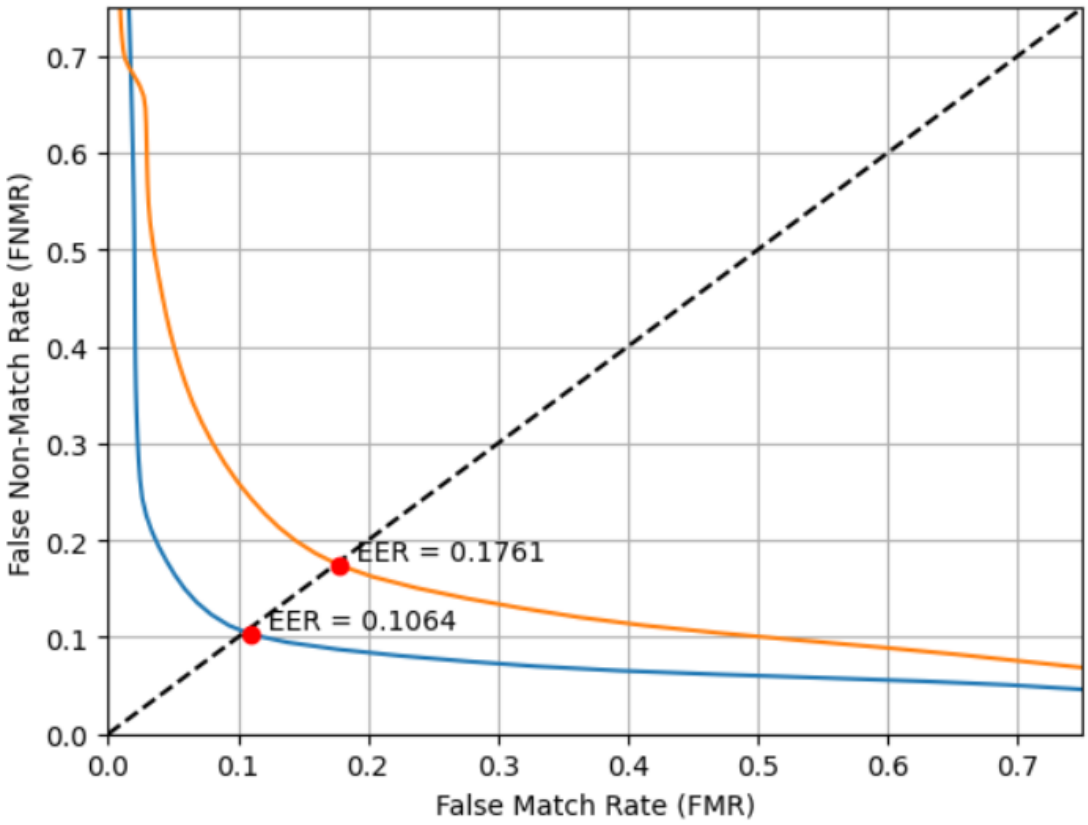


Figure 5-6: DET curve for the VGG16-SiaNet fusion model and VGG16-SiaNet on just ear photos test set

5.6. Conclusion

In this chapter, we introduced a feature fusion methodology specifically designed for multimodal biometric recognition using ear photos and ear-touch data. This approach represents an advancement in the field of biometric authentication, offering a robust solution that leverages the complementary strengths of visual and tactile modalities.

To develop and validate this method, we meticulously collected a comprehensive dataset comprising ear photos and ear-touch samples. The fusion model was then applied to this data, aiming to enhance the accuracy and reliability of biometric recognition. The experimental results demonstrate that our fusion-based approach outperforms unimodal recognition methods that rely solely on either ear photos or ear-touch data. Specifically, the fusion model not only reduces the error rates, as evidenced by lower False Match Rates (FMR) and False Non-Match Rates (FNMR), but also extends the recognition capabilities to scenarios where either modality alone might be insufficient.

The results clearly show that the combination of ear photo and ear-touch data leads to a more accurate and reliable biometric system. The VGG16-SiaNet fusion model, in particular,

exhibited a marked improvement in performance, achieving a lower Equal Error Rate (EER) and more favorable Detection Error Tradeoff (DET) curves compared to the unimodal approaches. This demonstrates the effectiveness of the fusion model in balancing and minimizing recognition errors across different thresholds, which is crucial for real-world biometric applications.

Moreover, this fusion methodology is particularly suited to non-intrusive biometric scenarios where traditional multimodal biometrics may not be available or practical. The natural combination of ear photo and ear-touch recognition offers a seamless user experience, integrating effortlessly into common usage patterns, such as holding a mobile device to the ear. This makes the proposed method not only technically effective but also user-friendly, enhancing its potential for widespread adoption.

6. Fusion ear-touch and ear-photo presentation attack detection

6.1. Introduction

Ear spoofing attacks include mask attacks, photo attacks, and synthetic 3D, with video and photo attacks being the most popular vision [19, 20]. The quality of captured images and videos of ear spoofing attacks might be because of the recognition device's resolution, noise signal, lack of medium, and image stripes. To confirm and identify the authenticity of ear data, PAD methods can be classified into image quality-based, texture-based, hardware-based, and deep feature-based approaches.

Recently, Alireza et al. proposed Lenslet Light Field Database for ear PAD [18]. They utilized conventional 2D solutions [60, 61] and light field-based solutions method [62] for classification. They showed that light field-based methods achieved very stable and effective PAD performance and they were very fast. To experiment, they used just 268 images; as a result, it was hard to prove the performance of a deep neural network by having a small number of images.

Previous research showed that light field-based methods achieved very stable and effective PAD performance. However, to further increase detection accuracy and make more trustworthy results, a new large ear PAD database is collected, and a PAD method based on a vision transformer fusion is proposed in this chapter.

The organization of the rest of the chapter includes a review and analysis in Section 5-2, a fusion method for ear PAD using ear photo and ear-touch in Section 5-3, experimental results and analysis in Section 5-4, and conclusions in Section 5-5.

6.2. Literature Review

This chapter investigates the Vision Transformers framework and its application in ear PAD. Previous studies have proposed several techniques, including multi-channel CNN [63], multi-level local binary patterns, specialized CNN architecture with additive operator splitting, LBPnet [64], multi-scale dynamic binarized statistical image features (MBSIFTOP) and

multiscale dynamic local phase quantization (MLPQ-TOP) [65], and deep belief network (DBN) with MobileNet-v1 CNN [66]. These techniques have been applied to different biometric features, and their performances have been evaluated on various databases.

In addition, the binarized statistical image features (BSIF) [67, 68] and Image Quality Assessment (IQA) techniques [69] have been employed for palm print PAD, and CNN-based techniques have been used for finger vein and voice PAD [70]. The experiments have been conducted on public databases, including AMI and UBEAR ear databases.

The proposed fusion technique for ear-touch and ear-photo PAD in this chapter aims to improve the accuracy and robustness of existing techniques. The fusion approach will combine the strengths of different techniques and provide a more comprehensive analysis of ear-based biometric features.

6.3. Fusion verification methodology

The proposed fusion-based methodology consists of three main components as shown in Figure 6-1: data, individual modality classification, and late fusion for final decision-making. This section outlines the design and implementation of each component in detail.

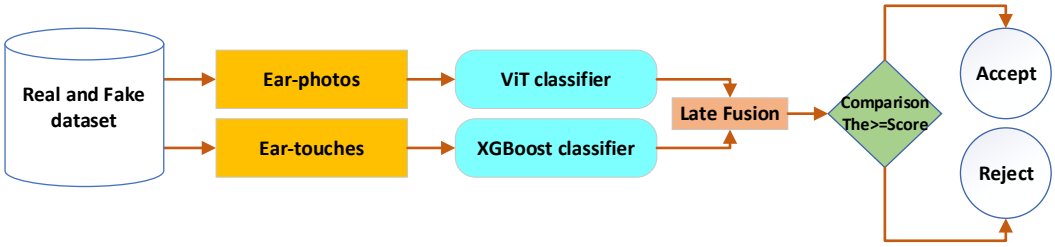


Figure 6-1: The proposed method for combining ear photo and ear-touch PAD system. It first inputs ear photos and ear-touches, and then two images are output by the RGB images and the set points are input into a self-attention fusion module for fusion. Finally, a classifier is considered for classification.

Acquire ear-photo and ear-touch data from the user: For the ear-photo modality, take a high-resolution photograph of the user's ear using a camera or smartphone. For the ear-touch modality, use a touch-sensitive device to record the pressure and shape of the user's ear when pressed against it.

Classification for each modality: Use appropriate classification techniques to classify real and fake images.

Combine the scores from each modality: Use a fusion method such as score-level fusion or decision-level fusion to combine the verification scores obtained from the ear-photo and ear-touch modalities.

Final decision: Compare the fused score to a preset threshold τ. If the score is $\geq\tau$, accept (verify) the user; otherwise, reject.

Performance evaluation: Assess the fusion method using standard PAD metrics—Bona Fide Presentation Classification Error Rate (BPCER) and Attack Presentation Classification Error Rate (APCER)—to quantify effectiveness.

6.3.1. Vision transformer model for ear photo PAD

The DeiT (Data-efficient Image Transformer) model [71] builds upon the Vision Transformer (ViT) architecture [72], which has demonstrated superior performance over traditional convolutional neural networks (CNNs) by utilizing a Transformer encoder . Unlike CNNs that rely on convolution operations to extract features, ViTs apply the Transformer architecture directly to images without using any convolutions. In the ViT approach, images are divided into patches, which are then treated as tokens similar to those used in Natural Language Processing (NLP). While this method effectively extracts features from image patches, it lacks the inherent translation and scaling invariance of CNNs, which can limit its generalization ability when trained on insufficient data.

The initial versions of ViTs faced challenges in generalizing well, particularly when trained on smaller datasets. To overcome this limitation, large-scale datasets containing millions of images (ranging from 14M to 300M) were employed, allowing ViTs to achieve performance comparable to state-of-the-art CNNs [72]. This success has led to the widespread adoption of ViTs in various image classification tasks.

However, training ViT models requires significant computational resources and extended training times, making them less accessible for applications with limited infrastructure. DeiT models address this issue by offering a more efficient training process, enabling them to achieve high performance with fewer computing resources and shorter training durations compared to the original ViTs. The key difference lies in DeiT's training strategy, which includes the introduction of a distillation token and the use of a linear classifier instead of a Multi-Layer Perceptron (MLP) head during pre-training.

Knowledge Distillation (KD) plays a crucial role in DeiT's training paradigm [73]. In KD, a student model learns from the "soft" labels provided by a more powerful teacher network. Instead of merely relying on the "hard" labels produced by the teacher's softmax function, the student model benefits from the nuanced information in the output vector, leading to improved performance. Several DeiT variants have been developed, leveraging strong teacher models through KD to enhance their effectiveness.

DeiT models incorporate a distillation token into the initial embeddings, which consist of image patches and a class token. This token interacts with other embeddings through self-attention mechanisms and is output by the network after the final layer. The distillation token is designed to optimize the distillation component of the loss function, allowing the DeiT model to learn from the teacher's output while maintaining the complementarity to the class embedding. For a more detailed exploration of DeiT and ViT models, readers are encouraged to consult the original works on DeiT [71, 73] and Vision Transformer (ViT) [72].

One of the most notable innovations in DeiTs is the introduction of a distillation token. This token is used during training alongside the class token. The distillation token is designed to capture additional information from a teacher model, typically a well-trained CNN, through a process known as Knowledge Distillation (KD).

Knowledge Distillation involves training a student model (in this case, the DeiT) using the "soft" labels (probabilities) provided by a more powerful teacher model. The distillation token learns from these soft labels, helping the DeiT model to achieve higher accuracy even when trained on smaller datasets.

The final classification layer in DeiTs still benefits from the rich feature representations learned by the Transformer encoder, enhanced by the distillation process.

DeiTs are particularly notable for their ability to perform well even with limited training data. This capability is crucial for many practical applications where large annotated datasets are not available.

By leveraging the distillation token and the efficient training process, DeiTs have been shown to match or exceed the performance of ViTs and even some CNN-based models on standard benchmarks like ImageNet, with far less computational cost.

Application to Ear Photo PAD:

Inspired by the exceptional performance of DeiTs using limited data and reduced computational resources, this thesis explores the application of DeiTs for Ear Photo PAD. To the best of our knowledge, this is the first attempt to apply DeiTs in the context of ear PAD. For the scope of this study, we focus on employing DeiT-small distilled, DeiT-base distilled with image sizes of 224 × 224, and DeiT-base distilled with image sizes of 384 × 384 in a transfer learning approach. These models were selected based on their strong performance on the ImageNet dataset.

- DeiT-small distilled (DeiT-small-distilled-224) achieves 81.2% top-1 accuracy on ImageNet.
- DeiT-base distilled (DeiT-base-distilled-224) achieves 82.9% top-1 accuracy.
- DeiT-base distilled with an input size of 384 × 384 (DeiT-base-distilled-384) achieves 85.2% top-1 accuracy.

The success of DeiTs in various image classification tasks has motivated their exploration in the context of Ear Photo PAD. This thesis represents the first known application of DeiTs for ear PAD, capitalizing on their data efficiency and strong performance on limited data.

In this work, several DeiT configurations are employed, including DeiT-base distilled models, with image resolutions of 384x384 pixels. These models are fine-tuned using transfer learning, leveraging pre-trained weights from ImageNet to adapt to the specific task of detecting presentation attacks in ear biometrics.

The choice of model and resolution is guided by performance benchmarks on ImageNet, where higher resolution models like DeiT-base-distilled-384 have demonstrated superior accuracy.

For all analyses in this thesis, the DeiT model is configured with a patch size of 16x16, and the Gaussian Error Linear Unit (GELU) activation function is used. The detailed parameters and configurations for the model is provided in the subsequent sections, along with a thorough evaluation of its performance in detecting presentation attacks on ear photos. The parameters for DeiT-base-distilled-384 are detailed in Table 2.1. These configurations form the foundation of our investigation into the effectiveness of DeiTs for Ear Photo PAD.

Table 6-1: Parameter configuration in DeiT for DeiT-base-distilled-384

Parameter	DeiT-base-distilled-384 (DeiT-EPAD)
Hidden activation	GELU
Hidden dropout probability	0.0
Hidden size	768
Image size	384
Initializer range	0.02
Intermediate size	3072
Layer normalization factor	1e-12
Number of attention head	12
Number of hidden layers	12
Patch size	16

6.3.2. XGBoost classifier for ear touch PAD

In this section, we delve into the application of the XGBoost classifier for detecting presentation attacks using ear touch data [74]. Ear touch data refers to a set of tactile measurements obtained when a subject interacts with a touchscreen smartphone by touching their ear. These measurements can capture unique patterns position, and contact points that vary between genuine and fake ears, making them valuable for PAD.

XGBoost, short for eXtreme Gradient Boosting, is an advanced implementation of the gradient boosting framework, which has gained significant popularity for its performance and efficiency in handling structured or tabular data. It is particularly well-suited for tasks that require high accuracy and where the data has a mix of numerical and categorical features. XGBoost works by sequentially adding weak learners (typically decision trees) to a model, each one correcting the errors of its predecessors. This iterative process results in a strong predictive model that can capture complex patterns in the data.

Why XGBoost for Ear Touch PAD?

Ear touch data consists of a set of discrete points or features that capture the interaction between the ear and the sensor. These features include the coordinates of touch points. Such data is

inherently structured and can exhibit complex, non-linear relationships, making it challenging to model with simpler methods.

XGBoost excels at identifying and modeling these intricate relationships through its use of decision trees, which can naturally handle interactions between features without requiring extensive data preprocessing or feature engineering.

One of the key advantages of XGBoost is its ability to provide insights into feature importance. After training, XGBoost can rank the importance of each feature based on how frequently it is used to make splits in the decision trees.

XGBoost includes several regularization techniques, such as L1 (lasso) and L2 (ridge) regularization, which help prevent overfitting, especially in cases where the model is trained on small or noisy datasets. These regularization methods penalize overly complex models, ensuring that the model generalizes well to unseen data.

Given that ear touch data might be limited or subject to variability, XGBoost's ability to mitigate overfitting is particularly beneficial for maintaining robust performance in PAD tasks.

Application to Ear Touch PAD

The first step in applying XGBoost to ear touch PAD is to prepare the dataset. Each instance in the dataset represents a touch interaction, with features corresponding to various measurements taken during the touch event. The dataset is split into training, validation, and test sets to ensure that the model is evaluated on unseen data, thereby providing a reliable measure of its performance.

The XGBoost classifier is then trained on the ear touch dataset. During training, the model iteratively builds decision trees, with each tree aiming to correct the errors made by the previous ones. This process continues until the model reaches a specified number of trees or until further improvements are negligible.

Hyperparameters, such as the learning rate, maximum depth of the trees, and the number of trees, are tuned using cross-validation. This tuning process helps optimize the model's performance and ensures that it is neither underfitting nor overfitting the training data.

After training, the XGBoost model provides a ranking of feature importance. This analysis reveals which features are most critical in distinguishing between genuine and fake ear touches.

6.3.3. Late fusion method for ear photo and ear touch PAD

In the late fusion approach for ear photo and ear touch PAD, the individual results from each modality are combined to produce a final decision regarding the authenticity of the biometric sample. The fusion process is designed to prioritize the more reliable modality, which in this case is the ear photo. This section outlines the methodology used for combining the match scores from both modalities and making the final authentication decision.

The weighted fusion method is employed to combine the match scores from the ear photo and ear touch data. The fusion strategy assigns different weights to the scores based on the relative importance of each modality in determining the authenticity of the biometric sample. Given the typically higher reliability and discriminative power of ear photos in PAD, the ear photo score is given a greater weight in the fusion process.

The fused score, S_{fused}, is calculated using a weighted sum of the individual modality scores, S_{photo} (from the ear photo) and S_{touch} (from the ear touch data). The formula for the fused score is as follows:

$$S_{fused} = w_{photo} \times S_{photo} + w_{touch} \times S_{touch}$$

Where:

w_{photo} is the weight assigned to the ear photo score.

w_{touch} is the weight assigned to the ear touch score.

S_{photo} is the match score obtained from the ear photo.

S_{touch} is the match score obtained from the ear touch data.

Given the higher importance of the ear photo in the authentication process, w_{photo} is set greater than w_{touch}. To ensure that the scores from both modalities are on a comparable scale, the ear touch score S_{touch} is normalized to a range between 0 and 1. In the experiments conducted for this study, the weights were empirically determined as:

$$w_{photo} = 0.7$$

$$w_{touch} = 0.3$$

This weighting reflects the greater trust placed in the ear photo modality while still incorporating valuable information from the ear touch data.

After computing the fused score S_{fused}, the system compares it with a preset threshold $\tau$to issue the final decision. This threshold controls sensitivity to presentation attacks: if $S_{\text{fused}} \geq \tau$, the sample is treated as bona fide and access is granted; otherwise, it is flagged as an attack and rejected. In our experiments, $\tau = 0.5$ was used to balance false acceptance and false rejection, with the value chosen from validation results to optimize overall performance.

6.4. Experimental results and analysis

In this section, we present and analyze the experimental results obtained from implementing the proposed fusion-based methodology for ear photo and ear touch PAD. The experiments were designed to evaluate the effectiveness of the individual modalities (ear photo and ear touch) as well as their combined performance using the late fusion strategy described in Section 6.3. The results provide insights into the relative strengths of each modality and the overall enhancement achieved through fusion.

To evaluate the performance of the proposed fusion ear-touch and ear-photo PAD, a dataset consisting of real and fake ear-touch and ear-photo images were created. The dataset was collected from 72 volunteers (52 males and 20 females) aged between 18 and 55 years old. The volunteers were asked to provide both ear-touch and ear-photo samples. The ear-touch images were captured using a mobile device touch screen, while the ear-photo images were captured using a mobile camera.

Table 6-2: Dataset of real and fake ear-touches and ear-photo.

Sex	**# of real ear-touches**	**# of fake ear-touches**	**# of real ear-photos**	**# of fake ear-photos**	**# of subjects**
Male	595	410	2352	3000	52
Female	345	110	1710	1000	20
Total	940	520	5062	4000	72

Table 6-2 presents a summary of the dataset used to evaluate the proposed fusion-based PAD system for ear photos and ear touch data.

The dataset comprises both genuine (real) and fake samples across two modalities: ear touch and ear photo. The ear touch data was gathered by capturing the interactions of the subjects with a mobile device's touchscreen, while the ear photo data was collected using the camera on the same mobile device. For each subject, multiple samples were collected to ensure a robust dataset capable of training and evaluating the PAD models effectively.

Genuine Samples: These include authentic interactions and images from the subjects, representing real-world use cases where the system should correctly authenticate the individual.

Fake Samples: These samples simulate potential presentation attacks. Fake ear touch data include interactions made with hand mimicking the ear, while fake ear photos could involve images or replicas designed to deceive the biometric system.

The dataset is carefully balanced to include a substantial number of both real and fake samples, ensuring that the PAD system is thoroughly tested against a variety of potential attack vectors. The fusion approach benefits from this diverse dataset by leveraging the complementary information provided by both modalities, ultimately leading to a more robust and accurate detection system.

This comprehensive dataset serves as a crucial foundation for training the PAD models, allowing for a detailed evaluation of their performance in distinguishing between genuine and fraudulent biometric samples.

6.4.1. Evaluation metrics

The function of various presentation attack detection technics was evaluated using the following metrics: bona fide presentation classification error rate (BPCER), attack presentation classification error rate (APCER), and average classification error rate (ACER) (the mean of APCER and BPCER). After that, half total error rate (HTER) was utilized as an evaluation indicator for showing a different view of evaluation.

6.4.2. Evaluation of ear photo PAD

The DeiT models were fine-tuned on a dataset of ear photos consisting of both genuine and fake samples. The dataset was divided into training, validation, and test sets, ensuring that the models were evaluated on unseen data to assess their generalization capabilities.

A comparison was done between General Image Quality Assessment (GIQA) [19], PADNet-1(Chapter 3), and the DeiT-base with input resolutions 384x384 (DeiT-EPAD). The DeiT-EPAD has obtained the most favorable results on presentation attack detection for the past three years. To have a comparative vision, the results of the method recommended in the chapter work were trained without additional private datasets.

The experimental results indicate near-perfect performance across all models tested, including PADNet-1, the DeiT-base model, and the GIQA model, each achieving a BPCER of 1%. While these results are highly promising, they may be partially attributed to the limitations of our dataset. The dataset's specific characteristics, such as its size or the controlled nature of the samples, could have contributed to the exceptionally low error rates. Consequently, while the findings demonstrate the robustness and effectiveness of the proposed detection algorithms, further evaluation on a more diverse and extensive dataset would be necessary to fully validate the generalizability and reliability of these results in broader, real-world applications.

Table 6-3 provides a comparison of the APCER across different models when evaluating ear-photo PAD under various attack scenarios. These scenarios involve different combinations of display devices and recognition devices, simulating a range of potential presentation attacks.

The results show a clear trend where the proposed DeiT-EPAD model consistently outperforms the other models, including PADNet-1 and GIQA, across nearly all attack types. For example, in the scenario involving a Dell Ultra Sharp Monitor and a Samsung Galaxy A7 as the recognition device, the DeiT-EPAD model achieves a significantly lower APCER of 4.16% compared to 12.2% for PADNet-1 and 19.6% for GIQA. This pattern is observed across various device combinations, indicating that the DeiT-EPAD model has a superior ability to detect presentation attacks, regardless of the display or recognition device used.

Notably, the DeiT-EPAD model excels in scenarios involving more challenging attacks, such as those using high-quality display devices like the Samsung Galaxy S9. In these cases, the APCER remains impressively low, with the DeiT-EPAD achieving as low as 3.11%, compared to significantly higher error rates for the other models.

The results also highlight the robustness of the DeiT-EPAD model against attacks involving different display and recognition device combinations. For instance, in attacks using the Nokia Lumia 1020, the APCER drops to as low as 0.39%, which is a substantial improvement over the other models. This suggests that the DeiT-EPAD model is particularly adept at handling variations in attack methods, making it a more reliable option for ear-photo PAD.

Table 6-3: The APCER for models on different type of attacks for ear-photo

Display device	Recognition device	Abbreviation of Display and Recognition device	APCER (%)		
			PADNet-1	GIQA	DeiT-EPAD
Dell Ultra Sharp Monitor	Samsung Galaxy A7	Dell-GA7	12.2	19.6	4.16
	Samsung Galaxy S9	Dell-GS9	8.68	9.7	3.11
	Nokia Lumia 1020	Dell-NL1020	0.91	2.7	0.39
SAMSUNG monitor	Samsung Galaxy A7	S3D-GA7	0.37	1.8	0.21
	Samsung Galaxy S9	S3D-GS9	7.4	9.56	3.2
	Nokia Lumia 1020	S3D-NL1020	0.7	2.3	0.42
Brother printer	Samsung Galaxy A7	Print-GA7	1.42	4.36	0.9

6.4.3. Evaluation of ear touch PAD

The ear touch PAD system was evaluated using a dataset comprising both real and fake ear-touch samples collected from 72 volunteers. The dataset was balanced across genders, with 63% of the participants being male and 36% female. In total, the dataset included 940 genuine ear-touch samples and 520 fake ear-touch samples, providing a comprehensive basis for training and evaluating the PAD system.

The XGBoost classifier was employed to detect presentation attacks based on the ear-touch data. XGBoost is well-suited for this task due to its ability to handle complex, non-linear relationships in structured data, such as the pressure points and contact patterns that characterize ear-touch interactions.

The XGBoost model achieved an APCER of 2.3%. This low APCER indicates that the model was highly effective in distinguishing between genuine and fake ear-touch samples.

The accuracy of the model suggests that it can reliably detect presentation attacks, even when the fake samples are designed to closely mimic genuine interactions.

The strong performance of the XGBoost model on this dataset highlights its capability in handling the nuanced differences between real and fake ear touches. The low APCER of 2.3% is particularly noteworthy, as it suggests that the model is highly sensitive to subtle variations in the touch data that may indicate a spoofing attempt.

The model's success across the gender-diverse dataset indicates that it performs consistently regardless of the participant's sex, further underscoring its robustness.

6.4.4. Evaluation of fusion PAD

In this section, we present the results and discuss the findings of our performance analysis for the fusion ear-touch and ear-photo presentation attack detection system. The system's effectiveness in distinguishing between real and fake ear instances is evaluated using various mentioned performance metrics.

The results presented in Table 6-4 provide a detailed comparison of the APCER for a fusion model that combines ear-photo and ear-touch data, compared to a model that uses only ear-photo data. This evaluation focuses on the model's ability to detect presentation attacks across various combinations of display devices, recognition devices for ear-photo, and touchscreens used to create fake ear-touch samples.

Table 6-4: The APCER for models on different type of attacks for ear-photo and ear-touch using fusion vision transformer model classification

Display device and recognition device for ear-photo	Touch-screen to make fake ear-touch	Abbreviation	APCER (%)	
			Fusion	Only photo
Dell Ultra Sharp 32 and Samsung Galaxy S9	Samsung Galaxy S10+	Dell-GS9-S10+	3.02	3.11
Dell Ultra Sharp 32 and Nokia Lumia 1020	Samsung Galaxy S10+	Dell-NL1020-S10+	2.9	3.2
Brother printer and Samsung Galaxy A7	Samsung Galaxy S10+	BP-GA7-S10+	0.9	0.9

In Dell Ultra Sharp 32 with Samsung Galaxy S9 and Samsung Galaxy S10+ scenario, the display device is a Dell Ultra Sharp 32 monitor paired with a Samsung Galaxy S9 as the recognition device for ear-photo, while a Samsung Galaxy S10+ is used to generate fake ear-touch data. The fusion model achieved an APCER of 3.02%, which is slightly lower than the 3.11% APCER observed when using only the ear-photo data. The slight improvement in APCER when using the fusion model indicates that the inclusion of ear-touch data provides additional discriminatory power, helping to detect subtle differences that might not be captured by the ear-photo data alone. This result suggests that the fusion model is more robust in this

attack scenario, likely because the tactile information from the fake ear-touch adds a layer of verification that helps to better distinguish between genuine and fake samples.

Dell Ultra Sharp 32 with Nokia Lumia 1020 and Samsung Galaxy S10+ scenario involves the Dell Ultra Sharp 32 monitor and a Nokia Lumia 1020 as the recognition device for ear-photo, with the same Samsung Galaxy S10+ used for creating fake ear-touch data. The fusion model records an APCER of 2.9%, compared to 3.2% when relying solely on the ear-photo data. The reduction in APCER when using the fusion model indicates a similar trend to the first scenario, where the combination of ear-photo and ear-touch data enhances the model's ability to detect presentation attacks. The tactile information likely contributes to identifying inconsistencies in the attack that the visual data alone might miss, leading to a more accurate detection.

In Brother Printer with Samsung Galaxy A7 and Samsung Galaxy S10+ scenario, a Brother printer is used to produce fake ear-photos, with a Samsung Galaxy A7 as the recognition device, while fake ear-touch data is generated using the Samsung Galaxy S10+. Both the fusion model and the model using only ear-photo data achieve the same APCER of 0.9%.

The equal APCER for both models suggests that in this particular setup, the ear-photo data alone is sufficient to detect the presentation attacks effectively. The high-quality of the printed fake images combined with the specific characteristics of the recognition device might result in an attack that is easier to detect visually, making the additional ear-touch data less impactful in this scenario. However, the consistency of the results across both models underscores the overall effectiveness of the detection system.

Across the different scenarios, the fusion model consistently shows a slight improvement in APCER compared to the model that relies solely on ear-photo data. This suggests that the fusion approach enhances the model's overall detection capabilities by incorporating complementary information from both modalities.

The effectiveness of the fusion model in reducing APCER across most scenarios suggests that it is a valuable approach for enhancing the security of biometric systems, particularly when facing increasingly sophisticated presentation attacks.

While the fusion model generally performs better, the marginal difference in APCER between the fusion and photo-only models in some scenarios indicates that the added value of ear-touch data may vary depending on the specific devices and attack methods used. In some cases, the

ear-photo data might be strong enough on its own, especially when the visual aspects of the attack are easily detectable.

6.5. Conclusion

In this chapter, we have explored the design, implementation, and evaluation of a novel fusion-based PAD system that integrates ear-photo and ear-touch modalities. The motivation behind this approach stems from the increasing sophistication of ear spoofing attacks, including mask, photo, and synthetic 3D attacks, where attackers leverage high-quality images and videos to deceive biometric systems. Traditional PAD methods—whether based on image quality, texture, hardware, or deep features—have shown varying degrees of effectiveness, but the evolving nature of attacks necessitates more robust and comprehensive solutions.

Previous research, such as the work by Alireza et al. on light field-based methods for ear PAD, has demonstrated that leveraging advanced imaging techniques can significantly improve detection stability and performance. However, the limitations posed by small datasets in these studies underscore the need for more extensive databases and enhanced detection strategies. In response to these challenges, we collected a larger, more diverse ear PAD database and proposed a fusion methodology based on vision transformers, specifically the DeiT model, combined with ear-touch data.

The fusion model's evaluation, which integrated ear-photo and ear-touch data, demonstrated promising results across various attack scenarios. The inclusion of ear-touch data consistently led to a slight reduction in the APCER compared to models relying solely on ear-photo data. This improvement, although marginal in some cases, highlights the added value of multimodal approaches in enhancing the overall security and reliability of biometric systems. The tactile information from ear-touch data proved to be particularly beneficial in scenarios where visual cues alone might not suffice to detect sophisticated attacks.

For example, in the scenario involving a Dell Ultra Sharp monitor paired with a Samsung Galaxy S9 and a Samsung Galaxy S10+ for generating fake ear-touch data, the fusion model achieved an APCER of 3.02%, slightly better than the 3.11% obtained from using only ear-photo data. Similarly, the fusion model showed an improved APCER of 2.9% versus 3.2% in a scenario involving a Nokia Lumia 1020 as the recognition device. These results suggest that the integration of tactile data can provide additional discriminatory power, helping to identify subtle inconsistencies that may be overlooked when relying solely on visual data.

However, it is important to note that the effectiveness of the fusion model varies depending on the attack scenario. In some cases, such as when using a Brother printer and Samsung Galaxy A7 for generating fake ear-photos, the APCER remained unchanged at 0.9% for both the fusion and photo-only models. This indicates that in certain scenarios, the ear-photo data alone might be sufficiently strong, especially when the quality of the printed images is high and the recognition device is effective in capturing distinguishing details.

7. Summary

This thesis delves into the rapidly evolving field of biometric authentication, focusing on the use of ear-based modalities—specifically ear photos and ear-touch—as viable methods for securing mobile devices. With the increase of mobile technology, there is an increasing demand for reliable, user-friendly, and secure biometric systems that can be smoothly and continuously integrated into everyday devices. Ear biometrics, due to their unique anatomical features and non-intrusive nature, present a promising solution to this demand. However, like all biometric systems, ear-based authentication is vulnerable to presentation attacks, where fraudulent representations of biometric traits are used to deceive the system. This thesis not only investigates the feasibility of using ear photos and ear-touch for authentication but also addresses the significant challenge of PAD in mobile contexts.

The research is grounded in the understanding that while ear biometrics offer unique advantages—such as being contactless and relatively stable over time—there are inherent challenges, particularly in ensuring that the system can reliably distinguish between genuine and fake inputs. This necessitates a robust framework that can handle the complexities of real-world applications, where variations in lighting, device quality, and user behavior can all impact the effectiveness of the biometric system.

The thesis is structured to provide a comprehensive exploration of these issues, starting with the foundational aspects of ear biometrics, moving through the development of specialized datasets, and culminating in the design and evaluation of advanced recognition and PAD algorithms. Each chapter contributes to building a holistic understanding of how ear-based biometrics can be effectively implemented in mobile devices, what challenges arise in this process, and how these challenges can be addressed through innovative solutions.

The introduction of this thesis delves into the rapidly growing field of biometric technologies, emphasizing their increasing importance in enhancing security and user convenience across a variety of platforms, particularly within the realm of mobile devices. Biometric systems have

become integral to modern security frameworks, offering unique advantages over traditional authentication methods such as passwords or PINs, which are vulnerable to being forgotten, stolen, or hacked. Among the wide array of biometric modalities, ear photo and ear-touch biometrics have emerged as particularly promising identifiers, owing to their distinctive and consistent characteristics.

The human ear, like fingerprints and irises, possesses unique physical attributes that can be leveraged for biometric recognition. Unlike other more commonly used biometric traits, such as facial recognition or fingerprint scanning, ear-based biometrics offer certain advantages. Ear shapes are less likely to change significantly over time and are not easily influenced by external factors like facial expressions or skin conditions.

This thesis focuses specifically on the use of ear photo and ear-touch biometrics in mobile device authentication. With the widespread adoption of smartphones and other portable devices, there is a growing need for secure, yet convenient, authentication methods. Ear biometrics present a non-intrusive, user-friendly option that can be smoothly and continuously integrated into the everyday use of mobile devices.

However, as with any biometric system, the implementation of ear-based biometrics is not without its challenges. One of the primary concerns is the potential for presentation attacks—attempts by unauthorized individuals to deceive the biometric system using fake or altered biometric data. These attacks, which can involve anything from printed images of ears to sophisticated 3D models, pose a significant threat to the reliability and security of biometric authentication systems. Addressing these concerns is critical for the development of robust biometric systems that can effectively protect against unauthorized access.

In response to these challenges, this thesis seeks to explore and advance the field of ear-based biometrics, with a particular focus on PAD. PAD is an essential aspect of biometric security, as it involves detecting and mitigating attempts to deceive the biometric system. By developing effective PAD techniques specifically tailored for ear biometrics in mobile contexts, this research aims to contribute to the creation of more secure and reliable biometric systems.

The introduction sets the stage for the subsequent chapters by outlining the key topics covered in this thesis. These include an in-depth examination of the uniqueness and reliability of ear photo and ear-touch as biometric identifiers, the critical importance of PAD in ensuring the security of these systems, and the new challenges that have emerged in the field of ear PAD

research. The thesis also discusses the broader implications of ear-based biometrics for the field of mobile device security, considering both the technical aspects of implementation and the ethical considerations related to the use of biometric data.

Moreover, the introduction provides a clear statement of the thesis's objectives, which include not only the exploration of ear photo and ear-touch biometrics but also the development of new datasets and algorithms to enhance the effectiveness of these methods. The scope of the research is carefully defined, include both the theoretical and practical aspects of ear biometric recognition and PAD. Additionally, the structure of the thesis is outlined, with each chapter building upon the previous one to provide a comprehensive examination of the topic.

One of the primary challenges in advancing ear biometric technologies has been the lack of comprehensive datasets tailored for this purpose, especially datasets that consider the mobile device environment, where such technologies are increasingly deployed. Chapter 2 delves into the creation and detailed exploration of the Multimodal Mobile Biometric Database, specifically named Warsaw University of Technology Ear Version 1.0 (WUT-Ear V1.0). The database was meticulously developed to fill this critical gap and to support the research and development of secure and reliable ear verification and PAD systems for mobile applications.

The motivation behind creating the WUT-Ear V1.0 database stems from the growing need for more reliable biometric authentication methods in mobile devices, where security concerns are the most important. Recognizing the absence of a dedicated PAD dataset in the realm of ear biometrics, this research undertook the task of gathering a comprehensive and diverse collection of ear images, along with data that includes both authentic samples and a variety of presentation attack instruments (PAIs). The goal was to simulate real-world conditions under which biometric systems operate and to provide a resource that could be used to rigorously test and improve PAD systems.

The WUT-Ear V1.0 database encompasses a wide range of ear images collected from diverse individuals using various mobile devices, ensuring a broad representation of the population and the conditions under which ear biometrics might be employed. This database is not limited to authentic ear images alone but also includes a significant number of fake samples, created using different techniques such as printed images and 3D models, which are common methods employed in presentation attacks. The inclusion of these fake samples is crucial for developing PAD systems that can accurately distinguish between legitimate users and potential attackers.

The creation of the WUT-Ear V1.0 database involved several stages, each designed to ensure the comprehensiveness and utility of the database for biometric research. The data collection process was meticulously planned to include a diverse set of participants, ensuring a balanced representation of different genders, ages, and ethnic backgrounds. This diversity is vital for creating biometric systems that are fair and effective across different demographic groups. The data collection was conducted using a variety of mobile devices, reflecting the range of environments in which ear biometrics might be deployed.

The database is composed of four key sub-datasets, each focusing on a different aspect of ear biometrics and PAD. These include the ear real photo dataset, the ear fake photo dataset, the ear real touch dataset, and the ear fake touch dataset. Each sub-dataset was carefully curated to address specific challenges in biometric recognition and PAD. For instance, the ear real photo dataset consists of high-quality images of participants' ears captured under controlled conditions, while the ear fake photo dataset includes images designed to mimic real ears using various spoofing techniques. Similarly, the ear real touch dataset comprises tactile data captured when participants pressed their ears against a mobile device's touchscreen, and the ear fake touch dataset includes tactile data generated using synthetic models designed to deceive the recognition system.

The methodologies employed in collecting these datasets were diverse, reflecting the need to simulate a wide range of real-world scenarios. The data collection for ear photos involved capturing images under various lighting conditions and angles, which is critical for developing robust recognition algorithms that can operate effectively in different environments. For the tactile data, multiple sessions were conducted to account for variations in how individuals press their ears against a device, ensuring that the dataset captures the natural variability present in real-world usage.

In addition to the data collection, significant effort was invested in preprocessing the data to make it suitable for use in machine learning models. This preprocessing involved standardizing the image resolutions, normalizing the tactile data, and ensuring that all data was anonymized to protect the privacy of the participants. Ethical considerations were paramount throughout the process, with strict measures implemented to ensure compliance with data protection regulations. This included obtaining informed consent from all participants, anonymizing data to prevent the identification of individuals, and implementing security measures to safeguard the dataset against unauthorized access.

The WUT-Ear V1.0 database represents a significant advancement in the field of ear biometrics and PAD. It provides a comprehensive resource for researchers and developers, enabling them to create and test more secure and reliable biometric systems. The inclusion of both real and fake samples, as well as the balanced representation of male and female participants, ensures that the dataset can be used to develop algorithms that are both effective and fair across different demographic groups.

Chapter 3 of the thesis delves into the complexity of ear authentication on mobile devices, with a particular focus on the development and assessment of PAD and recognition algorithms. The motivation behind this research stems from the growing interest in leveraging biometric modalities for secure and convenient user authentication on mobile platforms. The security of ear authentication systems faces significant challenges, particularly from presentation attacks, where attackers attempt to deceive the system by presenting fake or altered ear images.

The chapter begins by providing a comprehensive overview of the current state of ear authentication technology, emphasizing its advantages over other biometric modalities. The unique anatomical features of the ear, such as its shape, contours, and texture, make it an excellent candidate for biometric recognition. Unlike facial recognition, which can be affected by changes in expression or lighting conditions, ear recognition offers a higher degree of consistency, making it more reliable for mobile authentication purposes.

One of the primary concerns is the vulnerability to presentation attacks, where adversaries use various techniques to fool the system into accepting a fake ear image as genuine. These attacks can be executed using different methods, such as printed images of the ear, replay attacks with pre-recorded videos, or even more sophisticated techniques involving 3D models or masks. The chapter discusses these potential threats in detail, highlighting the need for robust PAD mechanisms to ensure the security and reliability of ear authentication systems on mobile devices.

To address these challenges, the research conducted in this chapter investigates several PAD and recognition algorithms specifically tailored for ear authentication on mobile devices. The newly introduced dataset, which contains a diverse set of ear images with various types of presentation attacks, serves as the foundation for evaluating the effectiveness of these algorithms. The dataset includes both genuine ear images and a wide array of presentation attack instruments, allowing for a comprehensive analysis of the system's ability to distinguish between real and fake inputs.

The research methodology involves a systematic evaluation of existing PAD and recognition algorithms on this new dataset. The performance of these algorithms is assessed based on several key metrics, including the APCER, which measures the system's ability to detect attacks, and the BPCER, which evaluates the accuracy of genuine recognition.

One of the critical findings of this research is the variability in algorithm performance depending on the type of presentation attack and the quality of the device used for recognition. For example, the results indicate that fine-tuned deep neural networks are particularly effective in detecting certain types of attacks, achieving APCER values as low as 1% in the best conditions. However, the effectiveness of these algorithms varies significantly across different devices, with lower-quality cameras or sensors leading to higher error rates. This highlights the importance of considering the hardware capabilities of mobile devices when developing and deploying ear authentication systems.

Chapter 4 of the thesis delves into the innovative concept of ear-touch as a biometric feature for mobile user authentication, presenting it as a novel and cost-effective alternative to more conventional biometrics like fingerprints, facial recognition, and iris scans. The chapter begins by discussing the growing need for secure and reliable authentication methods in mobile devices, which have become ubiquitous in daily life. Traditional biometric systems, while effective, often require additional hardware, driving up the cost and complexity of mobile devices. This chapter proposes ear-touch as a solution that leverages existing touchscreen technology without the need for specialized sensors, making it both accessible and economical.

The concept of ear-touch involves capturing the unique touch patterns and contours of an individual’s ear as it makes contact with a smartphone's multi-touch screen. Unlike earprints, which require high-resolution sensors and often involve detailed image capture, ear-touch relies on the simpler mechanism of recording multiple contact points across the screen. This method captures the essential features of an individual's ear through a series of touches, which are then used for authentication purposes. The process is similar to how fingerprints are used but offers a different set of advantages, particularly in terms of cost and ease of integration with existing mobile technology.

One of the key challenges addressed in this chapter is the issue of "missing points" in ear-touch data. Missing points occur due to the physical characteristics of the ear and the varying ways in which individuals press their ears against the screen. Unlike fingerprints, which are typically captured in a consistent manner, ear-touch data can vary significantly from one instance to the

next, depending on factors such as pressure, angle, and position. This variability presents a significant challenge for accurate recognition and requires sophisticated algorithms to account for and correct these inconsistencies.

The research introduces a method for capturing and processing ear-touch data, designed to address the variability and potential gaps in the captured points. By focusing on the consistent elements of the ear's structure, the proposed system can extract meaningful biometric data even when some points are missing. This method involves a series of preprocessing steps to normalize the data, followed by the application of an algorithm that matches the ear-touch data against stored templates. The system's ability to handle missing points effectively is critical for its success, ensuring that the authentication process remains reliable even under less-than-ideal conditions.

To validate the effectiveness of ear-touch as a biometric feature, the research includes the development of a comprehensive dataset comprising 92 subjects and 960 images of ear-touches. This dataset serves as the foundation for testing and refining the proposed method. The research applies a matching algorithm previously established in the literature, adapting it to the unique requirements of ear-touch recognition. The results demonstrate that the method achieves a remarkably low EER of 0.04, indicating a high level of accuracy and reliability.

Chapter 5 of the thesis explores the innovative approach of fusing ear-touch and ear-photo recognition to create a more robust and reliable biometric authentication system. This chapter addresses the limitations often encountered by unimodal biometric systems, such as those relying solely on a single type of biometric data. These systems can struggle with issues like noise, limited degrees of freedom, and higher error rates due to the variability and non-universality of biometric traits. The research presented in this chapter builds on the premise that by combining multiple biometric modalities, these limitations can be mitigated, resulting in a more accurate and effective recognition system.

The fusion of ear-touch and ear-photo recognition leverages the complementary strengths of these two biometric modalities. Ear photos capture the visual details of the ear's unique structure, while ear-touch data records the concrete patterns created when the ear makes contact with a smartphone's touchscreen. Individually, each modality has its advantages—ear photos provide rich visual information that is highly distinctive, while ear-touch offers an easily collectible biometric measure. However, when used in isolation, each modality also has inherent weaknesses. For example, ear photos can be vulnerable to variations in lighting, angle,

and image quality, while ear-touch data can be affected by inconsistencies in how users press their ears against the screen. By integrating these two modalities, the fusion approach aims to enhance overall recognition performance by compensating for the weaknesses of each individual modality.

To develop and validate the fusion model, the research involved the meticulous collection of a comprehensive dataset that includes both ear photos and ear-touch samples. This dataset forms the basis for training and testing the fusion algorithms. The research employed advanced deep learning techniques, specifically utilizing a VGG16-SiaNet fusion model, which was designed to optimize the integration of visual and sensible biometric features. The model was trained to recognize patterns and correlations between the two types of data, enabling it to make more accurate and reliable identity verifications.

The experimental results presented in the chapter demonstrate that the fusion-based approach significantly outperforms unimodal recognition methods. The fusion model consistently achieves lower error rates, as evidenced by reduced FMR and FNMR, compared to systems that rely solely on either ear-photo or ear-touch data. This improvement is particularly notable in scenarios where one modality might be compromised—such as when the quality of an ear photo is degraded due to poor lighting conditions or when ear-touch data is incomplete or inconsistent. In such cases, the complementary modality can provide the necessary information to correct or compensate, leading to a more accurate overall recognition outcome.

Moreover, the results of the research suggest that the fusion approach is particularly advantageous in enhancing the security and reliability of biometric systems in environments where traditional unimodal biometrics may fall short. The reduction in EER and the favorable DET curves observed in the experiments indicate that the fusion model is better equipped to handle the variability and unpredictability inherent in real-world biometric data. This robustness is crucial for deploying biometric systems in diverse and dynamic environments, such as those encountered in mobile device usage.

Chapter 6 of the thesis delves into the complex and critical issue of PAD by exploring a fusion-based approach that integrates both ear-touch and ear-photo modalities. As biometric systems become more widely adopted, they face increasingly sophisticated spoofing attempts, where attackers use high-quality images, videos, or even 3D models to deceive recognition systems. Traditional PAD methods have made significant strides in countering these threats, but the

evolving nature of presentation attacks necessitates the development of more advanced, robust, and comprehensive detection strategies.

This chapter begins by examining the different types of spoofing attacks that can target ear biometrics, such as mask attacks, photo attacks, and synthetic 3D models. These attacks exploit the visual and structural features of the ear, attempting to create replicas that are convincing enough to fool biometric systems. The quality of these spoofing attempts can vary significantly depending on the tools and techniques used, such as the resolution of the images or the precision of 3D models, making it challenging for traditional PAD methods to consistently detect and mitigate these threats.

In response to these challenges, the research in this chapter proposes a novel PAD system that fuses ear-photo and ear-touch data, leveraging the strengths of both modalities to improve detection accuracy. The motivation behind this fusion approach lies in the complementary nature of visual and tactile data. Ear photos provide detailed visual information that can be analyzed for inconsistencies in texture, lighting, and structure, while ear-touch data offers tactile information that reflects the physical interaction between the ear and the device. By combining these two data sources, the fusion model can detect subtle anomalies that might be missed when using a single modality.

The research introduces an advanced fusion methodology based on vision transformers, specifically utilizing the DeiT model, which is known for its effectiveness in processing visual data. The chapter details the process of collecting a new, larger ear PAD database, which includes a wide range of attack scenarios and devices, to train and test the fusion model. This comprehensive dataset is essential for ensuring that the model can generalize well to different types of attacks and perform reliably across diverse conditions.

The evaluation of the fusion model demonstrates its superiority over traditional, unimodal PAD methods. The inclusion of ear-touch data consistently leads to improved detection accuracy, as measured by the APCER. For example, in scenarios where high-quality fake ear photos are generated using sophisticated tools, the fusion model outperforms models that rely solely on ear-photo data, detecting subtle discrepancies that indicate a spoofing attempt. The sensible data from ear-touch interactions provides an additional layer of security, as it captures physical characteristics that are difficult to replicate accurately with synthetic materials.

The chapter also discusses specific case studies where the fusion model was tested against various attack scenarios. In one such scenario, a combination of a Dell UltraSharp monitor and a Samsung Galaxy S9 was used to generate fake ear photos. The fusion model achieved an APCER of 3.02%, which is a slight improvement over the 3.11% APCER obtained when using only ear-photo data. In another scenario involving a Nokia Lumia 1020, the fusion model further demonstrated its effectiveness by reducing the APCER to 2.9%, compared to 3.2% for the photo-only model. These results underscore the value of integrating tactile data to enhance the model's resilience against sophisticated spoofing techniques.

However, the chapter also acknowledges that the effectiveness of the fusion model can vary depending on the specific attack and the quality of the spoofing materials. For instance, in a scenario where high-quality printed images were used with a Brother printer and a Samsung Galaxy A7, the APCER remained unchanged at 0.9% for both the fusion and photo-only models. This suggests that in certain situations, the visual data alone may be sufficiently strong, particularly when the spoofing materials are of high quality and the recognition device is capable of capturing detailed and distinguishing features.

The thesis presents a comprehensive study of ear-based biometrics, particularly focusing on the challenges and advancements in ear PAD for mobile devices. Through the development of a new dataset and the exploration of innovative recognition and PAD algorithms, the research contributes to the field by offering robust solutions for mobile authentication. The integration of ear-touch and ear-photo modalities proves to be a promising approach, improving the accuracy, reliability, and security of biometric systems, while also addressing practical concerns such as cost and user accessibility.

Reference

[1] M. Choraś, "Ear biometrics based on geometrical feature extraction," in *Progress in Computer Vision and Image Analysis*: World Scientific, 2010, pp. 321-338.

[2] M. Choraś, "Perspective methods of human identification: ear biometrics," *Opto-electronics review,* vol. 16, no. 1, pp. 85-96, 2008.

[3] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Computer vision and image understanding,* vol. 189, p. 102805, 2019.

[4] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, and A. Ross, "Long range iris recognition: A survey," *Pattern Recognition,* vol. 72, pp. 123-143, 2017.

[5] Ž. Emeršič, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing,* vol. 255, pp. 26-39, 2017.

[6] Z. Zhao and A. Kumar, "Improving periocular recognition by explicit attention to critical regions in deep neural network," *IEEE Transactions on Information Forensics and Security,* vol. 13, no. 12, pp. 2937-2952, 2018.

[7] S. Zhao, J. Luo, and S. Wei, "A hybrid eye movement recognition system using ASM algorithm and Kalman filtering," *Journal of Intelligent & Fuzzy Systems,* no. Preprint, pp. 1-11.

[8] E. Vazquez-Fernandez and D. Gonzalez-Jimenez, "Face recognition for authentication on mobile devices," *Image and Vision Computing,* vol. 55, pp. 31-33, 2016.

[9] L. A. a. L. M. Esther Gonzalez. *AMI Ear Database*. [Online]. Available: https://webctim.ulpgc.es/research_works/ami_ear_database/

[10] Ž. Emeršič *et al.*, "The unconstrained ear recognition challenge 2019," in *2019 International Conference on Biometrics (ICB)*, 2019: IEEE, pp. 1-15.

[11] H. Nejati, L. Zhang, T. Sim, E. Martinez-Marroquin, and G. Dong, "Wonder ears: Identification of identical twins from ear images," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012: IEEE, pp. 1201-1204.

[12] X.-N. Xu, Z.-C. Mu, and L. Yuan, "Feature-level fusion method based on KFDA for multimodal recognition fusing ear and profile face," in *2007 International Conference on Wavelet Analysis and Pattern Recognition*, 2007, vol. 3: IEEE, pp. 1306-1310.

[13] T. Theoharis, G. Passalis, G. Toderici, and I. A. Kakadiaris, "Unified 3D face and ear recognition using wavelets on geometry images," *Pattern Recognition,* vol. 41, no. 3, pp. 796-804, 2008.

[14] S. Kokal, M. Vanamala, and R. Dave, "Deep Learning and Machine Learning, Better Together Than Apart: A Review on Biometrics Mobile Authentication," *Journal of Cybersecurity and Privacy,* vol. 3, no. 2, pp. 227-258, 2023.

[15] S. Dargan and M. Kumar, "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities," *Expert Systems with Applications,* vol. 143, p. 113114, 2020.

[16] Ž. Emeršič, V. Štruc, and P. J. N. Peer, "Ear recognition: More than a survey," vol. 255, pp. 26-39, 2017.

[17] M. Choras, "Image feature extraction methods for ear biometrics--a survey," in *6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM'07)*, 2007: IEEE, pp. 261-265.

[18] A. Sepas-Moghaddam, F. Pereira, and P. L. Correia, "Ear presentation attack detection: Benchmarking study with first lenslet light field database," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018: IEEE, pp. 2355-2359.
[19] J. Nourmohammadi-Khiarak and A. Pacut, "An ear anti-spoofing database with various attacks," in *2018 International Carnahan Conference on Security Technology (ICCST)*, 2018: IEEE, pp. 1-5.
[20] İ. Toprak, Ö. J. S. Toygar, Image, and V. Processing, "Ear anti-spoofing against print attacks using three-level fusion of image quality measures," pp. 1-8, 2019.
[21] K. Annapurani, M. Sadiq, and C. Malathy, "Fusion of shape of the ear and tragus–a unique feature extraction method for ear authentication system," *Expert Systems with Applications,* vol. 42, no. 1, pp. 649-656, 2015.
[22] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *Ieee Access,* vol. 2, pp. 1530-1552, 2014.
[23] C. Sousedik and C. Busch, "Presentation attack detection methods for fingerprint recognition systems: a survey," *Iet Biometrics,* vol. 3, no. 4, pp. 219-233, 2014.
[24] R. Raghavendra and C. Busch, "Presentation attack detection algorithm for face and iris biometrics," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, 2014: IEEE, pp. 1387-1391.
[25] M. Choras, "Human Identification Based on Ear Image Analysis," PhD, University of Technology and Life Sciences Bydgoszcz, Bydgoszcz, Poland, 2005.
[26] M. Choraś, "Image processing methods in person identification applications-ear biometrics," *Przegląd Elektrotechniczny,* vol. 81, no. 4, pp. 38-41, 2005.
[27] M. Choras and R. S. Choras, "Geometrical algorithms of ear contour shape representation and feature extraction," in *Sixth international conference on intelligent systems design and applications*, 2006, vol. 2: IEEE, pp. 451-456.
[28] Ö. T. İmren TOPRAK "Fusion of Full-Reference and No-Reference Anti-Spoofing Techniques for Ear Biometrics under Print Attacks," *International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES),* 2018.
[29] R. Raposo, E. Hoyle, A. Peixinho, and H. Proença, "UBEAR: A dataset of ear images captured on-the-move in uncontrolled conditions," in *2011 IEEE workshop on computational intelligence in biometrics and identity management (CIBIM)*, 2011: IEEE, pp. 84-90.
[30] H. Liu, "Force field convergence map and Log-Gabor filter based multi-view ear feature extraction," *Neurocomputing,* vol. 76, no. 1, pp. 2-8, 2012.
[31] C. R. Kumar, N. Saranya, M. Priyadharshini, and D. Gilchrist, "Face recognition using CNN and siamese network," *Measurement: Sensors,* vol. 27, p. 100800, 2023.
[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510-4520.
[33] K. Simonyan, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.
[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009: Ieee, pp. 248-255.
[35] J. He, Y. He, L. Zhai, and Y. Bi, "Self-supervised Siamese Networks with Squeeze-Excitation Attention for Ear Image Recognition," in *International Conference on Intelligent Computing*, 2024: Springer, pp. 122-133.
[36] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*, 2004: Springer, pp. 469-481.

[37] N. Boodoo-Jahangeer and S. Baichoo, "LBP-based ear recognition," in *13th IEEE International Conference on BioInformatics and BioEngineering*, 2013: IEEE, pp. 1-4.
[38] C. Li, W. Zhou, and S. J. T. V. C. Yuan, "Iris recognition based on a novel variation of local binary pattern," vol. 31, no. 10, pp. 1419-1429, 2015.
[39] C. Nwankpa, W. Ijomah, A. Gachagan, and S. J. a. p. a. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018.
[40] I. Goicoechea-Telleria, J. Liu-Jimenez, R. Sanchez-Reillo, and W. Ponce-Hemandez, "Vulnerabilities of Biometric Systems integrated in Mobile Devices: an evaluation," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, 2016: IEEE, pp. 1-8.
[41] B. Institute, "Information technology Biometric presentation attack detection Part 3: Testing and reporting," *INTERNATIONAL STANDARD ISO/IEC 30107-3,* vol. First edition, pp. 1-42, 2017.
[42] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, *Handbook of biometric anti-spoofing: Presentation attack detection*. Springer, 2019.
[43] L. Meijerman, A. Thean, G. J. F. s. Maat, medicine,, and pathology, "Earprints in forensic investigations," vol. 1, no. 4, pp. 247-256, 2005.
[44] A. Broeders, "Of earprints, fingerprints, scent dogs, cot deaths and cognitive contamination—a brief look at the present state of play in the forensic arena," *Forensic Science International,* vol. 159, no. 2-3, pp. 148-157, 2006.
[45] S. Halpin, "What have we got ear then: developments in forensic science: earprints as identification evidence at criminal trials," *UC Dublin L. Rev.,* vol. 8, p. 65, 2008.
[46] L. Meijerman *et al.*, "Individualization of earprints," vol. 2, no. 1, pp. 39-49, 2006.
[47] C. Holz, S. Buthpitiya, and M. Knaust, "Bodyprint: Biometric user identification on mobile devices using the capacitive touchscreen to scan body parts," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015: ACM, pp. 3011-3014.
[48] M. Maheshwari, S. Arora, A. M. Srivastava, A. Agrawal, M. Garg, and S. Prakash, "Earprint Based Mobile User Authentication Using Convolutional Neural Network and SIFT," in *International Conference on Intelligent Computing*, 2018: Springer, pp. 874-880.
[49] L. Meijerman *et al.*, "Exploratory study on classification and individualisation of earprints," vol. 140, no. 1, pp. 91-99, 2004.
[50] I. Alberink and A. J. F. s. i. Ruifrok, "Performance of the FearID earprint identification system," vol. 166, no. 2-3, pp. 145-154, 2007.
[51] A. Morales, M. Diaz, G. Llinas-Sanchez, and M. A. Ferrer, "Earprint recognition based on an ensemble of global and local features," in *2015 International Carnahan Conference on Security Technology (ICCST)*, 2015: IEEE, pp. 253-258.
[52] M. J. J. o. A. Atkinson, "An optimal algorithm for geometrical congruence," vol. 8, no. 2, pp. 159-172, 1987.
[53] H. Alt, K. Mehlhorn, H. Wagener, E. J. D. Welzl, and C. Geometry, "Congruence, similarity, and symmetries of geometric objects," vol. 3, no. 3, pp. 237-256, 1988.
[54] A. Pacut, "Alignment of the earprints," *Private,* Private pp. 1-14, 2021.
[55] A. Pacut, "Alignment of the earprints," *Private,* p. 14, 2021.
[56] Y. Ma, Z. Huang, X. Wang, and K. Huang, "An Overview of Multimodal Biometrics Using the Face and Ear," *Mathematical Problems in Engineering,* vol. 2020, 2020.
[57] M. Singh, R. Singh, and A. Ross, "A comprehensive overview of biometric fusion," *Information Fusion,* vol. 52, pp. 187-205, 2019.

[58] A. F. Abate, M. Nappi, S. J. I. T. o. S. Ricciardi, Man,, and C. Systems, "I-Am: Implicitly Authenticate Me—Person Authentication on Mobile Devices Through Ear Shape and Arm Gesture," vol. 49, no. 3, pp. 469-481, 2017.
[59] N. Hezil and A. Boukrouche, "Multimodal biometric recognition using human ear and palmprint," *IET Biometrics,* vol. 6, no. 5, pp. 351-359, 2017.
[60] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: an evaluation methodology and lessons learned," *IEEE Signal Processing Magazine,* vol. 32, no. 5, pp. 20-30, 2015.
[61] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Transactions on Information Forensics and Security,* vol. 11, no. 8, pp. 1818-1830, 2016.
[62] A. Sepas-Moghaddam, L. Malhadas, P. L. Correia, and F. Pereira, "Face spoofing detection using a light field imaging framework," *IET Biometrics,* vol. 7, no. 1, pp. 39-48, 2018.
[63] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," *IEEE transactions on information forensics and security,* vol. 15, pp. 42-55, 2019.
[64] G. B. De Souza, D. F. da Silva Santos, R. G. Pires, A. N. Marana, and J. P. Papa, "Deep texture features for robust face spoofing detection," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 64, no. 12, pp. 1397-1401, 2017.
[65] S. R. Arashloo, J. Kittler, and W. Christmas, "Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features," *IEEE Transactions on Information Forensics and Security,* vol. 10, no. 11, pp. 2396-2407, 2015.
[66] T. Chugh, K. Cao, and A. K. Jain, "Fingerprint spoof buster: Use of minutiae-centered patches," *IEEE Transactions on Information Forensics and Security,* vol. 13, no. 9, pp. 2190-2202, 2018.
[67] L. Ghiani, A. Hadid, G. L. Marcialis, and F. Roli, "Fingerprint liveness detection using binarized statistical image features," in *2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS)*, 2013: IEEE, pp. 1-6.
[68] Q. Li and P. P. Chan, "Fingerprint liveness detection based on binarized statistical image feature with sampling from Gaussian distribution," in *2014 International Conference on Wavelet Analysis and Pattern Recognition*, 2014: IEEE, pp. 13-17.
[69] X. Li, W. Bu, and X. Wu, "Palmprint liveness detection by combining binarized statistical image features and image quality assessment," in *Chinese Conference on Biometric Recognition*, 2015: Springer, pp. 275-283.
[70] H. Muckenhirn, P. Korshunov, M. Magimai-Doss, and S. Marcel, "Long-term spectral statistics for voice presentation attack detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 25, no. 11, pp. 2098-2111, 2017.
[71] T. Hugo, M. Cord, D. Matthijs, M. Francisco, S. Alexandre, and J. Herve, "Training data-efficient image transformers & distillation through attention," in *ICML*, 2021.
[72] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929,* 2020.
[73] B. Wu *et al.*, "Visual transformers: Token-based image representation and processing for computer vision," *arXiv preprint arXiv:2006.03677,* 2020.
[74] I.-Y. Kwak, J. H. Huh, S. T. Han, I. Kim, and J. Yoon, "Voice presentation attack detection through text-converted voice command analysis," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1-12.

Appendix A: Harmonized Biometrics Vocabulary

This Appendix aims at systematizing the biometrics-related terminology used in this Thesis. These terms are listed and defined below, with explanations provided from the ISO/IEC Information technology – Vocabulary – Part 37: Biometrics standard [67] with additional comments and extensions introduced by the Author where necessary.

Biometric characteristic: A biological or behavioral trait from which distinctive, repeatable features can be extracted for biometric recognition.

Biometric recognition: The automatic identification or verification of individuals using their biological or behavioral traits.

Biometric sample: An analog or digital capture of a biometric characteristic prior to feature extraction. *In this work, the biometric sample is an ear photo.*

Biometric feature: Numeric descriptors or labels derived from a biometric sample for comparison. *Here, "ear features" refer to information present in ear texture; when encoded for matching from an ear photo, we call this the biometric template.*

Biometric template: The stored feature representation used for direct comparison with probe features. *In this study, it is the mathematical encoding of an ear extracted from an ear photo by a specific processing method.*

Biometric data: A sample or any aggregation of samples and their derivatives (e.g., references, probes, features, properties). *In our context, this typically refers to sets of ear photos (and, where applicable, ear-touch data).*

Comparison decision: The outcome determining whether a probe and a reference originate from the same source, based on comparison scores, a decision policy (including a threshold), and possibly other inputs. A **match** is a positive decision; a **non-match** is negative. *Here, decisions are made by comparing two templates against a threshold.*

Threshold: A numerical value (or set of values) defining the decision boundary for acceptance versus rejection.

Comparison score: Numerical value (or set of values) resulting from a comparison.

Dissimilarity score, distance score: Comparison score that decreases with similarity.

Biometric enrolment: Act of creating and storing a biometric enrolment data record in accordance with an enrolment policy.

In this work, this term is understood as a process of creating the biometric template (a numerical representation) from a biometric sample (an ear photo) by performing biometric feature extraction.

Biometric feature extraction: Process applied to biometric sample with the intent of isolating and outputting repeatable and distinctive numbers or labels which can be compared to those extracted from other biometric samples.

Comparison: Estimation, calculation, or measurement of similarity or dissimilarity between biometric probe(s) and biometric reference(s).

In this work, a comparison is usually performed between two biometric templates, since we do not necessarily divide the templates into probe and reference.

Biometric presentation attack: Presentation to the biometric capture subsystem with the goal of interfering with the operation of the biometric system. In this work, a presentation attack usually refers to the presentation of fake printed ear photo.

Biometric identification: Process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual.

Biometric verification: Process of confirming a biometric claim through biometric comparison.

Failure to enroll, FTE: Failure to create and store a biometric enrolment data record.

In this work, failure to enroll refers to the inability to create a biometric template by a given ear recognition method.

Failure-to-enroll rate, FTE rate, FTER: Proportion of a specified set of biometric enrolment transactions that resulted in a failure to enroll.

False match: A match decision produced when the probe and reference come from different individuals. In this study, it refers to a match between two templates originating from different ears; also known as a false positive.

False Match Rate (FMR): The percentage of non-mated comparison attempts that incorrectly result in a match decision.

False non-match: A non-match decision produced when the probe and reference actually belong to the same individual. Here, it denotes a non-match between two templates of the same ear; also called a false negative.

False Non-Match Rate (FNMR): The percentage of genuine (mated) comparisons that are incorrectly classified as non-matches.

Appendix B: Pseudo-code

B.1: Pseudo-code 1: Alignments of ear-touch no missing points

Inputs: template T, ear-touch $X = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_{n-1} \\ x_n \end{bmatrix}$

Outputs:

mismatch between T and X based on Eq.2

best match of X w.r.t X (Eq.3)

Constraint:

$4 <= n <= 10$

mismatch = infinite

B = X

for each possible permutation P do

permute X according to P, and save it as S

use Pacut's method to solve $(R^*, l^*) = \underset{(R,l,P)\in\mathcal{R}\times\mathcal{L}}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| t_i - (Rx_{\pi_i} + l) \right\|^2$

$b_i^{aux} = R^* x_{\pi_i}^* + l^* \qquad i = 1,2,\ldots,N$

$B_{aux} \stackrel{\text{def}}{=} (b_1^{aux}, b_2^{aux}, \ldots, b_n^{aux})$

$min_{aux} = D_0(T, B_{aux})$

if $min_{aux} < mismatch$ then

mismatch = min_aux

$B = B_{aux}$

end for

return mismatch, B_{aux}

B.2: Pseudo-code 2: Template creation for no missing points

inputs: A series of N related ear-touches $Y = (Y^1, Y^2, \dots, Y^N)$

outputs: A template which minimizes the average mismatch to Y according to Eq.5

Constraints:

$4 <= N <= 10$

$k = 0$

$L_0 = infinite$

$T_0 = Y^1$

while $(k = 0)\ or\ (L_{k-1} - L_k > tol)$

$k = k + 1$

$B_k^j = BestMatch(T_{k-1}, Y^j) \quad j = 1,2,\dots,N$ // according to pseudo code 1

$t_{i,k} = \frac{1}{N}\sum_{j=1}^{N} b_{i,k}^j$

$L_k = \frac{1}{N} * \sum_{j=1}^{M} D_0(T_k, B_k^j)$

end while

return T_k

B.3: Pseudo-code 3: Matching in the presence of missing points

Inputs

- complete template $T = (t_1, t_2, \dots, t_N)$
- incomplete ear-touch $X' = (x_1, x_2, \dots, x_{N_{X'}})$

Outputs

- $D_2(T, X')$ based on Eq.6
- The pair (B, E) of incomplete best match that satisfies Eq.8

min = infinite

for each $P \in \mathcal{P}(N, N_{X'})$

Create a limited version of template as follows

$$s_i = t_{\pi_i} \quad i = 1,2, \dots, N_{X'}$$
$$S = (s_1, s_2, \dots, s_{N_{X'}})$$

find $(\bar{R}, \bar{l}) = \underset{(R,l)\in\mathcal{R}\times\mathcal{L}}{\operatorname{argmin}} \sum_{i=1}^{N_{X'}} \|s_i - (Rx_i + l)\|^2$ using Pacut's method

Store the virtually complete Best Match of X' w.r.t. S

$$c_i = \bar{R}x_i + l \qquad i = 1, 2, \dots, N_{X'}$$
$$C = (c_1, c_2, \dots, c_{N_{X'}})$$

mismatch = $D_0(S, C)$

if $mismatch < min$ then

min = mismatch

$$e_i = 0 \quad b_i = \binom{0}{0} \quad i = 1, 2, \dots, N$$

$$e_{\pi_i} = 1 \quad b_{\pi_i} = c_i \quad i = 1, 2, \dots, N$$

$$B = (b_1, b_2, \dots, b_{N_{X'}}) \quad E = (e_1, e_2, \dots, e_{N_{X'}})$$

end if

end for

return min, B, E

B.4: Pseudo-code 4: Creating template in presence of missing points

Inputs: A sequence of incomplete ear-touches $Y = (Y'_1, \dots, Y'_M)$ with corresponding existence sequences $E = (E'_1, E'_2, \dots, E'_M)$.

Outputs: An incomplete template which extends pseudo-code 2 for incomplete ear-touches.

Sort inputs (Y and E) by number of existing touch points in a descending order and Store them in original variables.

$k = 0 \quad L_0 = \infty \quad T_0 = Y_1' \quad E_0^T = E_1'$

while (k=0) or ($L_{k-1} - L_k > tol$)

k = k+1

$(B_k^j, E_k^j) = incBestMatch(T_{k-1}, E_0^T, Y_j', E_j') \quad j = 1,2, \dots, M$

$n_i = \sum_{j=1}^{M} e_{i,k}^j \quad t_i^k = \begin{cases} \frac{1}{n_i} \sum_{j=1}^{M} b_{i,k}^j & n_i > 0 \\ 0 & n_i = 0 \end{cases} \quad i = 1, 2, \dots, N$

$T_k = (t_{1,k}, t_{2,k}, \dots, t_{N,k})$

$L_k = \frac{1}{M} \sum_{j=1}^{M} D_0(T_k, B_k^j)$

end while

return T_k, E_0^T

B.5: Pseudo-code 5: Matching in the presence of missing points

Inputs

- N as the total number of existing touch points
- incomplete template $T' = (t_1, t_2, \dots, t_{N_{T'}})$
- incomplete ear-touch $X' = (x_1, x_2, \dots, x_{N_{X'}})$

Outputs

- The minimum mismatch between T' and X', considering all possible modes of common touch points

- The pair (B, E) of incomplete best match which satisfied the following:

$$attained\ mismatch = \sum_{i=1}^{N} e_i \|t_i - b_i\|^2$$

min $= \infty$ $N_c^{min} = N_{X'} + N_{T'} - N$ $N_c^{max} = N_{X'}$

for $N_c = N_c^{min} \dots N_c^{max}$

for each $Q \in \{X |\ X \subset \{1, 2, \dots, N_c^{max}\},\ N(X) = N_c\}$

Sort elements of Q in ascending order as $(q_1,\ q_2, \dots, q_{N_c})$

Create a limited version of X':

$$z_i = x\_q_i \qquad i = 1, 2, \dots, N_c$$
$$Z = (z_1, z_2, \dots, z_{N_c})$$

find the mismatch between Z and T' according to Pseudo-code 3

$$(mismatch, B', E') = Pseudocode3(T', Z)$$

if $mismatch < min$ then

$$e_i = e_i' \quad b_i = b_i' \quad i = 1, 2, \dots, N_c$$

$$e_i = 0 \quad b_i = \binom{0}{0} \quad i = N_c + 1\ , N_c + 2, \dots, N$$

$$B = (b_1, b_2, \dots, b_N) \quad E = (e_1, e_2, \dots, e_N)$$

end if

end for

end for

return min, B, E

Appendix C: List of Author's publications and achievements

This Appendix lists all journal and conference publications authored or co-authored by the Author of this Thesis. Percentage-wise contributions of each of the authors for each paper are given in blue.

To be submitted papers

Jalil Nourmohammadi Khiarak, Andrzej Pacut, "Transfer learning using deep neural networks for Ear Presentation Attack Detection: New Database for PAD", (To be submitted), 2025.

My contributions to this work include: a) Data collection for both real ear and fake ear, b) Writing c) Implementation d) Data analysis approx. 90% of the paper.

Jalil Nourmohammadi Khiarak, Andrzej Pacut, "Gaze Information and Pupil Dynamics: a survey of Liveness Detection and case study on PupilWUT database", (To be submitted), 2025.

My contributions to this work include: a) training the benchmark deep-learning-based model and generation of the occlusion masks from this method, b) writing approx. 95% of the paper, d) collecting data for this work.

Jalil Nourmohammadi Khiarak, Andrzej Pacut, "Combined Ear-touch and Ear Photo Verification and Presentation Attack Detection", (To be submitted), 2025.

My contributions to this work include: a) training the benchmark General Image Quality Assessment model and generation of the occlusion masks from this method, b) implementation of the ear normalization method, c) writing approx. 95% of the paper, d) collecting data for this work.

Jalil Nourmohammadi Khiarak, Andrzej Pacut, "Multimodal mobile biometric database (WUT-Ear V1.0): From Ear recognition system to Ear presentation attack detection", (To be submitted), 2025.

My contributions to this work include: a) writing approx. 95% of the paper, d) collecting data for this work.

Reviewed journal publications

Jalil Nourmohammadi Khiarak, Samaneh Mazaheri, and Rohollah Moosavi Tayebi. "Ear-Touch-Based Mobile User Authentication." Mathematics 12, no. 5 (2024): 752.

My contributions to this work include: a) data collection, b) writing, c) code implementations, d) data analysis, approx. 95% of the paper.

Jalil Nourmohammadi Khiarak, et al. "Exploring bias in sclera segmentation models: A group evaluation approach." IEEE Transactions on Information Forensics and Security 18 (2022): 190-205.

My contributions to this work include: a) data collection, c) code implementations, d) propose a method, approx. 5% of the paper.

Jalil Nourmohammadi Khiarak [70%], et al. "KartalOl: a new deep neural network framework based on transfer learning for iris segmentation and localization task—new dataset for iris segmentation." Iran Journal of Computer Science (2023): 1-13.

My contributions to this work include: a) training the benchmark deep-learning-based UNet neural network for iris recognition and iris localization, b) implementation of the iris normalization method, c) data collection for iris recognition.

Reviewed conference publications

Jalil Nourmohammadi Khiarak [90%], Andrzej Pacut [10%]. "An ear anti-spoofing database with various attacks." 2018 International Carnahan Conference on Security Technology (ICCST). IEEE, 2018.

My contributions to this work include: a) training the benchmark General Image Quality Assessment model and generation of the occlusion masks from this method, b) implementation of the ear normalization method, c) writing approx. 95% of the paper, d) collecting data for this work.

Jalil Nourmohammadi Khiarak [5%], et al. "Attacking a Smartphone Biometric Fingerprint System: A Novice's Approach." 2018 International Carnahan Conference on Security Technology (ICCST). IEEE, 2018.

My contributions to this work spoofing smartphone using fake fingerprint and reporting results.

Jalil Nourmohammadi Khiarak [5%], et al. (2020). "Ssbc 2020: Sclera segmentation benchmarking competition in the mobile environment". In 2020 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-10). IEEE.

My contributions to this work include: a) training the benchmark deep-learning-based UNet model and generation of the occlusion masks from this method, b) implementation of the sclera normalization method.

Jalil Nourmohammadi Khiarak [5%], et al. "The unconstrained ear recognition challenge 2019." 2019 International Conference on Biometrics (ICB). IEEE, 2019.

My contributions to this work include: a) training the benchmark deep-learning-based Siamese neural network for which the residual network architecture (ResNet-50) model, b) implementation of the ear normalization method.

Jalil Nourmohammadi Khiarak [5%], et al. "NIR Iris Challenge Evaluation in Non-cooperative Environments: Segmentation and Localization" 2021 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2021.

My contributions to this work include: a) training the benchmark deep-learning-based UNet neural network for iris recognition and iris localization, b) implementation of the iris normalization method, c) data collection for iris recognition.

Jalil Nourmohammadi Khiarak [5%], et al. "Exploring Bias in Sclera Segmentation Models: A Group Evaluation Approach." IEEE Transactions on Information Forensics and Security 18 (2022): 190-205.

My contributions to this work include: a) training the benchmark deep-learning-based UNet model and generation of the occlusion masks from this method, b) implementation of the sclera normalization method.

Appendix D: Grants and projects

European Union Horizon 2020 Grant: AMBER Project

- **Grant Title:** AMBER – Authentication Methodologies for Biometric Recognition
- **Grant Agreement No.:** 675087
- **Funding Programme:** European Union Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Actions (MSCA).
- **Project Duration:** January 1, 2017 – December 31, 2020
- **Consortium:** Five EU universities and seven industrial partners, fostering collaboration across academia and industry.
- **Objective:** To equip the next generation of researchers with the tools to design, investigate, and implement secure, efficient, and privacy-preserving biometric authentication systems.

My Role in AMBER: Early Stage Researcher (ESR 5)

- **Project Title:** Countermeasure Algorithms Against Subterfuge in Mobile Biometric Systems
- **Start Date:** December 2017
- **Objective:** To develop a robust set of countermeasures for mobile biometric systems operating in unsupervised environments, with a focus on PAD. The research addresses the security challenges posed by the increasing use of mobile devices for sensitive applications, such as mobile payment systems.
- **Research Focus Areas:**
    - Development of PAD methods tailored for mobile platforms, considering biometric modalities such as ear photo, and ear touches.
    - Fusion of PAD-related information to enable multimodal recognition and decision-making.
- **Expected Outcomes:**
    1. PAD methods optimized for single biometric modalities on mobile devices.
    2. A statistical framework for fusing PAD-related data and evaluating performance of single and fused biometric sources.

3. A working prototype demonstrating multimodal PAD capabilities on commercial mobile devices.

- **Secondments:**

University Group for Identification Technology (GUTI) at University Carlos III of Madrid: Two months dedicated to developing fusion methods for PAD information in multi-biometric mobile systems.

Impact of the Research:

This project directly contributed to advancing the field of biometric security by addressing critical challenges in mobile PAD. The results of this work have implications for enhancing the security and usability of mobile authentication systems, particularly in unsupervised environments.