Ph.D. Thesis

Mateusz Trokielewicz, M.Sc.

Iris recognition methods resistant to biological changes in the eye

Advisor
Adam Czajka, Ph.D., D.Sc.

WARSAW 2019
STATEMENT

The Author acknowledges contributions from co-authors of the papers relating to the topic of this doctoral dissertation: Adam Czajka and Piotr Maciejewicz. Precise descriptions of the contributions from each of the co-authors are included in the publication list in Appendix B.

This study had an institutional review board clearance (where applicable) and the ethical principles of the Helsinki Declaration were carefully followed.
AKNOWLEDGMENTS

Dr. Adam Czajka, University of Notre Dame, for invaluable counsel, support, and resoluteness as an Advisor

Dr. Piotr Maciejewicz, Medical University of Warsaw, for enthusiasm and contributions, which made this study possible

Prof. Andrzej Pacut, Warsaw University of Technology, for support and serving as formal Advisor during most of my PhD studies

Prof. Kevin Bowyer, University of Notre Dame, for his support and eagerness towards this work

and

my parents, Małgorzata Trokielewicz and Ireneusz Trokielewicz

Ewelina Bartuzi

and Marcel Młyńczak
Abstract

Iris recognition has served the society as a secure and fast way of personal authentication for more than 25 years. During this time, multiple new research challenges have been identified, one of them being the possibly degrading impact of biology-related changes in the human eye, such as those caused by ocular diseases, but also post-mortem decay processes. This doctoral dissertation aims at quantifying the negative influence that these changes may inflict on iris biometrics systems, learning what causes errors, and proposing countermeasures to neutralize them.

First, an examination of the reliability that current methodologies offer when confronted with difficult samples, analysis of errors, and their underlying reasons, are conducted. From this we learn that most of the erratic performance originate from an incorrectly executed image segmentation stage. However, even when segmentation is manually adjusted in post-mortem samples, not all of the matching accuracy is regained, indicating that biological processes can cause the iris features to be altered or even lost, especially for post-mortem samples collected over time horizon exceeding a couple of days.

The image segmentation problem is solved by introducing a novel method employing a deep convolutional neural network, which localizes the iris, while effectively masking out iris regions affected by severe disease or post-mortem changes. We show how to use such masks in a traditional iris recognition pipeline, which allows us to achieve matching performance superior to the existing state-of-the-art methodologies, allowing the reduction of equal error rate from 23.69% to 6.40% for post-mortem samples collected up to 60 hours after death, and from 18.73% to as low as 0.68% for samples collected no more than 24 hours after death for the OSIRIS method. A significant reduction of equal error rate from 8.90% to 1.73% was obtained for samples collected from disease-affected eyes.

In an attempt to regain some of the iris features that were altered during post-mortem decomposition, a new set of image filters is devised by combining typical Gabor wavelets with data-driven filters learned from the post-mortem images. This allows to further decrease the recognition error rates by even as much as one third – a reduction of equal error rate from 6.40% to 4.39% for the same capture time horizon.
Finally, the proposed system is complemented with the first known to the Author method for iris liveness detection in a post-mortem setting, offering 99% correct live and post-mortem presentations classification rate.

The Author hopes that the contributions of this doctoral dissertation will constitute a valuable addition to the state-of-the-art in iris recognition, and enable efficient and reliable identification in biologically challenging circumstances, while also possibly extending the applications of iris biometrics to new fields, such as forensic examinations.

**Keywords:** biometrics, iris recognition, reliability, ophthalmic disorders, post-mortem, presentation attack detection
Streszczenie

Biometria tęczówki już od ponad 25 lat buduje swoją pozycję wśród technik rozpoznawania i uwierzytelniania osób jako metoda bezpieczna i szybka. Okres ten pozwolił na identyfikację wielu nowych problemów badawczych, w tym ryzyka obniżonej skuteczności metod biometrii tęczówki w przypadku zmian biologicznych w oku, takie jak te wywołane chorobami narządu wzroku, jak również zmiany związane z procesami rozkładu tkank po śmierci człowieka. Przedstawiona rozprawa doktorska ma na celu zbadanie poziomu negatywnego wpływu wymienionych wyżej zmian na systemy biometrii tęczówki, analizę przyczyn błędów rozpoznawania oraz zaproponowanie metod przeciwdziałających spadkowi jakości działania algorytmów.

W ramach pierwszego z problemów badawczych wykonano analizę niezawodności istniejących metod w sytuacjach, gdy wykorzystywane są nietypowe próbki biometryczne oraz ocenę uzyskanych błędów i ich przyczyn. Pozwala to na wysunięcie wniosku, iż za większą część spadku jakości rozpoznania dla próbek pochodzących od oczu osób cierpiących na schorzenia okulistyczne oraz oczu osób zmarłych odpowiedzialny jest niepoprawny przebieg segmentacji obrazu. Jednakże, podczas eksperymentu obejmującego ręczne, poprawne wykonanie tego etapu przetwarzania obrazu tęczówki nie uzyskano stuprocentowej redukcji błędnego działania, co może sugerować, iż procesy biologiczne mogą zmieniać lub niszczyć cechy tęczówki, w szczególności dla próbek pobranych od zmarłych w wiele dni po śmierci.

W pracy zaproponowano sposób rozwiązania problemu błędnie działającej segmentacji z wykorzystaniem nowej metodyki opartej o głęboką splotową sić neuronową, która skutecznie lokalizuje tęczówkę przy jednoczesnym maskowaniu jej obszarów objętych zaawansowanymi zmianami chorobowymi lub pośmiertnymi. Pokazano sposób wykorzystania tego typu masek w tradycyjnym algorytmie rozpoznawania tęczówki, co pozwoliło na uzyskanie dokładności rozpoznawania przekraczającej tę uzyskiwaną przez dostępne obecnie metody, zarówno komercyjne, jak i o otwartym źródle. Uzyskano zmniejszenie częstości błędu zrównoważonego z 23.69% do 6.40% dla próbek pobranych w mniej niż 60 godzin po śmierci, oraz z 18.73% do zaledwie 0.68% dla próbek pobranych w mniej niż 24 godziny po śmierci.
Dla próbek pochodzących od oczu objętych procesami chorobowymi uzyskano spadek błędu zrównoważonego z 18.73% do 1.73%. Podane wartości dotyczą metody OSIRIS.

Podjęto również działania mające na celu odzyskanie części informacji zawartej w cechach tęczówki zmienionych na skutek działania procesów pośmiertnego rozkładu tkanek oka. W tym celu zaproponowano nowy zestaw filtrów, stanowiący połączenie tradycyjnie używanych w biometrii tęczówki filtrów Gabora z nowymi filtrami, uzyskanymi w procesie uczenia sieci neuronowej na podstawie obrazów tęczówek osób zmarłych. Ta modyfikacja etapu kodowania cech tęczówki pozwoliła na dalszą redukcję błędu zrównoważonego dla próbek pobranych w horyzontach czasowych dłuższych niż 24 godziny, dla przykładu dla wspomnianego wcześniej zestawu próbek pobranych do 60 godzin po śmierci błąd zmniejszono nawet o jedną trzecią – z uzyskanych wcześniej 6.40% do 4.39%.

Ostatnim elementem proponowanego systemu biometrii tęczówki jest pierwsza znana Autorowi pracy metodyka testowania żywotności w scenariuszu uwzględniającym prezentacje próbek pochodzących od osób zmarłych, która umożliwia odróżnianie takich próbek od tych pochodzących od osób żyjących z 99% skutecznością.

Autor wyraża nadzieję, że przedstawiona w poniższej rozprawie doktorskiej metodyka poprawy jakości działania systemów biometrii tęczówki dla trudnych próbek pochodzących od oczu objętych zmianami biologicznymi stanowić będzie wartościowy wkład w rozwój dziedziny, umożliwiając niezawodną identyfikację w nawet tak trudnych okolicznościach, jak również poszerzy możliwy krąg zastosowań rozpoznawania tęczówki na nowe obszary, takie jak wykorzystanie w kryminalistyce.

Słowa kluczowe: biometria, rozpoznawanie tęczówki, niezawodność biometrii, choroby oka, identyfikacja osób zmarłych, testowanie żywotności
## Contents

**List of Abbreviations** ................................................................. 15  

1. **Introduction** ................................................................. 17  
   1.1. Iris recognition and human biology ........................................ 17  
   1.2. Theses ............................................................................. 18  
   1.3. Iris as a biometric characteristic ........................................... 18  
   1.4. New challenges in iris biometrics research ............................ 21  
   1.5. Scope of this Thesis ......................................................... 24  
   1.6. Thesis layout ................................................................. 24  

2. **Influence of ophthalmic disorders** ........................................ 26  
   2.1. Iris recognition in the presence of ocular pathologies ............ 26  
   2.2. Related work ..................................................................... 28  
   2.3. Medical background ........................................................... 33  
   2.4. Dataset of iris images ........................................................... 36  
   2.5. Experimental methodology and tools .................................... 39  
   2.6. Results and discussion ....................................................... 43  
   2.7. Conclusions ................................................................. 51  

3. **Post-mortem iris recognition** ................................................. 53  
   3.1. Iris recognition after death ............................................... 53  
   3.2. Related work ................................................................. 54  
   3.3. Dataset of cadaver iris images .............................................. 59  
   3.4. Medical background ........................................................... 62  
   3.5. Visual inspection of post-mortem changes ........................... 65  
   3.6. Experimental methodology and tools .................................... 66  
   3.7. Results and discussion ....................................................... 69  
   3.8. Conclusions ................................................................. 79  

4. **Iris image segmentation resistant to biological changes** .......... 82  
   4.1. Iris image segmentation ..................................................... 82
Book chapters ................................................................. 161
Reviewed conference publications ........................................ 161

Appendix C: List of Author’s active conference participations .......... 167
Appendix D: List of Author’s awards and achievements ..................... 170
Appendix E: List of grants and projects in which the Author participated .......... 171
  As a Principal Investigator ................................................. 171
  As an Investigator .......................................................... 172
List of Abbreviations

- **EER**  Equal Error Rate
- **FTE**  Failure To Enroll
- **FMR**  False Match Rate
- **FNMR** False Non-Match Rate
- **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
- **K-S test** Kolmogorov-Smirnov statistical test
- **PAD**  Presentation Attack Detection
- **NN**  Neural Network
- **CNN**  Convolutional Neural Network
- **DCNN** Deep Convolutional Neural Network
- **FCEDN** Fully Convolutional Encoder-Decoder Network
- **SFS**  Sequential Forward Selection
- **SBS**  Sequential Backward Selection
- **CAM**  Class Activation Mapping
- **Grad-CAM** Gradient Backpropagation-Guided Class Activation Mapping
1. Introduction

1.1. Iris recognition and human biology

Well established position of iris recognition, including several large-scale applications, such as India’s Government program AADHAAR [1], or the CANPASS/NEXUS system maintained for efficient US-Canada border crossings [2], is attributed to a high uniqueness of the intricate pattern found in the iris tissue, enabling one of the lowest expected error rates among the typically employed biometric characteristics, as well as its asserted temporal stability and immutability. This assertion dates back to year 1987 with Safir and Flom’s patent, which first laid out theoretical ground for iris recognition: ‘significant features of the iris remain extremely stable and do not change over a period of many years’ [3]. This is later supported by John Daugman in his 1994 patent, in which he describes the iris pattern as ‘unique for each individual and stable over many years’ and ‘essentially immutable over a person’s life’ [4]. These claims, being cited throughout the iris biometrics literature, allowed a common belief to arise, that a single enrollment could be sufficient for a lifelong successful recognition of one’s identity.

Recent research in the field, however, including our own studies, has identified situations, in which the technology may fall short of its usually close-to-perfect accuracy, such as working with iris images obtained from people suffering from various ophthalmic disorders [5, 6, 7, 8, 9, 10, 11], and from the deceased subjects [12, 13, 14, 15, 16]. The first scenario proves that people with unhealthy eyes may have trouble using state-of-the-art iris recognition technologies, while the latter can be important from the forensic standpoint – if iris is used as a forensic modality, e.g., for identification of crime or accident victims, improvements to the existing methods must be developed to retain the reliability achieved on the living and healthy eye samples. And – if post-mortem iris biometrics can be reliably carried out – a need for means that enable discerning between live and dead eyes in iris recognition systems seems evident.
1.2. Theses

The aim of this doctoral study is to investigate biological processes that may influence existing iris recognition algorithms and to propose methods taking these changes into consideration and enabling correct recognition, including liveness detection.

The approach to improving iris recognition proposed in this doctoral dissertation can be enclosed in a form of the following theses:

1. Iris recognition is possible in cases of ophthalmic disorders and post-mortem changes to the eye, however the accuracy of the existing iris recognition algorithms is significantly degraded.


4. Liveness detection method can be proposed to discern post-mortem iris samples from those collected from living individuals, requiring only a static iris image.

1.3. Iris as a biometric characteristic

The iris is a colored, annular structure surrounding the pupil of the eye. It displays intrinsically complex patterns that are unique due to random epigenetic factors. The distinctive patterns, developed in the fetal stage, provide features that can be used with high confidence for biometric identification. Uniqueness, combined with the feasibility of fast template creation and matching, allows for the building of large-scale applications of iris recognition. Examples are India’s AADHAAR Unique
Identification Authority of India [1] with more than 1.2 billion people enrolled, and CANPASS system [2] maintained by the Canadian Border Services Agency (CBSA) which provides efficient entry into Canada for frequent travelers. Iris recognition is also being considered by the International Civil Aviation Organization (ICAO) as a candidate for inclusion in the next generation of biometric passports along with fingerprint and face biometrics [17]. According to the most recent NIST’s IREX IX report [18], which evaluates the current state-of-the-art in iris biometrics, the expected one-to-one error rates can be as low as FNMR=0.57% at FMR=0.001% obtained for the leading algorithm, and FNMRs of 0.66% to 0.70% obtained for four following submissions (False Non-Match Rate and False Match Rate, respectively, cf. Appendix A for exact definitions).

The temporal stability of iris patterns is strongly supported by Safir and Flom in a 1987 patent describing theoretical principles of iris recognition. The authors assert that the ‘significant features of the iris remain extremely stable and do not change over a period of many years’ [3]. Further support comes from John Daugman whose 1994 patent states: ‘the iris of the eye is used as an optical fingerprint, having a highly detailed pattern that is unique for each individual and stable over many years’ and ‘the iris of every human eye has a unique texture of high complexity, which proves to be essentially immutable over a person’s life’ [4]. Upon Daugman’s patent, the first commercial iris recognition method was created [19] and a first prototype of a fully functioning iris biometric system was created by Iridian Technologies in 1992. Since then, most commercially available products employed Daugman’s algorithm, roughly until the expiration of Safir and Flom’s patent in 2005. The method developed by John Daugman relies on an iris image acquired in near-infrared illumination, which was used to improve visibility of the iris texture regardless of its melanin concentration, as well as to minimize the user discomfort and reduce the pupillary constriction reflex. Then, the iris is extracted from the image by circle approximation and unwrapped onto a normalized rectangle of a fixed size, to account for differences in pupil size. Features of the iris are extracted by two-dimensional Gabor wavelet filtering of the resulting normalized image, and encoded in the form of a 4096-bit IrisCode© by complex plane phase quantization of the Gabor filtering outcome. The code comprises 2048 bits of the encoded iris information and 2048 bits denoting the occlusion mask.
Notably, in his recent IEEE T-IFS paper [20], Daugman wrote that in cases of the barrel shifting the iris codes, which is intended to compensate for the possible eyeball rotation during acquisition, 'the Re and Im parts (...) become redundant with each other, because their quadrature phase relationship allows each to predict the other’s value after a shift’, and later on that ‘for this reason, for more than a decade all public deployments of iris recognition have used only the Re part’. This may suggest that some of the wavelets currently found in iris matchers employing Daugman’s idea are redundant.

IrisCodes of two irises can then be efficiently compared using the logical exclusive disjunction operation (exclusive-or, XOR), yielding a fractional Hamming distance as a dissimilarity metric between the two respective codes.

Several alternatives to Daugman’s method were proposed over the course of years, a selection of which is briefly referenced here. Wildes et al. introduced a system employing iris texture descriptors in the form of Laplacian of Gaussian (LoG) pyramids of multiple scales [21]. In the work of Boles et al. the iris is represented at different resolutions by the zero-crossing function of the wavelet transform output, with the wavelet function being the first derivative of a cubic spline smoothing function [22]. In a patent application by Kim et al., the use of Haar transform for iris representation is depicted [23]. Li et al. propose a method taking advantage of irises’ local variations in intensity, which are extracted using dyadic Gabor wavelet transform and encoded into 1D intensity signals [24]. Czajka and Pacut devised an iris recognition algorithm employing Zak-Gabor wavelet packet transform with an effective adaptation of the transformation parameters to make it insensitive to varying frequencies in different irises [25].

Shen and Flynn explored the possibility of utilizing visible iris features, such as crypts – iris regions that are have strong edges and are oval- or disk-shaped, and darker than their surrounding when photographed in near infrared, and anti-crypts – more varied in shape and brighter than the surrounding iris tissue, for the purpose of machine-assisted human-interpretable iris recognition for forensic applications [26, 27, 28, 29]. Sunder and Ross [30] studied a possibility of employing pigmentation-related macro iris features, including: moles, freckles, nevi, melanoma, sectoral heterochromia, but not anatomical features such as iris crypts, using the
scale-invariant feature transform (SIFT). Rathgeb et al. [31] explored the concept of BSIF-based feature extractor for extracting iris features.

More recently, with enthusiasm towards deep learning applications in computer science, a feature-learning approaches to iris recognition began to surface [32, 33]. These methods use a lot of parameters and hyperparameters of the model that are learnt directly from the data, instead of being hand-engineered specifically for the model. Such approaches usually employ convolutional neural networks, and enable easier adaptation in difficult conditions, when data is far from perfect, such as working with images that come from two different sensors, are of different resolution and quality [34]. A solution with simultaneous iris masking and representation, trained with the use of a triplet loss function, employing both a positive and a negative sample in a single pass, was also proposed [35]. Nguyen et al. show that off-the-shelf CNN classifiers can offer equal or higher accuracy than the typical, Daugman method-based iris matcher [36]. On the other hand, deep learning-based methods usually require vast amounts of training data and lots of processing power, as compared to traditional approaches.

1.4. New challenges in iris biometrics research

Even though iris recognition has been consequently establishing its position as a secure, reliable, and fast means of biometric authentication over more than 25 years, the last decade has brought the attention of the biometrics community to several new challenges. This Section introduces the Reader to these novel aspects of iris recognition.

1.4.1. Influence of aging

The ISO/IEC biometrics vocabulary defines reference aging as ‘the changes in error rates with respect to a fixed reference caused by time-related changes in the biometric characteristic, its presentation, the sensor and other components of the biometric technology’. It consists of multiple aspects, including: biological aging of the eye and its structures; differences in sample presentation originating in pupil dilation, eyelid droop, acquisition conditions, etc.; sensor interoperability and aging - when gallery and probe samples are collected using different equipment and camera components wearing out, respectively.
Template non-stationarity is reported to play a vital role in decreasing over-the-years iris recognition performance in number of publications [37, 38, 39, 40, 41, 42, 43, 44, 45], including our own previous research [46, 47]. NIST’s IREX VI report, however, states the contrary [48], and was later criticized by Bowyer and Ortiz [49], and a response to that critique was also published [50]. Recently, more researchers have made efforts to better understand the non-stationarity of templates, namely by isolating as many factors as possible [51], studying the impact of segmentation quality [52], but also the influence of sensor aging [53]. This shows that despite research efforts having been put into solving these issues, template aging still presents challenges.

1.4.2. Influence of ophthalmic diseases

Numerous medical conditions affecting the eye structures, especially the iris, anterior chamber of the eye, and the cornea, have a potential of degrading its accuracy and reliability, which may be especially pronounced with mass-scale iris recognition deployments, where errors caused by such pathologies can surface. In 2012, 24 million Americans suffered from cataract (estimated to rise to 38.7 million in 2030) and 2.7 million suffered from glaucoma (4.3 million in 2030) [54]. In 2014, over 79 thousand corneal transplant surgeries were performed [55]. Such high volumes of pathology occurrences (especially for cataract, expected to affect over 10% of the US population by 2030) make it crucial to assess iris recognition performance in the presence of common pathologies.

Yet, due to the lack of appropriate datasets and difficulties in creating them, limited research is available, mostly centered around cataract and cataract extraction procedure influence on iris recognition performance: [6, 7, 8, 56, 57] (significant negative impact of cataract and cataract surgery reported by most researchers except for [8]), impact of refraction correction procedures [58] (no influence reported), but also studies regarding multiple disorders [5, 9, 10], and their impact on segmentation [59, 10].

1.4.3. Post-mortem iris recognition

Iris biometrics after death is important for at least two reasons. First, if post-mortem recognition is viable, it could prove useful in forensics, namely identification and verification of accident and crime victims, and even in the battlefield (when other
fast methods of identification are not accessible, say, victim has lost his fingers or face is disfigured). The latter reason connects with the use of iris biometrics for identity management and asset protection and an associated fear of identity theft - ‘will someone be able to steal my iris after I die, and use it to gain access to my identity?’ [60]. Several publications firmly mention that iris recognition after death cannot be performed due to pupil dilation and corneal cloudiness [61], ‘iris decay’ [62], ‘iris features vanishing with pupil dilation’ and ‘muscle relaxation’ [63, 64]. However, no experimental evidence is presented in either of those publications.

Due to the difficulties in data collection and the obvious unpleasantness of such experiments, very little research has been published on the topic of iris recognition in deceased subjects, especially when human eyes are concerned, with few exceptions, namely [65], a M.Sc. Thesis, which concludes that post-mortem iris recognition works fine in about 80% of the cases for samples acquired up to 2 days after death, and [14], which mostly focuses on post-mortem face and fingerprint recognition, with few conclusions regarding irises. Paper [66] presents a study of post-mortem iris recognition using cadaver eyes of a domestic pig, reporting that the eyes lose their capability to serve as a biometric identifier in 6 to 8 hours post-mortem.

1.4.4. Presentation attack detection and liveness detection

The ISO/IEC standard on biometrics-related vocabulary defines biometric presentation attack as a ‘presentation to the biometric capture subsystem with the goal of interfering with the operation of the biometric system’ [67]. Presentation attack may involve presenting the system with an artifact, such as a paper printout of the iris, eye prosthetics, computer displays, or using a genuine eye in a non-conformant scenario (e.g., use under coercion, or presentation improper enough to compromise the system). Thus, an important piece in every well-designed biometric system is a way to mitigate such attempts, i.e., a presentation attack detection (PAD) component. With post-mortem iris recognition gaining increasing attention in the biometrics and forensics communities, presentation attack detection in a form of liveness detection seems crucial in those cases when we don’t want our biometric traits to be used after death [68].
1.5. Scope of this Thesis

This doctoral dissertation focuses on the three out of the four problems introduced above, namely on analyzing the influence that ophthalmic disorders and post-mortem changes may inflict on iris recognition, proposing improvements to the iris recognition pipeline to make it more resilient to such changes, and, finally, including a presentation attack detection component designed to detect samples coming from cadaver eyes.

1.6. Thesis layout

This Thesis is organized as follows. After the introduction to iris biometrics, its recent trends and emerging challenges in Chapter 1, Chapters 2 and 3 relate to the first thesis stated in Sec. 1.2. In Chapter 2, a study that quantifies the influence of ocular disorders on the accuracy of iris biometrics is described. An appropriate database of iris images representing eyes suffering from various ocular pathologies is collected and described, together with an appropriate medical commentary. Experiments assessing the extent and reasons for performance degradation of existing iris recognition methods, when confronted with these difficult samples, are conducted and summarized with appropriate conclusions. Chapter 3 presents a similar experimentation, but carried out for samples coming from eyes of deceased subjects. Here as well, an appropriate database of post-mortem iris images is designed and collected, and a visual examination of the physical deterioration of the eyes is performed, supported with relevant medical knowledge. Experiments analogous to those done for disease-affected samples is carried out, quantifying the reliability of existing iris recognition methods in the context of cadaver iris images, and examining the errors and their potential causes.

The second thesis from Sec. 1.2 is implemented in Chapter 4, which introduces the first step in making iris recognition more reliable in the two scenarios described above, by proposing a data-driven segmentation method that is able to ignore the regions deformed by biological, disease-induced, or post-mortem changes. The newly designed resistant semantic segmentation method is then fitted into the existing iris recognition pipeline, showing a significant improvement in iris recognition accuracy for challenging post-mortem and disease-related datasets of iris images.
Chapter 5 describes the novel, post-mortem specific iris feature representation, which allows for further improvement in the recognition accuracy for samples collected from cadaver eyes, therefore constituting the third thesis introduced in Sec. 1.2.

Chapter 6 relates to the last, fourth thesis from Sec. 1.2, introducing a method for presentation attack detection, utilizing a static iris image and a feature-learnt model, enabling 99% accurate detection of cadaver iris samples. Finally, Chapter 7 contains a summary of this project.

Additionally, Appendix A contains the biometrics-related vocabulary, error metrics, and testing protocols that are used throughout this doctoral dissertation. Appendix B presents a full list of Author’s publications (including statements of his contributions), Appendix C lists active conference participations, Appendix D lists awards and achievements, whereas Appendix E compiles a list of grants and projects in which this Author participated, together with his roles within each project.
2. Influence of ophthalmic disorders

2.1. Iris recognition in the presence of ocular pathologies

Statements of excellent performance of iris recognition are founded upon the immutability of the iris pattern in a healthy eye. When an injury, disease or other ocular pathology is present, it is not a given that such a condition will not affect the visibility or appearance of the features of the iris. Such occurrences may degrade the performance of the iris recognition system either by altering the iris or obstructing its view. In certain circumstances, it may even render the eye unsuitable for use in authentication at all. This can be the case in occurrences of aniridia, a medical disorder, often bilateral, in which only a small, ring-shaped portion of tissue is present where the iris would normally be located. This often leaves a large and possibly irregularly shaped pupil [69].

Eye trauma and injuries are also contributing factors to degradation of iris recognition accuracy. Although in his patent regarding the first iris recognition method Daugman states that 'as an internal organ of the eye the iris is well protected from the external environment' [4], in some cases this kind of protection (i.e., by the cornea and the aqueous humor in the anterior chamber of the eye) may not be sufficient. Safir and Flom acknowledged it in their patent claim: 'A sudden or rapid change in such a feature [of iris pattern] may result in a failure to identify an individual, but this may alert the individual to the possibility of pathology of the eye' [3].

Various medical conditions affecting the structures of the eye, the iris in particular, may cause a deterioration of the reliability of iris recognition. The ISO/IEC 29794-6 standard on iris image quality lists some of the possible disorders and pathologies that can degrade or prevent automatic identity recognition by means of iris identification, such as abusing pupil-dilating drugs, iritis, micro- and megalocornea, keratitis, leukoma, aniridia, among others. Investigation of previous research in this field
(cf. Sec. 2.2) shows that we are still far from fully understanding how various eye conditions impact iris recognition. This can be attributed to the lack of large, heterogeneous, and publicly available databases appropriate for this subject. In this Chapter, we answer **five questions** related to ocular disorders and their impact on iris recognition:

1. **Do ocular pathologies impact the enrollment process?** If so, which structural impairments translate into an increase in the failure to enrol rate (FTE), i.e., the proportion of samples that could not be enrolled to the overall number of samples?

2. **Does iris recognition perform worse in eyes with a medical condition, but without visible impairments in comparison with healthy irises when photographed in near-infrared (NIR) light?**

3. **What kinds of visible impairments in unhealthy irises have the greatest impact on iris recognition?**

4. **Can cataract, the most prevalent ophthalmic disorder worldwide, influence iris recognition?**

5. **What are the main reasons for bad performance when iris recognition is applied to unhealthy eyes?**

To answer these questions, a dataset of iris images representing more than twenty different eye diseases was built with the use of a professional iris recognition camera operating in NIR light, along with an ophthalmological commentary (cf. Sec. 2.4 and 2.3). Most of the NIR samples are accompanied by color images to make possible independent ophthalmological interpretations. Experimental study done for four different and independent iris recognition algorithms is presented (cf. Sec. 2.5 and 2.6).

The experiments and results presented in this Chapter constitute a significant contribution over the initial experimentation carried out by the Author during his Master’s studies and described in [56, 9, 70]. These are briefly reviewed in Sec. 2.2. To our best knowledge, this Chapter describes the largest published dataset of NIR and color images for unhealthy eyes with a professional, ophthalmological commentary, prepared in cooperation with the Medical University of Warsaw, and presents the most extensive study to date of ways in which different types of diseases of the eye impact iris recognition. The original research included in this Chapter appeared in parts in the
following publications co-authored by the Author over the course of his PhD program: [10, 11, 71].

2.2. Related work

2.2.1. Cataract and cataract surgery

Probably the first experiment devoted to this field of research was conducted by Roizenblatt et al. [6] and involved 55 patients suffering from cataract. Each person was enrolled in the LG IrisAccess 2000 biometric system before cataract removal surgery, 30 days after cataract removal, and 7 days after stopping the administration of pupil-dilating drugs (when pupils had reverted to normal size and reaction to light). After the 30-day period, differences in the size of pupils were no larger than 1.5mm when compared with images obtained prior to the treatment. Each eye that underwent a surgery was also given a score between 0 and 4 based on visual inspection performed by an ophthalmologist. One point was given for each of the following ocular pathologies: depigmentation, pupil ovalization, focal atrophy with and without trans-illumination. A correlation was revealed between the visual inspection score and change in Hamming distances (HD) yielded by the IrisAccess method between templates created using pre- and post-surgery samples. 6 out of 55 eyes were no longer recognized, thus yielding FNMR of about 11%. For the remaining irises there were significant shifts towards worse HD (11.3% increase in average HD when scores between gallery samples and post-surgery samples are compared against scores between gallery samples and pre-surgery samples) and worse visual scores (11.13% increase in average visual score for images collected post-surgery, as compared to average visual score for those obtained pre-surgery). Surgeries in this study were performed with a procedure called phacoemulsification, which involves insertion of a small probe through an incision in the side of the cornea. The probe emits ultrasound waves that break up the opacified lens which is later removed using suction [56], and authors hypothesize that the energy released inside the eyeball during the cataract surgery may be a cause of atrophic changes to the iris tissue. Re-enrollment is suggested as a countermeasure in cases with significant, visual alteration to the iris visible during a slit-lamp examination.
Dhir et al. [8] studied the influence of the effects of mydriatics accompanying cataract surgery. A group of 15 patients had their eyes enrolled before surgery. A verification was performed at 5, 10, and 15 minute intervals after application of the drug and again two weeks following the procedure itself. None of the eyes was falsely rejected after this two-week period. One must, however, keep in mind that the authors excluded from the dataset eyes with pre-existent corneal and iris pathologies, or those with iris tissue damaged during the surgery. The study suggests that recognition deterioration may originate from a slight shift of the iris towards the center of the eyeball resulting from implanting an artificial lens that is thinner than the natural lens. Specular reflections from the implant may also contribute to erroneous segmentation. However, increase in pupil diameter induced by mydriatics led to FNMR of 13.3%, as 6 out of 45 verification attempts failed. In addition, Hamming distances increased with the elapse of time after the instillation of the drug. The authors warn that this phenomenon may be exploited by criminals in order to enroll in a biometric system under multiple identities to deceive law enforcement organizations.

Another scenario regarding the impact of cataract surgery was explored by Seyed-dain et al. [7] who performed an experiment to establish the effects of phacoemulsification and pharmacologically induced mydriasis on the iris. The experiment aimed to determine whether the irises, following phacoemulsification or drug induced mydriasis (preventing the dilated pupil from reacting to light stimulation) perform worse when compared to the same irises before the procedure or before the drug-induced pupil dilation. They revealed that 5.2% of the eyes that were subject to cataract surgery could no longer be recognized after the procedure. In the pupil dilation group, this portion reached as high as 11.9%. In both cases the authors suggest re-enrollment for patients whose eyes were not successfully identified after the surgery or instillation of mydriatics. No false acceptances were observed in either case.

In [56] we aimed at quantifying the impact of cataracts on iris recognition performance. An experiment involving three different iris recognition methods revealed differences in system performance when comparison scores calculated before surgery from cataract-affected eyes are used instead of those obtained from healthy eyes. For all three methods there was a perceivable degradation in average genuine comparison scores with differences reaching from 12% of genuine score degradation for an
academic BiomIrisSDK matcher, up to 175% of genuine score degradation for one of the commercial matchers (MIRLIN). For two out of three matchers, these changes also affected the final false non-match rate.

Later, in [70], we expanded the cataract-related experiments with a study assessing an influence of cataract surgery, in which a change in similarity scores has been observed when scores calculated with images taken after this procedure are compared to those yielded by images taken prior to the surgery. During testing, academic matchers performed better than their commercial counterparts, producing average genuine scores 8.2% (BiomIrisSDK) and 21.7% (Daugman’s method implementation) worse, while for commercial matchers those values were 29.8% (MIRLIN) and 43.6% (VeriEye).

Ramachandra et al. [57] conducted experiments regarding iris biometrics in the context of cataract surgery, when verification is performed using pre-surgery gallery samples and post-surgery probe samples, coming from the same individual. Studies were carried out using a database of iris images acquired from 24 hours pre-surgery and 36-42 hours post-surgery from 84 subjects. Recognition accuracy is reported to drop significantly, reaching genuine match rate of 85.19% @ FMR=0.1 and EER=7%, when compared to performance achieved using pre-surgery images only.

2.2.2. Refractive surgeries

Yuan et al. [58] examine another type of medical procedure – laser-assisted refractive correction surgery – and its possible impact on iris recognition. These procedures take advantage of laser radiation to ablate the corneal tissue and compensate for refractive pathologies such as myopia, hypermetropia and astigmatism. Researchers carried out an experiment to find out whether such manipulation may result in increased FNMR of an iris biometric system. Using Masek’s algorithm for encoding, 13 eyes (out of 14) were correctly recognized after a procedure had been performed. However, the one eye that was falsely non-matched had a significant deviation in circularity of the pupil and increased pupil diameter. Therefore, the authors argue that refractive correction procedures have little effect on iris recognition.
2.2.3. Other ocular pathologies

In a study by Aslam et al. [5], 54 patients suffering from several different eye conditions were enrolled in a biometric system using the IrisGuard H100 camera during their first visit. Their eyes were again photographed after treatment. Researchers calculated Hamming distances between codes obtained before and after the treatment to determine whether treatment had any impact on recognition accuracy. Tested methodology turned out to be resilient for most illnesses, i.e., glaucoma treated using laser iridotomy, infective and non-infective corneal pathologies, episcleritis, scleritis and conjunctivitis. However, 5 out of 24 irises affected by the anterior uveitis\footnote{a condition in which the middle layer of the eye, the uvea, which includes the iris and the ciliary body, becomes inflamed [72]} were falsely non-matched after treatment, producing an FNMR rate of about 21%. It is worth noting that each of the eyes that yielded a false non-match had earlier been administered mydriatics; therefore, the pupil was significantly dilated. In addition, two eyes suffered from high corneal and anterior chamber activity, while the remaining three had posterior synechiae that caused deviation from the pupil circularity. The hypothesis stating that the mean Hamming distance in the anterior uveitis subset is equal to that of the control group (consisting of healthy eyes) has been rejected with $p-value < 10^{-4}$, while there were no statistically significant differences between scores obtained from other disease subsets when compared to the control group (thus the null hypotheses could not be rejected). As for the pathologies related to the corneal opacities, Aslam tries to explain lack of recognition performance deterioration by the fact that the NIR illumination used in iris biometrics is more easily transmitted through such obstructions and therefore allows correct imaging of underlying iris details. Laser iridotomy also showed little influence, as the puncture in the iris tissue made by laser radiation appears to be too small to significantly alter the iris pattern. However, certain combinations of synechiae and pupil dilation can affect the look of the iris texture sufficient to produce recognition errors. A deviation in the pupil's circularity caused by the synechiae may also contribute to segmentation errors.

Borgen et al. [73] conducted a study focusing on iris and retinal biometrics in which they take advantage of 17 images selected from the UBIRIS database. Those images

\footnote{a condition in which the middle layer of the eye, the uvea, which includes the iris and the ciliary body, becomes inflamed [72]}
were then digitally modified to resemble changes to the eye structures caused by various ocular illnesses: keratitis and corneal infiltrates, blurring and dulling of the cornea, corneal scarring and surgery, angiogenesis, tumors and melanoma. High FNMR values (32.8% – 86.8%) are reported for all modifications except for the pathological vascularization (6.6%), changes in iris color (0.5%) and iridectomy-derived damage in the iris, for which FNMR=0. Faulty segmentation is suspected to be the main cause, especially in cases involving clouding of the cornea. The authors, however, do not acknowledge the fact that NIR illumination enables correct imaging even in eyes with corneal pathologies such as clouding or other illness-related occlusions.

In a study by McConnon et al. [59] three groups of medical disorders (conditions causing pupil/iris deformation, pupil/iris occlusion and eyes having no iris or a very small iris) were distinguished to estimate the impact they may have on the reliability of iris segmentation. Due to lack of publicly available datasets, the database used in this work consisted of images drawn from the Atlas of Ophthalmology, making them imperfectly suited for iris recognition (i.e., for having been captured in visible light). Those images have been resampled to $320 \times 240$ resolution and manually segmented to obtain the ground truth iris localization. Automatic segmentation, performed using Masek’s algorithm, deviated by two or more pixels in 46% and 55% of images for the limbic and pupillary boundaries, respectively.

In [70], we divided the data from our database into subsets in respect to the type of impact generated by the disease and those are compared against a reference subset containing healthy eyes. Disease influence turned out to be most prominent in subsets containing eyes with distorted pupils, altered iris patterns and obstructed iris. The degradation of scores was 792.7% for the worst performing method. We have introduced the first, publicly available dataset of iris images obtained from ophthalmology patients in [9], together with the experiments regarding cataract-related effects and attempts to assess which types of eye damage caused by disease have the greatest impact on the accuracy of a given biometric system.

2.2.4. Conclusions from the literature review

Table 2.1 summarizes the most important works reviewed in this Section. Very high FNMR values are reported in most of these papers, all of them being significantly
higher than those reported for healthy irises in the latest IREX IX report, in which 30 out of 46 submitted algorithms were able to keep the FNMR below the 2% mark [18], although these results are not directly comparable with the ones reported in this literature review because of different experimental data. A lack of appropriate datasets encompassing samples representing a variety of disorders for a large number of subjects is a major problem. Also, most of the works referenced above focus on only one disease - typically the cataract. Notably, none of these studies involve proposing new methods in iris recognition that would counteract the negative changes introduced by diseases. We may conclude that three most important research problems in this field are:

— the collection of an appropriate, extensive database of iris images, cf. Sec. 2.4,
— carrying out a comprehensive analysis of influence of different ocular pathologies, cf. the remainder of this Chapter,
— and proposing methods that would serve as a countermeasure against disease-induced degradation in iris biometrics’ reliability, cf. Chapter 4.

2.3. Medical background

The successful performance of an iris biometric system depends on the capacity to correctly image details of the texture of the iris. Imaging of irises afflicted by diseases can be problematic for several reasons. Heavily distorted pupils, deviating severely from their usual circular shapes, can affect image segmentation algorithms approximating pupillary and limbic boundaries with circles. Eyes with changes to the cornea may perform worse when segmentation methods utilizing image local gradients are used. Severe damage to the iris tissue itself may change the pattern significantly, sufficient to make correct identification impossible. The following subsection presents a brief characterization of medical conditions represented in the database created for this study. The potential impact on the performance of a biometric system, depending on the category of these disorders, is discussed as well.
### Table 2.1: Summary of the relevant scientific literature.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roizenblatt et al. [6]</td>
<td>2004</td>
<td>decreased recognition accuracy post-cataract-surgery (FNMR=11%)</td>
</tr>
<tr>
<td>Yuan et al. [58]</td>
<td>2007</td>
<td>refractive surgeries have little to no impact on iris recognition</td>
</tr>
<tr>
<td>Aslam et al. [5]</td>
<td>2009</td>
<td>iris recognition showed resilience against iridotomy, corneal pathologies,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>episcleritis, scleritis, conjunctivitis, but anterior uveitis can degrade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>performance (FNMR=21%)</td>
</tr>
<tr>
<td>Borgen et al. [73]</td>
<td>2009</td>
<td>no actual disease-related iris samples used</td>
</tr>
<tr>
<td>Dhir et al. [8]</td>
<td>2010</td>
<td>decreased recognition accuracy after use of pupil-dilating drugs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(FNMR=13.3%)</td>
</tr>
<tr>
<td>McConnon et al. [59]</td>
<td>2012</td>
<td>ocular disorders shown to degrade image segmentation,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low-resolution, visible light samples used</td>
</tr>
<tr>
<td>Seyeddain et al. [7]</td>
<td>2014</td>
<td>decreased post-cataract-surgery accuracy (FNMR=5.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>decreased post-pupil-dilating drugs accuracy (FNMR=11.9%)</td>
</tr>
<tr>
<td>Trokielewicz et al. [56]</td>
<td>2014</td>
<td>decreased genuine comparison scores for cataract eyes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(up to 8-fold decrease, compared to healthy eyes)</td>
</tr>
<tr>
<td>Trokielewicz [70]</td>
<td>2014</td>
<td>lower genuine comparison scores for post-cataract-surgery eyes</td>
</tr>
<tr>
<td>Trokielewicz et al. [9]</td>
<td>2015</td>
<td>changed pupil-iris geometry and obstructions decrease recognition accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(initial results on the first version of the dataset)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>database offered to the community</td>
</tr>
<tr>
<td>Trokielewicz et al. [10]</td>
<td>2015</td>
<td>experiments extended over [9] to a larger database of iris images,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>including enrollment failures analysis</td>
</tr>
<tr>
<td>Ramachandra et al. [57]</td>
<td>2016</td>
<td>decreased recognition accuracy post-cataract-surgery (FNMR=14.8%)</td>
</tr>
<tr>
<td>Trokielewicz et al. [11]</td>
<td>2016</td>
<td>this paper publishes the most recent findings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>included also in this doctoral dissertation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>significantly extended database offered to the community</td>
</tr>
<tr>
<td>Trokielewicz et al. [71]</td>
<td>2019</td>
<td>a review study in the form of a book chapter</td>
</tr>
</tbody>
</table>

#### 2.3.1. The cornea

As the outermost part of the eye, the cornea, despite being fairly durable, can still suffer from numerous factors. Chemical injury can deal extensive damage to the ocular surface epithelium, the cornea, and the anterior segment of the eye. It can lead to opacification and neovascularization of the cornea, formation of a symblepharon and cicatrical ectropion or entropion. If significant corneal scarring is present, a corneal transplant may be required to restore vision. A benign growth of the conjunctiva – pterygium – commonly forms from the nasal side of the sclera to the center of the cornea. This fibrovascular proliferation often occludes a part of the iris. Bacterial
keratitis is an erosion or an open sore in the outer layer of the cornea with stromal infiltration, edema and hypopyon. Common pathogens that may lead to corneal ulcers include *Streptococcus pyogenes, Acanthamoeba, Herpes simplex,* or fungal infections mainly caused by use of non-sterilized contact lenses. Acute glaucoma, with increased pressure inside the eye, can occur suddenly when the iris is pushed or pulled forward. High intra-ocular pressure produces symptoms such as corneal edema, shallowness of the anterior chamber and dilation of the pupil which may become oval in shape. Corneal laceration usually requires placement of corneal sutures. These disorders usually impact the look and clarity of the corneal area, partially or totally covering the iris.

2.3.2. The anterior chamber

Hyphema is a condition characterized by the presence of blood in the anterior chamber of the eye that can partially or completely obstruct the view of the iris. Hyphemas are frequently caused by injuries but may also occur spontaneously. A long-standing hyphema may result in hemosiderosis and heterochromia in a form of partial changes to iris coloration. Hypopyon is a leukocytic exudate present in the anterior chamber of the eye, usually accompanied by redness of the conjunctiva. It is a sign of an iridial inflammation. Both conditions may significantly obstruct the view of the iris by obscuring the iris texture, thus causing problems with segmentation.

2.3.3. The iris

Rubeosis iridis is a medical condition of the iris in which new, abnormal blood vessels are found on the surface of the iris. It is usually associated with disease processes in the retina. Iris sphincter tear is a frequent concomitant of both laceration and blunt trauma of the anterior segment. Iridodialysis is defined as a rupture of the iris at its thinnest area, the iris root, manifested as a separation or tearing of the iris from its attachment to the ciliary body. It is usually caused by blunt trauma to the eye but may also be caused by penetrating eye injuries or as a complication of an intraocular surgery. Iridodialyses can often be repaired using suturing techniques. Synechiae are adhesions between the iris and other structures in the eye. Iris bombe occurs when there is a complete adhesion (posterior synechiae) between the iris and the anterior
capsule of the lens creating a 360-degree area of adhesion. All of the aforementioned are capable of introducing severe distortions or damage to the iris region.

2.3.4. The lens

Anterior lens luxation (wherein the lens enters the anterior chamber of the eye) can cause damage to the cornea, swelling, and progressive lens opacity, blurring the iris image. Phacolytic glaucoma is an inflammatory glaucoma caused by the leakage of the lens protein through the capsule of a hyper-mature cataract. Escalating corneal edema and milky aqueous humor in the anterior chamber also blur the iris image.

2.3.5. Pars plana vitrectomy

This is a general term used to describe a group of surgical procedures performed in the deeper part of the eye and behind the lens. Silicone oil is used as an intraocular tamponade in the repair of retinal detachment or diabetic retinopathy. Sometimes it may relocate itself to the front of the iris, causing an obstruction that prevents quality iris imaging.

2.4. Dataset of iris images

2.4.1. Data collection protocol

For the purpose of the studies carried out by the Author, a new database was designed and collected specifically for the assessment of how iris recognition is immune or prone to ocular pathologies. The dataset comprises images collected from patients during routine ophthalmological examinations. All patients participating in the study were provided with detailed information on the research and an informed consent has been obtained from each volunteer. The collection of this novel database was carried out in close collaboration with Dr. Piotr Maciejewicz from the Department of Ophthalmology at the Medical University of Warsaw, who performed the collection of iris photographs. The contribution of the Author of this Thesis was the design of the database collection protocol and the acquisition camera setup (together with Dr. Adam Czajka), as well as data curation: sorting, censoring, and processing of the samples. As of the time of this Thesis’ preparation, there were no other databases of iris
images collected from ophthalmology patients that would be available to the research community.

Data collection lasted approximately 16 months, out of which the database design process and first 8 months of data collection took place during the Master’s course of this Author (summarized by publication of the first version of the dataset in [9, 74]) and the remaining 8 months during the course of his PhD program (summarized by the publication of the second, extended version of the dataset in [10, 75]). During each patient visit, both NIR-illuminated images (compliant with the ISO/IEC 19794-6:2011) and color photographs (for selected cases). The data was acquired with three commercial cameras: 1) the IrisGuard AD100 for NIR images, 2) Canon EOS1000D with EF-S 18-55 mm f/3.5-5.6 lens equipped with a Raynox DCR-250 macro converter and a ring flashlight suited for macrophotography, and 3) an ophthalmology slit-lamp camera Topcon DC3.

Table 2.2: Format characterization and numbers of collected images for each sensor.

<table>
<thead>
<tr>
<th>Device</th>
<th>Image format</th>
<th>Number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>IrisGuard AD100</td>
<td>grayscale, 640x480 BMP</td>
<td>1793</td>
</tr>
<tr>
<td>Canon EOS1000D</td>
<td>color, 10 Mpixel JPEG</td>
<td>868</td>
</tr>
<tr>
<td>Topcon DC3</td>
<td>color, 8 Mpixel JPEG</td>
<td>335</td>
</tr>
</tbody>
</table>

2.4.2. Database statistics

The entire dataset comprises 2996 images of 230 distinct irises, Tab. 2.2. Every class contains NIR-illuminated images, while for some of them visible light photographs were also taken (in cases when visual inspection revealed significant changes to the structures of the eye). Fig. 2.1 shows sample images of five different eyes obtained using all three devices.

Images for 184 irises were captured during one acquisition session; for 38 irises this was possible during two sessions; for 6 irises, in three sessions. Finally, for 2 irises, there were four different acquisition sessions. Typically, the second and subsequent sessions contain images obtained after some kind of medical procedure, e.g., a cataract surgery. Detailed information, including a precise description of medical conditions and procedures performed in each case, is disclosed in the metadata that accompanies the published dataset. No data censoring was performed when collecting the data,
Figure 2.1: Samples of 5 different eyes acquired using three different imaging systems: IrisGuard AD-100 (top row), Canon EOS 1000D (middle row), and Topcon DC3 slit-lamp camera (bottom row). Each column includes samples corresponding to a different group used further in our experimental study, namely: healthy eye (H1-H3), unhealthy eye but with a clearly visible iris pattern (C1-C3), eye with geometrical deviations (G1-G3), eye with iris tissue impairments (T1-T3), and eye with obstructions in front of the iris (O1-O3). Figure reprinted from [11].

except for immediate removal of images that did not show an eye at all. This database is publicly available in version 1 and an expanded version 2 [9, 11].

2.4.3. Disease representation in the database

While the database images represent more than 20 different medical conditions, most of them can be roughly classified into three main groups of related disorders or conditions and their effects on the eye:

1. Cataract and related conditions, which can be found in different stages of the treatment process:
   — opacified lens,
   — lens implant after the surgery,
   — aphakia – lack of lens after complicated surgeries,
   — capsulotomies – incisions in the lens to correct capsule opacity.

2. Glaucoma and associated medical procedures:
   — trabeculectomy – surgical removal of a part of the upper portion of the iris tissue,
   — iridotomy – puncturing the iris using laser radiation.
3. Cornea pathologies leading to opacification, making it more difficult to obtain a good image of the iris beneath it because of:
   — corneal inflammation and ulcers,
   — trauma and chemical burns.

2.5. Experimental methodology and tools

2.5.1. Iris recognition methods

In this work four iris recognition methods were employed, namely OSIRIS, MIRLIN, VeriEye, and IriCore. OSIRIS comes from an academic community, while the three remaining algorithms are commercially available products. These four methods are used as baseline iris comparators throughout this Thesis. The goal of this section is to briefly characterize these solutions.

OSIRIS (Open Source for IRIS) [76] is an open source implementation of the Daugman’s iris recognition concept. The OSIRIS software comprises four independent operations: a) image segmentation employing the Viterbi algorithm, b) image normalization, c) iris coding by quantization of the Gabor filtering outcomes, and d) iris code comparison based on fractional Hamming distance. Original implementation used in this work calculates the iris code for three different resolutions of the complex Gabor filter kernel. Phase of only 256 unique and equidistantly located points is quantized to one of four possible quadrants of the complex plane, employing only two bits per point. This results in the iris code length of 1536 bits (3 resolutions × 256 points × 2 bits for the coding point’s phase). The occlusion mask, calculated at the segmentation stage, eliminates iris code bits corresponding to non-iris areas. As in Daugman’s solution, we should expect a dissimilarity score close to zero when comparing samples of the same eye, and close to 0.5 when comparing different irises (as in the comparison of two sequences of heads and tails obtained in independent coin tosses). Due to rotation compensation realized by shifting the iris code and finding the best match, the distribution of impostor comparison scores is typically skewed towards smaller values (about 0.4).
MIRLIN (Monro Iris Recognition Library) has been offered on the market as a Software Development Kit (SDK)\(^2\) [77]. It employs a discrete cosine transform calculated for local iris image patches to deliver the binary iris code [78]. Similarly to Daugman’s approach, the iris codes are compared to each other by calculating the fractional Hamming distance, normalized by the number of valid iris code bits, \(i.e.,\) originating from non-occluded iris regions. As for the OSIRIS method, comparing two images of the same eye should yield a fractional Hamming distance close to zero, while the distance for two different eyes should oscillate around 0.5. The advantage of MIRLIN is the visualization of automatic segmentation results, helpful when analyzing sources of possible errors when processing images of unhealthy eyes.

VeriEye, the third matcher involved in this study, is another commercial product offered by Neurotechnology for more than a decade [79]. It incorporates an unpublished iris encoding methodology, although thoroughly evaluated in numerous applications and scientific projects, \(e.g.,\) in NIST ICE 2005 project [80]. The manufacturer claims to employ an off-axis iris localization with the use of active shape modeling. In contrast to OSIRIS and MIRLIN, the VeriEye delivers a similarity score between two iris images. The higher the score, the more similar the images. A zero score denotes a perfect non-match.

IriCore is the fourth iris recognition system employed in this study and offered on the market as an SDK by IriTech Inc. [81]. As with the VeriEye solution, scientific papers divulging the implemented methodology have not been published. Nonetheless, the IriCore implementation was placed in a narrow set of the best solutions tested by NIST in 2005 [80]. The manufacturer claims conformance with two editions of the ISO/IEC 19794-6 standard (one issued in 2005 and the most current one published in 2011). Using the IriCore system to compare two same-eye images should result in a near-zero dissimilarity score, while scores between 1.1 and 2.0 are typically observed when comparing images of two different eyes.

2.5.2. Data curation

Our data shows that most unhealthy eyes suffer from more than one condition, often unrelated and impacting the eye in different ways. While some illnesses cause

\(^2\) However, MIRLIN has been discontinued by its current proprietor Fotonation Ltd. and as of 2019 is no longer available on the market.
the pupil to distort and deviate from its usual circular shape, other pathologies impact the iris directly or cause changes to other parts of the eyeball, such as uvea, cornea, anterior chamber, or even retina. Hence, conducting an insightful analysis, particularly a separate analysis for each individual impairment, may be challenging or even impossible. The data was, therefore, partitioned respectively according to the type of visible influence that a given ocular pathology inflicts on the eye.

This allowed us to devise five different subsets: Healthy, comprising healthy eyes only; Clear, made up of eyes with a disease present, but having no perceivable effect on the eye structures; Geometry (eyes whose pupil geometry has been distorted by the pathology); Tissue (eyes with damage inflicted on the iris tissue) and Obstructions, encompassing the eyes with obstructions present in front of the iris. Figure 2.1 shows sample images belonging to each subset. Table 2.3 presents the numbers of classes (i.e., different eyes) and images in each subset. All of the five subsets created here are eye-disjoint, i.e., eyes belonging to one of the subsets cannot be found in another one.

For the purpose of this study, we have selected a subset of the original dataset that comprises only NIR images obtained during the first acquisition session for each eye. Images representing eyes to which pupil dilating drugs were administered have also been excluded from the dataset. This was done in order to leverage disease-induced changes only, and to eliminate any changes caused by the examiner’s actions during the patient’s visit or the effects of treatment that patients may have undergone between individual visits. For this reason, the number of distinct irises and images (shown in Tab. 2.3) do not total 230 (the number of all unique irises in the dataset) and 1793 (the number of all NIR iris images in the dataset), respectively.

Additionally, since cataract is the most represented single illness in the database, a separate subsets of images representing this disease only could be created. This one is later referred to as the Cataract subset. This subset is comprised of samples that already belong to one of the five subsets defined above, and therefore is not sample- or subject-disjoint.
Table 2.3: Numbers of unique irises (classes) and numbers of unique samples in each of the five data subsets, and the total number of classes and samples in the data subset selected for this particular study.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Number of irises</th>
<th>Number of near-infrared images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>35</td>
<td>216</td>
</tr>
<tr>
<td>Clear</td>
<td>87</td>
<td>568</td>
</tr>
<tr>
<td>Geometry</td>
<td>53</td>
<td>312</td>
</tr>
<tr>
<td>Tissue</td>
<td>8</td>
<td>50</td>
</tr>
<tr>
<td>Obstructions</td>
<td>36</td>
<td>207</td>
</tr>
<tr>
<td>Total</td>
<td>219</td>
<td>1353</td>
</tr>
</tbody>
</table>

2.5.3. Evaluation methodology

To answer the first question formulated in Sec. 2.1, namely whether ocular pathologies impact the enrollment process, failure-to-enroll error rates (FTE) are calculated for each database subset and using all four iris recognition methods introduced in Sec. 2.5.1.

To answer questions two and three, namely what kinds of impairments in unhealthy irises have the greatest impact on iris recognition, all possible genuine and all possible impostor comparison scores were obtained for each database subset. To judge whether the observed differences in comparison scores across subsets can be considered as samples drawn from the same distribution, a two-sample Kolmogorov-Smirnov test is applied with the significance level $\alpha = 0.05$ (further referred to as K-S test). One-sided variant of the test is used when comparing genuine score distributions, and a two-sided variant for the impostor score distributions. The K-S test makes no assumptions on the distributions (apart from their continuity) and the test statistics simply quantifies the distance between two empirical cumulative distribution functions $F(x_1)$ and $F(x_2)$ of the random variables $x_1, x_2$ being compared. The K-S test statistic for the two-sample variant is:

$$D = \sup_x |F(x_1) - F(x_2)|$$

and the null hypothesis is rejected at significance level $\alpha$ when

$$D > c(\alpha)\sqrt{\frac{n + m}{nm}}$$
where \( c(\alpha) = \sqrt{-\frac{1}{2}\ln(\alpha)} \) is the inverted Kolmogorov distribution at \( \alpha \). To alleviate the issue of statistical dependencies between comparison scores that are introduced when performing all possible comparisons between samples, we resample with replacement each set of comparison scores 1,000 times for genuine scores and 10,000 times for impostor scores, providing sets of statistically independent samples.

**For answering question four**, related to the impact cataract can cause, similarly to the methodology employed for questions two and three, all possible genuine and impostor comparisons were performed in both the Healthy and Cataract subsets. These are later represented in the form of empirical cumulative distribution functions and compared using the analogous resampling and K-S testing procedure.

Finally, **to answer the fifth question** regarding reasons for erratic performance, an analysis of segmentation errors and visual inspection of eye samples resulting in the worst comparison scores were performed for selected matchers.

### 2.6. Results and discussion

#### 2.6.1. Enrollment performance

FTE rates obtained in each subset (Tab. 2.4) suggest that iris recognition performs the worst for samples included in the **Geometry** and **Obstructions** subsets, and this behavior is observed across all matchers, albeit to a different extent (0.32% and 0.97% for IriCore and even 16.03% and 18.36% for MIRLIN). Those subsets comprise images in which the pupil is either distorted or not visible at all due to various types of occlusions. It is worth noting that the enrollment process realized for different iris recognition methods is affected unevenly across the methods. This can be explained by different quality metrics implemented in the employed algorithms and their varying sensitivity to quality issues generated by eye disorders. To summarize, the enrollment stage is sensitive to those conditions that distort pupil geometry or that obstruct, partially or completely, the iris pattern. The observed impact on the enrollment process, however, is uneven across algorithms.
Table 2.4: FTE rates obtained in each subset for four iris recognition methods used in this work. The worst result for each method is in **bold type**. The second column provides reference to sample images from each subset, shown in Fig. 2.1

<table>
<thead>
<tr>
<th>Subset</th>
<th>Samples in Fig. 2.1</th>
<th>MIRLIN</th>
<th>VeriEye</th>
<th>OSIRIS</th>
<th>IriCore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>H1 - H3</td>
<td>1.85%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Clear</td>
<td>C1 - C3</td>
<td>4.40%</td>
<td>0%</td>
<td>1.23%</td>
<td>0%</td>
</tr>
<tr>
<td>Geometry</td>
<td>G1 - G3</td>
<td>16.03%</td>
<td>5.13%</td>
<td>5.45%</td>
<td>0.32%</td>
</tr>
<tr>
<td>Tissue</td>
<td>T1 - T3</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Obstructions</td>
<td>O1 - O3</td>
<td>18.36%</td>
<td>3.86%</td>
<td>8.21%</td>
<td>0.97%</td>
</tr>
</tbody>
</table>

2.6.2. Influence in respect to type of illness

Cumulative distributions $F$ of all possible comparison scores calculated for four iris recognition algorithms are shown in Figs. 2.2 and 2.3. In each figure we collectively plot five $F$ graphs (for all considered subsets) along with the mean values of the comparison scores to visualize possible differences among subsets. $F$ graphs highlight the differences in genuine and impostor scores independently, which brings the first observation that eye disorders impact same-eye comparisons to a higher extent when compared to different-eye matching, since differences between $F$ graphs for genuine comparisons are significantly larger than for impostor comparisons.

Similarly across all methods, the Geometry and Obstructions subsets present the worst scores, with the $F$ graph shifted to the right for VeriEye, and to the left for the remaining matchers, when compared to the Healthy subset serving as a control group. Uneven behavior can be observed for the Clear subset, which gives worse scores for most matchers, except for the OSIRIS matcher, in which its $F$ graph intertwines with the $F$ graph corresponding to the Healthy subset. Surprisingly, the Tissue subset displays behavior that is either similar to this of the Healthy subset, or even slightly better. These fluctuations can be, however, attributed to the small number of samples in the Tissue subset, which makes it easier for given samples to influence the performance of the whole subset.

Table 2.5 presents the results of K-S testing for genuine comparisons, which confirms that the differences between the Clear, Geometry, and Obstructions subsets when compared against the Healthy subset are statistically significant. As for the impostor-related $F$ graphs, again rather large differences may be observed for the IriCore matcher, and smaller for the remaining three methods. The K-S statistical
testing again confirms that there are statistically significant differences between each of the four subsets comprising diseased eyes and the Healthy subset, Tab. 2.6).

In addition to $F$ graphs, we also present Receiver Operating Characteristic (ROC) curves which demonstrate the accuracy of these iris recognition systems when they are presented with eyes afflicted by cataract, compared to a scenario, in which healthy eyes are used. The ROC curve presents a relation of true positive ratio to false positive ratio obtained by a given decision system and is therefore helpful for assessing its expected behavior. Fig. 2.4 present ROC curves plotted collectively for all five data subsets representing different types of damage inflicted to the eye. This is repeated for all four
of the iris matchers involved in this study. Here as well, the Geometry and Obstructions subsets are giving the worst performance. Surprisingly, for the IriCore and VeriEye matchers, the Clear subset also performs poorly, while the Tissue subset is behaving similarly or better than the Healthy subset.

2.6.3. Cataract influence

F graphs for the Cataract and Healthy subsets shown in Figs. 2.5 and 2.6 reveal that cataract-afflicted eyes perform worse than their healthy counterparts: for each of the four iris matchers we can observe a visible shift of the genuine score distribution
the quantifying a system’s performance. For all four employed iris recognition methods, are shown in Fig. 2.7. Equal Error Rate values are also provided as another metric for The results of these tests are presented in Table 2.7. ROCs for respective iris matchers
dure, involving Kolmogorov-Smirnov (K-S) tests, as described previously in Sec. 2.5.3. in score distributions are smaller and uneven across matchers (highest for IriCore, higher for the remaining three methods). Regarding impostor scores, the differences obtained from the Cataract subset towards worse scores (i.e., lower for VeriEye, and higher for the remaining three methods). Regarding impostor scores, the differences in score distributions are smaller and uneven across matchers (highest for IriCore, negligible for MIRLIN and OSIRIS).

To formally confirm these observations of the $F$ graphs, a statistical testing procedure, involving Kolmogorov-Smirnov (K-S) tests, as described previously in Sec. 2.5.3. The results of these tests are presented in Table 2.7. ROCs for respective iris matchers are shown in Fig. 2.7. Equal Error Rate values are also provided as another metric for quantifying a system’s performance. For all four employed iris recognition methods, the Cataract subset gives worse ROC-wise performance that the Healthy subset except

### Table 2.5: Kolmogorov-Smirnov statistical testing results for the disease influence type experiment, genuine comparisons. The null hypotheses $H_0$ in all tests state that the samples originating from two compared subsets are drawn from the same distribution. Alternative hypotheses are detailed in rows labeled $H_1$. One-sided test is used. $F(g_k)$ denotes the cumulative distribution function of $g_k$, where $g_k$ denotes the genuine scores calculated to the $k$-th subset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Clear ($g_c$) vs. Healthy ($g_h$)</th>
<th>Geometry ($g_y$) vs. Healthy ($g_h$)</th>
<th>Tissue ($g_t$) vs. Healthy ($g_h$)</th>
<th>Obstructions ($g_o$) vs. Healthy ($g_h$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSIRIS</td>
<td>$H_1$ $F(g_c) &lt; F(g_h)$</td>
<td>$F(g_y) &lt; F(g_h)$</td>
<td>$F(g_t) &lt; F(g_h)$</td>
<td>$F(g_o) &lt; F(g_h)$</td>
</tr>
<tr>
<td>p-value</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$0.1071$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>MIRLIN</td>
<td>$H_1$ $F(g_c) &lt; F(g_h)$</td>
<td>$F(g_y) &lt; F(g_h)$</td>
<td>$F(g_t) &lt; F(g_h)$</td>
<td>$F(g_o) &lt; F(g_h)$</td>
</tr>
<tr>
<td>p-value</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$0.0039$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>IriCore</td>
<td>$H_1$ $F(g_c) &lt; F(g_h)$</td>
<td>$F(g_y) &lt; F(g_h)$</td>
<td>$F(g_t) &lt; F(g_h)$</td>
<td>$F(g_o) &lt; F(g_h)$</td>
</tr>
<tr>
<td>p-value</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$0.3325$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>VeriEye</td>
<td>$H_1$ $F(g_c) &gt; F(g_h)$</td>
<td>$F(g_y) &gt; F(g_h)$</td>
<td>$F(g_t) &gt; F(g_h)$</td>
<td>$F(g_o) &gt; F(g_h)$</td>
</tr>
<tr>
<td>p-value</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$0.6942$</td>
<td>$&lt; 0.0001$</td>
</tr>
</tbody>
</table>

### Table 2.6: Same as in Table 2.5, except that impostor comparison scores are analyzed and two-sided Kolmogorov-Smirnov test (for the resampled data) was applied. $F(i_k)$ denotes the cumulative distribution function of $i_k$, where $i_k$ denotes the impostor scores calculated to $k$-th subset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Clear ($i_c$) vs. Healthy ($i_h$)</th>
<th>Geometry ($i_y$) vs. Healthy ($i_h$)</th>
<th>Tissue ($i_t$) vs. Healthy ($i_h$)</th>
<th>Obstructions ($i_o$) vs. Healthy ($i_h$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSIRIS</td>
<td>$H_1 : F(i_c) \sim F(i_h)$</td>
<td>$H_1 : F(i_y) \sim F(i_h)$</td>
<td>$H_1 : F(i_t) \sim F(i_h)$</td>
<td>$H_1 : F(i_o) \sim F(i_h)$</td>
</tr>
<tr>
<td>p-value</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>MIRLIN</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>IriCore</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>VeriEye</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
</tr>
</tbody>
</table>
Figure 2.4: Receiver Operating Characteristic (ROC) curves obtained for all four iris recognition methods denoting the performance of these systems for five data subsets. Equal Error Rate values are provided in brackets. Figures reprinted from [71].

The performance of these systems for five data subsets is shown in the ROC curves. For the MIRLIN matcher, which gives similar EER values for both subsets. We can thus draw a conclusion that cataract can significantly degrade the accuracy of existing iris recognition methods, up to as much as 8 percentage points of the EER metric. The differences, however, are uneven across iris matchers, and the degradation of scores applies to genuine comparisons only, thus no additional false-matches can be expected.

2.6.4. Likely sources of errors

To reveal the actual reasons behind degraded accuracy, we performed a careful visual inspection of the samples that generated the worst comparison scores. As
impostor comparison scores are not impacted in a significant way, this is done only for the genuine scores. Since bad performance in iris recognition typically originates from incorrect execution of the segmentation stage, we employed two of the iris matchers that are capable of showing image segmentation results, MIRLIN and OSIRIS, to generate iris images with denoted iris localization results. This analysis confirmed that
Table 2.7: Kolmogorov-Smirnov statistical testing results for the cataract experiment. The null hypotheses $H_0$ in all tests state that the scores originating from two compared subsets are drawn from the same distribution. Alternative hypotheses $H_1$ for genuine scores state that scores obtained from Cataract set are worse than those obtained from Healthy subset, while for impostor scores $H_1$ state that scores obtained from Cataract set are different than those obtained from Healthy subset. One-sided test is used for genuine comparisons and two-sided test for impostor comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Genuine comparisons</th>
<th></th>
<th>Improver comparisons</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cataract ($g_c$) vs. Healthy ($g_h$)</td>
<td>$H_1: F(g_c) &lt; F(g_h)$</td>
<td>Cataract ($i_c$) vs. Healthy ($i_h$)</td>
<td>$H_1: F(i_c) \neq F(i_h)$</td>
</tr>
<tr>
<td>OSIRIS</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIRLIN</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IriCore</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cataract ($g_c$) vs. Healthy ($g_h$)</td>
<td>$H_1: F(g_c) &gt; F(g_h)$</td>
<td>Cataract ($i_c$) vs. Healthy ($i_h$)</td>
<td>$H_1: F(i_c) \neq F(i_h)$</td>
<td></td>
</tr>
<tr>
<td>VeriEye</td>
<td>$&lt; 0.0001$</td>
<td>$&lt; 0.0001$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.7: Receiver Operating Characteristic (ROC) curves obtained for all four iris recognition methods denoting the performance of these systems when cataract eyes are enrolled compared to a control group of healthy eyes. Equal Error Rate values are provided in brackets. Figures reprinted from [71].

failed iris localization is the most prevalent source of bad iris matcher performance. Segmentation errors that we have come across were most likely caused by some artifacts, such as distortions in the pupil boundary, obstructions such as corneal hazes, or damages to the iris tissue being interpreted by image segmentation algorithms as the pupil itself. Thus, the following matching stage, which is executed after the
segmentation stage, could not be performed correctly, but instead was performed using the non-iris portions of the image. This is especially true for Geometry and Obstructions subsets of the data, which is coherent with exceptionally poor ROC-wise performance of the data belonging to these subsets. VeriEye and IriCore algorithms do not provide a way to read the segmentation results, however, an examination of those samples that perform the worst when using these method identifies conditions that may be responsible for errors, namely: significant geometrical distortions, severe corneal hazes, blurred boundary between the iris and the pupil, letting us hazard a guess that segmentation issues are responsible for errors here as well.

2.7. Conclusions

The results presented in this Chapter prove the first part of the first thesis formulated in Sec. 1.2, namely that iris recognition is possible in case of ophthalmic disorders, however the accuracy of the existing iris recognition algorithms is significantly degraded.

In the first experiment the Author evaluated the recognition accuracy deterioration with respect to the type of damage inflicted by pathological processes in the eye, regardless of the actual medical origin and disease taxonomy. Deterioration in accuracy begins manifesting itself as early as at the enrollment stage, with FTE rates being significantly higher for eyes with geometrical distortions in the pupillary area, or those with pathology-induced objects interfering with correct imaging of the iris. The same two types of impairments are also expected to have the most profound effect on the reliability of the comparison stage, since such eyes perform significantly worse than their healthy counterparts, mostly for same-eye comparisons. With large differences between genuine comparison scores, a substantial influence on the within-class variability can be expected, with possibility of additional false-non matches happening. However, since the impostor comparison scores for disease-affected images do not substantially differ from those obtained from healthy irises, a probability of additional false matches is low.

For all algorithms employed, there are also slight changes in comparison scores obtained from healthy eyes and those afflicted with, but not visibly affected by disease.
While the observed differences are statistically significant, they are rather small and do not show a potential for generating false rejections or false acceptances.

An attempt to explain the underlying reasons of such poor performance is carried out, pointing to image segmentation errors as the predominant source of performance deterioration. However, knowing which types of eye damage are the ones most likely to cause recognition errors, one can employ visual inspection of the eyes of a person under enrollment to assess whether iris recognition can be reliably used to manage this person’s identity.

The latter of the two experiments deals with probably the most proliferated eye illness worldwide – the cataract, proving that despite usually not affecting the eye and the iris in a significant way, this pathology is capable of causing serious negative impact on the performance of state-of-the-art iris recognition technologies used today. With Equal Error Rates for the cataract-affected eyes being a few percent higher than those obtained with data corresponding to healthy eyes, one may expect recognition accuracy to be noticeably lower for people suffering from this illness. Combining this with high number of cataract occurrences, especially in third-world countries, leads to a conclusion that this issue should be seriously taken into consideration when building future, large-scale biometric applications employing iris recognition.
3. Post-mortem iris recognition

3.1. Iris recognition after death

Identification of deceased individuals through their biometric traits has long been used for forensic purposes, exploiting characteristics such as fingerprints, DNA, or dental records to recognize victims of accidents, or natural disasters and crimes [82, 83]. Post-mortem iris recognition, however, has not received considerable attention, despite excellent performance of this biometrics when applied to live eyes. Studying this area has at least two important goals:

To aid forensics: Can iris biometrics be a fast and accurate complement or alternative method to the existing approaches to post-mortem identification? If the answer is affirmative, it could be useful in cases when other methods cannot be applied, such as for victims of accidents with severed fingers or disfigured faces.

To improve security: Can dead iris be effectively used in presentation attack detection? Understanding the dynamics and reasons for post-mortem iris performance degradation allows to provide more precise answer to this question, and may help in development of better countermeasures against forgeries with cadaver eyes.

To come up with as many answers as possible, this Chapter presents a comprehensive study of post-mortem iris recognition involving iris images acquired from 5 hours to almost 34 days after death in near-infrared (NIR) and visible light (VIS). It aims at answering the following six questions:

1. Is automatic iris recognition possible after death?
2. What are the dynamics of deterioration in iris recognition performance?
3. What type of images are the most favorable for post-mortem iris recognition?
4. What are the main reasons for errors when comparing post-mortem iris samples?
5. Which factors influence post-mortem iris recognition performance?

6. What are the false-match risks when post-mortem samples are compared against databases of live iris images?

To answer the above questions, 2,294 NIR and 2,572 VIS images from 79 cadavers were acquired in cooperation with the Medical University of Warsaw, during multiple sessions organized from 5 to 814 hours after death. The bodies were kept in controlled mortuary conditions and stable temperature of 6° Celsius (42.8° Fahrenheit). Four independent iris recognition methods were used to show that automatic iris recognition stays occasionally viable even 21 days after death, and is close to perfect approximately 5 to 7 hours post-mortem. This allows to reject prior hypotheses that the iris cannot be used as a biometrics after death [61, 62, 63, 64]. We also show that using the red channel of VIS post-mortem iris images can be considered as a good alternative to NIR samples. Images composed of only red channel of the VIS samples will be later referred to as ‘R images’ in the paper. We also show that the performance of cross-wavelength post-mortem iris matching (NIR vs R) is significantly worse than same-wavelength (NIR vs NIR and R vs R) matching. We analyze possible reasons for false match and false non-match instances, and by manual correction of the segmentation for the whole dataset we assess the impact of erroneous segmentation on the post-mortem iris recognition performance. Medical commentary on these post-mortem metamorphoses observed in the eye that degrade the recognition reliability the most is provided.

The original research included in this Chapter appeared in parts in the following publications co-authored by the Author over the course of his PhD program: [12, 13] and most recently [16].

3.2. Related work

3.2.1. Common beliefs and opinions

A belief that iris recognition is difficult or impossible after a person’s death has been hypothesized for a long time in both scientific and industry communities. In 2001, John Daugman, who without doubt can be referred to as ‘the father of iris recognition’, stated the following in his interview for the BBC: ‘soon after death, the pupil dilates considerably, and the cornea becomes cloudy’. While this statement is fairly moderate,
others put forward far stronger claims, for instance, Szczepanski et al. write that 'the iris (...) decays only a few minutes after death’ [62]. References to post-mortem iris recognition can be also found commercial materials, for instance: (...) the notion of stealing someone’s iris after death is scientifically impossible. The iris is a muscle; it completely relaxes after death and results in a fully dilated pupil with no visible iris at all. A dead person simply does not have a usable iris!’ [63], or ‘after death, a person’s iris features will vanish along with pupil’s dilation’ [64]. However, none of these assertions are backed by any scientific argumentation or experimentation.

3.2.2. Experimental studies

Due to technical and ethical difficulties in collecting biometric samples from cadavers, only a small number of researchers have studied the post-mortem iris recognition problem using scientific methods. Sansola [65] used IriShield M2120U iris recognition camera together with IriCore matching software in her experiments involving 43 subjects who had their irises photographed at different post-mortem time intervals. Depending on the post-mortem interval, the method yielded 19-30% of false non-matches and no false matches. She reported a relationship between eye color and post-mortem comparison scores, with blue/gray eyes yielding lower correct match rates (59%) than brown (82%) or green/hazel eyes (88%). Saripalle et al. [66] used ex-vivo eyes of domestic pigs and they came to the conclusion that irises are slowly degrading after being taken out of the body, and lose their biometric capabilities 6 to 8 hours after death (2015). However, ex-vivo eye degradation is expected to be much faster than the same processes occurring while the eye is still a part of the cadaver. Ross [84] observed a fadeout of the pupillary and limbic boundaries found in post-mortem iris images, as well as corneal opacity, which developed in all of the samples under observation.

In our first study [12], we showed that despite popular claims, the iris can still successfully serve as a biometric identifier for 27 hours after death. The pupils were found to remain in the so called ‘cadaveric position’, meaning that no excessive dilation or constriction is present, and hence the iris structure remains well visible. The data were collected in three acquisition sessions:

- session 1: 5-7 hours post-mortem (S1 in Fig. 3.1)
- session 2: 16.5-21 hours post-mortem (S2 in Fig. 3.1)
Figure 3.1: Cumulative distribution functions of genuine comparison scores for four iris recognition methods employed in [12]. Same-session results (S1 vs S1 samples) are shown in violet, whereas inter-session results are shown in blue (S2 vs S1 samples) and red (S3 vs S1 samples). FNMR is shown for the default acceptance threshold. Mean comparison scores are shown in brackets. Plots reprinted from [12].

— session 3: 27 hours post-mortem (S3 in Fig. 3.1)

When observing same-session comparisons (violet lines in Fig. 3.1) it is evident that dead irises can be encoded and recognized in more than 90% of the cases, reaching even a perfect recognition for one method (IriCore, cf. Fig. 3.1a). The analysis of the following two cumulative distributions (blue and red lines in Fig. 3.1) shows that progressing iris degradation can be observed. FNMR increases significantly to 48.88%
for the academic OSIRIS method, however, two commercial matchers – IriTech and MIRLIN – still present fairly good ability to recognize the samples (approx. 94.96% and 82.76% of correct verifications achieved for IriCore and MIRLIN methods, respectively, cf. Figs. 3.1a and 3.1b). Surprisingly, we are still able to correctly recognize 73.33% of irises 27 hours after death using IriCore method (cf. red graph in Fig. 3.1a). Performance of the remaining methods is uneven, since they are able to recognize from 13.33% (OSIRIS, cf. Fig. 3.1d) to 60% (VeriEye, cf. Fig. 3.1c) of cadaver irises 27 hours after death. These results show that the dynamics of post-mortem changes are much lower than commonly believed.

Later in [13], these experiments were extended to a larger database, encompassing samples collected up to 17 days post-mortem. A short-term analysis, concentrating on samples acquired up to 60 hours after death, revealed that the best performing IriCore method can still offer EER as low as 13%, which indicates that iris recognition can still be a reasonable identification tool after such a period of time. Long-term analysis taking into account all samples collected over a period of 17 days, however, shows that the iris deterioration progresses fast over such a long time horizon, and although correct matches can still be expected even after these 17 days, they are sparse and cannot be considered a reliable proof of one’s identity post-demise. In this work, we have also offered the first publicly available database of 1330 near-infrared and visible light post-mortem iris images acquired from 17 cadavers [85].

Bolme et al. [14] attempted to track biometric capabilities of face, fingerprint and iris during human decomposition. Twelve subjects were placed in the outdoor conditions to assess how the environment and time affect the biometric performance. Although fingerprints and face are shown to be moderately resilient to decomposition, the irises degraded quickly regardless of the temperature. The authors state that irises typically became useless from the recognition viewpoint only a few days after exposition to outdoor conditions, and if the bodies are kept outside for 14 days the correct verification rate was only 0.6% in their study. The real-life chance of recognizing an iris is estimated by the authors to be less than 0.1%. The most recent paper in this field by Sauerwein et al. [15] showed that irises stay readable for up to 34 days after death, when cadavers were kept in outdoor conditions during winter. The readability was assessed by human experts acquiring the samples and no iris recognition algorithms were used
in this study, however it suggests that low temperatures increase the chances to see an iris even in a cadaver left outside for a longer time.

### 3.2.3. Conclusions from the literature review

Table 3.1 summarizes the scientific literature reviewed in this Section. Similarly to the disease-related studies, cf. Chapter 2, the FNMR values reported in these works highly exceed those expected for healthy irises, namely less than 2% according to the IREX IX report [18]. Again, such evaluations are usually suffering from very limited data available to researchers. Until 2016, there were no publicly available databases comprising iris images collected from deceased subjects. We have introduced the first database of such kind in [13], which is offered to researchers. Notably, none of these papers offer any kind of method that is aimed for making iris biometrics more robust against post-mortem changes to the human eye. Therefore, the question of how to get iris recognition to work reliably in such scenarios was still open. This dissertation proposes the first such method.

Table 3.1: Summary of the relevant scientific literature.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sansola [65]</td>
<td>2015</td>
<td>19% to 30% of false non-matches reported for post-mortem iris samples depending on the time since death (FNMR=19-30%)</td>
</tr>
<tr>
<td>Saripalle et al. [66]</td>
<td>2015</td>
<td>ex-vivo pig eyes lose biometric capabilities approx. 8 hours after being taken out of the cadaver</td>
</tr>
<tr>
<td>Trokielewicz et al. [12]</td>
<td>2016</td>
<td>post-mortem iris verification can work up to 27 hours after death no excessive dilation found, as previously believed (FNMR=26.7% for IriCore method, 27 hours since death)</td>
</tr>
<tr>
<td>Trokielewicz et al. [13]</td>
<td>2016</td>
<td>EER=13% up to 60 hours post-mortem (IriCore) occasional matches up to 17 days, but tissue deterioration extensive database offered to the community</td>
</tr>
<tr>
<td>Bolme et al. [14]</td>
<td>2016</td>
<td>correct verification rate 0.6% for bodies kept outdoor for 14 days (assessment by automatic methods)</td>
</tr>
<tr>
<td>Sauerwein et al. [15]</td>
<td>2017</td>
<td>some irises readable after 34 days outdoor during winter (assessment only by a human examiner)</td>
</tr>
<tr>
<td>Trokielewicz et al. [16]</td>
<td>2018</td>
<td>this paper publishes the most recent findings included also in this doctoral dissertation significantly extended database offered to the community</td>
</tr>
</tbody>
</table>
3.3. Dataset of cadaver iris images

3.3.1. Data collection

A crucial part of this study was to create a new database of iris images, which would represent eye regions of recently deceased persons. We had a rare opportunity to collect iris scans from hospital mortuary subjects. Similarly to the procedure described in Sec. 2.4, the collection of this novel database was carried out in close collaboration with Dr. Piotr Maciejewicz, MD from the Department of Ophthalmology at the Medical University of Warsaw, who performed the collection of iris photographs. The contribution of the Author was again to design the database collection protocol and the acquisition camera setup, as well as the curation of the database. The following section briefly characterizes the acquisition methodology and timeline of acquisition sessions.

**Equipment.** Two different sensors were used for image acquisition: a commercial NIR iris sensor IriShield M2120U, and a consumer-grade color camera Olympus TG-3. Color images were collected simultaneously with NIR ones and each subject and each acquisition session are represented by at least one image of each type. The IriShield sensor is equipped with a near-infrared illuminant, whose irradiance falls into the 710-870 nm band, with a peak at 810 nm [86].

**Environmental conditions** All acquisition sessions were conducted in the hospital mortuary. The temperature in the mortuary room was approximately 6°C (42.8°Fahrenheit). Other conditions, such as air pressure and humidity were unknown, yet stable. The environmental conditions, in which the cadavers were kept prior to entering the cold storage are unknown.

**Acquisition timeframe** Depending on the subject, 1 to 13 acquisition sessions could be organized during data collection. In each session at least one NIR and one VIS image were acquired. Subjects were not available prior to passing, therefore no ante-mortem samples could be collected. The first session for each subject was thus always organized as soon after death as possible, typically 5 to 7 hours. The following sessions were organized based on the availability of deceased persons, who were
subject to medical or police investigations, and were retained in the mortuary during varying time slots. The overview of acquisition sessions for all subjects is shown in Fig. 3.2. For 12 subjects, only a single acquisition session was possible.

**Within-session acquisition protocol** When collecting images within a single acquisition session, all samples can be considered separate presentations as recommended by the ISO/IEC 19795-2, *i.e.*, after taking a photograph, the camera was moved away from the subject and then positioned for the next acquisition.

### 3.3.2. Databases of iris images

**Near-infrared and visible-light samples** During data acquisition, we had the opportunity to collect images using two types of cameras: one producing near-infrared images of VGA size (640 × 480 pixels), and the second producing color photographs of high resolution. Using only the red channel of color iris samples has been found to offer high recognition accuracy [87] even when being matched with near-infrared samples [88]. Thus, in this study, two types of samples are used: a) original near-infrared and compliant to ISO/IEC 19794-6 standard, and b) red channel of visible-light images manually center-cropped to conform the VGA image type, as defined in ISO/IEC 19794-6, Fig. 3.3. The resolution of all images is 640 × 480 pixels. This cropping of visible-light sample additionally protects the identity of donors, as original images contained portions of the face region.

**Manually-annotated ground-truth iris masks** The unstable execution of the segmentation stage is usually a main cause of drops in iris recognition performance, when samples presented to the algorithms are of challenging nature. To test whether this is also the case for post-mortem data, we have taken the effort to prepare manually annotated iris masks for all of the iris images involved in this study, including both NIR and visible light samples, as depicted in Fig. 3.3.

**Dataset of live iris images** For the purpose of assessing false-match risks in scenarios when post-mortem samples are expected to be compared in an open-set scenario with the existing datasets of live iris images, we have collected a complementary dataset of iris images from 74 living subjects. To minimize the bias in the data, we have used the
Figure 3.2: Hours post-mortem for each acquisition session plotted independently for each deceased subject.
same sensor (IriShield MK 2120U) as used for the collection of post-mortem data. Two subject-disjoint subsets gathering data from 37 subjects each were created, containing 557 and 611 images, respectively.

3.3.3. Statistics

The post-mortem dataset comprises 2,294 NIR images, accompanied by 2,572 color images. These images represent eye regions of 79 different subjects (157 different irises, since only one eye was imaged for one cadaver). Age of the deceased ranged from 19 to 77 years old. 11 subjects were female and 68 were male. Causes of death included heart failure (42 subjects), car or train crash (12), suicide by hanging (11), suicide by jumping (2), unspecified type of murder (2), shooting (1), poisoning (4), and head trauma (5). The eye colors were blue/gray/light green (61 cadavers), light brown/hazel (11) and dark brown (7). This database is publicly available in version 1 and an expanded version 2 [13, 16].

3.4. Medical background

3.4.1. General overview

Initially, the post-mortem decomposition of human organs may not be visible to the naked eye, since these processes start at the cellular level and then slowly progress to a macroscopic level. Early changes include algor mortis (body cooling), rigor mortis (desiccation with stiffening of the body), pallor mortis (paleness) and livor mortis (lividity), while late ones comprise of progressing decomposition caused by autolysis and putrefaction. Autolysis is a cellular self-destruction process caused by hydrolytic enzymes that were originally contained within cells. Putrefaction is a degradation of
tissue caused by microorganism (e.g., bacterial) activity, and is visible macroscopically as discoloration or bloating of the skin.

3.4.2. The cornea

The most prominent metamorphoses observed in the eyes after death, and possibly the most troubling for iris recognition, are the changes to the cornea. A live cornea is a clear, transparent, dome-shaped structure in front of the eyeball. It is responsible for about 2/3 of the total eye optical power because of its curvature and the resulting refractive index. The cornea must remain transparent to refract light properly, but also to allow good quality iris image capturing. Its transparency is maintained by a controlled hydration with the tear film, produced by lacrimal glands and distributed by eyelids. As secretion stops, anoxia, dehydration and acidosis lead to progressing autolysis of the cells. Corneal thickness decreases immediately after death and increases thereafter. This results in opacification that increases with time. Upon death the cornea slowly becomes hazy. The change in corneal opacity is believed to be secondary to the change in hydration and architectural destruction of the collagen fiber network, functional alteration of corneal endothelium, disregulation of proteoglycan hydration and ion concentration in corneal stroma. It was confirmed that temperature has significant influence on protein degradation. Another effect associated with these mechanisms is the wrinkling of the corneal surface, manifesting itself with difficulties to obtain a good visibility of the underlying iris pattern. The progression of these effects is influenced by multiple factors, such as closure of the eyelids, environment humidity, temperature, and air movement. It is also dependent on the age and general medical condition of the deceased person. Due to reduced intraocular pressure, we can also notice central depression of the globe, flaccidity of the eyeball, and loss in its firmness [89].

3.4.3. The iris

There are no evident changes to the iris surface observed after death. After demise, pupils are usually mid-dilated in the so-called ‘cadaveric position’, and in some cases they can be slightly dilated, because of the relaxation of the iris muscles and later they can become slightly constricted with the onset of rigor mortis of the constrictor muscles. In other cases, we may observe initial myosis within the first few hours after
death with strong variations between individual cases. If rigor mortis affects ciliary
muscles of two irises unequally, pupils in both eyes may have different apertures.
Sometimes, if different segments of the same iris are unequally affected then the pupil
may be irregularly oval or have an off-center position. Shape and size of the pupils can
also depend on the medical history of the subject, including treatment with drugs and
eye surgeries.

3.4.4. Muscles of the iris

Death was once defined as the cessation of heartbeat and breathing, but with the
development of cardiopulmonary resuscitation these can be restarted in some cases.
Thus, we now typically rely upon the concept of brain death to define whether a person
is clinically dead. Supravitality – sensitivity to excitation – relates to survival rates of
tissue after complete irreversible ischemia (restriction of blood flow). The supravital
reaction is the response of muscles to stimulation in the early period after death, as
some cells do not die immediately after the brain death. Within the first couple of hours
we can notice a decreasing pupillary reaction for pupillomotoric drugs, for example to
pilocarpine and atropine [90]. Iris tissue response to myotic (pupil-constricting) and
mydriatics (pupil-dilating) agents can be observed.

3.4.5. Other aspects

There are other changes that the eye undergoes after death. The loss of intraocular
lens transparency with time is due to the metabolic processes that take place between
the lens and the aqueous and vitreous humor, and the aggregation of crystalline
proteins in the fibers of the lens nucleus. It has been hypothesized that lens proteins
aggregate to large particles that scatter light, causing lens opacity [89]. After death we
may observe a black spot in the sclera, referred to as ‘tache noire’, caused by desiccation
of the sclera with open eyelids, usually symmetrical corresponding to the position
of the eyelids. Also, the vitreous humor – a gelatinous substance contained in the
posterior chamber of the eye, keeping the retina in place and maintaining the spherical
shape of the eyeball – tends to liquefy and later to dry, starting the process of eyeball
collapse [91].
3.5. Visual inspection of post-mortem changes

We have taken the effort to carefully examine the samples throughout the time period since death for all subjects, and confront the observed changes with medical knowledge. This yielded a qualitative evaluation of post-mortem changes to the iris reported in this Section. Having both NIR and VIS images is crucial for such assessment, as these two types of illumination often reveal different appearance of the iris when changes to the cornea and the anterior chamber are present. This is shown in Fig. 3, where visible-light samples are compared against near-infrared samples for the same eye. NIR illumination typically used in iris recognition cameras is capable of alleviating corneal opacification effects to some extent. Such differences are also reported in [5] and [11].

A summary of example post-mortem changes is presented in Fig. 3.4, together with a timeframe for a selected subject. It must be noted, however, that the dynamics of these changes are heavily subject-dependent and can happen with different rapidity, intensity and prevalence on the appearance of iris tissue.

First, a corneal opacification progresses with time since death, and it becomes visible after a few days post-mortem (e.g., 95 hours, or 4 days, after death, as depicted...
in Fig. 3.4). Second, a wrinkling of the corneal surface is expected to appear (e.g., 359 hours, or 15 days, Fig. 3.4). At this point, a strong influence on the automatic image segmentation procedures can be anticipated, as the iris tissue becomes less visible and additional patterns and light reflections emerge. Third, a loss of intraocular pressure in the eyeball due to post-mortem biochemical changes can be observed (e.g., 574 hours, or 24 days, Fig. 3.4), causing the eye to slowly collapse into the eye socket. At this point in time, iris recognition methods are expected to seldom work, as the iris pattern is severely obstructed and thus challenging for iris image segmentation. Finally, after about a month, the eyeball was observed to dry out completely, leaving no traces of a healthy iris structure.

Contrary to initial predictions, we did not come across any sample that would be affected by *tache noire*. Also, the severe corneal opacification was visible in original VIS samples only, while NIR and R images worked in favor of exposing post-mortem iris texture better than original VIS samples, as depicted in Fig. 3.5.

3.6. Experimental methodology and tools

3.6.1. Iris recognition methods

For a comprehensive analysis of how iris recognition can perform when used with post-mortem samples, we have employed four independent iris recognition methods. Three of them are commercially available products, and one is an open source solution. These are the same methods that were previously used for the assessment of ophthalmic disease-relates iris recognition in Chapter 2, namely: OSIRIS, MIRLIN, VeriEye, and IriCore, with their descriptions already having been provided in Sec. 2.5.1. A modification of the original OSIRIS method was performed by the Author to include score normalization as proposed by Daugman [92]:

\[
HD_{norm} = 0.5 - (0.5 - HD_{raw})\sqrt{\frac{n}{N}}
\]  

(3.1)

This transforms the samples of scores obtained when comparing different eyes into samples drawn from the same binomial distribution, as opposed to drawing sample scores from different binomial distributions with \( \sigma \) dependent on the number of bits \( n \).
that were available for comparison (commonly unmasked bits). $N$ is the typical number of bits compared (unmasked) between two different irises, being said to equal 911 or 960, depending on the data. The $N$ parameter is estimated for a particular database of iris images. Since post-mortem iris samples are different from images of live iris images, and the OSIRIS does not necessarily use identical Gabor wavelets as used in [29], we estimated $N$ with our post-mortem samples and for the OSIRIS coding to be 1416 and 1446 for the automatic and manual segmentation, respectively. The total number of bits in the OSIRIS code is 1536.

3.6.2. Types of analyses

Due to difficult data acquisition resulting in sparse and irregular image capture moments, there are four experiments conducted in this part of the dissertation:
Table 3.2: Number of failed comparisons (genuine and impostor) and total number of comparisons calculated for all variants of analyses done in this study. Numbers are presented separately for each iris recognition method and suggest that iris image quality control mechanisms differ significantly among methods when post-mortem samples are used.

<table>
<thead>
<tr>
<th>Total number of comparisons (genuine + impostor)</th>
<th>Short-term analysis</th>
<th>Long-term analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NIR images</td>
<td>R images</td>
</tr>
<tr>
<td></td>
<td>39,621</td>
<td>76,245</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of failures to compute a comparison score</th>
<th>IrICore</th>
<th>MIRLIN</th>
<th>OSIRIS</th>
<th>OSIRIS (manual)</th>
<th>VeriEye</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>4,376</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>779</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>282 (0.11%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>106,989 (43.51%)</td>
<td>179,641 (40.11%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30,178 (12.27%)</td>
<td>55,484 (5.69%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2,538 (1.03%)</td>
<td>19,999 (4.47%)</td>
<td></td>
</tr>
</tbody>
</table>

1. a **failed comparisons analysis**, which determines the number of comparisons that could not be performed by each of the biometric methods,

2. a **short-term analysis**, which assesses the post-mortem recognition accuracy of the methods, and is based on samples collected in the first acquisition session for each subject, and consists of all possible same-session comparisons,

3. a **long-term analysis** that employs the entire data representing a maximum period of 814 hours after death, including all possible same- and inter-session comparisons,

4. an **open-set recognition** to assess the false match probability when post-mortem samples are compared to live samples.

Without having both ante-mortem and then post-mortem iris scans from the same individuals, and after visual inspection of samples (cf. Sec. 3.5), we assume that iris scans obtained shortly (*i.e.*, a few hours) after death would not differ much from those obtained ante-mortem. Images acquired in the first session therefore serve as gallery samples in all subsequent analyses.
3.7. Results and discussion

3.7.1. Failed comparisons statistics

Iris recognition methods are often equipped with image quality control mechanisms that prevent comparing samples with unacceptable properties. That is, if at least one of two samples being compared presents a less-than-accepted quality, the comparison fails and no valid comparison score is returned.

The required iris image quality seems to be defined differently in matchers used in this study, since the numbers of failed comparisons are not uniform. It is also difficult to hypothesize on what the exact reasons of these failures are, since out of four recognition algorithms, only one of them, IriCore, is explicitly said by the manufacturer to be compliant with the ISO/IEC 29794-6 standard on iris image quality. As for the open-source OSIRIS matcher, neither ISO/IEC metrics nor any other methods for discarding bad samples are implemented. For this method, samples denoted as those that failed to produce a template did so because of errors returned by low-level OpenCV routines used in OSIRIS: a ‘bad ROI’ error likely occurring when the segmentation stage fails to localize the iris region. For the remaining two matchers, VeriEye and MIRLIN, there is no information how the iris image quality is assessed in these commercial products.

Table 3.2 summarizes numbers of all comparisons in short-term and long-term analyses along with numbers of the failed comparisons. For instance, the most restrictive method, MIRLIN, fails to compute almost 44% of comparison scores when comparing NIR samples, and fails to compute over 40% of comparison scores when comparing R images. In turn, the IriCore software was able to compute all short-term comparison scores and failed to compute only 0.11% of long-term comparison scores for NIR images. Analyses presented in the remainder of this Section are based only on comparison scores that were correctly calculated. We should expect that methods controlling the iris image quality more restrictively should present a higher recognition accuracy (e.g., lower EER) than the methods implementing weaker quality discrimination.
Figure 3.6: ROC curves for scores obtained when matching intra-session samples within the first acquisition session (NIR images: blue line, R images: red line, cross-wavelength matching: dashed black line). Equal error rate (EER) is also shown for each case. Plots generated only for successful comparisons (i.e., not rejected automatically due to low image quality by the software). Plots reprinted from [16].
3.7.2. Baseline post-mortem performance: short-term analysis

Fig. 3.6 illustrates Receiver Operating Characteristic (ROC) curves obtained in short-term analysis, when matching samples obtained within the first acquisition session (5-7 hours post-mortem). Iris recognition performs reasonably well on average in short-term post-mortem horizon, which conforms with earlier works on a smaller data sample [12, 13]. The best performing method, IriCore, achieved EER=4% for NIR images, and EER=2% for R images.

Interestingly, using red channel of iris images acquired in visible light with a high-resolution camera, allows for a better performance of all of the involved iris recognition methods when compared to the performance achieved for VGA NIR images. We can observe even six-fold gains in recognition accuracy for the VeriEye matcher, measured EER-wise (2% error with R images versus 13% error for NIR images). One of the contributing factors for such a favorable performance of red-channel (R) images is likely a dominant proportion of lightly-colored eyes in the dataset, for which iris recognition is known to work well even without a specialized NIR camera, provided that the images are of good quality [87]. For completeness, we also include an analysis of cross-spectral matching scenarios, where R images are matched against NIR images (dashed black lines in Fig. 3.6), however, the results of such matching are discouraging and are not considered in the long-term analysis. Apparently, for post-mortem images, the differences in sample presentation under different illumination are large enough to make such a scenario unusable.

Note that all algorithms, except for the OSIRIS, did not reject any sample due to low quality (cf. ‘Short-term analysis’ columns in Tab. 3.2), which makes these methods perfectly viable for samples collected a few hours after death. The worst method in short-term analysis, MIRLIN, presents equal error rate of 7% for R images and 20% for NIR images. However, this could possibly be compensated by more restrictive quality control since MIRLIN, as IriCore, did not reject any sample prior to matching. Even so, the worst result (EER=20%) is still better than 50% expected for a random chance classifier, indicating a possibility to correctly recognize a large subset of cadaver irises.

Finally, when examining results obtained when matching manually segmented samples using the OSIRIS matcher (bottom-right plot in Fig. 3.6), we may draw a
conclusion that most of the recognition errors can be associated with erroneous execution of the automatic segmentation stage, as this method, when processing manually segmented images, presents EER=1% for R images and 2% for NIR images. This shows that when post-mortem samples are carefully segmented to represent the correct iris region, iris recognition is a viable method for identification.

3.7.3. Baseline post-mortem performance: long-term analysis

Samples acquired in the second and subsequent sessions are sparsely distributed in time and across the subjects. Thus, Figs. 3.7 through 3.15 present genuine and impostor scores calculated by all the methods, between session 1 images (5-7 hours after death) and all the samples acquired in later sessions, together with close-up analysis of example false matches and false non-matches.

Each method, except for OSIRIS with Daugman’s score normalization, generated false matches and all matchers generated false non-matches. Thus, we selected the worst pair of images in each case, i.e., the most similar images of different eyes and the most distinct images of the same eye, both in terms of the comparison score. We were also able to read the segmentation results in two methods (MIRLIN and OSIRIS), which helped to find a reason behind a given error.

For the OSIRIS method, when employing automatic image segmentation, we can expect correct matches for samples obtained up to 150 hours post-mortem for NIR images, and 263 hours for R images, Fig. 3.7. However, when the manual corrections to the segmentation stage are introduced, Fig. 3.9, the recognition horizon for NIR samples extends to 263 hours, while for R samples it declines to only 215 hours post-mortem, which may indicate that some of the correct matches obtained for R images were a result of badly segmented irises, and not a correct match of iris codes. Plots of impostor comparisons shown in Figs. 3.7 and 3.9 suggest that there is no clear trend in comparison scores when time after death increases, also in the case when manual segmentation was applied.

Figure 3.8 presents the worst results obtained from OSIRIS prior to applying score normalization. Incorrect segmentation explains the reason for a false non-match obtained for the left pair of samples. The right pair and the segmentation results show a very interesting case of a false match. Even with incorrect segmentation, as
Figure 3.7: Long-term analysis for all subjects for the OSIRIS matcher using automatic iris segmentation, for both NIR and R samples. Green dots represent correct behavior, i.e., a match for genuine pair, and a non-match for impostor pair, while red dots correspond to incorrect behavior, i.e., a false non-match for a genuine pair, and a false match for impostor pair between samples acquired in the first session (after 5-7 hours) and samples acquired in the following sessions. Graphics reprinted from [16].

(a) False non-match (0.60)  (b) False match (0.08)

Figure 3.8: Example image pairs generating a false non-match and a false match for the OSIRIS matcher prior to score normalization. The comparison scores are shown in brackets. Left image of each pair is the gallery sample. We assume that OSIRIS returns a match when the comparison score is below 0.32. After score normalization, this false match is rectified. Graphics reprinted from [16].

observed in this case, we still have some unoccluded image areas that are significantly different (part of the skin on the first image and part of the wrinkled iris/cornea on the second image). So why the false match is observed? The reason for that may be a
very small number of bits being compared due to application of occlusion mask during calculation of the Hamming distance. Note that the occlusions found by the OSIRIS in this image pair are almost mutually exclusive: the unoccluded part is mostly located on the left part of the hypothetical iris on the first image, while the unoccluded part of the iris shown on the second image is located mostly on the right. This false match does not happen after applying score normalization that penalizes low numbers of mutually un-occluded bits in OSIRIS codes, Eq. 3.1.

As for the remaining matchers, IriCore and MIRLIN were able to deliver correct matches for samples acquired even 503 hours post-mortem. VeriEye occasionally recognizes samples acquired up to 260 hours after demise, at the assumed acceptance thresholds. IriCore was again a method that rejected a very small number of comparisons (0.11% of NIR images and 0% of R images) when compared to other methods (cf. ‘Long-term analysis’ column in Tab. 3.2). This means that in favorable conditions (IriCore software and IriShield sensor come from the same manufacturer) iris recognition may still be possible almost 21 days after death. Unfortunately, as this method does not allow for generating segmentation results, we are not able to
investigate whether these correct matches are indeed a result of genuinely matching the corresponding iris features, or of an incorrect segmentation causing the similar portions of the image to be matched.

Figure 3.11 presents iris image pairs yielding false non-match (left pair) and false match (right pair) when the MIRLIN method is used. In this case we can also observe the segmentation results, which give a clear explanation of the observed errors in both cases. A false non-match is caused by comparing non-matching iris areas due
to incorrect iris localization. A false-match is probably due to comparing sclera that is very similar in both samples, rather than actual iris texture.

Figure 3.12: Same as in Fig. 3.7, but for the **IriCore** matcher. Graphics reprinted from [16].

(a) False non-match (1.39)

(b) False match (1.09)

Figure 3.13: Same as in Fig. 3.8, except for the **IriCore** matcher. We assume that IriCore method returns a match when the comparison score is below 1.1. Graphics reprinted from [16].

Figure 3.13 presents the worst cases leading to false non-match (left pair) and false match (right pair) when the IriCore method is used. The reason for a false non-match is the collapse of the eyeball and severe cornea wrinkling observed in the second sample (acquired 622 hours post-mortem) of the left pair. The right pair of images, however, does not provide any obvious clue for the observed false match.
Figure 3.14: Same as in Fig. 3.7, but for the VeriEye matcher. Graphics reprinted from [16].

(a) False non-match (0)  
(b) False match (46)

Figure 3.15: Same as in Fig. 3.8, except for the VeriEye matcher. We assume that VeriEye returns a match when the comparison score is above 40. Graphics reprinted from [16].

Figure 3.15 shows the worst pairs of images from the VeriEye method point of view. The possible cause of a false non-match (left pair) is a compensation of lower intraocular pressure by manually pressing the eyeball (a finger of a personnel pressing the eyeball is visible in the second image). This made the cornea less wrinkled (when compared to the first image of the left pair), but simultaneously changed the visible texture, ending up with creation of different iris features for those samples. However, the right pair of images again does not provide a clear explanation for a false-match.
3.7.4. Assessment of false match risk when comparing post-mortem samples against live iris samples

All the above experiments were performed for a closed set of deceased subjects. A question arises: what happens if post-mortem samples are compared in an open-set scenario with live irises of other subjects? To answer this, we used the IriShield MK 2120U sensor to collect an additional database of iris images from 74 living persons, which was split into subject-disjoint live-test and live-reference sets, both comprising data from 37 subjects and approximately 600 images. The most accurate (in this study) IriCore method was used to generate and compare two impostor score distributions: (a) scores obtained when matching post-mortem iris images (from 37 cadavers, a subset of the entire database, selected to balance the number of classes with the available living subjects data) against live-test iris images, and (b) scores obtained when matching the same live-test images with samples from the subject-disjoint live-reference set. The post-mortem set was created by randomly choosing images in a way that from each acquisition session half of the samples are selected. This ends up with balanced post-mortem and live iris sets, Fig. 3.16.

Fig. 3.17 (left) presents the False Match Rate calculated for different acceptance thresholds and for two above scenarios (a) and (b). For the threshold equal to 1.1, as recommended by the manufacturer, we observed small FMR $< 0.1\%$ in both scenarios. However, when the threshold is relaxed, the chances of getting a false match when live
Iris images are compared to post-mortem images are higher than in scenario when only live iris images are compared. At the same time, it is hard to see any clear dependency between false match rate and time after death when the iris was photographed, as depicted in Fig. 3.17 (right).

![Figure 3.17: Left: FMR as a function of IriCore acceptance threshold for comparisons between live irises and live vs post-mortem irises. Right: FMR as a function of post-mortem acquisition time calculated for the IriCore acceptance threshold 1.2, when live irises are compared with post-mortem irises. Image reprinted from [16].](image)

### 3.8. Conclusions

This Chapter delivers a comprehensive analysis of post-mortem iris recognition as performed using current state-of-the-art iris matchers and academic solutions, which proves the second part of the first thesis introduced in Sec. 1.2, namely that iris recognition is possible in cases of post-mortem changes to the eye, however the accuracy of the existing iris recognition algorithms is significantly degraded.

**Dynamics of iris deterioration.** Due to inevitable decomposition of human body, recognition accuracy becomes progressively worse. We did not find any iris image taken later than 503 hours after death that would result in a correct match in all four iris recognition methods used in this work. On the other hand, **503 hours (almost 21 days)** is a horizon giving an ample amount of time to make post-mortem forensic analysis and suggests that the iris may deliver actual biometric features for a longer period than initially believed. One should also note that four methods used in this study presented highly heterogenous performance on the same set of post-mortem samples. For instance, IriCore failed to compute only 0.11% comparison scores, while MIRLIN
failed to compute 43.51% comparison scores for NIR images. This may suggest that the algorithms implement significantly different mechanisms to control the quality of iris image pair used to compute the score.

**The most favorable type of iris imaging.** In the short-term analysis, when matching samples are obtained shortly after death, we have shown that employing high-resolution R images offers much higher recognition rates when compared to those obtained with NIR images (2% EER for R images versus 4% for NIR images, obtained for the IriCore matcher). Both types of images offer close-to-perfect performance when the image segmentation stage is executed manually, with small difference in favor of the R images (1% ERR versus 2% for NIR images, obtained for the OSIRIS matcher). This may prove important for forensic applications, which usually involve a human expert, who could then perform the necessary segmentation stage. In addition, we have shown that, probably due to significant differences in the appearance of iris features in post-mortem samples under different wavelengths, a cross-spectral scenario cannot be recommended, with ERRs not dropping below 15%, even with manual image segmentation in place.

In the long-term analysis, there is no single conclusion on whether NIR or R images are better for post-mortem iris recognition. For the method OSIRIS with manually corrected image segmentation, NIR images seem to offer better chance of getting a correct match. For MIRLIN, R images are generating far more false matches than NIR images, while for IriCore it is the opposite, with NIR images causing more false matches. However, NIR images also allow for correct matches during a longer period post-mortem, compared to R images. VeriEye, on the other hand, seems to work better in general, when R images are used, namely offering less false-matches and more correct matches. Therefore, the matching performance in regard to the image type is heavily matcher-dependent.

**Main reasons for errors.** After visual inspection of the automatic segmentation results for falsely matched and non-matched samples, one of the most obvious reasons for failures is incorrect localization of iris texture due to post-mortem decomposition processes. However, most of these failed examples still show the iris texture that
is partially or fully useful. It means that new methods in iris segmentation that are insensitive to post-mortem changes might increase the matching accuracy.

**Factors influencing post-mortem iris recognition performance.** The data collected for the purpose of this study represents images acquired from subjects who passed away due to various reasons, of different age, and of both genders. This gives a rare opportunity to examine whether cause of death, age and gender can give a priori insights on the expected performance of post-mortem iris recognition. Significantly worse comparison scores were observed for subjects who were either poisoned or murdered. Significant differences were observed neither between males and females, nor among groups of comparison scores sorted by the age in the moment of death. However, due to a relatively small number of available genuine comparison scores, these results should be considered as qualitative assessment, rather than formal statistical analysis.

**False-match risks when post-mortem samples are compared against databases of live iris images.** As shown in Sec. 3.7.4, false-match probability may be higher when live iris images are compared with post-mortem samples than when only live samples are used in comparisons. This translates to a higher chance of observing a false match when post-mortem probe sample is compared to a gallery of live iris images, and it calls for post-mortem-specific iris matching strategies to address a possibly of higher false match probabilities in post-mortem iris recognition.

**Limitations.** The most important limitation that we are aware of is a relatively small dataset of 79 deceased subjects. This data were collected in a very difficult environment of the hospital mortuary and acquisition moments could not interfere with various examinations, including criminal proceedings. Time spent in the mortuary was beyond our control, hence irregular acquisition sessions. The lack of ante-mortem samples limits the calculation of the reference templates to the earliest post-mortem images. However, visual inspection suggests that the first-session samples do not exhibit any visible deterioration when compared to living irises. Also, it would be valuable to repeat these analyses for samples collected under varying outdoor conditions, as done by Bolme et al. [14]. However, we are not aware of any other database of post-mortem iris samples that would be available to the researchers at the moment of preparation of this Thesis.
4. Iris image segmentation resistant to biological changes

4.1. Iris image segmentation

Erratic image segmentation is often put forward as a potential cause of degraded performance of iris recognition algorithms when they are made to work with difficult samples, with disease-affected and cadaver samples being no exception, as they usually pose trouble to image segmentation algorithms due to the non-circular appearance of the iris and the pupil, cf. Chapter 2. Post-mortem decay leads to macroscopic changes in the eye, such as deviations from the pupil’s circularity, cornea wrinkles causing additional specular reflections, and changes in the iris texture, cf. Chapter 3. Hence, for making iris recognition more reliable, a segmentation method should be designed specifically for post-mortem and pathology-induced deformations. This Chapter introduces such an algorithm, based on a deep convolutional neural network, and experimental results showing that it offers a considerable improvement over the segmentation results produced on the same data by a conventional segmentation method.

4.2. Related work

4.2.1. Deep convolutional networks for image segmentation

Deep convolutional neural networks (DCNN) are useful in solving selected groups of computer vision problems, such as image classification [93, 94, 95], automatic image captioning [96], visual question answering [97], and semantic segmentation by dense labeling, which has been reviewed extensively in [98]. These approaches are often named *data-driven*, as they learn the correct solution from the data itself, with minimum
use of prior knowledge and with a lot of parameters (usually a few million) and hyperparameters to be guessed directly from samples. This opposes to hand-crafted approaches that use the prior knowledge on the subject and the training encompasses fine-tuning of not-so-many (usually no more than a few hundred) hyperparameters. Both approaches have upsides and downsides, and data-driven models are often used when our prior knowledge on the subject is limited or difficult to be transformed into formulas possible to be applied in hand-crafted algorithms. Segmentation of post-mortem iris images is an example of such problems. One of the most successful DCNN architectures built for semantic segmentation tasks is SegNet, comprising a fully convolutional encoder-decoder architecture [99]. The encoder stage of SegNet is composed of the VGG-16 model graph introduced earlier by the Oxford-based Visual Geometry Group in [94]. The decoder stage comprises several sets of convolution and upsampling layers, whose target is to retrieve spatial information from the encoder output, and produce a dense pixel-wise classification output of the softmax layer that is of the same size as the input image.

4.2.2. Applications of convolutional networks to iris segmentation

Regarding the applications of iris segmentation utilizing neural networks, several attempts at this task have been made, mostly aiming to improve segmentation of difficult, noisy iris images, such as those collected in visible spectrum, using low quality equipment, and pictures captured on-the-move and at-a-distance.

Liu et al. [100] explored hierarchical convolutional neural networks (HCNNs) and multi-scale fully convolutional neural networks (MFCNs) for the purpose of improving segmentation of noisy iris images, e.g., visible light images with light reflections, blurry images captured on-the-move, at-a-distance, gaze-away eyes, etc., with iris pixels being located without any a priori knowledge or hand-crafted rules. HCNNs constructed by the authors employ hierarchical patches as input, ranging from scales small to large for capturing both local and global iris information. However, this approach is said to lack efficiency due to the sliding of the path window, which increases the computational overhead and limits the receptive field of neurons by the patch size. MFCNs, on the other hand, are reported to overcome these limitations with no sliding window (all pixel labels predicted simultaneously) and no limitation of the receptive
field size. Experiments were performed on the UBIRIS.v2 and CASIA.v4-distance databases, comprising noisy color images acquired in unconstrained conditions and NIR at-a-distance images, respectively. MFCNs use the VGG-21 model [94], trained for natural image classification, which is later fine-tuned using iris images with annotated masks. Segmentation errors defined as deviation from the ground truth segmentation by the proportion of disagreeing pixels are 0.9% on the UBIRIS.v2 and 0.59% on the CASIA.v4-distance dataset.

He et al. [101] approached the challenge of segmenting noisy iris images obtained in the visible spectrum with a modified DeepLab CNN model which is similar to VGG-16, but with fully connected layers replaced with feature pooling layers of kernel size $1 \times 1$ and an additional upscaling layer to match the output size to this of the input. The authors trained their solution on the visible spectrum iris dataset consisting of low quality samples, and reported an accuracy of 92% IoU (Intersection over Union, cf. Eq. 4.6), which outperforms the traditional Hough transform method applied to the same data. Similar problem is studied by Arsalan et al. [102], where two-stage method comprising of initial approximation of the iris boundary and finer localization with a fine-tuned VGG-face model. The solution is shown to achieve good accuracy in segmenting irregular specular reflections.

Jalilian and Uhl employed fully convolutional encoder-decoder networks (FCEDNs) to benchmark their performance on several iris datasets, including both good and poor quality samples [103]. These FCEDNs, based on the SegNet architecture, are reported to offer segmentation accuracy comparable with traditional approaches for good quality samples, and better for those of low quality.

4.2.3. Challenges in post-mortem and disease-affected iris image processing

An important conclusion from the existing literature is that DCNNs built for semantic segmentation tasks are a promising solution for dealing with poor quality iris images. Post-mortem iris images represent another category of difficult iris samples since they are heavily impacted by biological decay processes. In addition, metal retractors used to open the eyelids are often visible in the image as well, see Fig. 4.1. Similarly, samples coming from disease-affected eyes are often very different from what can be expected to see in a healthy eye, with possible distortions in the
pupillary region, obstructions of varying origin in front of the iris, or damage to the iris tissue itself, Fig. 4.2. These make conventional iris segmentation methods, e.g., those based on circular approximations of the iris inner and outer boundaries, inaccurate and therefore ineffective.

Since the FCEDN-based methods seem to offer state-of-the-art segmentation performance for difficult iris tasks, such as noisy images described above, we chose this group of methods for implementation in our task. One of the most popular, and receiving the most support on different platforms, architectures of FCDNs is the SegNet architecture [99]. The DCNN model has been retrained with ground truth iris masks excluding iris portions significantly affected by post-mortem changes.

4.3. Experimental methodology

4.3.1. Convolutional neural networks: background

Convolutional neural networks are a modification of the multi-layer perceptron designed to efficiently describe local features of the input, but also higher-order features in the deeper layers by merging the local features from distant parts of the input. In particular, CNNs differ from MLPs by two important properties:

— the receptive field of each neuron is limited to only selected inputs,
— the parameters in all neurons implementing the same convolution kernel are shared [104].

This section briefly introduces the main components of a typical CNN architecture, as well as briefly discusses the core idea behind the model optimization and techniques used during training.

Model building blocks

Convolutional layers extract local features of the input by convolving the input with a kernel:

$$z(i, j) = \sum_{k,l} f(i - k, j - l) h(k, l)$$ (4.1)
for every \( i, j \), where \( f \) is the input image and \( h \) is the filter kernel. Kernels are learnt directly from data.

**Activation layers** introduce non-linearity by providing an activation function; the activation types include sigmoid \( \sigma(x) = \frac{1}{1+e^{-x}} \), hyperbolic tangent \( f(x) = \tanh(x) \), and recently most commonly ReLU (Rectified Linear Unit) \( f(x) = \max(0, x) \), and its variations: leaky ReLU and ELU (Exponential Linear Unit).

**Pooling layers** are intended to increase the invariance to local translations and reduce spatial dimensions of the input by replacing the output of a layer at a certain location with a statistic of a nearby regions [104], such as max pooling, which outputs max value of the region, or average pooling, which outputs mean value of the region.

**Upsampling layers** allow to retrieve the original resolution of the feature maps that were subject to pooling. This can be done for example by storing the indices of the pooling performed earlier.

**Fully-connected layers** are equivalent to the multi-layer perceptron discussed earlier, and are usually employed at the top of the convolutional part of the network, serving as a classifier for features extracted by the convolutional stage.

**Softmax layers** provide class-wise probabilities in the final step of the classifier.

**Model training**

Training the model involves minimizing the cost function \( c(w) \) with respect to free parameters of the model \( w \), denoting neuron trainable weights and biases, as well as the input \( x \). To satisfy the necessary requirement for optimality, namely when the gradient vector of the cost function is zero:

\[
\nabla c(w; x) = 0
\]

(4.2)

an iterative updating of the parameters \( w \) is performed

\[
w(n + 1) = w(n) - \eta \nabla c(w(n); x)
\]

(4.3)

where \( n \) denotes an optimization step and \( \eta \) denotes learning rate – a factor by which the network weights are adjusted. To stabilize the gradient descent and prevent the trajectory of \( w \) from following an oscillatory path in the parameter space,
a modification of the gradient descent is commonly introduced, called \textit{momentum}, which involves replacing the currently calculated gradient with a weighted sum of past gradients, so instead of the weights update

\[ \Delta w(n) = -\eta \nabla c(w(n); x) \]  

we use

\[ \Delta w(n) = -\eta \sum_{i=0}^{n} \alpha_i \nabla c(w(n - i); x) \]  

where \( \alpha_i \) is the weighting factor for the gradient calculated in the \((n - i)\)-th step.

Other practical techniques for improving the training include \textbf{batch learning}, which involves only adjusting the weights of the model after a small subset of training examples has been presented, requiring less storage (usually in the graphics processor memory), when compared to gradient adaptation after presenting all training samples. \textbf{Batch normalization} is often applied after each pooling layer and normalizes neuron activations in training mini-batches, to reduce the changes in the distribution of each layer’s inputs, reducing \textit{internal covariate shift} [105]. This provides faster convergence of the network. \textbf{Dropout} is a technique used to alleviate network over-adaptation due to neuron co-adaptation, \textit{i.e.}, units’ reliance on other units. This involves random removal of a selected portion of the fully connected layer neurons and their connections during training [106].

\subsection*{4.3.2. Datasets of iris images}

In this Thesis, we consider three subject-disjoint post-mortem iris image databases:

- \textbf{Warsaw-BioBase-Postmortem-Iris-v1.1}, comprising 574 near-infrared (NIR) and 1023 visible light (VIS) images collected from 17 cadavers over a period of up to 34 days [85, 13],
- \textbf{Warsaw-BioBase-Postmortem-Iris-v2}, a subject-disjoint extension of the v1.1 of the database, adding 626 NIR and 764 VIS data from 20 more cadavers [107, 16],
- \textbf{Warsaw-BioBase-Postmortem-Iris-v3}, a new set of images collected for the purpose of this study, adding data from 40 more subjects with a total of 1094 NIR images and 785 VIS images, collected up to 370 hours since demise.
In addition to the post-mortem data, we use another database of challenging iris images for further re-training of the segmentation models and also for testing of the proposed approach:

— Warsaw-BioBase-Disease-Iris-v2.1, comprising iris images collected mostly from elderly patients of an ophthalmology clinic, including subjects with various ophthalmic conditions [74, 9].

### 4.3.3. Preparing ground truth data

For every sample in the dataset, we have carefully annotated the corresponding ground truth binary mask, which denotes regions of iris that are unaffected by post-mortem and disease-induced changes, together with specular reflections, regardless of their origin. Example images from the dataset and ground truth masks for cadaver samples are shown in Fig. 4.1, whereas Fig. 4.2 presents the same for selected disease-affected samples. To expedite the training process and reduce memory overhead, the images were downsampled to the size of $120 \times 160$ pixels, and the mask predictions produced by the network are later upscaled to match the original size of $480 \times 640$ pixels.
Figure 4.2: Example images from the Warsaw-BioBase-Disease-Iris-v2.1 dataset and their corresponding manually annotated masks, which remove the iris portions affected by disease-induced changes to the iris tissue (left), related to obstructions of the iris (middle), and to geometrical distortions of the pupillary area (right).

4.3.4. Evaluation strategies

Evaluation of the proposed approach consists of two methodologies. In the first, we only assess the accuracy of the segmentation predictions produced by our models, whereas in the second, the segmentation results are injected into OSIRIS iris recognition pipeline, whose recognition accuracy is then evaluated in comparison to the unmodified OSIRIS performance, as well as the best performing commercial matcher IriCore.

Segmentation accuracy

During testing, a prediction in the form of binary mask is obtained from the network for each of the images. For each predicted mask, Intersection over Union (IoU) is calculated between the prediction and the ground truth mask:

$$IoU = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} \land G_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} \lor G_{ij}}$$

(4.6)

where $P_{ij}$ and $G_{ij}$ denote the logical values of prediction mask and ground truth mask for the $ij$-th pixel, respectively, $m, n$ is the image size in pixels, whereas $\land$ and $\lor$ denote the AND and OR bitwise logical operators, respectively.
IoU scores obtained individually for each image are then averaged to get the mean IoU for each test split. To compare the DCNN-based method developed in this work with a conventional segmentation method, we did exactly the same evaluations on the train/test splits using the OSIRIS, for which IoUs are also calculated and averaged within each test split.

For this part of the experimentation, a subset of 1308 images from the Warsaw-BioBase-Postmortem-Iris-v1 dataset was used, excluding images for which a mask denoting undeformed iris regions could not be created due to a total decompose of the tissues. From the Warsaw-BioBase-Disease-Iris-v2.1 dataset, a subset of 743 images representing those eyes, which were the most affected by disease-related processes, and not the eyes that can be easily segmented using a conventional approach, so that the potential errors are not underestimated by diluting the dataset with ‘easy’ samples. In both the post-mortem and disease-related parts of the experiments, these data were used to generate train/test splits, as detailed in the following sections.

Matching accuracy

For evaluation of the matching performance that the proposed segmentation methods can yield, the results obtained in the first part of the evaluation, namely binary iris masks for test images, are injected into the OSIRIS. For the matching accuracy evaluation, we use all available post-mortem samples:

- **training data:** Warsaw-BioBase-Postmortem-Iris-v1.1 and v2 (ground truth masks are available for these samples),
- **testing data:** Warsaw-BioBase-Postmortem-Iris-v3.

As for the disease data, we use two subject-disjoint subsets Warsaw-BioBase-Disease-Iris-v2.1 of the with similar image and subject count:

- **training data:** 553 NIR images from 74 randomly chosen eyes (ground truth masks are available for these samples),
- **testing data:** 551 NIR images from 76 remaining eyes.

4.3.5. Model architecture and training

For our solution, we use the SegNet model for semantic segmentation [99], which is a modified VGG-16 network with fully connected layers removed, and added decoder stage. The resulting architecture is a coupled encoder-decoder network with five sets of
convolutional and pooling and upsampling layers in each half of the network, Fig. 4.3. SegNet performs the non-linear upsampling of the encoded data by employing stored indices from the max-pooling layers in a corresponding decoder. The softmax layer is then followed by a pixel-level classification layer, which yields a binary decision for each pixel (in our case: iris or non-iris/non-healthy iris).

Figure 4.3: Encoder-decoder architecture of SegNet. Inference takes place from left to right. Size of the Softmax layer is equal to the size of the input image. Drawing reprinted from [99].

For training and testing procedure, 10 subject-disjoint train/test data splits were created by randomly choosing the data from 14 (out of 17) subjects to the train subset, and the data from the remaining 3 (out of 17) subjects to the test subset, for the Post-mortem data. For the Disease data, similar 10 subject-disjoint train/test data splits were created by randomly choosing the data from 62 out of 78 subjects to the train subset, and the data from the remaining 16 subjects to the test subset (an approximation of the 80/20 split). All ten splits were made with replacement, making them statistically independent. The network is then trained with each train subset independently for each split, and evaluated on the corresponding test subset. This procedure gives 10 statistically independent evaluations and allows to assess the variance of the obtained results. The training, encompassing 60 epochs in each experiment, was accomplished with stochastic gradient descent as the minimization method. We applied momentum of 0.9, learning rate of 0.001, and $L_2$ regularization of 0.0005.
4.4. Segmentation accuracy evaluation

4.4.1. Results: *Post-mortem* data

Fig. 4.4 summarizes average IoU offered in all 10 splits by both solutions, and Table 4.1 details the results obtained in each split.

![Boxplots representing Intersection over Union in 10 test splits, for both OSIRIS and DCNN-based approaches. Median values are shown in red, height of each boxes corresponds to an inter-quartile range (IQR) spanning from the first (Q1) to the third (Q3) quartile, whiskers span from Q1-1.5*IQR to Q3+1.5*IQR, and outliers are shown as crosses. Figure reprinted from [108].](image)

The proposed DCNN-based solution clearly outperforms the conventional method, with IoU=88.53%, whereas OSIRIS offers IoU=73.58% on average in identical evaluation. Looking at the results obtained in each data split, the DCNN-based solution always outperforms the OSIRIS, even by as much as 40.9% (split 7). Example segmentation results for both DCNN-based and conventional algorithms are shown in Fig. 4.5, to discuss potential reasons of failures when:

— both algorithms performed well (achieved simultaneously the highest IoU),
— both algorithms failed (achieved simultaneously the lowest IoU),
Table 4.1: Mean Intersection over Union in each test split obtained for OSIRIS and DCNN-based post-mortem iris segmentation. The best and the worst results are bolded. The third column shows the improvement in performance split-wise. Averaged results are shown in the last line. Table adapted from [108].

<table>
<thead>
<tr>
<th></th>
<th>Mean IoU (OSIRIS)</th>
<th>Mean IoU (CNN)</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Split 1</td>
<td>0.7793</td>
<td>0.8587</td>
<td>10.2%</td>
</tr>
<tr>
<td>Split 2</td>
<td>0.6002</td>
<td>0.7657</td>
<td>27.6%</td>
</tr>
<tr>
<td>Split 3</td>
<td>0.7786</td>
<td>0.8681</td>
<td>11.5%</td>
</tr>
<tr>
<td>Split 4</td>
<td>0.7533</td>
<td>0.8427</td>
<td>11.9%</td>
</tr>
<tr>
<td>Split 5</td>
<td>0.8715</td>
<td>0.8853</td>
<td>1.6%</td>
</tr>
<tr>
<td>Split 6</td>
<td>0.6203</td>
<td>0.7986</td>
<td>28.7%</td>
</tr>
<tr>
<td>Split 7</td>
<td>0.4823</td>
<td>0.6794</td>
<td>40.9%</td>
</tr>
<tr>
<td>Split 8</td>
<td>0.8032</td>
<td>0.87</td>
<td>8.3%</td>
</tr>
<tr>
<td>Split 9</td>
<td>0.8078</td>
<td>0.8564</td>
<td>6.0%</td>
</tr>
<tr>
<td>Split 10</td>
<td>0.8621</td>
<td>0.8822</td>
<td>2.3%</td>
</tr>
<tr>
<td>Average</td>
<td>0.7358</td>
<td>0.8303</td>
<td>12.8%</td>
</tr>
</tbody>
</table>

— DCNN-based solution failed (achieved the lowest IoU) when the conventional method succeeded (achieved the highest IoU),

— DCNN-based solution succeeded (achieved the highest IoU) when the conventional method failed (achieved the lowest IoU).

As expected, both methods perform well for post-mortem iris images compliant with ISO/IEC 19794-6, cf. top row in Fig. 4.5. Both methods, however, failed to accurately recognize a small portion of the non-deformed iris texture in the iris that underwent heavy post-mortem processes, second row in Fig. 4.5. The DCNN-based method was not able to localize any iris portion in this difficult sample acquired 574 hours (almost 24 days) post-mortem. However, this behavior is still more favorable than finding the iris in the incorrect region.

There are samples which are easier to process by conventional segmentation method. Third row in Fig. 4.5 presents a post-mortem sample that displays a regularly shaped iris with good contrast between the iris and the background. Hence, this sample was relatively easy to process by OSIRIS software, which presents a high IoU in this case. However, the intensity and texture of the iris region departed from what the DCNN saw in the training samples, and thus our solution was very selective in annotating the iris areas, ending up with low IoU.
both methods present a good-quality outcome

failure of both the DCNN-based and the conventional segmentation

failure of the DCNN-based method, success of the conventional method

success of the DCNN-based method, failure of the conventional method

(a) DCNN (proposed)  (b) OSIRIS  (c) Ground truth

Figure 4.5: Segmentation results illustrated for four different post-mortem samples, for the DCNN-based segmentation (a), the conventional segmentation (b) compared with ground truth (c). Graphics adapted from [108].

However, one can observe an opposite result more frequently: the DCNN-based segmentation was able to detect irregular specular reflections and wrinkles, offering way better result than the conventional algorithm, cf. bottom row in Fig. 4.5, especially
when neither the pupil nor the iris are perfectly circular, and the iris texture started to be muddy due to cornea opacification.

4.4.2. Results: Disease data

Here as well, a comparison with OSIRIS is done with the same evaluations on the 10 train/test splits obtained analogously to those in Sec. 4.3.5, i.e., by sampling approx. 80% of the classes for the training subset and the assigning the remaining classes for the test subset. Fig. 4.6 summarizes average IoU offered in the 10 splits by both solutions, and Table 4.2 details the results obtained in each split. Our DCNN-based model performed better than the competing traditional method in 9 out of 10 train/test splits, usually by a wide margin. The averaged improvement over OSIRIS is 23.5%, although the overall performance of the model remains lower than this obtained for the Post-mortem data. This can be attributed to the inherent difficulty in the data, especially with samples that needed to be masked out entirely due to the lack of healthy iris texture.

![Figure 4.6: Boxplots representing Intersection over Union in 10 test splits, for both OSIRIS and DCNN-based approaches, same as in Fig. 6.6, but for the Disease subset of data and a respective DCNN-based algorithm.](image-url)
Table 4.2: Same as in Table 4.1, but for the Disease subset of data and a respective DCNN-based segmentation method.

<table>
<thead>
<tr>
<th></th>
<th>Mean IoU (OSIRIS)</th>
<th>Mean IoU (CNN)</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Split 1</td>
<td>0.5001</td>
<td>0.6598</td>
<td>31.9%</td>
</tr>
<tr>
<td>Split 2</td>
<td>0.6154</td>
<td>0.5998</td>
<td>-2.5%</td>
</tr>
<tr>
<td>Split 3</td>
<td>0.6061</td>
<td>0.7323</td>
<td>20.8%</td>
</tr>
<tr>
<td>Split 4</td>
<td>0.3874</td>
<td>0.6259</td>
<td>61.6%</td>
</tr>
<tr>
<td>Split 5</td>
<td>0.5997</td>
<td>0.6572</td>
<td>9.6%</td>
</tr>
<tr>
<td>Split 6</td>
<td>0.5762</td>
<td>0.6733</td>
<td>16.9%</td>
</tr>
<tr>
<td>Split 7</td>
<td>0.5280</td>
<td>0.6564</td>
<td>24.3%</td>
</tr>
<tr>
<td>Split 8</td>
<td>0.5512</td>
<td>0.6925</td>
<td>25.6%</td>
</tr>
<tr>
<td>Split 9</td>
<td>0.4888</td>
<td>0.6190</td>
<td>26.7%</td>
</tr>
<tr>
<td>Split 10</td>
<td>0.5374</td>
<td>0.7426</td>
<td>38.2%</td>
</tr>
<tr>
<td>Average</td>
<td>0.5390</td>
<td>0.6659</td>
<td>23.5%</td>
</tr>
</tbody>
</table>

Below we present a close-up analysis of selected samples in the same four categories as for the post-mortem analysis, namely:

— both algorithms performed well,
— both algorithms gave unsatisfactory results,
— DCNN-based solution performed worse than the conventional method,
— DCNN-based solution succeeded when the conventional method was inaccurate.

Similarly to the behavior observed for a model trained on Post-mortem data, when a sample does not differ much from these of healthy irises, both methods offer accurate predictions, Fig. 4.7, top row.

However, a different situation occurs for a sample that is very difficult to process, Fig. 4.7, second row. This image has had its ground truth mask denoted as empty, as the iris is heavily degraded and shows none of the typical iris features originating from its trabecular meshwork. Hence, such an image should be discarded prior to the encoding stage. Our DCNN-based method was very selective in annotating iris regions suitable for encoding. The traditional method, however, produced an incorrect output.

Some samples prove difficult to be processed by the DCNN method, cf. third row in Fig. 4.7. Here, a fairly accurate prediction is obtained from OSIRIS, while the DCNN-based algorithm was again too selective in annotating iris portions of the
both methods present a good-quality outcome

failure of both the DCNN-based and the conventional segmentation

failure of the DCNN-based method, success of the conventional method

success of the DCNN-based method, failure of the conventional method

(a) DCNN (proposed)  (b) OSIRIS  (c) Ground truth

Figure 4.7: Segmentation results illustrated for four different disease samples, for the DCNN-based segmentation (a), the conventional segmentation (b) compared with ground truth (c).

image. However, again similarly to what we saw with the Post-mortem model, an opposite situation is more frequent – the DCNN-based method is able to deal with non-standard iris images, such as those with irregularly shaped pupils, cf. Fig. 4.7, bottom row.
4.5. Iris recognition with DCNN-based image segmentation

In this section we show how to use segmentation masks predicted by neural networks in conventional, Gabor-based iris recognition method, which employs circular approximations of the pupillary and limbic iris boundaries. As a whole, this method allows for a significant improvement in post-mortem iris recognition accuracy over the methods designed only for live irises. For this part of the research, we only consider the OSIRIS and IriCore methods – the former being a baseline matcher which is subject to the modifications introduced by the Author of this Thesis, and the second serving as an additional commercial benchmark method, which was found to offer the best performance out of the three commercial methods evaluated in Chapters 2 and 3.

4.5.1. Data-driven image segmentation: fine-tuning

Over the course of the experiments, we have found that the model from [108] can be significantly improved, as more experimental data became available. We therefore have trained several additional variations, which are listed in this Section. This is done to assess the quantity and quality of data that is required during the training stage to ensure accurate performance of the new image segmentation component.

All of the new models are presented pair-wise, i.e., for each model trained with fine-grained manually annotated masks (denoting only the iris regions unaffected by biological changes), a second model is also built, trained with coarse ground truth masks (denoting only the inner and outer iris boundaries and eyelids). This is necessary because the highly irregular predictions yielded by the fine-grained model, albeit being correct from the subsequent feature encoding standpoint, are difficult to automatically approximate the iris boundaries. The coarse model is used for the image normalization stage, as its smoother predictions enable accurate circular approximation of the iris boundaries.

In all of the further experiments, the image normalization stage is performed using the coarse model, whereas the iris matching results are obtained with occlusion masks yielded by both the fine and the coarse models, and are evaluated separately.
Models trained with just the post-mortem data:

1. **fine v1** segmentation model: trained with data from the Warsaw-Biobase-Postmortem-Iris-v1 database with fine-grained ground truth masks for 60 epochs, and yielding predictions of $120 \times 160$ pixels [108],

2. **fine v2** segmentation model: similar to the **fine** model, but trained with data from Warsaw-Biobase-Postmortem-Iris-v1 database together with NIR samples from the Warsaw-Biobase-Postmortem-Iris-v2 with fine-grained ground truth masks for 120 epochs, yielding predictions of $240 \times 320$ pixels,

3. **coarse v2** segmentation model: trained with both NIR and VIS data from the v1 and v2 versions of the Warsaw-Biobase-Postmortem-Iris database for 120 epochs, but this time with coarse ground truth masks, producing masks in $240 \times 320$ size.

Models trained with both the post-mortem and disease data:

1. **fine v3** segmentation model: similar to the **fine v2** model, but with training encompassing also the training set of the Warsaw-BioBase-Disease-Iris-v1 iris samples,

2. **coarse v3** segmentation model: similar to the **coarse v2** model, but with training including the training set of the Warsaw-BioBase-Disease-Iris-v1 iris samples.

In [109] we have introduced a data-driven iris segmentation model, based on the same SegNet architecture [99], which is re-trained with several standard iris biometrics datasets their corresponding ground truth masks, including the Biosec baseline corpus [110] (1200 images), the BATH database$^1$ [111] (148 images), the ND0405 database$^2$ (1283 images), the UBiRIS database [112] (1923 images), and the CASIA-V4-Iris-Interval database$^3$ (2639 images). The result has been shown to achieve superior performance when evaluated on high quality, healthy iris samples, allowing to obtain almost perfect recognition accuracy on the novel dataset of near-infrared iris videos collected from 42 subjects [109]. Here, we use the same model as another benchmark method, as well as train yet another model, this time by aggregating all available

---

$^1$ http://www.bath.ac.uk/elec-eng/research/sipg/irisweb/
$^2$ https://cvrl.nd.edu/projects/data/
$^3$ http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp
data (healthy, post-mortem, and elderly/disease related) during the training stage, again with fine-grained and coarse ground truths of the post-mortem data variants. We thus end up with three models employing healthy data in total:

**Model trained with healthy irises:**

1. **healthy** segmentation model: trained only with healthy iris images, same as the one introduced in [109].

**Models trained with all available data:**

1. **fine v4** segmentation model: trained with healthy, post-mortem, and elderly/disease data, with fine-grained post-mortem ground truths,
2. **coarse v4** segmentation model: a counterpart to the **fine v4** model, trained with coarse post-mortem ground truths.

### 4.5.2. Iris image normalization

For the proposed method to serve as a drop-in replacement for the Daugman method recognition pipeline, the prediction obtained from the DCNN-based segmentation has to be correctly normalized onto a dimensionless polar-coordinate rectangle. For this stage, a method for localizing pupillary and limbic iris boundaries has been devised, which employs a Hough transform that is applied to the prediction obtained from the **coarse** segmentation model. These boundary parameters are then used in all subsequent experiments, including those involving the **fine** and **fine v2** models. The segmentation methodology employing the **coarse** segmentation model is visualized in Fig. 4.8. Similar results are then presented in Fig. 4.9 for the two remaining fine models. Notably, the original **fine** model from [108] seems to over-aggressively mask out some portions of the iris, while at the same time mistakenly denoting some portions of the pupil as iris. This is fixed in the **fine v2** model, which was trained with twice the iterations count, cf. right side of Fig. 4.9.
OSIRIS segmentation, normalized images, and normalized masks

coarse CNN binary predictions with fitted Hough circles, segmented images, normalized images, and normalized masks

Figure 4.8: Example segmentation results obtained from the unmodified OSIRIS method and the coarse segmentation model. Warsaw-BioBase-Postmortem-Iris-v3 (left), Warsaw-BioBase-Disease-Iris-v1 (right).
fine and fine \textit{v2} CNN segmented images, normalized images, and normalized masks

Figure 4.9: Segmentation results for post-mortem sample from Fig. 4.8, obtained from the \textit{fine} and \textit{fine v2} segmentation models. Both \textit{fine-grained} predictions shown here use the circular approximations fitted with the prediction obtained from the \textit{coarse} model.

4.6. Matching accuracy evaluation

4.6.1. Comparison score generation

We evaluate the new segmentation approach by applying the segmentation results obtained from the DCNN-based models into the OSIRIS recognition pipeline, and comparing their performance with unmodified OSIRIS algorithm, as well as with the accuracy offered by the IriCore. Generation of genuine and impostor comparison score distributions was done by performing comparisons between all possible iris image pairs. One comparison per a given image pair was performed, \textit{i.e.}, if image A was being matched against image B, then image B is not being matched against image A.

For the post-mortem data, 9 separate score distributions are generated for both the genuine and impostor comparisons, taking into account samples collected during different, progressing observation horizons in respect to time that has elapsed since a subject’s death: scores obtained from samples collected fewer than 12 hours after death, fewer than 24 hours, 48, 60, 72, 110, 160, 210, and finally 370 hours (all available samples in the test dataset). This is to evaluate the recognition accuracy in respect to the increasing post-mortem interval. All possible comparisons between samples collected during a given time horizon were performed.
4.6.2. Results: Post-mortem data

The results are presented in the form of receiver operating characteristic (ROC) curves, which plot the correct match rates against false match rates. In addition, equal error rates (EER) are presented. Fig. 4.10 shows ROC curves denoting the performance of the baseline methods (OSIRIS and IriCore) compared to the performance obtained by employing image segmentation results from the DCNN-based models for three shortest post-mortem time horizons, comprising samples collected during the first three capture time intervals. Fig. 4.11 and 4.12 contain analogous graphs, but for the remaining longer time horizons of up to 370 hours post-mortem.

Of all the tested variants, the fine v4 model offers consistently the best matching accuracy determined both by the shape of the ROC curve, as well as the Equal Error Rate, therefore all further discussions of the results refer to this model.

Notably, very good performance can be achieved for samples collected up to 12 hours after death, with EER as low as 0.76%, which even decreases to 0.68% when samples collected up to 24 hours (one day) post-mortem are added to the database. For samples collected up to 48 hours (two days) after death, the lowest EER offered by the fine v4 model is 2.57%, which means that the method can still be usable. As more difficult samples are being added to the database, the errors increase, reaching 21.36% for the best performing fine v4 model. Original OSIRIS, on the other hand, is incomparable to the proposed approach from the very beginning, yielding EER of almost 17% for the ‘easiest’ samples, and up to 33.59% for the entire database (samples collected up to 370 hours post-mortem). The IriCore’s performance starts with EER=5.37% for the easiest samples, which is still more than 4.5 percentage points behind our proposed solution, and decreases steadily with the increasing post-mortem time horizon, reaching 25.38% for the most challenging set of samples.

4.6.3. Results: Disease data

Fig. 4.13 presents ROC curves obtained when matching samples from the Warsaw-BioBase-Disease-Iris-v1 databases using a baseline OSIRIS approach and the proposed segmentation approach based on the v2, v3, healthy, and v4 models.

The original OSIRIS offers a performance of just 8.9% EER on this dataset. Better, yet still far from satisfactory is IriCore, with EER=3.97%. Each of the proposed models
Figure 4.10: Receiver Operating Characteristic curves obtained when comparing post-mortem samples with different observation time horizons: 12, 24, and 48 hours post-mortem. Results for the baseline OSIRIS performance, IriCore benchmark method, and modifications introduced by the Authors are presented. Plots on the right are close-ups of the ones on the left in each pair.

offers recognition accuracy that is better than this of the two existing matchers. With the model trained solely on the post-mortem data, we can expect the error to drop to 2.55%, whereas including similar, elderly- or disease-related iris images during the training stage of the segmentation model allows the EER to be as low as 2.19%.

Surprisingly, however, the best accuracy is obtained using iris masks generated by the segmentation model that was trained with high quality images of healthy
Figure 4.11: Same as in Fig. 4.10, but plotted for comparison scores obtained from samples collected up to 60, 72, and 110 hours post-mortem.

irises, with EER of only 1.73%. This behavior may be attributed to the fact that the healthy model was trained with more data (more than twice as many) than the models tuned with post-mortem and disease data, and that the images from the Warsaw-BioBase-Disease-Iris database are not as challenging as those from the post-mortem test dataset, and thus the network trained on healthy data can generalize well enough for the disease data.

Finally, the last two models, during training of which all of the available training data is used – healthy, post-mortem, and disease, achieve slightly worse matching
Figure 4.12: Same as in Fig. 4.10, but plotted for comparison scores obtained from samples collected up to 160, 210, and 370 hours post-mortem.

performance than the model trained solely on healthy iris data. The EERs obtained here are 2.34% and 2.24% for the coarse and fine model, respectively.
4.7. Conclusions

This Chapter introduces the first known to us attempt at improving iris recognition reliability for post-mortem samples, by proposing a robust image segmentation method that can be used as a drop-in replacement for the OSIRIS image segmentation. The proposed solution, employing DCNN-based semantic segmentation, allows close-to-perfect recognition accuracy on the subject-disjoint evaluation database of post-mortem samples, as well as on another challenging database of iris images, collected from elderly individuals with various ophthalmic conditions. This proves the flexibility of the proposed approach in solving problems associated with non-standard iris images, which the models introduced in this work can reliably and consistently localize and segment in most of the cases. Despite the expected drop in recognition accuracy with the increasing time horizon of post-mortem sample collection, the proposed approach consistently outperforms both the academic and the commercial, state-of-the-art iris recognition methods. This proves the second thesis formulated in Sec. 1.2, namely that the proposed iris segmentation methods employing convolutional neural networks and counteracting disease and post-mortem changes improve the recognition accuracy.
5. Iris representation with iris-driven image filtering

5.1. Introduction

As shown in the previous Chapter, applying deep-learning based iris image segmentation methodology, coupled with existing iris representation and matching pipeline of the OSIRIS software, allows for a significant improvement in recognition accuracy for challenging post-mortem samples, with recognition errors decreased not only beyond those offered by the original OSIRIS, but also better than the performance of the commercial, state-of-the-art methodologies – IriCore.

However, despite the progress that has been made, the accuracy still remains far from perfect for capture horizons longer than 48 hours post-mortem. Since the image segmentation stage is already being executed without major errors for almost all of the samples, achieving even higher recognition rates requires a novel approach to iris encoding.

Widely used in iris recognition, as introduced by Daugman more than 25 years ago, is the quantization of filter response over the iris image, most commonly using Gabor wavelets [113] or their approximations, as used in the OSIRIS algorithm [76]. To date, several alternatives were proposed, such as LoG pyramids [21], Haar wavelets [23], or Zak-Gabor wavelet packets [25]. A review of existing approaches to iris representation is delivered in Sec. 1.3.

Despite recent advancements in field of deep-learning models for iris discrimination, these ‘traditional’ methods offer the interpretability, as opposed to the predictions provided by a DCNN-based algorithm. On the other hand, a data-driven model can offer performance that is superior to a hand-crafted model because of its ability to learn the most efficient way to represent the data from the data itself.
We make an attempt at improving the efficiency of the iris representation by introducing data-driven filters that are learnt from post-mortem iris images. A shallow – i.e., with one convolutional layer – Siamese network is employed for learning the novel iris descriptor in a form of two dimensional filter kernels that can be further used in any conventional (i.e., Daugman’s) iris recognition pipeline. Such set of filters is then used to optimize the original OSIRIS filter bank to improve the method’s accuracy for post-mortem iris images.

5.2. Related work

5.2.1. One-shot recognition and Siamese networks

Recent advancements in deep learning allowed DCNN-based image classifiers to achieve performance superior to any other class of methods. However, their one important drawback is the need for large quantities of labeled data for supervised learning. This becomes a problem in applications where a prediction must be obtained about the data belonging to a class that the model had never seen during the training.

Siamese networks, on the other hand, perform well in the so called one-shot recognition tasks, being able to give reliable similarity prediction about the samples from classes that were not included in the model training phase. Koch et al. [114] introduces a deep convolutional model architecture consisting of two convolutional branches sharing weights and joined by a merge layer with $L_1$ cost function describing distance between the two inputs $x_1, x_2$:

$$L_1(x_1, x_2) = |f(x_1) - f(x_2)|$$  \hspace{1cm} (5.1)

where $f$ denotes the encoding function. This is combined with a sigmoid activation of the single neuron in the last layer, which maps the output to the range of $[0, 1]$. This architecture is trained and tested in a class-disjoint manner on the Omniglot dataset, which comprises handwritten characters belonging to 50 different alphabets, achieving accuracy of 92%, which is almost on par with human performance.

The applications of siamese networks include many areas, most importantly one-shot image recognition, with good benchmark performances achieved on
well-known datasets such as Omniglot (written characters recognition for multiple alphabets, 95% accuracy) and ImageNet (natural image recognition with 20000 classes, 87.8% accuracy) [115], but also object co-segmentation [116], object tracking in video scenes [117], signature verification [118], and even matching resumes to job offers [119].

5.2.2. Data-driven image descriptors

Several approaches to learning feature descriptors for image matching have been explored, mostly in the field of visual geometry and mapping for image stitching, orientation detection, and similar general-purpose approaches.

Simo-Serra et al. [120] present a novel point descriptor, whose discriminative feature descriptors are learnt from the real-world, large datasets of corresponding and non-corresponding image patches from the MVS dataset, containing image patches sampled from 3D reconstructions of the Statue of Liberty, Notre Dame cathedral, and Half Dome in Yosemite. The approach is reported to outperform SIFT, while being able to serve as a drop-in replacement for it. The method employs a Siamese architecture of two coupled CNNs with three convolutional layers each, whose outputs are patch descriptors, and an $L_2$ norm of the output difference is minimized between positive patches and maximized otherwise.

A similar approach is demonstrated by Zagoruyko and Komodakis [121], who train a similarity function for comparing image patches directly from the data employing several methods, one of them being a Siamese model with two CNN branches sharing weights, connected at the top by two fully connected layers.

Yi et al. introduce a method that is intended to serve as a full SIFT replacement, not only as a drop-in descriptor replacement [122]. The deep-learnt approach consists of a full pipeline with keypoint detection, orientation estimation, and feature description, trained in a form of a Siamese quadruple network with two positive (corresponding) input patches, one negative (non-corresponding) patch, and one patch without any keypoints in it. Hard mining of difficult keypoint pairs is employed, similarly to [120].

DeTone et al. [123] introduce a so-called SuperPoint network and a framework for self-supervised training of interest point detectors and descriptors that are able to operate on the full image as an input, instead of image patches. The method is able to compute both interest points and their descriptors in a single network pass.
Moving to the field of biometrics, Czajka et al. [124] have recently employed human-inspired, iris-specific binarized statistical image features (BSIF) filters from iris image patches derived from an eye-tracking experiment, during which human iris examiners were asked to classify iris pairs [124]. Data-driven BSIF filters were also studied by Bartuzi et al. for the purpose of person recognition based on thermal hand representations [125].

5.3. Proposed methodology

5.3.1. Preprocessing of the training data

To train our new, post-mortem-aware iris feature descriptor, we use NIR iris images from the Warsaw-BioBase-Postmortem-Iris-v1.1 and Warsaw-BioBase-Postmortem-Iris-v2, which were then processed with the best performing segmentation model from the previous experiment (fine v4) and normalized using the algorithm introduced in Sec. 4.5.2 to come up with polar iris images 512 × 64 pixels in size.

The first step in preparing the training data is to ensure that the polar iris images are aligned within any given class, i.e., that there are no vertical shifts in the images that reflect eyeball rotation in the cartesian coordinate system. Since the acquisition process could not be controlled, such rotations are present in the dataset, as shown in Fig. 5.1. To reduce the resulting shift in the polar iris images, we performed the image alignment procedure, which involved manual annotations of the eye corners. This allowed to calculate a relative rotation of the eyeball represented in the two images, and in turn the amount of pixels, by which the polar image must be shifted, see Fig. 5.1. Typical iris recognition methods shift iris codes in the matching stage to compensate for eyeball rotation, instead of shifting images, including the OSIRIS used in our experiments. Therefore there is no justification to make the neural network learn how to discriminate between irises that are not ideally spatially aligned.

The resulting aligned polar iris images were then subject to examination in respect to the amount of occlusions caused by eyelids or eyelashes. During iris verification a binary occlusion mask is usually employed to discard regions with noise present. On the other hand, introducing noisy data during the training phase could bias the model
into considering noise (here: eyelids) as iris features, which is not desired. To ensure good quality of the training data, we divide each polar iris image into two patches. Since in our data eyelid occlusions are only present either on the left or on the right portion of the polar iris image, this enables discarding such samples while at the same time saving the other, unaffected patch, Fig. 5.2. 1801 patches in total were extracted for training.
5.3.2. Model architecture and filter learning

For learning the iris-specific filters a shallow Siamese architecture is used, Fig. 5.3, composed of two branches, each responsible for encoding of one image from the image pair being compared, comprising a single convolutional layer with 6 kernels of size $9 \times 15$, to reflect the number of filters found in the OSIRIS, and the size of the smallest OSIRIS filters. The initial calculations revealed that $9 \times 15$ kernels allows better results than $9 \times 27$ or $9 \times 51$ (also found in OSIRIS), hence the decision. The weights are shared between the two branches. Following the convolutional layers is a merge layer calculating the $L_1$ distance (see Eq. 5.1), or the absolute mean, between the two sets of features from the convolutional layer. A single neuron with a sigmoid activation function is then applied to yield a prediction from the range $[0, 1]$, with 0 being a perfect match between the two images, and 1 being a perfect non-match.

The training data is passed to the network in batches containing 32 pairs of iris patches, out of which 16 are genuine, and 16 are impostor pairs, randomly sampled without replacement from the dataset during each training iteration, with a total of 20000 iterations. ADAM optimizer – a first-order gradient-based optimization method, based on adaptive estimates of lower-order moments: an exponentially decaying average of past gradients and squared past gradients [126], is used with $lr = 0.0006$. 

Figure 5.3: Siamese network used for iris feature representation learning.
Filter kernels: Example iris codes:

Mean iris code value distributions:

Figure 5.4: Learnt kernels from the Siamese network, an example set of iris codes they produce, and distributions of mean code values. Kernel 6 is discarded.

5.3.3. Feature selection

The learnt filter kernels, together with example iris codes that they produce, as well as distributions of mean iris code values produced by each of them are illustrated in Fig. 5.4. By analyzing the distributions of mean iris code values obtained by each of the new filters, we see that codes produced by the sixth filter do not represent the iris well, as most of the texture information is lost during encoding, resulting in a mostly zeroed iris code. This filter is discarded from all further experiments.

Notably, employing only the iris-specific filter kernels instead of those found in OSIRIS did not yield better results – perhaps due to the fact that regular iris texture is well represented using the conventional Gabor wavelets, and the newly learnt filters
are necessary to boost the performance for difficult, decay-affected samples. To utilize these new filters, and to offer an advantage over the baseline method, a modification of the OSIRIS Gabor filter bank was performed to propose a hybrid filter bank comprising a combination of Gabor wavelets and the iris-specific kernels.

This procedure employed by Sequential Feature Selection, with a combination of Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS), which involve adding the most discriminant features to the classifier, while removing the least discriminatory ones. During the feature selection procedure, post-mortem-specific filters were added to the original OSIRIS filter bank, whereas those OSIRIS filters, which do not contribute to decreasing the error rates were removed. This estimation is performed on the Warsaw-BioBase-Postmortem-Iris-v3, which is subject-disjoint with the databases used for network training. The error metric minimized during the feature selection is the EER obtained for samples acquired up to 60 hours post-mortem, in the same way as done earlier in Sec. 4.6.

The feature selection procedure can be enclosed in the following steps, starting with the original, unmodified OSIRIS filter bank comprising six Gabor wavelets:

**Step 1.** Calculate the performance obtained using the current filter bank and each of the siamese filters added independently.

**Step 2.** **Perform SFS by adding the most contributing filter to the filter bank.**

**Step 3.** Calculate the performance obtained using the filter bank obtained in the previous step with each of the OSIRIS filters removed independently.

**Step 4.** **Perform SBS by removing the least contributing filter from the filter bank.**

**Step 5.** → Go back to Step 1 and repeat until the error metric stops improving.

After the SFS/SBS feature selection procedure involving four iterations of SFS and SBS, Fig. 5.5, the EER was decreased by almost a third, from 6.40% obtained for the 60 hours post-mortem time horizon for the original OSIRIS filter bank, to 4.39% obtained for the new, hybrid filter bank, Fig. 5.6. Although performing an additional feature selection iteration does give a small decrease in EER, we observed an increase in EER for capture time horizons larger than 60 hours, and adding yet another iteration (SBS iteration 3 in Fig. 5.6) significantly increases the error.
Figure 5.5: Illustration of the filter selection for the new filter bank using Sequential Forward Selection and Sequential Backward Selection. Iris codes for an example iris produced by the proposed hybrid filter bank at different stages of filter selection.

Figure 5.6: SFS/SBS filter selection: Equal Error Rates in subsequent iterations of the procedure are shown. The filter selection is stopped at SBS iteration 2.

This filter bank is then used in the final testing of our iris recognition pipeline in the following section. The final set of filter kernels is shown in Fig. 5.5.
5.4. Results and discussion

5.4.1. Testing data and methodology

For testing, the same database is used as in Chapter 4, namely the Warsaw-BioBase-Postmortem-Iris-v3. This way, the results obtained here are directly comparable to those obtained during the evaluation of the proposed segmentation method. During testing, the same protocol as described in Sec. 4.6 applies, we generate all possible genuine and impostor scores between images that were captured up to a certain time horizon.

5.4.2. Recognition accuracy

Figures 5.7-5.9 present ROC curves obtained using the newly introduced filter bank, compared against ROCs corresponding to the best results obtained when only the segmentation stage is replaced with the proposed modifications.

When analyzing graphs presented in Fig. 5.7, which correspond to samples collected up to 12, 24, and 48 hours post-mortem, we do not see improvements in recognition performance measured by the EER, that is EER=0.56%→0.76%, 0.69%→0.68%, and 2.45%→2.57%, respectively. However, the shapes of the red graphs corresponding to the scores obtained with the new filter bank show an improvement over the black graphs in the low FMR registers, meaning that the proposed system offers higher recognition rates in situations, when very few false matches are allowed.

Moving to more distant post-mortem sample capture time horizons, Fig. 5.8, the advantage of the proposed method becomes clearly visible in both the decreased EER values, as well as in the shapes of the ROC curves. Applying domain-specific filters allowed to reduce EER from 6.40% to 4.39%, from 8.12% to 5.86%, and from 9.99% to 7.78%, for samples acquired less than 60, 72, and 110 hours post-mortem, respectively.

Finally, Fig. 5.9 presents ROC curves obtained for samples collected during the three longest subject observation time horizons, namely up to 160, 210, and 369 hours after death. Here as well, a visible improvement offered by the new feature representation scheme is reflected in the decreased EER values – 14.59%→11.88% for samples collected up to 160 hours, 17.09%→14.98% for those captured up to 210 hours, and
21.36% → 19.27% for the longest and most difficult set, encompassing images acquired up to 369 hours (more than 15 days).

Figure 5.7: Receiver Operating Characteristic curves obtained when comparing post-mortem samples with different observation time horizons: 12, 24, and 48 hours post-mortem, plotted for two baseline iris recognition methods OSIRIS (blue) and IriCore (green), OSIRIS with new segmentation introduced in Chapter 4 (black), as well as OSIRIS with both the improved segmentation and new filter set (red). Plots on the right are close-ups of the ones on the left in each pair.
5.4.3. False Non-Match Rate dynamics

In addition to the ROCs, we have also calculated the False Non-Match Rate (FNMR) values at acceptance thresholds which allow the False Match Rate (FMR) values to stay below certain values, namely 0.1%, 1% and 5%, Figs. 5.10, 5.11, 5.12, respectively. This is done to reveal the dynamics of the FNMR as a function of post-mortem sample capture horizon, and therefore to know the chances for a false non-match as time since death progresses. We plot this dynamics for the two baseline methods: original
Figure 5.9: Same as in Fig. 5.7, but plotted for comparison scores obtained from samples collected up to 160, 210, and 369 hours post-mortem.

OSIRIS and IriCore, as well as for the best DCNN-based segmentation modification – the fine v4 model, and the proposed iris representation, again coupled with the fine v4 segmentation.

While acceptance thresholds allowing FMR of 5% or even 1% can be considered as very relaxed for large-scale iris recognition systems, such criteria make sense in a forensic scenario. In such, the goal of the automatic system is typically to aid a human expert by proposing a candidate list, while minimizing the chances of missing the
Figure 5.10: Dynamics of False Non-Match Rates (FNMR) in the function of post-mortem sample capture horizon for a set False Match Rate (FMR) of 0.1%, plotted for OSIRIS (blue) and IriCore (red), OSIRIS with proposed segmentation (yellow), and OSIRIS with both the proposed segmentation and the new filter set (violet).

correct hit. Therefore, allowing a higher False Match Rate will make it more likely for the correct hit to appear within the candidate list.

The proposed modification to the iris feature representation is again able to outperform the traditional OSIRIS filter bank consisting of only Gabor kernels during each moment of the increasing sample acquisition time horizon for the two most restrictive FMR thresholds of 0.1% and 1%, Figs. 5.10 and 5.11, and is only slightly worse for the first three shortest time horizons up to 48 hours, when FMR of 5% is allowed, Fig. 5.12.

Notably, for each moment during the increasing post-mortem sample capture time horizon, our proposed approach consistently offers an advantage over the other two algorithms, allowing to reach nearly perfect recognition accuracy for samples collected up to a day after a subject’s death. This difference in favor of our proposed solution is even larger when the acceptance threshold is relaxed to allow for 5% False Match Rate - in such scenario, we can expect no false non-matches in the first 24 hours, and only approx. 1.5% chance of a false non-match in the first 48 hours. The new method for iris feature representation, on the other hand, shows its greatest advantage in the acquisition horizons that are longer than 48 hours, being able to reduce the errors by as much as a third in the 60-hour and 72-hour capture moments.
5.5. Conclusions

In this Chapter we have introduced the novel iris feature representation method, employing iris-specific image filters that are learnt directly from the data, and are thus optimized to be resilient against post-mortem changes affecting the eye during the increasing sample capture time since death. By finding the optimal combination of typical Gabor-wavelets-based iris encoding with the new post-mortem aware
encoding we are able to further reduce the recognition errors by as much as one third, proving the third thesis formulated in Sec. 1.2, namely that the proposed post-mortem-specific iris feature representation further improves recognition accuracy for cadaver samples.
6. Presentation attack detection for cadaver iris

6.1. Is that eyeball dead or alive?

With increasing importance that biometric authentication gains in our daily lives, fears fueled by Hollywood’s depictions of severed thumbs and plucked eyeballs are increasingly common among users, regarding the possibility of unauthorized access to one’s data, identity, or assets after demise. Law enforcement officers in the U.S. have been already using the fingerprints of the deceased suspects to unlock their iPhones [127], which immediately brings up the topic of whether liveness detection should be implemented in such devices. With a constantly growing market share of iris recognition, and recent research, including results presented in this Thesis, proving that iris biometrics in a post-mortem scenario can be viable [12, 13, 14, 16], these concerns may soon become a reality for iris as well. In a recent interview for the IEEE Spectrum online magazine, Czajka discussed the issue of liveness detection, which is crucial in cases when we don’t want our biometric traits to be used after death [68].

To the Author’s best knowledge, there are no prior papers or published research regarding the topic of discerning live irises from dead ones. This Chapter presents a method for iris liveness detection in a post-mortem setting, using a static iris image and based on a deep convolutional neural network fine-tuned with a dataset of post-mortem and live iris images.

6.2. Related work

6.2.1. Presentation Attack Detection in iris recognition

Presentation attack detection is already a well established research area in the field of biometrics. Existing methods include detection of fake representations of irises (paper printouts, textured contact lenses, prosthetic eyes, displays), or a non-conformant
use of an actual eye. The most popular techniques used in iris PAD use various image
texture descriptors (Binarized Statistical Image Features (BSIF) [128], Local Binary
Patterns (LBP) [129], Binary Gabor Patterns (BGP) [130], Local Contrast-Phase Descrip-
tor (LCPD) [131], Local Phase Quantization (LPQ) [132], Scale Invariant Descriptor
(SID) [133], Scale Invariant Feature Transform (SIFT) and DAISY [134], Weber Local
Descriptor (WLD) [131], or Wavelet Packet Transform (WPT) [135]), image quality
descriptors [136], or deep-learning-based techniques [137, 138, 134, 139]. If hardware
adaptations are possible one may consider multi-spectral analysis [140] or estimation
of three-dimensional iris features [141, 142] for PAD. Making the PAD more complex,
one may consider measuring micro-movements of an eyeball, either using Eulerian
video magnification [143] or by using an eye-tracking device [144], or measuring pupil
dynamics [145]. An extensive review of the state of the art in PAD for iris recognition,
including a systematization of attack methodologies and countermeasures, can be
found in [146]. Independent evaluations of algorithms detecting iris paper printouts
and textured contact lenses can be studied from the LivDet-Iris competition series¹,
which has had its last edition in 2017 [147].

Despite the abundance of proposed methods, there are still no published papers
that would explore the concept of presentation attack detection in a scenario when
cadaver eyes are used to perform a presentation attack. However, one can still envisage
such situation, in which a dead eye is used with an unsupervised biometric system to
gain an unauthorized access.

6.2.2. A feature-learning approach

We have already shown that although post-mortem iris recognition is, to some
extent, possible with current software, it also poses challenges that we do not yet
have solutions to [13]. Most importantly, there currently are no mathematical models
that would explain the iris’ behavior over the course of post-mortem time horizon,
\textit{i.e.}, quantify and predict the changes that the iris may undergo after one’s demise.
Therefore, when aiming at discerning live irises from dead ones, a potentially promis-
ing way of solving this problem is to rely on the feature-learnt approach that utilizes
post-mortem and live iris images to develop a model directly from the data.

¹ http://livdet.org/
6.2.3. Visual explanations from deep networks

DCNNs in their basic designs do not provide a human-interpretable explanation for their decisions. This makes such methods badly suited as a tool for assisting human experts in a courtroom scenario, for instance, because a softmax layer output cannot be expected to convince the jury of a person’s innocence or guilt.

To alleviate these issues, several techniques have recently been proposed, including class activation mapping (CAM), first introduced by Zhou et al. [148] for identification of discriminative image regions, decisive for the model prediction. The authors achieve this by removing fully-connected layers in the popular network architectures (AlexNet, GoogLeNet, VGG), and replacing them with global average pooling layers followed only by a softmax layer. As a result, image regions that are important for discrimination are highlighted in a heatmap-like manner. Selvaraju et al. introduce improvement over the Zhou’s method with Grad-CAM [149], which does not require any changes to the network’s architecture, making it easier to use and more flexible. The solution yields coarse localization heat-maps highlighting the regions that contribute the most to the model’s prediction, but also high-resolution visualization of features learned by the network, obtained from guided back-propagation introduced by Springenberg et al. in [150]. By combining these two, the authors obtain fine-grained importance maps, which apart from highlighting a coarse region of the image that is considered discriminatory, also allow insight into which features are important. An example visualization of these techniques is shown in Fig. 6.1.
These methods can be important for building a robust liveness detection for two reasons, both of them being explored in this Chapter. First, class activation mapping can help analyze the potential bias in the raw data that interfere with the network training, causing the model to learn features that are not directly related to the task. For instance, learning the presence of metal retractors used to open cadaver eyes, and missing in live eyes, ends up with perfect accuracy albeit with no relation to PAD accuracy. Second, we hope to gain some knowledge regarding the iris/eye features being employed by the network for discriminating between live and dead irises.

6.3. Experimental dataset

For the purpose of this study, we used the Warsaw-BioBase-PostMortem-Iris-v1 dataset, which gathers 1,330 iris images collected from 17 cadavers.

Since this database does not offer any ante-mortem samples, we had to collect a complementary dataset of iris images collected from live people. To mimic the original acquisition protocol as closely as possible, and thus to minimize a possible bias in training the DCNN, we have employed the same iris camera as was used in the post-mortem acquisition – IriShield M2120U. Example iris images used for the development of the PAD are shown in Fig. 6.2.

6.4. Proposed methodology and evaluation

6.4.1. Model architecture

For our solution, we employed the VGG-16 model pre-trained on natural images from the ImageNet database [94], which has been shown to repeatedly achieve excellent results in various classification tasks after minor adaptation and re-training. We performed a modification to the last three layers of the original network to fit the binary classification into live and post-mortem types of images, and fine-tuned the original network weights with our dataset of iris images representing live and post-mortem classes.
6.4.2. Training and evaluation procedure

For the network training and testing procedure, 20 subject-disjoint train/test data splits were created by randomly assigning the data from 3 subjects to the test subset, and the data from the remaining subjects to the train subset, both for the live and post-mortem parts of the database. These twenty splits were made with replacement, making them statistically independent. The network was then trained with each train subset independently for each split, and evaluated on the corresponding test subset. This procedure gives 20 statistically independent evaluations and allows to assess the variance of the estimated error rates. The training encompassing 10 epochs was performed with stochastic gradient descent as the minimization method with momentum $m = 0.9$ and learning rate of 0.0001, with the data being passed through the network in mini batches of 16 images.

During testing, a prediction of the live or post-mortem class-wise probability was obtained from the softmax layer, together with a corresponding predicted categorical label. The probabilities of post-mortem samples belonging to their class are also associated with a time that elapsed since death until the sample acquisition. This allows for the analysis of classification accuracy in respect to the post-mortem time horizon.

6.5. Assessing the bias in post-mortem and live samples

We are aware that despite our efforts to keep the acquisition protocols as alike as possible, there will be some bias in the data. Dataset bias has been blamed for limiting the progress and scope of object recognition research due to problematic generalization of deep-learnt methods from one dataset to another [151]. The aim of this Section is thus to carefully examine these differences, discuss their importance and impact on the experiments, and to propose countermeasures, where possible. For this purpose, we employ two types of bias evaluations - a qualitative one, employing class activation maps (cf. Sec. 6.2.3), and a quantitative one, utilizing ISO-recommended iris image quality metrics.
6.5.1. Qualitative bias assessment

There are differences between the samples that originate from different presentations of the biometric characteristics in live and post-mortem scenarios. This is most notably related to the appearance of eyelids, which in the post-mortem data are often pulled apart with a metal retracting device to keep the eye open for image acquisition. To partially mitigate these differences, the subjects participating in the collection of the reference data were asked to open their eyes as widely as possible, cf. Fig. 6.2.

To examine whether the metal retractors will serve as cues for the DCNN when it is trained to discern post-mortem irises from the live ones, we have employed the class activation mapping technique described in Section 6.2.3 with the original, uncropped images. This implementation was done using the Keras framework [152], employing an adapted code from [153].

Example predictions were then obtained for both the live and post-mortem classes, Figs. 6.3 and 6.5. This shows that the metal parts of the medical equipment used to open the eyelids of deceased subjects indeed provide class discriminatory cues (Fig. 6.3), which in this case is undesirable, as we want our network to recognize post-mortem irises, and not the equipment used for post-mortem data collection. Also, for samples with no metal parts visible, but with heavily distorted eyelids, the network pays attention not to the iris itself, but rather to its surrounding (Fig. 6.3). In none of these cases can we see strong activations by the iris region. However, when distinctive features such as metal retractors or heavily distorted eyelids are absent from the image, and its overall resemblance to an image of a live iris is strong, the model does focus on the iris region, but fails to make a correct prediction, Fig. 6.4.
Figure 6.3: Example class activation maps obtained using the Grad-CAM technique for samples from the original, unmodified dataset, with a model trained on the original dataset. Correct classification of *post-mortem* is shown. Graphics adapted from [154].

Figure 6.4: Same as in Fig. 6.3, but with *post-mortem* sample misclassified as *live*. Graphics adapted from [154].

Figure 6.5: Same as in Fig. 6.3, but for *live* samples. Graphics adapted from [154].

Notably, a similar behavior can be observed for images of live irises. When analyzing example activation maps for the *live* class, we see that it is the iris region that produces the strongest activations, while the iris surroundings remain mostly unused, as depicted in Fig. 6.5. This behavior, contrary to what we observe with post-mortem samples, can be considered desirable.
6.5.2. Quantitative bias assessment

For a quantitative evaluation of the variations between images coming from the two datasets that may originate in difference between camera operators or the environment, we have calculated three covariates related to iris quality, namely: average intensity, grayscale utilization (image histogram entropy), and image sharpness (the latter two being suggested by the ISO [155]):

— **average image intensity**: 

\[
AI = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij}
\]

(6.1)

where \(I_{ij}\) is the pixel intensity and \(n, m\) is the image size in pixels;

— **grayscale utilization**, or the entropy of the iris image histogram \(H\), measured in bits, examines pixel values of an iris image to calculate a spread of intensity values and assess whether the image is properly exposed:

\[
H = -\sum_{i=1}^{256} p_i \log_2 p_i
\]

(6.2)

where \(p_i\) is the probability of each gray level \(i\) occurring in the image, hence, the total count of pixels at gray level \(i\), divided by the total number of pixels;

— **sharpness**, determined by the power resulting from convolving the image with a Laplacian of Gaussian kernel with \(\sigma = 1.4\). For convenience, we repeat this formula from the standard. Let

\[
F = \begin{bmatrix}
0 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 0 \\
1 & 2 & 4 & 5 & 5 & 4 & 2 & 1 \\
1 & 4 & 5 & 3 & 0 & 3 & 5 & 4 & 1 \\
2 & 5 & 3 & -12 & -24 & -12 & 3 & 5 & 2 \\
2 & 5 & 0 & -24 & -40 & -24 & 0 & 5 & 2 \\
2 & 5 & 3 & -12 & -24 & -12 & 3 & 5 & 2 \\
1 & 4 & 5 & 3 & 0 & 3 & 5 & 4 & 1 \\
1 & 2 & 4 & 5 & 5 & 4 & 2 & 1 \\
0 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 0
\end{bmatrix}
\]

(6.3)
be the convolutional kernel, with which the image \( I(x, y) \) is convolved and a weighted sum of the filter response is computed for every fourth row and column location in \( I(x, y) \) to produce filtered image \( I_F(x, y) \):

\[
I_F(x, y) = \sum_{i=-4}^{4} \sum_{j=-4}^{4} I(x+i, y+j)F(i+5, j+5)\forall x \in [1, 5, ..., w], y \in [1, 5, ..., h]
\] (6.4)

where \( w \) and \( h \) are the width and height of \( I(x, y) \), respectively. Then the squared sum \( SS \) of the \( I_F(x, y) \) is computed:

\[
SS = \sum_{\forall x,y \in I_F(x,y)} I_F(x, y)^2
\] (6.5)

and the power \( P \) in \( I_F \):

\[
P = \frac{SS}{w_F \times h_F}
\] (6.6)

where \( w_F \) and \( h_F \) are the width and height of \( I_F(x, y) \), respectively. Finally, the sharpness covariate \( SH \) is given by:

\[
SH = 100 \times \frac{P^2}{P^2 + c^2}
\] (6.7)

where the \( c \) constant has been chosen empirically to be 1800000.

Results of these calculations are shown in Fig. 6.6. Notably, only the sharpness covariate differs largely between the two datasets, as post-mortem iris images have lower sharpness on average. This can be a result of a combination of factors, such as: more difficult collection environment (e.g., a hospital mortuary), a less experienced operator (e.g., medical staff), limited time, and such. For completeness, to provide formal statistical analysis, we ran a Wilcoxon rank-sum test for each pair of covariates, which revealed that there are statistically significant differences between the subsets of live and post-mortem iris images, as the null hypothesis stating that the compared scores are samples from continuous distributions with equal median was rejected at significance level \( \alpha = 0.05 \) in all three cases. However, all three pairs of covariates do not provide enough differentiation between the images coming from different datasets, to themselves serve as features for presentation attack detection detection.
Figure 6.6: Boxplots representing differences in image quality metrics between the two datasets. Median values are shown in red, height of each boxes corresponds to an inter-quartile range (IQR) spanning from the first (Q1) to the third (Q3) quartile, whiskers span from Q1-1.5*IQR to Q3+1.5*IQR, and outliers are shown as crosses. Figures reprinted from [154].

6.5.3. Dataset modification to counteract the bias

To force-shift the network’s attention to the iris, and not its neighborhood, we have manually segmented all images in both datasets, approximating the outer iris boundary with a circle with a radius \( R_i \) and then cropping and masking the image to the size of \( 1.2R_i \), see Fig. 6.7. This margin of \( 0.2R_i \) is preserved purposefully, to represent the differences in the iris surroundings and in the iris-sclera boundary, and not only those present in the tissue itself. The same reasoning is behind leaving the pupillary region unmasked, as the appearance of the pupil can bear liveness information as well. The raw images obtained from the sensor are thus referred to as the original dataset, while the modified version is called the cropped_masked dataset.

Figure 6.7: Cropped_masked versions of the images from Fig. 6.2.

To validate this reasoning, we train the same model that was employed for assessing class activation maps in the unmodified data, but with the cropped_masked images instead. Activation maps drawn for example cropped_masked iris images are shown in Figs. 6.8, 6.9, and 6.10. As for the post-mortem samples, the new model now mostly
Figure 6.8: Example class activation maps for samples from the cropped_masked dataset, with a model trained on the cropped_masked dataset as well. Correct classification of post-mortem samples. Graphics adapted from [154].

Figure 6.9: Cropped version of the post-mortem sample misclassified earlier in Fig. 6.4. The classification is now correct. Graphics adapted from [154].

focuses on the iris and its boundary, cf. Fig. 6.8, which seems reasonable, as the iris-sclera boundary is quickly getting blurry as time since death progresses. This is different for the problematic sample examined earlier in Fig. 6.4, for which the activation map is centered near the pupillary region of the eye. This sample, however, is now correctly classified as a post-mortem one.

When it comes to samples representing live eyes, the network also seems to assess their PAD score by analyzing the iris-boundary region, but to some degree it also brings its attention to the iris itself, and to the specular reflection found in the middle of the pupil, as depicted in Fig. 6.10. These features seem to offer enough discriminatory power to successfully differentiate between the post-mortem and live samples, as the validation accuracy on the cropped_masked subset of subject-disjoint iris images reaches 100%, compared to less than 95% obtained for the unmodified set of the same iris images. This also shows that we have managed to successfully shift the attention of the network towards the iris features and discriminative information they offer in recognizing cadaver samples.
Figure 6.10: Same as in Fig. 6.8, but for live samples. Graphics reprinted from [154].

Figure 6.11: (Right) Accuracy of classification into live and post-mortem classes, achieved in 20 independent, subject-disjoint train/test data splits. (Left) ROC curves showing classification accuracy averaged over 20 independent, subject-disjoint train/test data splits. AUC denotes the Area Under Curve metric. Graphs reprinted from [154].

6.6. Results and discussion

6.6.1. Averaged classification accuracy

Several metrics can be utilized to evaluate the classification accuracy of our solution. First, we average the accuracy achieved in each of the twenty train/test splits, measured as a share of correctly assigned labels to the overall number of trials in a given split, Fig. 6.11. Notably, in most of the splits, the solution achieves a 100% classification accuracy on the test subset, with the average of 98.94%.

6.6.2. Global performance by Receiver Operating Characteristic

Receiver Operating Characteristic curves (ROC) are often employed for visualizing performance of classification systems, especially when binary classification is in place. Thus, we present ROC graphs with Areas Under Curve (AUC) calculated for each
Figure 6.12: Boxplots representing differences in liveness scores earned by the samples acquired in different moments after death. **The lower the score, the more likely the sample represents a post-mortem eye.** Samples denoted as acquired zero hours post-mortem are those collected from live irises. Figures reprinted from [154].

Having post-mortem samples acquired for different subjects at multiple time points after death, makes it possible to analyze the performance of our solution in respect to the time that has passed since a subject’s demise, Fig. 6.12. Interestingly, and perhaps accordingly to a common sense, the probability of a sample being post-mortem increases as time since death elapses. We can expect a few false matches (post-mortem samples being classified as live samples) with images obtained 5 hours after death, regardless of the chosen threshold. This can be attributed to the fact that these images are very similar to those obtained from live individuals, as post-mortem changes to the eye are still not pronounced enough to allow for a perfect classification accuracy. However, the already good accuracy is getting close-to-perfect when these samples are not
taken into consideration, Fig. 6.11, red dotted line. Attack Presentation Classification Error Rate (APCER, misclassifying post-mortem samples as live ones) equal to zero can be achieved in such scenario, provided that an appropriate acceptance threshold is established. Bona Fide Presentation Classification Error Rate (BPCER, misclassifying live samples as post-mortem ones) in such case is only $\approx 1\%$.

6.7. Conclusions

This Chapter introduces the first known to us method for iris liveness detection in respect to the post-mortem setting. The proposed PAD component is able to correctly classify nearly 99% of the samples, assigning alive or post-mortem labels, respectively. An attempt to explain the reasoning behind the decisions provided by our DCNN-based solution is also made, especially with respect to the iris regions that are considered when making these decisions. By employing the Grad-CAM class activation mapping technique, we managed to show that the image regions bearing the most useful discriminatory cues for liveness detection are those containing iris-sclera boundary, and to some extent also the pupillary region.

This proves the last, fourth thesis formulated in Sec. 1.2, namely that liveness detection method can be proposed to discern post-mortem iris samples from those collected from living individuals, requiring only a static iris image.
This doctoral dissertation is built around four theses, as stated in the introductory part, which correspond to four main contributions of this work. After the introduction to iris biometrics and formulation of the research problems in Chapter 1, the most comprehensive evaluation of the impact of biological processes in the eye on the accuracy and reliability of iris recognition, that we are aware of, is carried out in Chapters 2 and 3. The experiments revealed that disease-borne changes to the iris texture, to its geometry, and to the cornea, as well as post-mortem processes have a large impact on recognition performance.

We have identified an incorrectly executed segmentation stage as the prevalent source of errors, which was a starting point to develop a new, robust iris image segmentation algorithm, which is then introduced in Chapter 4. The proposed solution incorporates a deep convolutional neural network and is able to effectively learn deformations of the iris specific to post-mortem and disease-induced biological processes. We then show how to use the DCNN-based segmentation results in the typical iris recognition pipeline, which allows to achieve superior matching performance when compared against two conventional approaches, including one of the state-of-the-art commercial iris recognition method. This new method outperforms the two mentioned algorithms by a significant margin on the most challenging database of post-mortem iris images, enabling close-to-perfect matching accuracy and no missed matches for samples collected up to a day after death, and only a 1.5% of missing a match during the first 48 hours, when allowing a 5% false match rate. Recognition accuracy for disease-affected samples is also improved, with EER=1.73% compared to 3.97% and 8.90% obtained from the commercial and academic method, respectively.

Although the proposed segmentation method allows to decrease the recognition error rates by a wide margin, some of the post-mortem processes in the eye seem to alter the iris features themselves, and as such they cannot be neutralized only by applying the robust image segmentation. Therefore, in the next step of improving
post-mortem iris biometrics, we have devised a novel feature representation scheme, which employs combining the traditional Gabor-based image filtration with new filters that were learnt from the data. This allows to further rectify the error rates by as much as one third, depending on the considered post-mortem time interval.

Finally, Chapter 6 introduces the last component of the proposed biometric system, namely iris liveness detection capable of discerning live from post-mortem iris presentations with 99% correct classification rate.

The Author hopes that this doctoral dissertation will constitute a valuable addition to the state-of-the-art in iris recognition carried out under biologically challenging situations. The advancements to post-mortem iris biometrics proposed in this work may offer an important contribution to the field of forensics, possibly extending the range of post-mortem human identification tools. With new data processing techniques, rapidly accelerating advancements in machine learning, and more focus from the scientific community, iris recognition can not only keep its leading role in secure and reliable authentication, but also move into previously unexplored fields of knowledge and applications.


[60] Science Focus, “Would an iris scan work if you were unconscious or your eyeball was detached?” accessed: April 10, 2016. [Online]. Available: http://www.sciencefocus.com/qa/would-iris-scan-work-if-you-were-unconscious-or-your-eyeball-was-detached


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.

7.1. General concept terms

7.1.1. biometric characteristic

*Biological and behavioral characteristic of an individual from which distinguishing, repeatable *biometric features* can be extracted for the purpose of *biometric recognition*.

7.1.2. biometric recognition

*Automated recognition of individuals based on their biological and behavioral characteristics.*

7.2. Terms for data in biometric systems

7.2.1. biometric sample

*Analog or digital representation of *biometric characteristics* prior to *biometric feature extraction*.*

In this work, *biometric samples* are equal to iris images.

7.2.2. biometric feature

*Numbers or labels extracted from *biometric samples* and used for *comparison*. 
In this work, **iris features** are understood rather as personal information that exists within the iris texture, than a mathematical representation of an iris extracted from an iris image. For the latter, the term **biometric template** is used.

### 7.2.3. biometric template

*Set of stored biometric features comparable directly to probe biometric features.*

In this work, a **biometric template** is a mathematical representation of an iris extracted from an iris image by a particular iris processing method.

### 7.2.4. biometric data

**biometric sample** or aggregation of **biometric samples** at any stage of processing, e.g. biometric reference, biometric probe, **biometric feature**, or biometric property.

In this work, **biometric data** usually refers to an aggregation of biometric samples, in this case – iris images.

### 7.2.5. comparison decision

*Determination of whether the biometric probe and biometric reference have the same biometric source, based on comparison score(s), a decision policy including a threshold, and possibly other inputs. A match is a positive comparison decisions, whereas a non-match is a negative one.*

In this work, where applicable, the **comparison decision** is issued based on a **comparison score** between two **biometric templates** and an appropriate **threshold**.

### 7.2.6. threshold

*Numerical value (or set of values) at which a decision boundary exists.*

### 7.2.7. comparison score

*Numerical value (or set of values) resulting from a comparison.*

### 7.2.8. dissimilarity score, distance score

**Comparison score** that decreases with similarity.
In this work, three iris recognition methods used by the Author yield dissimilarity scores: in the form of a fractional Hamming distance (OSIRIS, MIRLIN), or in the form of a proprietary metric (IriCore).

7.2.9. similarity score

Comparison score that increases with similarity.

In this work, the only iris recognition method used by the Author yields a similarity score in the form of a proprietary metric (VeriEye).

7.3. Functioning terms

7.3.1. 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 iris image) by performing biometric feature extraction.

7.3.2. 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.

7.3.3. 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.
7.4. Interacting terms

7.4.1. 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 a cadaver eye to the iris recognition camera.

7.5. Applications terms

7.5.1. biometric identification

_Process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual._

7.5.2. biometric verification

_Process of confirming a biometric claim through biometric comparison._

7.6. Performance terms

7.6.1. failure to enrol, FTE

_Failure to create and store a biometric enrolment data record._

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

7.6.2. failure-to-enrol rate, FTE rate, FTER

_Proportion of a specified set of biometric enrolment transactions that resulted in a failure to enrol._

7.6.3. false match

_Comparison decision of match for a biometric probe and a biometric reference that are from different biometric capture subjects._
In this work, this means a **match** decision for two **biometric templates** that are from different irises. Also: a **false positive**.

### 7.6.4. false match rate, FMR

Proportion of the completed biometric non-mated comparison trials that result in a **false match**.

### 7.6.5. false non-match

*Comparison decision of non-match* for a biometric probe and biometric reference that are from the same biometric capture subject and of the same biometric characteristic.

In this work, this means a **non-match** decision for two **biometric templates** that are from the same iris. Also: a **false negative**.

### 7.6.6. false non-match rate, FNMR

Proportion of the completed biometric mated comparison trials that result in a **false non-match**.
Appendix B: 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.

Journals with Impact Factor

[MNiSW list A, 40 points]
My contributions to this work include: a) co-designing the methodology and experiments, b) initial analysis and pre-processing of the biometric data, c) preparation of the extended biometric database to be shared with the research community, d) review and summary of the related works, e) conducting most of the experiments, except for Section V-F (done together with Adam Czajka) and Section V-G (done by Adam Czajka), f) creating Figures 1, 4, 5, 6, 8, 9, 11, and 13, g) writing approx. 70% of the manuscript.

[MNiSW list A, 35 points]
My contributions to this work include: a) reviewing and summarizing related works, b) initial analysis, censoring, and categorization of the biometric data, c) preparation of the extended biometric database to be shared with the research community, d) calculation of the comparison
scores between irises using four different biometric methods, e) preparation of approx. 50% of the manuscript.

Other refereed journals


[MNiSW list B, 12 points]
My contributions to this work include: a) co-designing the methodology and experiments, b) designing data collection protocol, c) preparing the graphical representation of the results and conducting statistical analyses, d) writing approx. 70% of the manuscript.

Book chapters


[4 points MNiSW]
My contributions to this work include: a) re-running all the necessary experiments, b) conducting statistical analysis of the results, c) creating all plots and graphics, d) writing all of the chapter manuscript.

Reviewed conference publications

My contributions to this work include: a) training the benchmark deep-learning-based SegNet
model and generation of the occlusion masks from this method, b) implementation of the iris normalization method, c) writing approx. 15% of the paper.


My contributions to this work include: a) designing (together with Adam Czajka) parts of the experimentation related to the iris matching accuracy assessment, b) implementation of the iris normalization method, c) analysis of the segmentation accuracy results, d) running the experiments related to iris matching accuracy, e) creating all plots and graphics except for Fig. 1, f) writing approx. 60% of the paper.


My contributions to this work include: a) designing (together with Adam Czajka) and conducting the gaze-tracking experiments at NASK, b) statistical analysis of their results, c) assessment of human iris verification performance in comparison with a machine solution, d) writing approx. 20% of the paper.


My contributions to this work include: a) co-designing the methodology and experiments, b) implementation of the models and conducting all of the experiments, c) statistical analysis of the results, d) creating all plots and graphics, e) writing most of the manuscript except for Sec. 2.1.
My contributions to this work include: a) discussion of the methodology and concepts, b) drafting most of the paper manuscript, except for the review of related works, c) edits to the plots and graphics.

My contributions to this work include: writing approx. 90% of the initial manuscript version and most of the final revised submission.

My contributions to this work include: a) co-designing the methodology and experiments, b) implementation of the models and conducting all of the experiments, c) statistical analysis of the results, d) creating plots and graphics, except for Figs. 4-7, e) writing most of the manuscript except for Sec. III-F, f) participation in the preparation of the database of manually annotated ground truth masks.

My contributions to this work include: a) designing the methodology and experiments, b) review of the related works, c) implementation of the model and conducting the experiments, d)
statistical analysis of the results, e) writing approx. 70% of the manuscript, f) preparation of the database to be shared with the research community.

My contributions to this work include: a) designing the methodology and experiments, b) review of the related works, c) implementation of the model and conducting the experiments, d) statistical analysis of the results, e) writing approx. 70% of the manuscript.

My contributions to this work include: a) initial analysis, censoring, and categorization of the biometric data, b) calculation of the comparison scores, c) preparation of the entire manuscript.

My contributions to this work include: a) reviewing and summarizing related works, b) initial analysis, censoring, and categorization of the biometric data, c) preparation of the biometric database to be shared with the research community, d) calculation of the comparison scores between irises using four different biometric methods, e) preparation of approx. 45% of the manuscript.
My contributions to this work include: a) reviewing and summarizing related works, b) initial analysis, censoring, and categorization of the biometric data, c) calculation of the comparison scores between irises using four different biometric methods, d) preparation of approx. 45% of the manuscript.

My contributions to this work include: a) reviewing and summarizing related works, b) participation in the design and development of the data collection device, c) initial analysis, censoring, and categorization of the biometric data, d) preparation of the biometric database to be shared with the research community, e) calculation of the comparison scores between irises using four different biometric methods, f) preparation of approx. 65% of the manuscript.

My contributions to this work include: a) reviewing and summarizing related works, b) Database of iris images acquired in the presence of ocular pathologies and assessment of iris recognition reliability for disease-affected eyes", 2nd IEEE International Conference on Cybernetics CYBCONF 2015, Special Session on Reliable Biometrics BIORELIABILITY 2015, June 24-26, 2015, Gdynia, Poland, DOI: 10.1109/CYBConf.2015.7175984
My contributions to this work include: a) reviewing and summarizing related works, b) participation in the design and development of the data collection device, c) initial analysis, censoring, and categorization of the biometric data, d) preparation of the biometric database to be shared with the research community, e) calculation of the comparison scores between irises using four different biometric methods, f) preparation of approx. 60% of the manuscript.


My contributions to this work include: a) designing the methodology and experiments, b) designing the data collection protocol, c) analysis of the results, d) writing most parts of the manuscript.


My contributions to this work include: a) reviewing and summarizing related works, b) participation in the design and development of the data collection device, c) initial analysis, censoring, and categorization of the biometric data, d) calculation of the comparison scores between irises using three different biometric methods, f) preparation of approx. 60% of the manuscript.
Appendix C: List of Author’s active conference participations

"Presentation Attack Detection for Cadaver Iris", 9th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2018), October 22-25, 2018, Los Angeles, USA
oral and poster presentations of paper [Ref8]

"Data-Driven Segmentation of Post-mortem Iris Images", 6th IAPR/IEEE International Workshop on Biometrics and Forensics (IWBF 2018), June 7-8, 2018, Sassari, Italy
oral presentation of paper [Ref11]

"Iris recognition with visible light on mobile devices", European Biometrics Symposium, November 8, 2017, Canterbury, UK
invited talk

oral and poster presentations of paper [Ref12]

oral presentation of paper [Ref13]
oral presentation of paper [Ref14]

"Post-mortem Iris Recognition in Human Subjects: Study and Database", 8th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2016), September 6-9, 2016, Niagara Falls, USA
oral presentation of paper [Ref15]

poster presentation of recent research results

"Iris Recognition Reliability in the Context of Template Aging, Ocular Disorders, and Death", EAB European Biometrics Research and Industry Awards Ceremony, September 21, 2016, Darmstadt, Germany
oral presentation of recent research results

"Post-mortem Human Iris Recognition", 9th IAPR International Conference on Biometrics (ICB 2016), June 13-16, 2016, Halmstad, Sweden
oral presentation of paper [Ref16]

"Iris Recognition with a Database of Iris Images Obtained in Visible Spectrum Using Smartphone Camera", IEEE International Conference on Identity, Security and Behavior Analysis ISBA 2016, February 29 - March 2, 2016, Sendai, Japan
oral and poster presentation of paper [Ref17]

"Assessment of iris recognition reliability for eyes affected by ocular pathologies", IEEE 7th International Conference on Biometrics: Theory, Applications and Systems (BTAS 2015), September 8-11, 2015, Arlington, USA (awarded with the Best Paper Award)
oral presentation of paper [Ref18]
"Database of iris images acquired in the presence of ocular pathologies and assessment of iris recognition reliability for disease-affected eyes", 2nd IEEE International Conference on Cybernetics CYBCONF 2015, Special Session on Reliable Biometrics BIORELIABILITY 2015, June 24-26, 2015, Gdynia, Poland
oral presentation of paper [Ref19]

oral presentation of paper [Ref20]

"Linear regression analysis of template aging in iris recognition", IEEE 3rd International Workshop On Biometrics and Forensics, 3-4 March 2015, Gjøvik, Norway
oral presentation of paper [Ref21]

oral and poster presentations of paper [Ref21]

oral presentation of paper [Ref22]

oral presentation of recent research results
Appendix D: List of Author’s awards and achievements

EAB European Biometrics Research Award (First Prize), 21 Sep 2016, Darmstadt, Germany, awarded for the research results on *Iris Recognition Reliability in the Context of Template Aging, Ocular Disorders, and Death* [156]

**Best Paper Award (Gold)** at the IEEE 7th International Conference on Biometrics: Theory, Applications and Systems (BTAS2015), 8-11 Sep 2015, Arlington, USA, awarded for the paper *Assessment of iris recognition reliability for eyes affected by ocular pathologies* [10] (first author)

**Best Student Paper Competition (Second Prize)** in the Competition for the best PhD student paper or patent in the academic year 2015/2016, Faculty of Electronics and Information Technology, Warsaw University of Technology, awarded for the paper *Implications of Ocular Pathologies for Iris Recognition Reliability* [11] (first author)
Appendix E: List of grants and projects in which the Author participated

As a Principal Investigator

[1/9/2018 – 7/12/2018] Zbadanie charakteru i dynamiki zmian częstotliwości przestrzennych cech tęczówki w przypadku oczu objętych zaawansowanymi procesami starzenia i oczu osób zmarłych (Eng. Dynamics of iris texture frequency changes in elderly and deceased subjects), NASK Grant Agreement, 2018 Grants for Young Scientists

[1/8/2018 – 31/12/2018] Metody segmentacji obrazów termicznych dłoni dla celów rozpoznawania tożsamości na podstawie rozkładu temperatury dłoni (Eng. Methods for thermal hand images segmentation for heat-distribution-based biometric recognition), Faculty of Electronics and Information Technology, Warsaw University of Technology, Grant Agreement under the 2018 Dean’s Grant Program

[1/5/2017 – 31/3/2018] Rozpoznawanie tęczówki z wykorzystaniem splotowych sieci neuronowych (Eng. Iris recognition using convolutional neural networks), Faculty of Electronics and Information Technology, Warsaw University of Technology, Grant Agreement under the 2017 Dean’s Grant Program


As an Investigator


[1/8/2017 – 31/8/2017] A Tool Supporting Human Examination of Post-mortem Iris Images (17TSHEPII), visit at the University of Notre Dame, Project for the FBI Biometrics Center of Excellence, University of Notre Dame, USA, Lead Principal Investigator: Dr Adam Czajka


[20/12/2015 – 30/9/2018] BIOWIZ – Opracowanie i stworzenie systemu mającego na celu identyfikację osób (sprawców przestępstw) na podstawie wizerunku utrwalonego na zdjęciu lub materiałach wideo (Eng. BIOWIZ – Design and construction of a system for recognition of persons (offenders) based on face images captured on photograph or video material), NCBiR (National Center for Research and Development) - Research and development projects for state defence and security 7/2015, contract no. DOB-BIO7/18/02/2015, Principal Investigator: Prof. Andrzej Pacut