UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

NEW METHODS FOR 3D IRIS SCANNING FROM MULTIPLE 2D VISIBLE-LIGHT IMAGES

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN INGENIERÍA ELÉCTRICA

DANIEL PATRICIO BENALCÁZAR VILLAVICENCIO

PROFESOR GUÍA:
CLAUDIO PÉREZ FLORES
PROFESOR CO-GUÍA:
KEVIN W. BOWYER

MIEMBROS DE LA COMISIÓN:
PABLO ESTEVEZ VALENCIA
DOMINGO MERY QUIROZ
GONZALO RUZ HEREDIA

Este trabajo ha sido parcialmente financiado por ANID (CONICYT) a través de los proyectos FONDECYT 1161034 y 1191610, por el proyecto basal AMTC-AFB180004, así como por el Departamento de Ingeniería Eléctrica (DIE), Universidad de Chile.

SANTIAGO DE CHILE
2020

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE DOCTOR EN INGENIERÍA ELÉCTRICA
POR: DANIEL PATRICIO BENALCÁZAR VILLAVICENCIO
FECHA: SEPTIEMBRE 2020
PROF. GUIA: PH.D. CLAUDIO ANDRÉS PÉREZ FLORES

NUEVOS MÉTODOS PARA EL ESCANEO 3D DEL IRIS A PARTIR DE MULTIPLES IMÁGENES 2D EN LUZ VISIBLE

El reconocimiento de iris es uno de los métodos biométricos más exitosos; sin embargo, emplea imágenes 2D para el análisis, cuando el iris es en realidad una estructura muscular tridimensional. Esas fibras musculares producen un relieve en la superficie del iris, lo cual queremos caracterizar en un modelo 3D. La información adicional de profundidad permitiría incrementar precisión en reconocimiento de iris y tendría aplicaciones potenciales en oftalmología. En esta tesis doctoral desarrollamos y comparamos dos distintas aproximaciones al escaneo 3D de iris, empleando por separado "Structure from Motion" (SfM) y Redes Neuronales Convolucionales (CNN). Para entrenar la arquitectura propuesta, irisDepth, capturamos 26.520 imágenes de 120 sujetos. El método SfM produjo 11.105 puntos 3D en promedio mientras el método CNN produjo 65.535, lo cual es 6 veces superior. La resolución es 11µm y 17.7µm para SfM y CNN respectivamente. El error medio entre una Tomografía de Coherencia Óptica y el corte correspondiente en el modelo 3D es 123µm para SfM y 77µm para CNN. De manera que CNN incrementa la exactitud en 60% respecto a SfM. Finalmente, los modelos 3D incrementaron el desempeño en 68% respecto al código de iris en un conjunto de 50 sujetos y 2.000 imágenes.

SUMMARY OF THE THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR IN ELECTRICAL ENGINEERING
BY: DANIEL PATRICIO BENALCÁZAR VILLAVICENCIO
DATE: SEPTEMBER 2020
ADVISOR: PROF. CLAUDIO ANDRES PEREZ FLORES

NEW METHODS FOR 3D IRIS SCANNING FROM MULTIPLE 2D VISIBLE-LIGHT IMAGES

Iris recognition is one of the most successful biometric methods; however, it uses 2D images for the analysis when the iris is in fact a 3D muscular structure. Those muscular fibers create a relief in the iris surface iris, which is what we propose to characterize in a 3D model. The additional depth information aims to increase iris recognition performance and has potential applications in ophthalmology. In this Doctoral thesis, we developed and compared the performances of two different approaches to 3D iris scanning using separately Structure from Motion (SfM) and Convolutional Neural Networks (CNN). To train the proposed CNN architecture, irisDepth, we captured 26,520 images from 120 subjects. The SfM method produced 11,105 3D points in average while the CNN method produced 6 more at 65,536. The resolution of SfM and CNN were 11µm and 17.7µm respectively. The average error between a ground-truth Optic Coherence Tomography and the corresponding slice in the 3D model was 123µm for SfM and 77µm for CNN. Thus, the CNN method increased accuracy in 60% with respect to SfM. Finally, the 3D models increased iris recognition performance 68% with respect to the standard iris code in a dataset of 50 subjects and 2,000 images.

“Whether you think you can, or you think you can't--you're right.”

— Henry Ford

To my parents Patricio and Julia. All of this is thanks to your constant support, as well as the values you taught me as a child.

Thank you both!

Acknowledgements

This research has been funded by CONICYT through FONDECYT projects 1161034 and 1191610, as well as by the Department of Electrical Engineering, Universidad de Chile.

Finishing a doctorate thesis within four years has been by far the hardest accomplishment in my life. More than intelligence alone, it required integrity, constancy, tenacity, patience, prudence, and boldness. This accomplishment would have never been possible without the help and support of wonderful people during these years. People who also instilled those values in me. I would like to thank all of them for their invaluable contribution in my professional and personal growth.

To my mother Julia, my father Patricio and my brother David. Thanks for your unconditional love and support, for raising me to be the man I am today, and for showing me the bright side in my darkest times.

To my advisor Claudio Perez. Thanks for receiving me with open arms to work in this project, and for showing me the ways in the research world. Thanks for challenging me to improve myself, and for pushing me to set a higher bar with our methods and results; but specially thank you and Prof. Kevin W. Bowyer for the hard work and long nights you both put in this research project.

To my family and friends. Thanks for always being there for me when I needed advice or a good laugh. Special thanks to my dear friends who helped me polish my presentations.

Finally, thanks to family, friends, and students at Universidad de Chile for lending me your eyes when I needed iris pictures for this thesis. I am in great debt to you for this work would not have been possible without all of you.

Daniel Patricio Benalcazar-Villavicencio

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Motivation

The need for accurate and unequivocal identity verification systems has driven the research on iris recognition techniques. These techniques exploit the texture of the human iris because it exhibits a distinctive pattern with a high variability between individuals [1][2]. Iris recognition is nowadays the most reliable biometric technology in the market because of its non-invasive, accurate and robust methodology [3][4]. Iris recognition has succeeded in both small and large scale applications. Some large scale uses include United Arab Emirates border-crossing [5] and India's Unique ID program [6]. Furthermore, iris recognition resolved disputes in the registration list for elections in Somaliland [7].

The main reason behind the growth of iris recognition as a biometric measure is that a high level of accuracy can be obtained. For instance, Daugman demonstrated that a false match rate of 1 in 4 million can be achieved with a Hamming Distance (HD) criterion of 0.33 [1][8]. Therefore, the human iris has a variability that is statistically significant in large databases. That is true even between the left and right eyes of the same person, and between identical twins [1][9]. In contrast, the variability of fingerprint analysis would produce inadmissible false matches in large databases [2], which is why India, for example, migrated from fingerprints to Iris recognition for its iris-based Unique ID program [6].

Although iris recognition is a well-developed technique with various applications, it presents some problems. First, iris recognition might encounter similar limitations as fingerprint analysis if the database is large enough. For instance, in a database of one billon individuals, a false match rate of 1 in 4 million indicates that 250 people have similar iris codes. In addition, the iris is usually partially occluded by eyelids, eyelashes and reflections, and therefore the amount of extracted information decreases. Moreover, iris recognition techniques assume that the iris is flat, when in fact it is a three-dimensional organic structure.

In order to control the amount of light that hits the retina, the human iris dilates and contracts using two muscle systems, a sphincter and several radial dilator muscles [10]. Those structures create a relief in the surface of the iris that can be evidenced in an Optical Coherence Tomography (OCT) [11]. However, when this complex 3D structure is mapped onto a 2D plane, due to the image acquisition process, the depth information is lost. Therefore, capturing depth information could, in principle, enhance the amount of information we can extract from the iris.

That is why in this doctoral thesis a new method is proposed to develop a 3D iris scanning system, i.e., the reconstruction of a 3D model of the iris surface.

In the literature, two of the main methods used to reconstruct 3D models from 2D images are Structure from Motion (SfM) [12][13], and Convolutional Neutral Networks (CNN) [14][15]. SfM treats a 3D reconstruction as an optimization problem: given a group of images, find the camera poses and 3D model that minimize the total re-projection error [12][16]. The 3D model in this case consists of a set of 3D points resulting from keypoints in the input images [17][18]. The other main method is based on CNNs, making use of their great abstraction power that can estimate a depthmap from a single view of a scene [14][19]. After training with thousands of image depthmap pairs, encoder-decoder architectures are able to predict the depth value of every pixel in the input image [20][21]. Some CNN architectures even make use of synthetic images to achieve state-of-the-art results in scenarios where a large dataset is hard to obtain [22][15]. Both SfM and CNN methods expand the information present in 2D images by the addition of a depth axis.

With the additional depth information, extra degrees of freedom might allow iris recognition to be applied in even larger databases than what is currently possible, even when only a portion of the iris is available for analysis. Thus, the scope of iris recognition would be extended. Moreover, 3D iris scanning could be a promising screening tool for Acute Angle Glaucoma [23][24][25], a disease currently diagnosed with an OCT [26]. In addition, an iris 3D model will enable studying the deformation of this organ due to pupil dilation in greater detail [23][24][25]. This in turn could lead to the formulation of more accurate rubber sheet models [1][8] for iris normalization even under extreme dilation differences [27].

1.2. Problem Statement

We are interested in extracting a 3D representation of the iris surface in order to incorporate additional depth information in the Iris Recognition paradigm. We aim to adapt 3D-reconstruction methods, such as SfM and CNN, into a robust methodology capable of generating a 3D model of the human iris, which closely resembles the structural features of the original. Since both SfM and CNN rely on 2D images alone, the resulting method of this research would produce a non-invasive way to study the relief in the surface of the human iris in greater detail. Additionally, a 3D iris scanner may enable studying, in the near future, other phenomena involving the iris such as: 3D iris surface deformation due to pupil dilation, and Acute Angle Glaucoma.

Despite the fact that SfM and CNN produce detailed 3D reconstructions of static scenes, the human iris is a complex dynamic organ enclosed in the anterior chamber of the eye. Iris dilation introduces deformation on iris images which neither SfM nor CNN are capable of handling directly. Furthermore, close range pictures of the iris require macro lenses, which introduce additional levels of distortion [24]. Therefore, the nature of the iris introduces additional challenges for 3D iris scanning, which we tackled in this thesis.

Part of the challenge in 3D iris scanning is developing a suitable image acquisition device and methodology to build the 3D model. Although traditional iris recognition devices use as a standard frontal NIR illumination with a rather low spatial resolution (generally 360×240 pixels), it might not be optimal to extract anatomical information from the surface of the iris. Additionally, the quality and resolution of the sensor plays an important role in 3D reconstruction [23]. That is why we have also explored alternative illumination methods and cameras, in the Visible Light (VL) spectrum, in order to obtain the best possible performance.

1.3. General Objective

Develop a reliable and robust methodology for 3D Iris Scanning using multiple 2D VL images of the iris to obtain 3D models of the iris.

1.4. Specific Objectives

- Develop a device to capture close range 2D images of the iris from several perspectives in a controlled VL illumination environment.
- Compare the iris recognition performance of lateral and frontal VL with respect to the standard NIR illumination.
- Develop 3D iris scanning algorithms, which account for pupil dilation and lens distortion, using separately SfM, and CNNs.
- Evaluate the scanning resolution of SfM, and CNN methods by measuring features in scanned real 3D patterns of known dimensions.
- Evaluate the iris recognition performance of the iris 3D model in comparison with the standard 2D iris code when both methods use the same dataset of 2D iris images.

1.5. Hypotheses

- The surface of the human iris can be scanned and reconstructed in a 3D model from 2D VL images using methods such as SfM, and CNN.
- Lateral VL illumination enhances the texture of iris images compared to frontal VL illumination alone. That texture should produce similar iris recognition results as images captured under the standard NIR illumination.
- The resolution of the 3D iris scanning method can be assessed by measuring features in 3D scans of physical patterns that have known dimensions.

- The additional degree of freedom that provides depth information is expected to increase the discriminative information that can be extracted from the iris, and by extension increase iris recognition performance.

1.6. Contributions

The following are the main contributions of this thesis:

- Improving the scanning device introduced in [23] with better resolution, less noise, precise control over the illumination level of each eye, gaze control of both eyes, the inclusion of a close range lens, the use of VL instead of NIR light, and the use of a combination of Lateral and Frontal Visible Light (LFVL) illumination, which highlights structural information of the iris.

- Acquiring two datasets of iris images from 96 and 120 subjects with 1,920 and 26,520 images, respectively. Both datasets were acquired using the proposed device.

- Creating a dataset of 72,000 synthetic iris images which simulate the LFVL illumination conditions of the proposed device. We made this dataset freely available on GitHub. https://github.com/dpbenalcazar/irisDepth

- Adding a preprocessing stage to correct lens-distortion [28][29] and to enhance the quality of the iris input images by means of local-laplacian contrast enhancement [30].

- Adapting the SfM pipeline for this specific application to make the system more robust and less computationally expensive. This includes creating pupil-dilation handling methodologies, obtaining sparse and dense 3D point-cloud models, and computing the 3D mesh that best adapts the point-cloud model.

- Re-training depth estimation CNNs using real and synthetic iris images, in order to estimate the depthmap representation of the iris surface. We also proposed a novel architecture, irisDepth, which combines the training scheme of T^2Net [15] with the sophisticated encoder-decoder of DenseDepth [20] to increase performance.

- Evaluating the performance of our 3D iris scanner in terms of spatial resolution by scanning objects of known dimensions.

- Comparing the performance of both SfM and CNN in 3D iris scanning.

- Developing preliminary methods that use iris 3D models for iris recognition and comparing their performances to that of the standard 2D iris code.

1.7. Thesis Organization and Papers Published

This doctoral thesis is organized according to "Format 2: New optional doctoral thesis format based on 2 accepted/published ISI journal papers." https://www.die.cl/sitio/proceso-de-doctorado/

Therefore, Chapter 1, the Introduction, contains the following sections: the research motivation, the problem statement, the general and specific objectives, the hypotheses, the contributions, the thesis organization and papers published.

As a result of this doctoral thesis two journal papers (WOS) were published and also one paper was published on an international conference. A fourth paper was submitted to an international conference. As Format 2 requires, the contents of Chapters 2 through 5 are the published papers in chronological order.

Chapter 2 [24] contains a published international conference paper, presented at the 2019 Winter Conference on Applications of Computer Vision (WACV 2019) in January 2019. In this research we developed the LFVL acquisition device, and we compared the iris-recognition performance of the standard NIR illumination against that of combinations of lateral and frontal VL. For that purpose, we acquired a dataset of 1,920 images from 96 subjects at the FCFM, Universidad de Chile. This research is the foundation of the rest of the thesis because it proves the efficacy of the proposed illumination method in texture enhancement and iris recognition. The device developed in this chapter was then used in all the following chapters to acquire iris images from several perspectives in a controlled illumination environment.

Chapter 3 [25] contains the first journal paper, published in IEEE Access. In this research, we developed a 3D iris scanning method based on SfM. The device developed in Chapter 2 allowed capturing sixteen-megapixel iris images, from different perspectives, and with enhanced texture. This texture increased the number of keypoints and 3D points this method can obtain from the iris. In this research, we produced 10 iris 3D models from 10 subjects, and compared one of them with the OCT scan of the same subject. We also measured the scanning resolution of the SfM method using 3D patterns of known dimensions.

Chapter 4 [31] contains the second journal paper, published in IEEE Access. It describes the development of a 3D iris scanning method based on CNN. Lateral illumination, produced by the device developed in Chapter 2, enhanced the shadows casted by iris features. This allowed CNN-based encoder-decoder architectures to relate shadows in RGB images with depth information. For this research we trained several state-of-the art networks to solve the iris-surface depth-prediction task. We also proposed our own architecture, irisDepth, to increase performance. In order to train those networks, we acquired both a real iris dataset of 26,520 images and a synthetic dataset of 72,000 images. In this research we evaluated the performance of the trained networks, we compared the 3D reconstruction performance of SfM and CNN using OCT scans, and we measured the resolution of the CNN method using 3D patterns of known dimensions. We also introduced a 3D iris recognition technique, as a proof of concept, with great accuracy in a test set of 480 iris 3D models.

Chapter 5 [32] contains a paper submitted for publication at the 2020 IEEE International Joint Conference on Biometrics (IJCB 2020). It proposes a new method to exploit the iris 3D models produced by SfM and CNN for iris recognition. This method uses 3D local descriptors [33][34] to find the best transformation that aligns two iris 3D models. Then we compute the similarity of the two models as a function of the transformation residual distance and the inlayer ratio. This method outperformed the preliminary technique presented in Chapter 4, the standard 2D iris code [8], and alternative 3D iris recognition methods [35][36]. This chapter illustrates the potential of 3D iris scanning as a biometrics technique.

Finally, Chapter 6 contains the general conclusions of this thesis, as well as the discussion and future work of 3D iris scanning.

Chapter 2

Iris Recognition: Comparing Visible-Light Lateral and Frontal Illumination to NIR Frontal Illumination

In most iris recognition systems, the texture of the iris image is either the result of the interaction between the iris and Near Infrared (NIR) light, or between the iris pigmentation and visible-light. The iris, however, is a three-dimensional organ, and the information contained on its relief is not being exploited completely. In this article, we present an image acquisition method that enhances viewing the structural information of the iris. Our method consists of adding lateral illumination to the visible light frontal illumination to capture the structural information of the muscle fibers of the iris on the resulting image. These resulting images contain highly textured patterns of the iris. To test our method, we collected a database of 1,920 iris images using both a conventional NIR device, and a custom-made device that illuminates the eye in lateral and frontal angles with visible-light (LFVL). Then, we compared the iris recognition performance of both devices by means of a Hamming distance distribution analysis among the corresponding binary iris codes. The ROC curves show that our method produced more separable distributions than those of the NIR device, and much better distribution than using frontal visible-light alone. Eliminating errors produced by images captured with different iris dilation (13 cases), the NIR produced inter-class and intra-class distributions that are completely separable as in the case of LFVL. This acquisition method could also be useful for 3D iris scanning.

2.1. Introduction

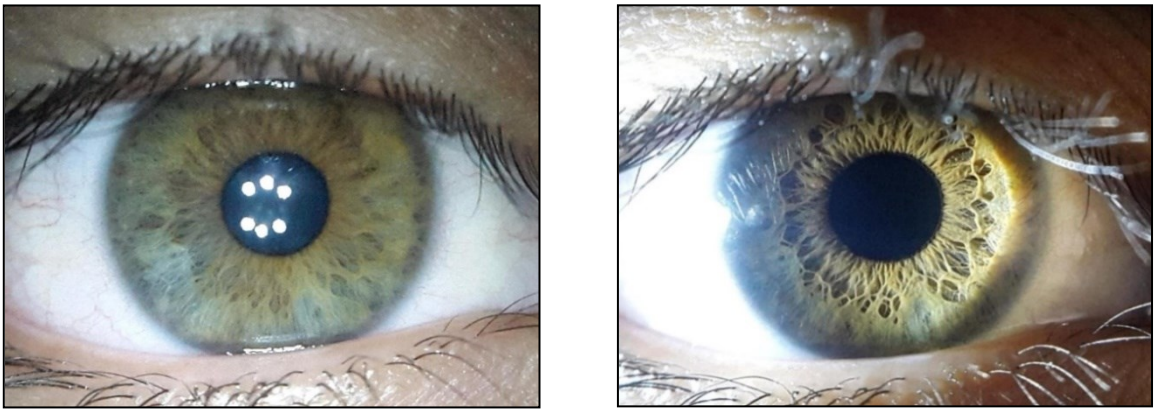
The human iris has a complex structure that varies from eye to eye [1][9]. The unique iris texture enables accurate recognition with very low false match rates in large databases under constrained scenarios [1][2][37][38]. Iris recognition systems have been used for several years in a variety of applications, and over one billion people have had their iris images electronically enrolled in various databases across the world [39]. The largest programs using iris recognition include the Unique IDentification Authority of India (UIDAI) program [6], the national ID program of Indonesia, and the US Department of Defense program [39]. Iris recognition systems have been used over a decade in the case of the Canadian border-crossing application and the UAE entry application [5].

The most widely used method for iris recognition is Daugman's approach with some variations [3][8][40][41]. The following are the main steps of this method. The acquisition process consists first of capturing a near infrared range (NIR) image of the eye. After that, the iris is segmented, leaving out eyelids, eyelashes, and reflections. The next step is transforming the circular iris to a rectangle of fixed dimensions so that it can be compared to other irises. This is achieved by mapping the Cartesian coordinates of the iris to polar coordinates [1][8]. This resulting normalized rectangular region is sometimes called an "unwrapped" or "rectangular" iris image [3][42][43] and is achieved by applying the Daugman rubber sheet model [8].

NIR illumination is preferred for iris recognition over visible-light illumination. The manner in which NIR light is absorbed or scattered by an organic tissue gives information about its structure [44]. Under NIR illumination, iris pictures highlight details that are not present under visible-light (VL) illumination [3][8][45]. This is especially true in dark color irises because melanin absorbs most wavelengths in the visible-light, and therefore the iris will appear dark [3][46]. Additionally, NIR light is non-invasive and non-intrusive; thus, it is comfortable for the user [8]. Visible light, on the other hand is intrusive, and can cause discomfort in the user, as well as pupil and eyelid reactions. That is why NIR imaging is the current standard in iris recognition. Previous attempts to use VL illumination in iris recognition include those of Proenca [47] who studied the feasibility of performing reliable biometric recognition with data acquired in unconstrained scenarios, and Hosseini et al. [45] who argued that NIR images lose pigment melanin features. They therefore developed a methodology to include these features in iris recognition. They extracted features from both NIR and VL images of the same individual and fused that information to improve recognition performance [45]. Abdullah et al. [48] compared the iris biometric performance of VL and NIR images of the same individual. They also proposed the possibility of cross-channel matches between the red channel of VL and the NIR images. Finally, there has been a growing interest in creating reliable iris recognition algorithms in a less restrictive environment. For this purpose, VL benchmarks such as UBIRIS [49], NICE [50], MICHE [51] have been created.

For both types of illumination, the texture of the iris image is the result of the interaction of light with the iris structure and pigmentation [45]. However, the spatial features of the iris relief are not completely exploited with frontal illumination. The iris has a sphincter and several radial dilator muscles, which control the size of the iris to adjust the amount of light entering the pupil [3][10][37]. All these structures produce a 3D relief in the surface of the iris which can be seen in an Optical Coherence Tomography (OCT) scan [11]. The depth information of this relief can also be captured with a 3D iris scanner [23]. A method that captures a more detailed structure in a 2D image of the iris would incorporate additional information into the iris recognition paradigm producing an even more accurate method.

In order to capture the structural information of the iris relief in a 2D digital image, we propose a method that consists of including more than one direction of the incoming light. The muscle fibers on the iris produce a relief that is comparable to a landscape with peaks and valleys. If a light source were to impact the surface of the iris from the side, the contrast between those peaks and valleys would be enhanced on the iris surface. For example, Figure 2.1 shows two images of the same eye. In Figure 2.1a, the eye is illuminated with frontal VL, while in Figure 2.1b, the eye is illuminated with lateral VL. Light distribution in Figure 2.1b produces a



(a) (b)

Figure 2.1: Comparison of high resolution images of the right eye of subject 55 in our database. Both images were captured with our device, under visible-light-white LED light, with no image pre-processing. (a) Only frontal illumination was applied. (b) Only lateral illumination was. The lateral LED is almost parallel to the iris plane.

richer texture in the iris image, enhancing the anatomical features of the iris. Similar results are found on both dark and light-colored irises, as shown in Figure 2.3.

In this article, we report a study of whether or not visible light iris images, under lateral and frontal illumination, contain sufficient texture information for iris recognition. We compare this illumination to the standard NIR frontal illumination and to frontal VL illumination. For this purpose, we captured a dataset of 1,920 images using both a conventional NIR frontal-light iris camera, and a custom VL device which illuminates the iris with white-light LEDs from the front and from the side. To differentiate the effects of the frontal LEDs from the lateral one, we acquired three sets of images with our device: lateral illumination only, frontal illumination only, and a combination of lateral and frontal illumination. Finally, we compared iris recognition performance under the four lighting conditions to determine if the results of the visible-light cases are comparable to those of the NIR lightning condition in terms of iris recognition performance. The analysis is based on the Hamming distance distribution of iris-code comparisons.

2.2. Methodology

2.2.1. Image Acquisition

The iris images for this study were acquired from the right eyes of 96 volunteers using two devices. The first device is the binocular camera: Irishield BK-2121-U of Iritech Inc. This device illuminates the iris from the front with NIR light at a distance of 15 cm. The NIR images of the iris have a resolution of 420×315 pixels. The second device was developed by us and is similar to that used in [23]. The device illuminates the right eye with white LED light from both the front and the side using a custom-made frame and printed circuits. This device uses the main camera of a Samsung S5 cellphone with a macro lens to capture 16 Mpx VL images at a distance of 4 cm. However, in order to create a fair comparison, those images are cropped and resized to the same resolution as the first device.

Five images were captured from each subject for each lighting condition. Therefore, each condition has a total of 480 images, and the database contains 1,920 iris images. All pictures were taken indoors, and the time between consecutive images of the same illumination type was less than two seconds. Using both devices, four different lighting conditions were explored. The first condition was given by the Iritech device while the other three are from our VL device with white illumination as follows:

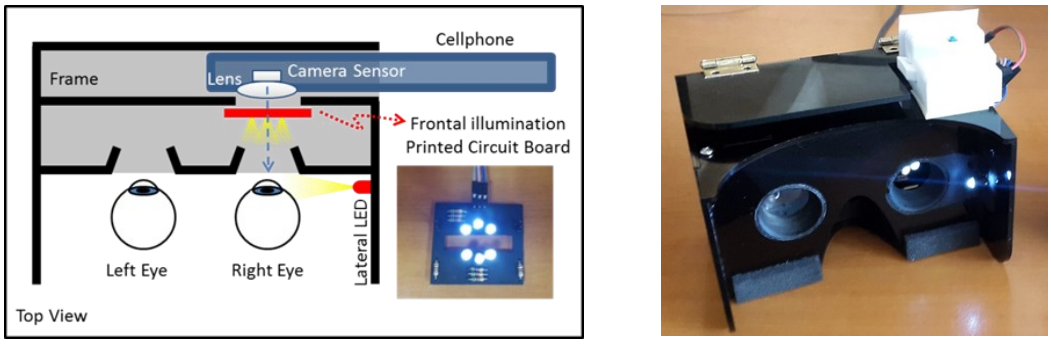
- Frontal NIR illumination (FNIR).
- Frontal Visible-Light (FVL).
- Lateral Visible-Light (LVL).
- Lateral + Frontal Visible-Light (LFVL).

In the database there are images of 39 brown irises, 40 dark brown irises, 11 green irises, and 6 greenish-brown irises collected from 75 males and 21 females. The average age of the subjects is 24.0 (STD=3.95).

2.2.2. Device Description

Our device is based on the Google-Cardboard VR [52], and it is inspired by the designs of Bastias [23] and Mariakakis [53]. The frame was laser cut in acrylic, and the inner surfaces were painted opaque black to avoid reflections. A sketch of the inner chamber of the device is presented in Figure 2.2a. The device allows the camera to capture close range images of the right eye in a controlled illumination environment. A picture of the device is shown in Figure 2.2b.

The illumination setup consists of an array of six white-light LEDs for frontal illumination, one white-light LED for lateral illumination, and an illumination controller. The frontal LEDs are placed in a PCB with a hole in the middle, as proposed by Bastias et al. [23]. This PCB is carefully constructed so that the specular reflections of these lights are mapped in the pupil region of the image. Light rays from the lateral LED are almost parallel with the iris plane to maximize the length of the shadows they produce. The device additionally has a PWM voltage regulator with two knobs for adjusting the amount of frontal and lateral illumination individually. This component is the white box on the top of the device shown in Figure 2.2b.



(a) (b)

Figure 2.2: (a) Top view of the VL device schematics and a picture of the frontal illumination PCB. Images of the right eye are acquired with a 16Mpx Cellphone camera with the help of an external lens. The distance between the camera and the eye is approximately 4cm depending on the subject. The frontal illumination circuit and the lateral illumination LED can be controlled independently. (b) Picture of the visible-light device.

2.2.3. Iris Image Preprocessing

The preprocessing stage of the database consisted of the following steps in cascade:

Lens Distortion Correction. The visible-light images were captured using a macro lens and the lens of a cell phone. The captured images have distortion that can be observed on straight lines that show slight curvatures in the image. The lens distortion was corrected using the Matlab App [54] of the Camera Calibration process [28][29]. The process consists of acquiring a set greater than 10 images of a checkerboard pattern in different locations in front of the camera [54]. The camera was placed in front of the pattern at the same distance as the iris. The algorithm then finds the positions of the corners on the 10 checkerboard images and computes the amount of radial and tangential distortion of the lens, as well as the central point. With this information, lens distortion correction is performed on the iris images.

Cropping and Resizing. The NIR images were neither cropped nor resized, as they already complied with NIST standards [55]. In contrast, the visible-light images were adjusted to the same resolution as the NIR images. First, the center (Ci) and the radius (Ri) of the irises were localized in the 16Mps images using the Matlab implementation of Masek's segmentation algorithm [56]. This algorithm is explained in more detail in section 2.4. Then, the images were cropped, around Ci, using NIST specifications [55], i.e., the iris is centered in the image, a margin of 0.2×Ri is provided on the vertical axis, and a margin of 0.6×Ri is provided on the horizontal axis. The resulting image has an aspect ratio of 4:3 with a dimension that depends on the size of the iris in the image. Finally, the images were resized to a fixed resolution of 420x315 pixels, which is the resolution of the NIR images.

Contrast Enhancement. In order to maximize the dynamic range of the brightness in the iris region, contrast enhancement was applied to all the images in the database. The contrast enhancement was performed using a local Laplacian filter as described in [30]. This filter makes use of the different levels of the Laplacian pyramid to produce edge aware detail enhancement [30]. Only neighboring pixels are therefore involved in the process, and removing highlights or the pupil is not required in order to enhance the details of the iris as would be necessary in a histogram equalization method [57].

2.2.4. Iris Recognition

Iris recognition has been noted to be very successful when a compact and robust representation of the iris information, called the Iris Code [3][8], is used. The Iris Code comparisons are robust to most changes in brightness, size, dilation, and rotation [3][8]. The generation of the iris code requires the steps of pupil-and-iris segmentation, polar-coordinate normalization, and Gabor-filter application [3]. In order to use this robust method, we employed the publicly-available Matlab implementation by Libor Masek [56]. Although it is a week implementation by modern standards, Masek's algorithm has been widely used [3][40][58]. The following is a description of each step in the algorithm.

Segmentation. The segmentation process consists of identifying the iris, pupil, eyelids, eyelashes, and reflections in the image [3][56]. To find the pupil/iris and the iris/sclera boundaries, Masek first uses a Canny edge detector [56][59]. After that, a circular Hough transformation is applied to detect the circular shapes, as well as their centers and radii [37][56][60]. In this way, the pupil and iris are found, and an annular mask that highlights the iris is created. After that, a linear Hough transformation is used to identify eyelids. The pixels of the mask that are above and below the upper and lower eyelids are removed [56]. Finally, highlights are also removed using a simple thresholding operation. A mask with the position of the iris pixels in the image is generated in this manner [56].

Normalization. Masek uses Daugman´s Rubber Sheet model for iris normalization [8][56]. This method consists of mapping all the points in the annular region of the iris (x,y) onto a rectangle of fixed dimensions, through polar coordinates (r,θ) [8]. The implementation also works when the pupil and iris circles are not concentric [56]. The center of the pupil is considered to be the origin, and the angle θ is swiped in a constant interval. The segment between the pupil/iris boundary and the iris/sclera boundary is evenly sampled for each value of θ [8][3][56]. All the samples are placed in the normalized or rectangular image, where the horizontal axis corresponds to the angle θ, and the vertical axis represents the radius r [8]. As is commonly done, we used 20 samples in the radial direction while the circumference was divided into 240 parts. The normalized image, therefore, has a resolution of 240×20 pixels. The mask from the segmentation process was normalized with the same procedure. Thus, rectangular versions of the iris and its mask are the outputs at the end of the normalization stage.

Iris Code. The iris code is a binary representation of the information in the iris. Daugman proposed filtering the image with sinusoidal signals modulated by Gaussians in order to capture both spatial and spectral information from each pixel in the normalized iris [8]. Oppenheim demonstrated that phase information is more robust than magnitude information [56][61]. Therefore, an even-symmetric filter is used to compute the real part, and an odd-symmetric filter is used to calculate the imaginary part [8]. Then, the phase information is encoded in two bits as the quadrant of the resulting complex number [8][9]. Each pixel in the normalized image therefore contributes two bits to the iris code [56]. As a result, the iris codes in our database have a dimension of 480x20. In the Matlab implementation, Masek uses Log-Gabor filters to account for the DC component that the even-symmetric Gabor filter has [56]. Finally, the segmentation mask is also adjusted to the resolution of 480×20 pixels.

2.2.5. Data Base Analysis

Once the iris codes and masks were computed for all the iris images of the database, each iris code was compared with all the others. The comparisons were carried out by means of a modified Hamming Distance (HD) which scores how similar two iris codes are [8][56]. This comparison procedure uses the mask to remove the eyelids and reflections from the final score [8]. The modified Hamming distance is computed in (2.1):

$$\mathrm{HD} = \frac{\|(\mathrm{codeA} \otimes \mathrm{codeB}) \cap \mathrm{maskA} \cap \mathrm{maskB}\|}{\|\mathrm{maskA} \cap \mathrm{maskB}\|} \text{ [8]}, \tag{2.1}$$

where codeA and codeB are two iris codes being compared, maskA and maskB are their respective masks, $\otimes$ is a bitwise XOR operator, $\cap$ is a bitwise AND operator, and the norm ($\|\quad\|$) is the number of logical ones [8].

In order to make the iris-code comparisons robust to rotation, (2.1) is applied several times for different translations in the θ direction of one of the iris codes [8][56]. Then, only the smallest HD is stored as the comparison between the two codes. In this way, the HD is evaluated in only the best relative rotations between the two irises. In our study, we tested variations of ±5°. Using (2.1), each of the 480 iris codes was compared with the others of the same illumination condition. A total of 480×479/2 = 114,960 comparisons were made per lighting condition with neither repeated comparisons, nor comparisons of a code against itself. Therefore, a total of 114,960*4 = 459,840 iris code comparisons were made in this study.

Then, the comparisons were classified as intra-class, or "Match", when they came from the same subject, or inter-class, or "Non-Match", when they came from different subjects. Since 5 images were taken per iris, the number of comparisons per subject was 10. As there were 96 subjects, the number of intra-class comparisons was 960 for each lighting condition. There were, therefore, 114,960 - 960 = 114,000 inter-class comparisons per lighting condition. The histograms of the Hamming distance distribution of each lighting condition are presented in Figure 2.4 in the Results section.

In order to assess the results, we studied how separable the two classes are. Daugman uses a method that accounts for the distances between means (*d*) and the shapes of the distributions [8]. The decidability index (d') is computed below in (2.2):

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{0.5*(\sigma_1{}^2 + \sigma_2{}^2)}} \text{ [8]}, \tag{2.2}$$

where μ_1 and μ_2 are the mean HD values and σ_1 and σ_2 are standard deviations of the two classes. For both d and *d'*, the larger the value, the greater the distance between the two distributions.

2.2.6. FNIR and LFVL Iris Codes Comparison

One of the hypotheses of this study is that, compared to NIR illumination, our method collects different information from the iris structure. A test that supports this hypothesis is to compare iris recognition with FNIR and LFVL cases. If both groups of images had the same kind of information, we would expect iris codes from the same individual to be similar. Thus, the intra and inter class distributions would be separable. However, if FNIR and LFVL extract different information from the same iris, then the iris codes would be different, and those distributions would not be easily separable. The following experiment was conducted on our database. First, we computed the HD for each FNIR image with all the LFVL images. Then we computed the HD for each LFVL image with all NIR images. Finally, we selected the HD that corresponds to comparisons of the same person (intra-class or match) in one vector, and all the rest in another vector (inter-class or non-match). Both histograms with the distributions are shown in Figure 2.6.

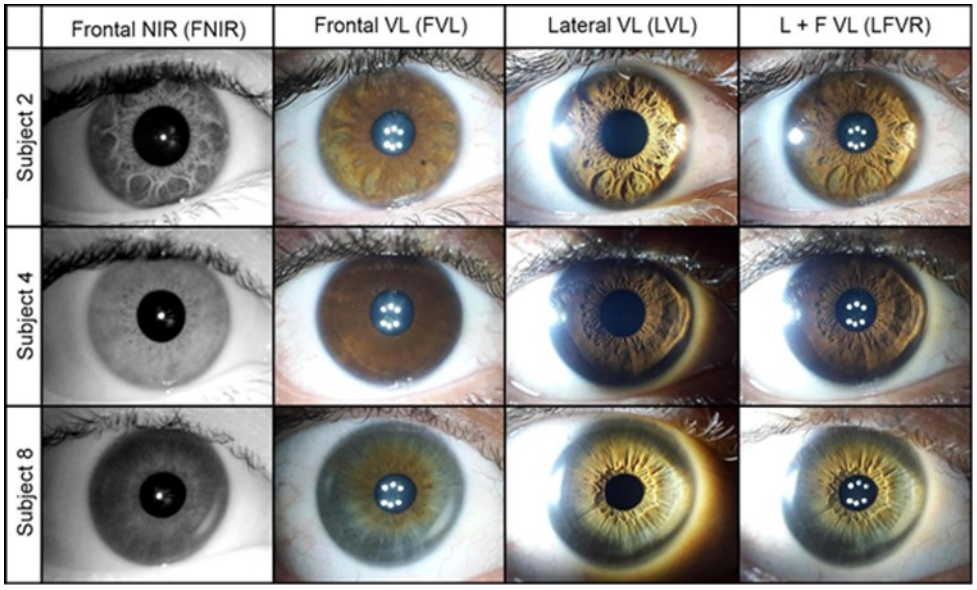


Figure 2.3: Iris images from the database without contrast enhancement. Images in the first column were captured with an Irishield BK-2121-U camera under NIR light (FNIR). The second column shows frontal visible- light (FVL) images captured with our device. The third column shows lateral visible-light (LVL) images, and the fourth column shows images captures with a combination of lateral and frontal visible-light (LFVL).

2.3. Results

The database contains four sets of 480 images for each of the four illumination cases (FNIR, FVL, LVL, LFVL). Figure 2.3 shows one of the five images of the same iris in the four data sets. In general, NIR images show a similar shape to those of visible-light images; however, different types of details appear in the texture of each image [1][45][60]. The intensity of the pixels in the frontally illuminated images depends on the amount of light that is reflected to the camera. Bright spots reflect more light, while dark spots absorb more light [1][8][45]. In contrast, images under lateral illumination reveal details that come from the 3D surface of the iris. Dark spots are shadows of the peaks and valleys on the surface of the iris showing the actual structure in more detail.

Observing Figure 2.3 verifies that images under NIR illumination show more texture than those created under frontal white illumination, as is described in the literature [1][8][45]. This is especially true for dark color irises as is shown in the second row of Figure 2.3. For example, the iris of subject 4 shows texture in the FNIR image (first column) while that iris appears to have almost constant intensity in the FVL image (second column). Additionally, comparing the third (LVL) and fourth (LFVL) columns, the iris region for each subject appears to be very similar. However, a strong gradient under lateral illumination can be observed on the sclera. Nevertheless, with a combination of lateral and frontal illumination (LFVL), this gradient is less intensive, which is important for the segmentation process.

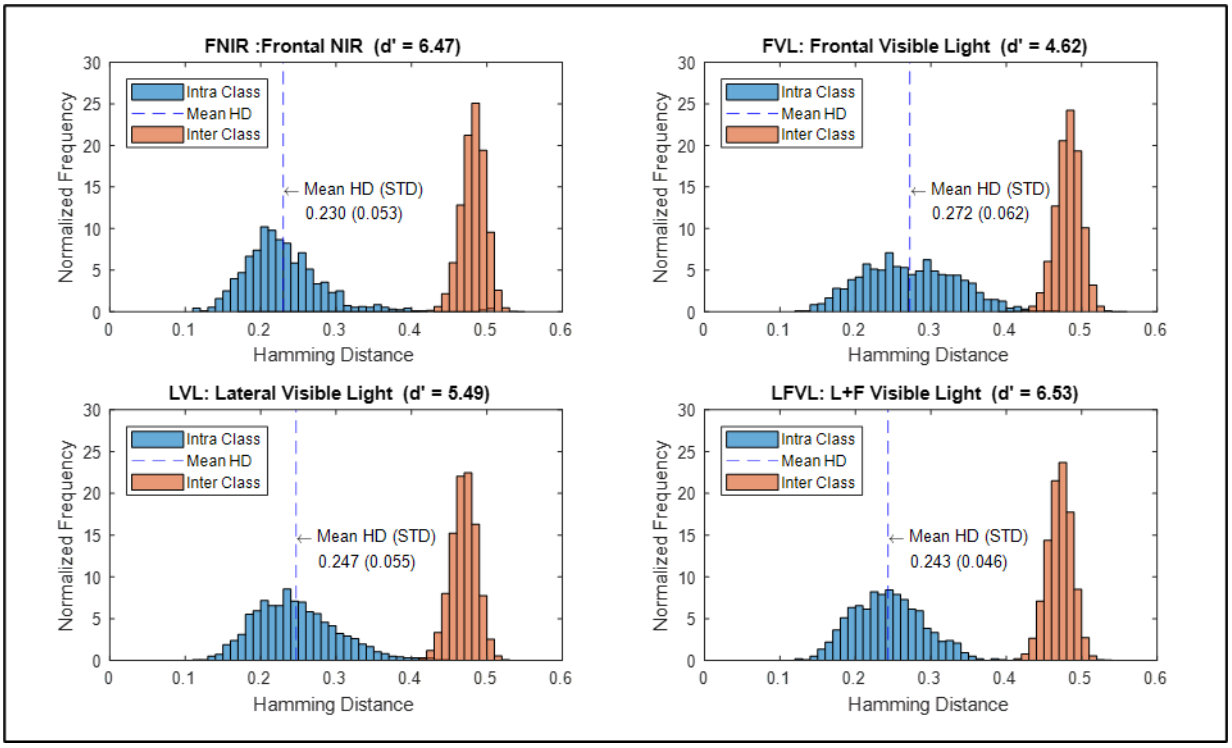


Figure 2.4: Hamming Distance distributions of intra-class (match) and inter-class (non-match) of iris code comparisons for our database with 1,920 irises. The intra-class histogram is constructed from 960 comparisons while the inter-class histogram was made from 114,000. In order to account for the difference in the number of comparisons, the PDFs were normalized to have an area of 1.

Figure 2.4 shows the distributions of the intra-class and inter-class comparisons of the iris code for the four different illumination conditions. The histograms in Figure 2.4 were normalized so that their PDFs have an area of 1. The main parameters for the distributions are shown in Table 2.1. The *d'* value was computed using (2.2). The shapes of the inter-class distributions are similar in all the illuminations, and their mean values are all about 0.475. In contrast, the shapes and means of the intra-class distributions differ from case to case.

Table 2.1: Parameters of the Iris Code comparison distributions of Figure 2.4.

Method	Intra-Class Mean HD (STD)	Inter-Class Mean HD (STD)	*d'*
FNIR	0.230 (0.053)	0.482 (0.016)	6.47
FVL	0.272 (0.062)	0.482 (0.017)	4.62
LVL	0.247 (0.055)	0.469 (0.017)	5.49
LFVL	0.243 (0.046)	0.471 (0.017)	6.53

The intra-class distribution for the FVL illumination appears to be more flat than that of the FNIR illumination. This is also indicated by the fact that FVL STD is 18% larger than that of NIR. This contributes to the overlapping of the intra and inter-class distributions in the FVL case.

Additionally, the decidability index, *d'*, for the FVL is smaller than that of the FNIR. As a consequence, the performance of FVL is worse than that of NIR, which confirms what has been

described in the literature with respect to FNIR light being much more suitable for iris recognition than frontal VL illumination [3][8][9][37].

NIR, LVL, and LFVL cases have more compressed distributions, and their mean values are similar. However, LVL is skewed to the left, and its intra and inter-class distributions have mode overlap. Therefore, the best performing cases are FNIR and FLVL. Despite the fact that NIR has the smallest mean HD, the decidability d' index of LFVL is larger than that of FNIR. This indicates that the distributions of the LFVL case are easier to separate. Therefore, the results for the combination of lateral and frontal illumination are better than those of NIR in this database.

Among the VL distributions, the lateral illumination cases have a larger *d'* index than that of the FVL. In fact, FVL had the worst performance of all the methods. Therefore, it is possible to conclude that the lateral LED is responsible for the good performance of the LFVL method.

An analysis that focuses on the small discrepancies of the separability of the distributions is the Receiver Operator Characteristics (ROC) curve [3]. All possible thresholds that could be used to separate the inter-class and intra-class distributions are applied, and the performance of each threshold is computed. In this study, we evaluated the performance in terms of the True Positive Rate (TPR or Sensitivity), and the False Positive Rate (FPR or 1-Speciphicity). The ROC curves are presented in Figure 2.5(a). The ROC curves in Figure 2.5(a) reveal that the LFVL distributions are completely separable, since this case can have 100% sensitivity and 100% specificity. The other VL cases also have good ROC curves. The FNIR case had the worst performance according to this test, but its score is still high with 99.4% sensitivity and 99.9% specificity with a threshold of 0.425. A closer look at the dataset reveals that the false positive cases of FNIR belonged to 13 subjects with a high discrepancy in iris dilation levels among captured images. Therefore, a decrease in performance is expected from those cases [10]. When those cases are removed from the database, the ROC curve of FNIR is identical to the LFVL case, as shown in Figure 2.5(b).

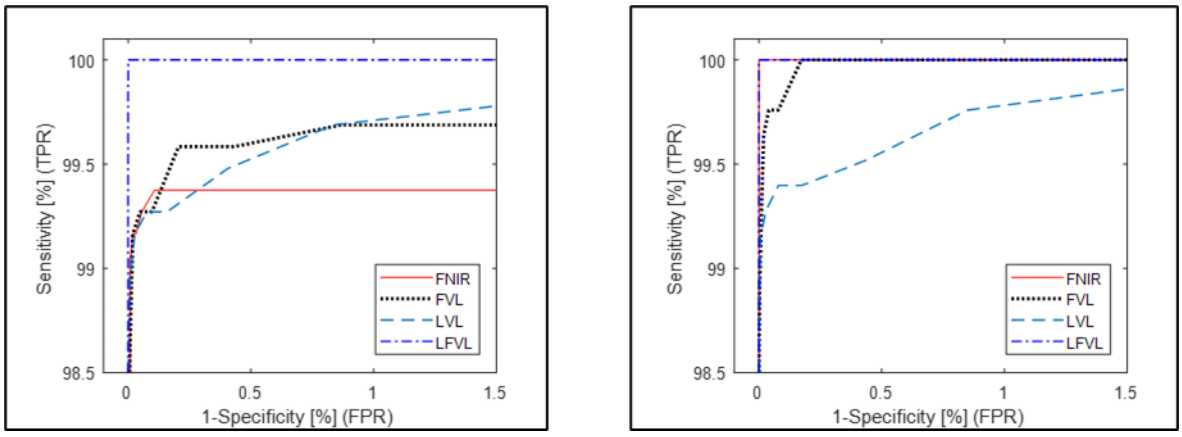


Figure 2.5: ROC curve of the different methods. (a) Using all the 96 subjects of the database. (b) Discarding 13 subjects with large dilation differences between captured images.

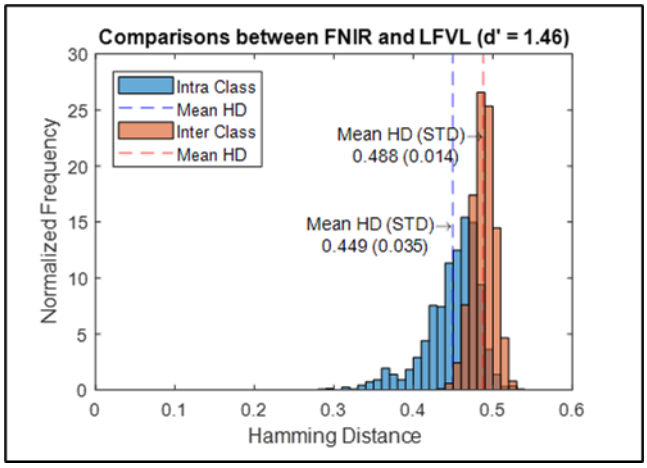


Figure 2.6: The iris codes of FNIR were compared against those of FLVL, and vice versa. An important overlap in the distributions indicates a small similarity between images of the same eye under different illumination techniques.

The results of the FNIR and LFVL iris code comparisons are presented in Figure 2.6. The comparisons excluded the 13 subjects with dilation problems. Figure 2.6 shows that the intra and inter class distributions are closely entangled. The decidability index d' is 1.46, which is less than 1/3 of the worst case (FVR) in the previous test. This means that iris images from the same individual under FNIR and LFVL are sufficiently different to be scored as impostor. This is compatible with the hypothesis that both methods capture different information from the same iris.

2.4. Discussion and Conclusions

The iris is a three-dimensional organ, and the information contained on its 3D surface is not being exploited completely with current methods. A new method for improving iris information extraction based on visible-light illumination by adding lateral illumination to the frontal illumination is proposed here. The proposed method enhances the visibility of the structural information of the iris. The method could be used to improve iris recognition, as well as in novel applications to generate 3D models of the iris for possible use in characterizing acute angle glaucoma [23], a disease that currently uses OCT [26] to measure iris thickness. The method was tested with a database of 1,920 irises captured from 96 subjects. Four illumination types were compared: FNIR, FVL, LVL, and LFVL.

The results based on the intra-class and inter-class distributions, as well as those based on the ROC curves, show that LFVL illumination obtained the best decidability together with similar results achieved with NIR illumination. The results confirm that VL iris images using frontal and lateral illumination are suitable for iris recognition and other applications. The improved performance of the LFVL illumination is probably due to the rich texture that lateral illumination reveals. Lateral illumination enhances the contrast on the muscular fibers in the iris 3D surface thus producing a richer texture. As expected, Figure 2.3 shows that FVL illumination of the iris results in poorly textured images for dark irises. This may explain the poorer results in iris recognition reported in the literature with this type of illumination. It can therefore be concluded

that adding lateral visible-light illumination is a key factor in improving the performance of iris recognition when VL illumination is used.

Additionally, this new method could be useful in 3D iris reconstruction methodology [23] since richer textures from the iris are produced. Richer textures increase the number of keypoints that can be extracted from the iris, which in turn produces more accuracy and better reconstruction of 3D iris models [12]. Since LFVL images highlight the structure of the iris, and they can be captured at 16M px, a photogrammetry algorithm should be able to reconstruct a 3D model with enhanced spatial resolution [12][23].

The results of the NIR and LFVL iris code comparisons indicated that iris images from the same individual under NIR and LFVL are sufficiently different to be scored as impostor. This is compatible with the hypothesis that both methods capture different information from the same iris.

An aspect that can be improved in the future is to reduce the specular reflection on the left side of LVL and LFVL images. This reflection is a result of the lateral LED. Frontal reflections fall into the pupil area and therefore do not produce artifacts.

Convolutional Neural Networks (CNNs) have been applied successfully in many areas of computer vision. In iris recognition several applications have been proposed [62][63][64][65]. We performed a preliminary test using CNNs for feature extraction in our database. First, we obtained a 112×448 rubber sheet after Masek´s segmentation [58]. Then, the rubber sheet is split into two images of 112×224, and concatenated vertically to form one image of 224×224, as in DeepIrisNet [64]. The VGG-19 CNN was applied to obtain a feature vector [65][66]. The 4096-feature vector consists of the output of the penultimate layer of VGG-19 [66]. The CNN results of comparing distributions of intra-class (match) and inter-class (non-match) of iris code comparisons coincide with our results of this article, i.e., the lateral illumination improves the decidability index. The CNN results show the existence of significantly more overlap in the distributions than those of our approach. The decidability index reached a maximum of 3.65 for the CNN applied to the rubber sheet model, which is much lower than those shown in Table 2.1 for our proposed method.

Another line of research will be followed for the CNN approach modifying the inputs to the CNN and using pretrained CNNs with iris images. Future work will also include extending testing to investigate whether NIR lateral illumination improves iris recognition performance.

2.5. Acknowledgements

This research has been funded by CONICYT through project FONDECYT 1161034 and by the Department of Electrical Engineering, Universidad de Chile. We would like to acknowledge the valuable contributions of Dr. Juan Tapia in the application of the iris segmentation, codification, and recognition method. We would also like to thank the students at the School of Engineering, Universidad de Chile, who participated enthusiastically as volunteers for iris image acquisition.

Chapter 3

A 3D Iris Scanner from Multiple 2D Visible Light Images

The development of new methods for biometrics using the 3D surface of the iris could be useful in various applications, such as reliable identity verification of people when only segments of the iris are available, the study of how the iris code changes with pupil dilation, and studying acute angle glaucoma and its relation to the 3D iris structure. The goal of this research was to build a 3D model of the iris surface from several 2D iris images, adding depth information to the iris model. We developed a 3D iris scanner which reconstructs a 3D mesh model of the iris surface from several 2D visible light images. First, a smartphone camera captures visible-light iris images from different angles in a controlled illumination environment. Then, a Structure-from-Motion algorithm reconstructs a point-cloud 3D model. Finally, the best-fitting 3D-mesh model is obtained using the Screened Poisson Surface Reconstruction technique. Our results include the reconstruction of the 3D iris models of seven subjects. These models contain an average of 11,000 3D points. The spatial resolution of our device was measured as 11 μm by scanning a 3D pattern of known dimensions. The 3D model of the iris is compared with the results from an Optical Tomography (OCT) performed on one iris. Our results show that our new 3D iris scanning method produces a model with potential applications in biometrics and ophthalmology.

3.1. Introduction

The need for accurate identity verification systems has driven the research on iris recognition techniques. These techniques exploit the texture of the human iris because it exhibits a distinctive pattern with great variability among individuals [1][2]. Iris recognition is currently the most reliable biometric technology on the market because of its non-invasive, accurate, and robust methodology [3][4]. Furthermore, iris recognition has succeeded in both small- and large-scale applications. Some large-scale uses include United Arab Emirates border-crossing [5] and India's Unique ID program [6]. Iris recognition was also used to create a voter registration list in Somaliland [7].

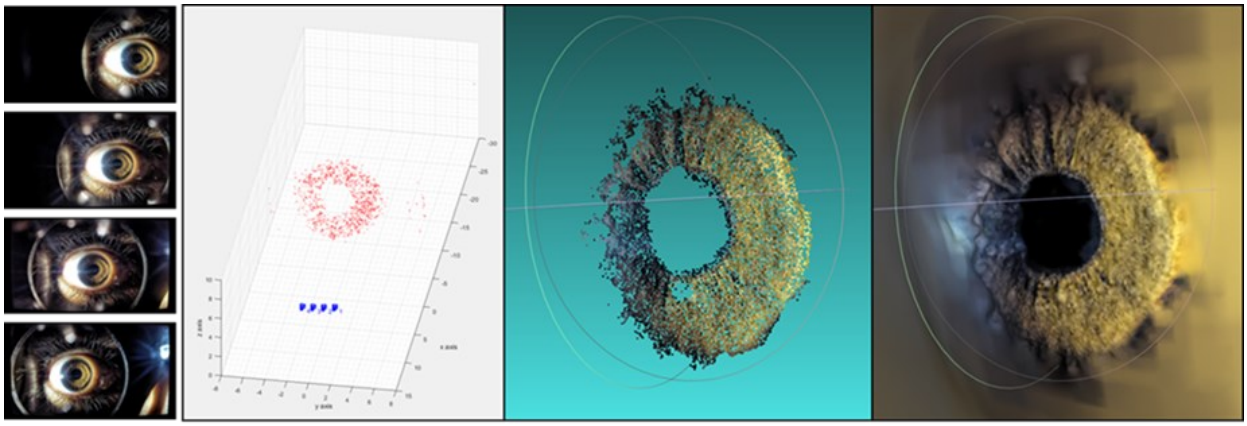
One of the main reasons for the success of iris recognition technology is its high accuracy. Daugman demonstrated that a false match rate of 1 in 4 million could be achieved when using a fractional Hamming Distance (HD) of 0.33 [1][8]. The variability in iris texture as analyzed by Daugman's method is as great for left and right eyes of the same person, and between identical twins, as it is for irises of unrelated persons [1][9]. By contrast, the variability in fingerprint

analysis would produce inadmissible false matches in large databases [2], which is why India, for example, uses iris recognition as well as fingerprints for its Unique ID program [6].

Although iris recognition is a well-developed technique with successful applications, there are still areas for improvement. First, iris recognition could encounter the same issue as fingerprint analysis if the database were large enough. For instance, in a database of one billion individuals, a false match rate of 1 in 4 million indicates that 250 people have similar iris codes. In addition, it is normal to have partial occlusions of the iris due to eyelashes, eyelids, and specular reflections. When there is less iris available for matching, the accuracy that can be achieved is lower. Moreover, iris recognition techniques assume that the iris is flat, when, in fact, it is a three-dimensional organic structure. The human iris dilates and contracts using two muscle systems, a sphincter and several radial dilator muscles, in order to control the amount of light that hits the retina [10]. These muscular systems cause the surface of the iris to have a 3D relief that can be visualized in Optical Coherence Tomography (OCT) [11]. However, when this complex 3D structure is mapped onto a 2D plane, all depth information is lost. Recapturing depth information could, in principle, enhance the amount of information we can extract from the iris. Our 3D scanner allows studying, in detail, the impact of pupil dilation on Daugman's rubber sheet model of iris deformation. The results of this research will allow us to measure the displacement of the iris structure as a function of the pupil dilation in 3D, and to relate the iris displacements to the rubber sheet model used for identification. This, in turn, could lead to the formulation of more accurate models for iris normalization, even with extreme dilation differences [27].

Previous work in this area includes the following. Bastias et al. [23] reconstructed a 3D model of the iris from several NIR images in order to capture the depth information of the iris surface. A raspberry-pi V2 camera was used to capture NIR iris images along a circular arch [23]. Then, a photogrammetry algorithm was used to reconstruct a 3D model, thus creating a new method for human iris analysis [23]. Issues with that method include the noisy nature of the camera, its resolution, and the plane correction needed to integrate all the images [23]. In recent work, Benalcazar et al. [24] used lateral visible-light (VL) illumination to capture highly textured images of the iris at a resolution of 16Mpx. Lateral illumination allowed VL images to capture structural details of the iris [24]. In fact, a combination of lateral and frontal illumination showed slightly superior iris-recognition performance than that of frontal NIR illumination [24]. Our work builds upon these previous studies to create a robust and reliable method of 3D iris scanning that is not only useful for biometrics, but also is promising as a less expensive screening tool for Acute Angle Glaucoma, a disease currently diagnosed with an OCT [23][24][26].

The proposed 3D iris scanning methodology consists of four main stages, as is illustrated in Figure 3.1. First, VL images of the iris are captured from different perspectives, as shown in Figure 3.1a. Then, we use the Structure from Motion (SfM) [13] to jointly estimate the camera pose of every image and a Sparse 3D model of the iris. Our method applies several constraints to to adapt the SfM algorithm specifically for the iris. Figure 3.1b shows camera positions in blue, and the initial point-cloud model of the iris in red. Then, a dense 3D reconstruction is performed by extracting more keypoints from each image, as shown in Figure 3.1c. Finally, the point-cloud model is converted into a mesh surface by the Screen Poisson Surface Reconstruction technique [67]. This method produces a mesh representation of the iris surface that integrates the depth information of a large number of points on the iris. Figure 3.1d shows the details that are captured



(a) (b) (c) (d)

Figure 3.1: Process of 3D Iris Reconstruction: (a) Image acquisition from different views. (b) Camera pose estimation (blue) and sparse 3D reconstruction (red). (c) Dense 3D point-cloud reconstruction. (d) Mesh surface reconstruction.

from the iris in the final mesh representation. With the additional depth information, extra degrees of freedom can allow iris recognition to be applied in even larger databases than is currently possible, thus extending the scope of iris recognition.

The following are the main contributions of this study: First, we improve the scanning device so that iris images present better resolution, less noise, and additional color information. Second, we improve the preprocessing stage by adding lens-distortion correction [28][29] and local Laplacian contrast enhancement [30]. Third, we adapt the SfM pipeline for this specific application to make the system more robust and less computationally expensive. Fourth, we reconstructed 3D iris models for 10 subjects. Fifth, we evaluate the performance of our 3D iris scanner in terms of spatial resolution by scanning objects of known dimension depths. Sixth, we compared our 3D reconstruction with the OCT of one iris. The OCT is a paid medical exam, and therefore, it was performed for just one volunteer to provide a qualitative comparison with our method, to show the promise of our approach as an alternative to OCT.

3.2. Review on 3D Reconstruction

Methods to produce a detailed reconstruction of objects or scenes from images and video have improved significantly over time [12]. Application of these technologies varies from object scanning to robotic vision and city mapping. New algorithms, as well as increasing CPU and GPU power, allow processing thousands of images in a linear-time incremental manner [13]. For example, a collection of internet images has been used to reconstruct buildings and even city blocks [17]. Other research has focused on real time implementations for environment mapping and robotic navigation [68][69]. In this section, we present a brief survey of 3D reconstruction methodologies that have inspired our pipeline.

Methods for 3D representation of the environment from digital images can be grouped into two areas [12]. Both obtain 3D points from keypoints in images from various perspectives; thus, these methods need several highly textured images to produce an optimal reconstruction [12][70]. In the first group is the Visual Simultaneous Localization and Mapping (SLAM), which

processes image frames, from video or real time, one by one [68][71]. This method aims to reduce the uncertainty of the measurements with each new observation, and incorporates new 3D points into the model from each new image [68][69][71]. In the second group the Multi-View Stereo (MVS) processes all the unsorted images in a batch [13][18]. MVS treats 3D reconstruction as an optimization problem: given a group of images, find the camera poses and 3D model that minimize the total re-projection error [12]. The re-projection error is the distance between a keypoint and the pixel where the associated 3D point is projected back onto the image plane [12].

Visual SLAM was conceived as a method for autonomous robotic navigation without the need for sophisticated sensors such as LIDAR [68][69][71]. This method utilizes a Kalman filter to jointly estimate the camera trajectory and the position of landmarks in the environment [68][71]. Visual Slam is an active area of research where state-of-the-art implementations can run in real time. For instance, the ORB-SLAM system utilizes ORB features, which are fast and robust, to generate a map of the environment for robotic navigation [69]. ORB-SLAM has a robust mechanism that stores only important frames, manages automatic model initialization, and detects traversing on the same path for loop closing [69]. Fang et al. [72] also implemented Visual SLAM with ORB features, but with an FPGA architecture to reduce power consumption. The PL-SLAM system extracts point and line features to work even with low textured images [73]. The Visual SLAM paradigm can also be used for augmented reality systems in real time [74].

MVS estimates all the camera poses and an initial sparse 3D model using a process known as Structure from Motion [12][17]. SfM uses epipolar geometry to find the relative spatial motion, in terms of the Essential Matrix, between two cameras observing the same scene [13]. The Essential Matrix represents the transformation (rotation and translation) of coordinate systems from one camera to the next [70]. If one camera is fixed, the absolute position and orientation of the other cameras can be found [16][70]. After finding all camera poses, the coordinates of the 3D points in the model are triangulated [14][70]. SfM has a final refinement step which is called Bundle Adjustment [75]. This is the step where the camera poses and 3D model points are jointly optimized to minimize the re-projection error. The re-projection error can be lowered even to sub-pixel values in this step [12].

MVS builds upon the output of the SfM process to obtain a more detailed model [18]. This model is called a dense reconstruction, and can be represented in different formats: point-cloud [13][18], depth-map [76], voxel [77], and deformable polygonal mesh [78]. Furukawa et al. proposed a versatile point-cloud representation, in which each 3D point is represented by a small rectangular oriented plane [18]. This MVS implementation is based on a repeated process of match, expand, and filter [18]. This process incorporates new patches, expands them filling uncharacterized areas, and then trims the patches that result in a faulty geometry [18]. At the end, this implementation is capable of finding detail even in low texture regions from non-keypoint pixels [18]. The best known implementation of MVS and SfM is that of Agarwal et al.: “Building Rome in a day” [17]. They utilized 150,000 assorted images of Rome from the internet, and produced a point-cloud model of the city, which includes the most famous buildings [17]. Melow et al. improved the MVS algorithm by identifying and removing shading and reflectance from the

model [79]. Lim et al. adapted the MVS paradigm to scan the digestive system of human beings [80]. Their endoscope system proves that MVS can work in biomedical applications [80].

3.3. Methods

3.3.1. Iris Imaging Device

The first step towards reconstructing a 3D model of the iris is capturing sharp well-illuminated images of the iris from different viewpoints. These images need to have high texture and high resolution in order to obtain as many keypoints as possible. Additionally, as we observed in our previous research [23], those images need to have a consistent pupil dilation level; thus, a controlled illumination environment is also needed. Finally, all the components of the device have to be mounted on an ergonomic structure that minimizes motion related artifacts. With all of these considerations in mind, the proposed image acquisition device consists of the following components: a cellphone camera, a rail on which the cellphone moves freely in front of the iris, an LED illumination setup, an illumination control circuit, and a frame on which all these components are mounted. Figure 3.2 shows the device schematics. The device was inspired by the designs of Bastias [23] and Mariakakis [53], and presents slight improvements on the one presented in [24]

The frame consists of two parts: a main part and a mobile part. The main part is a chamber based on Google Cardboard VR glasses [52]. This chamber produces a steady structure to capture images of the eye while blocking external light sources. The structure was laser-cut in acrylic for stability and painted opaque black to reduce specular highlights [24]. The illumination setup and circuitry are allocated to this structure. The mobile part holds the camera and a close-range lens. This part was 3D printed and also painted opaque black. The 3D-printed piece includes a handle and a rail that allows the two parts to move freely along the axis shown in Figure 3.2. The subject remains fixed with respect to the main part; thus, when the mobile part moves along the axis, the camera can take several pictures of the right eye from different perspectives. Figure 3.3 shows a picture of both parts of the device.

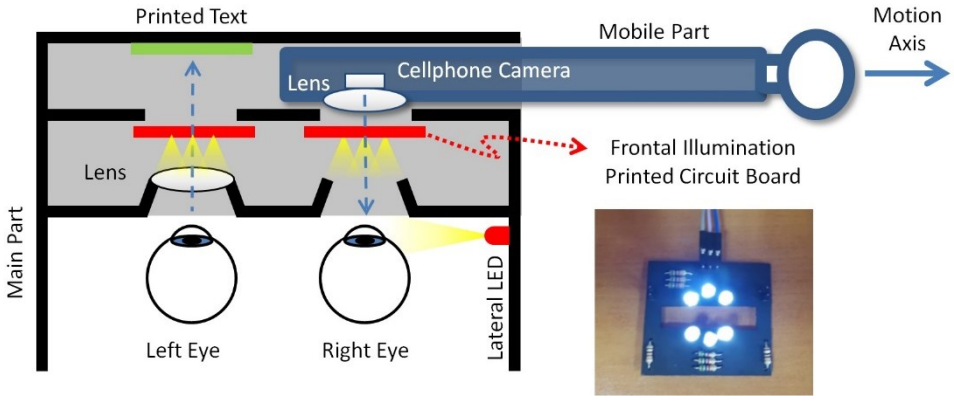


Figure 3.2: Top view of the device schematics and a picture of the frontal illumination PCB. The left half of the device allows the users to focus their gaze on printed letters, while the right side allows the camera to capture close range images of the right eye in a controlled illumination environment. The cellphone camera moves in a linear trajectory while taking pictures of the right eye from different viewpoints (angles).

Figure 3.3: Picture of the image acquisition device.

Image acquisition is done with a Samsung S6 cellphone with a macro lens. This macro lens is a biconvex lens, 25mm diameter and focal distance of 45mm. It captures 16Mpx VL images at 4 cm from the iris [24]. The eye is illuminated with white LEDs from the front and from the side. The lateral LED enhances the contrast in the images since it produces shadows from the relief on the surface of the iris [24]. The frontal LEDs maintain uniform illumination across the iris and the sclera [24]. The frontal illumination PCB was based on the design in [23]. This PCB allows the illumination to be fixed while the camera moves behind it. The device was designed so that the specular reflections from the frontal LEDs fall on the pupil region avoiding interference with the iris texture. Another feature included to reduce eye motion, is a printed text of the first seven letters of the alphabet, as is shown schematically in Figure 3.2. This helps to fix the gaze of the subject's left eye, by the operator asking the subject to focus their attention on a specific letter. Also, to avoid variations in pupil dilation, the left eye is illuminated with a duplicate of the frontal illumination PCB. This helps in maintaining constant pupil dilation between frames because both eyes receive the same amount of light. Light intensity of both circuits, as well as that of the lateral LED, is controlled independently using pulse width modulation. The illumination controller is the white box on top of the device shown in Figure 3.3.

3.3.2. Image Processing

- Lens Distortion Correction

As is usual in the first step of an image processing pipeline, it was necessary to correct the lens distortion between the cellphone camera and the macro lens to improve the image processing results, reduce the complexity of the camera model, and the processing time [12]. To estimate both lens distortion coefficients and the camera model, we used the Matlab implementation [54] of the camera calibration process [28][29]. This process uses a set of between 10 and 20 images of a checkerboard pattern in different locations of the image [54]. The pattern is placed at the same distance from the camera as that of the iris. With the location of the corners in the checkerboard images, this algorithm computes the intrinsic parameters of the camera model such as: focal length, principal point, skew, radial distortion, and tangential distortion. With this information, lens distortion correction is performed on the iris images.

The camera model is the mathematical transformation that represents the way in which 3D points of the real world are projected onto the pixels of the image. Equation 1 describes the pin-hole camera projection model [21][69]

$$\tilde{p} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}^{-1} \times P \quad [70], \tag{3.1}$$

$$\tilde{p} = K \times T^{-1} \times P \quad [70], \tag{3.2}$$

where P is a 3D world point, $\tilde{p}$ is the coordinates of the pixel in which P is projected in homogeneous coordinates, f_x and f_y are the horizontal and vertical focal lengths in pixels, s is the skew, (c_x, c_y) are the coordinates of the principal point which is the intersection of the optical axis with the image plane, r_{ij} are rotation coefficients of the camera pose, and the vector (t_x, t_y, t_z) is the translation of the camera position from the origin of the 3D world coordinates. K is the intrinsic matrix, which is the result of the camera calibration process. Finally, T is the extrinsic matrix, which is computed in later steps.

- Dilation Based Image Selection

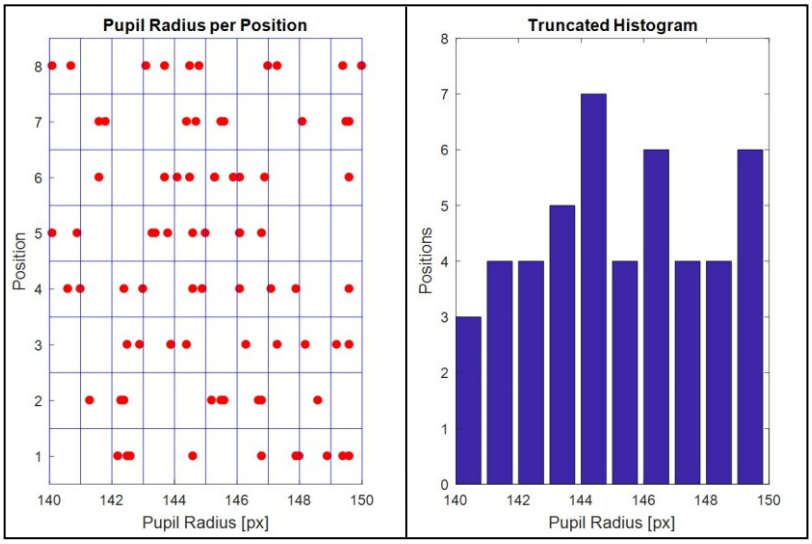
SfM and MVS methods assume that the scanned object does not deform between frames. However, the human iris changes because of dilation, which is a form of distortion [10]. The greater the variation in iris dilation in the images used in the 3D reconstruction process, the greater the noise level in the results. Illuminating both eyes with the same amount of light improves the stability in the dilation; however, small dilation changes naturally still occur. To further reduce dilation related problems, we propose an image selection method that identifies a set of images with the same dilation to generate a 3D model. First, we take a burst of 10 or more images of the eye at each camera position. Then, we identify the pupil in each image and measure its radius. Finally, we compute a histogram of the number of positions available per pupil radius. The dilation level which is present in the greatest number of positions is then chosen, and all the images that form that histogram bin are used in the reconstruction of the 3D model. Figure 3.4 illustrates the amount of dilation present in a typical image set, and the image selection method. At the end of this step, a group of images having small dilation variation and coming from the largest number of camera positions available is selected.

- Iris Segmentation

We segmented the iris in the selected images by masking out the pupil, the sclera, eyelids and eyelashes [8][56]. Although the 3D model can be reconstructed without iris segmentation, this process removes irrelevant 3D points from eyelids and eyelashes in the final 3D model.

- Contrast Enhancement

The final preprocessing step is to enhance the dynamic range of brightness in the iris region. This step improves the number of keypoints that can be extracted from the iris. A Local Laplacian



(a) (b)

Figure 3.4: Pupil dilation in an image set. (a) Pupil radius of 10 burst images from 8 camera positions. (b) Histogram of positions available per pupil radius. A radius of 144.5 pixels was present at 7 different camera positions; those images were therefore selected to reconstruct the 3D model.

(a) (b)

Figure 3.5: Contrast Enhancement of iris images by means of Local Laplacian [20]. (a) Original Image. (b) Enhanced Image.

filter was used for contrast enhancement [30]. This filter uses information of only local patches and, thus, eliminating extreme bright and/or extreme dark zones is not required. Figure 3.5 shows the results of this stage.

3.3.3. Keypoint Extraction and Matching

- Keypoints And Descriptors

The starting step in SfM to produce the 3D model is identifying distinctive features across all the images. The coordinates of those features help to infer the pose of the camera for each image, as well as establishing an initial sparse 3D model. The development of robust and fast feature detectors and descriptors, such as SIFT, SURF, and ORB, was a crucial factor in developing the modern SfM and MVS algorithms [12]. We use SURF keypoints, as well as its 64-element descriptor, because it is efficient and invariant to lighting, scale, and rotation [70][81][82].

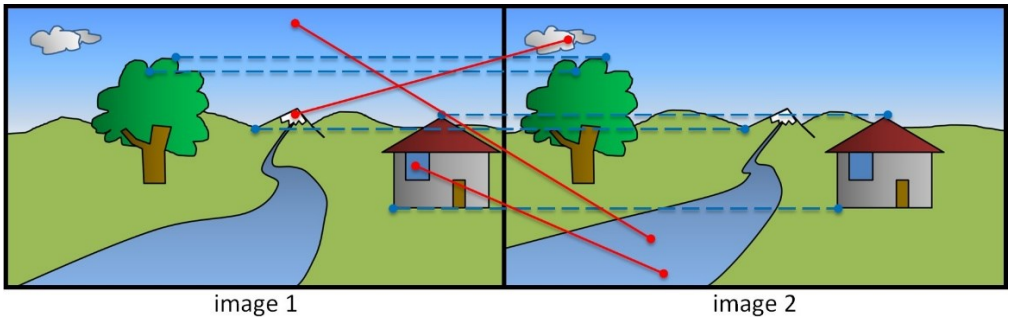


Figure 3.6: Slope-based match filter. The two images are placed side by side horizontally and line segments join the matched keypoints. Since the camera moves on the *x* axis, line segments are expected to have a slope of 0 (blue lines). However, mismatched keypoints have similar descriptors, but produce slopes different from 0 (red lines).

- Keypoint Matching and Slope-Based Filter

Keypoints of two images can be matched by finding the Euclidian distance of their descriptors. However, if the texture is similar in different regions of the object, some points can be matched incorrectly. Those mismatches diminish the precision of the essential matrix and the object reconstruction. In order to identify those outliers, we make use of the fact that the camera moves in a linear trajectory. That linear motion creates a predominant effect of iris translation on the x axis from image to image. Figure 3.6 shows a graphic representation of our slope-based filter. First, for each keypoint in image 1, we rank all the keypoints in image 2 according to the distance between their descriptors. Then, for the keypoint pair with the smallest distance, we compute the slope of the line that joins them when the two images are placed side by side horizontally. If the slope is between ±5º, the match is accepted; otherwise, the match is rejected, and the pair of points with the second best distance is tested. This process is repeated until each keypoint from image 1 is mapped with one keypoint of image 2. In this way, incorrect matches are removed.

At the end of this step, keypoints and descriptors are extracted from the images, and they are matched among consecutive images.

3.3.4. Sparse 3D Reconstruction

- Camera Pose Estimation

The camera pose consists of the position and orientation of the camera in space, which can be determined from the essential matrix between two images. The essential matrix is a 3x3 matrix which encodes the Epipolar geometry that results from projecting one point in space P onto two different cameras or views to produce the corresponding points $\tilde{p}_1$ and $\tilde{p}_2$ [12][70]. The two corresponding points from this projection are related with the essential matrix [83] given by:

$$\tilde{x}_2 \times E \times \tilde{x}_1 = 0, [70] \quad (2.3)$$

where $\tilde{x}_1$ and $\tilde{x}_2$ are the normalized coordinates of the corresponding points $\tilde{p}_1$ and $\tilde{p}_2$, and E is the Essential Matrix [70][83]. The essential matrix can be resolved with a minimum number of 8 corresponding points [83][84]. To reduce the error caused by incorrect matches, a RANSAC

process is used in the computation of the essential matrix [70][83]. This process identifies and removes the outliers that might have passed through the previous filtering stage. The essential matrix encodes the relative camera pose for two different angles (views). If one of them is known, we can compute all the r_{ij} coefficients, as well as (t_x,t_y,t_z), in the pin-hole camera projection model in (3.1). In our algorithm, the pose of the first camera is fixed in advance, and the pose of the subsequent views is determined with the described method.

- Point-Cloud 3D Reconstruction

Once the pose of the camera is found for each image, all the matches can be projected onto the 3D space to form a Sparse 3D reconstruction. The projection process consists of finding the 3D line that joins the center of the camera with a keypoint, and extending it. Such a 3D line is called a bundle [21]. The process is repeated for each camera position. The point in space where the bundles of corresponding keypoints meet is considered to be a 3D point in the point-cloud model. Since a keypoint can usually be tracked in more than two views, the bundles might not intersect at the same exact point. Therefore, the 3D point is the one that produces the least error in the intersection. The process is repeated for each keypoint to obtain all the 3D points in the point-cloud model.

Data of the matching points and views are managed using the track format to make the process more efficient [12]. A track j is conformed by a 3D point P_j, the camera models of all the images I on which P_j is visible, and the keypoints $\tilde{p}_{ij}$ which generated the 3D point [12].

- Bundle Adjustment

A refinement process in the 3D reconstruction is called Bundle Adjustment. This process consists of jointly optimizing the camera poses and the point-cloud model to minimize the re-projection error [12][75]. The re-projection consists of projecting the point-cloud model back to each image using the calibrated camera model (3.1) [12][70]. In this sense, the error is the distance between a projected point and the keypoint that originated it. The goal is to find the camera poses and 3D points that minimize the total error given by:

$$\text{error}(T,P) = \sum_j \sum_{i \in V(j)} \left(\alpha_{ij} K {T_i}^{-1} P_j - \tilde{p}_{ij}\right)^2, [70] \quad (3.4)$$

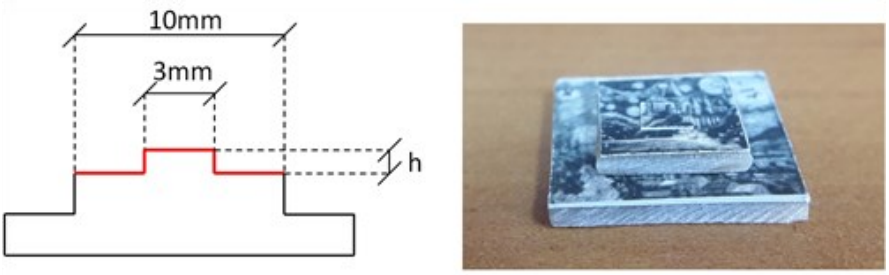
where j is the track, i is the camera view, K is the intrinsic matrix, T_i is the extrinsic matrix, P_j is a 3D point in Cartesian coordinates, α_{ij} is a constant that represents the normalization from homogeneous coordinates to Euclidean coordinates, and $\tilde{p}_{ij}$ is a keypoint.

At the end of this step, the accurate position of the camera is found for each image, as well as the initial point-cloud model of the iris tissue. On average, the reconstructed models have 2000 3D points at this step.

3.3.5. Dense 3D Reconstruction

With the purpose of increasing the number of keypoints and 3D points, we acquired new keypoints using the minimum eigenvalue algorithm by Shi and Tomasi [85]. This technique extracted five times more keypoints than SURF from our images. Then, the 3D points were computed using the calibrated camera models of each view from the sparse reconstruction. After that, the bundle adjustment was executed a second time to refine the model. Finally, the normal to each 3D point was initialized to be perpendicular to the iris plane, and they were refined using Meshlab, an open-source software for 3D model processing [86]. At this stage, the number of 3D points in the point-cloud model is over 10,000 on average.

The final step in our 3D reconstruction pipeline was the estimation of a mesh model for the iris. To accomplish this, we used the Meshlab implementation of the Screen Poisson Surface Reconstruction algorithm [67][86]. This algorithm estimates the best-fitting mesh for the point cloud, and is robust to clutter [67]. This mesh helps to fill patches where the iris texture is uniform and an insufficient number of keypoints is available.



(a) (b)

Figure 3.7: Pattern of known spatial dimensions to test 3D scanning accuracy. Only the red surfaces of the diagram were used in the analysis. Three patterns were created with final step height h of 75 μm, 150 μm, and 375 μm.

3.3.6. Scanning Resolution Measurement

After developing the 3D iris scanning methodology, it is necessary to determine the smallest spatial detail the proposed method is able to resolve. For this purpose, we scanned 3D patterns of known dimensions and analyzed the Signal to Noise Ratio (SNR). The shape of the pattern is a flat surface with a step in the middle, as is shown in red in Figure 3.7. The pattern needs to have a rich texture so that our proposed methodology can be applied. That is why the flat surfaces contain a high texture picture printed on sticker paper over a rigid copper/acrylic sheet (PCB material). The height of step h is controlled by varying the number of paper layers. Each paper layer has a thickness of 75 μm. In total, three patterns were created using one, two, and five layers. The patterns were scanned with the same device and same illumination settings as used for the iris.

We compute the signal to noise ratio, SNR [85], using:

$$\text{SNR} = \frac{A}{\sigma}, \tag{3.5}$$

where A is the amplitude of the signal and σ is the standard deviation of the noise. We take the measured value of h in the 3D model as the amplitude A.

We compared the performance of our SfM pipeline against that of two general-purpose SfM implementations available online. The first SfM software is VisualSFM [13] for the sparse reconstruction and CMVS [18] for the dense reconstruction. The second software is Regard 3D, available online on: http://www.regard3d.org/. For the test, we compare the reconstruction error in the 3D patterns of Figure 3.7. Since the ground truth is the height of the step h, the reconstruction error can be calculated as:

$$\text{error} = \frac{1}{N+M}\left(\sum_{i=1}^{N}|Z\text{up}_i - h| + \sum_{i=1}^{M}|Z\text{low}_i|\right), \tag{3.6}$$

where $Z\text{up}_i$ is a 3D point in the upper level, $Z\text{low}_i$ is a point in the lower level, and N and M are the total number of points in the upper level and the lower level respectively.

3.3.7. Subjects for 3D Iris Reconstruction

Using the proposed method, we scanned 10 subjects to build the 3D iris models. The subjects' irises are shown in Figure 3.8 for frontal VL, and the typical iris appearances are shown for both dark and light colored eyes. Within this group, there is one green iris, three dark brown irises, and six light brown irises. We imaged their right eyes with the proposed device using a combination of lateral and frontal VL to enhance iris texture. The image acquisition time is approximately 30 s for 6 images of the iris. In the future, this time could be reduced by using an automatic scanning procedure such as controlling camera motion with a linear actuator. Additionally, an OCT scan was performed on the right eye of subject 1. In the results section, we show the reconstructed 3D model for all 10 subjects, as well as a cross-section comparison between the 3D model and the OCT scan for subject 1. To compensate for the pupil dilation differences between the OCT scan and the 3D model, a linear deformation was applied to the 3D model slice until both methods have the same dilation level [10].

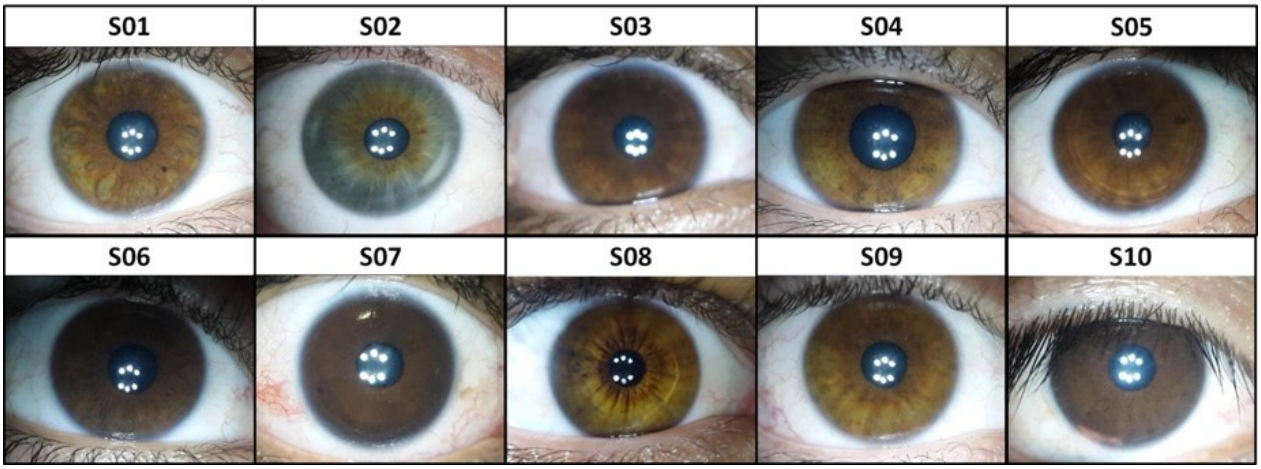
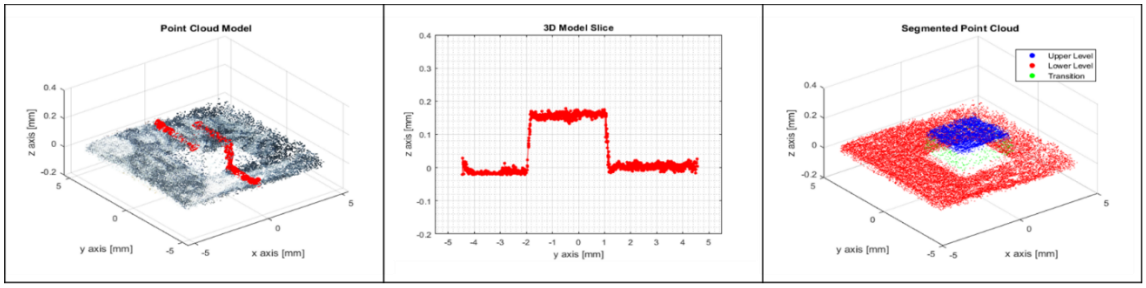


Figure 3.8: Irises captured with frontal VL for the 10 subjects included in the 3D iris model reconstruction.

3.4. Results

3.4.1. Results for the Proposed Method Spatial Resolution

We scanned the three spatial 3D patterns described in section 3.3.6 and measured the amplitude of step A, as well as the standard deviation of the noise σ. Figure 3.9 shows an example of a scanned 3D pattern. The lower and upper levels of the pattern were segmented using two thresholds on the z axis. The cross-section shown in Figure 3.9b is for illustrative purposes only, since the analysis was made using the 3D segmented data that is shown in Figure 3.9c. Since we are interested in measuring the height of the step as the difference between the lower and upper levels, green points in Figure 3.9c were not used in the analysis.



(a) (b) (c)

Figure 3.9: Example of the scanned 3D spatial pattern for h=150 µm. (a) Point cloud model and extracted cross section. (b) Corresponding 2D cross section. (c) Segmented point cloud.

Table 3.1 shows the results of this test for the three spatial patterns 75 µm, 150 µm, and 375 µm. The height of each step was measured with an error smaller or equal to 5 µm in all of the three samples. As expected, the SNR increases proportionally to the height of the step. The level of the noise present in our method is consistent for the three models with a value of 11 µm ± 1 µm. This implies that the smallest detail we can detect is 11 µm.

TABLE 3.1
SIGNAL TO NOISE RATIO IN 3D SPATIAL PATTERNS OF KNOWN DIMENSIONS

h [µm]	3D Points	A [µm]	σ [µm]	SNR
75	28431	79	10	7.9
150	30034	155	12	12.9
375	35023	378	12	31.5

Table 3.2 shows that our SfM algorithm produces reconstrunction errors smaller than 10µm. This results are consistent with the test in Table 3.1. Additionally, our method produced smaller errors than the other tested algorithms.

TABLE 3.2
SFM METHOD COMPARISON: AVERAGE RECONSTRUCTION ERROR IN PATTERNS OF KNOWN DIMENSIONS

SfM Method	75 µm	150 µm	375 µm
VisualSFM / CMVS	13.3 µm	16.1 µm	12.8 µm
Regard 3D	10.4 µm	13.7 µm	13.2 µm
Ours	7.8 µm	9.5 µm	9.2 µm

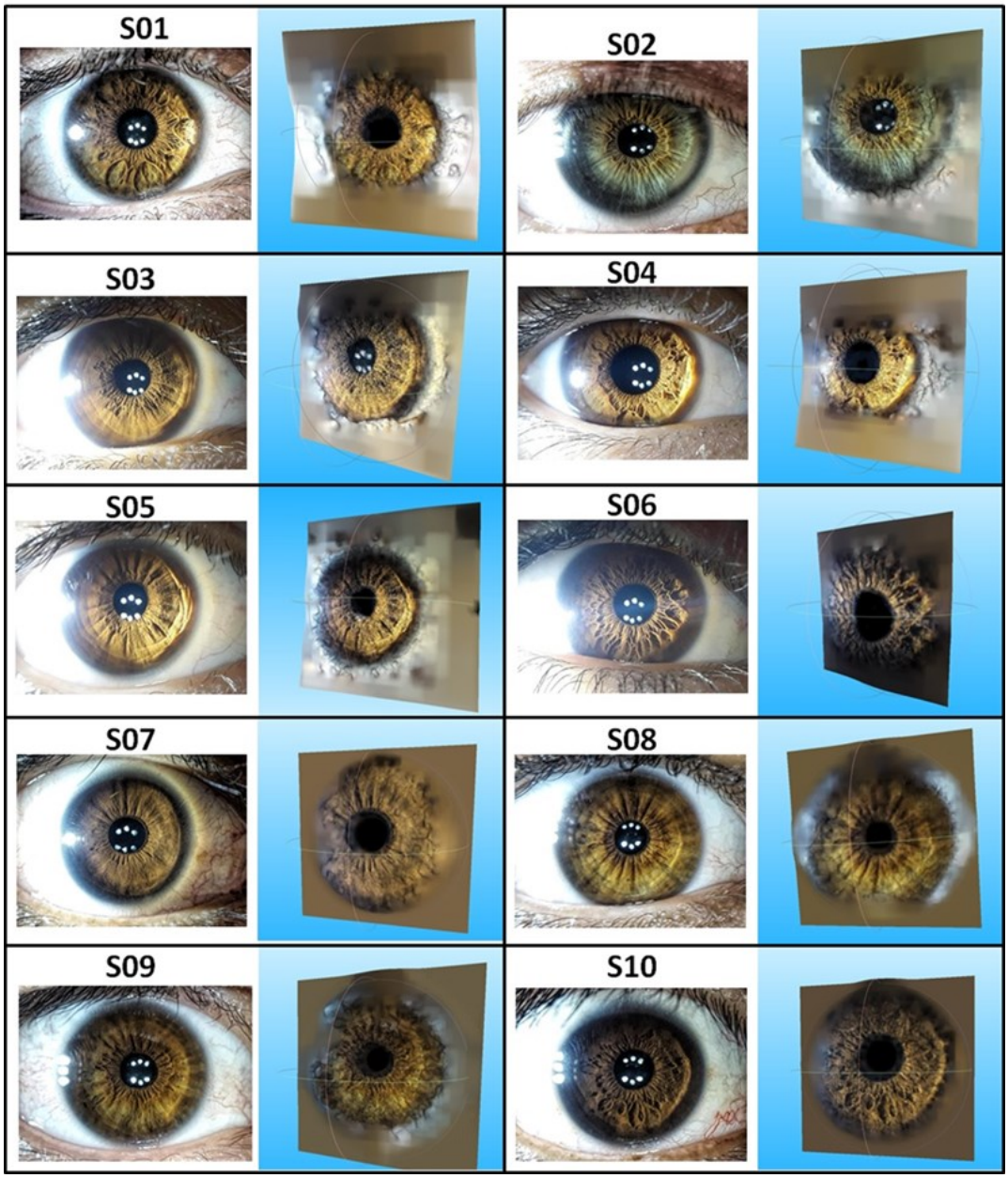


Figure 3.10: Reconstructed 3D iris models for the 10 subjects. The 3D reconstructed mesh is presented next to one of the iris images from which the model was generated. Images were captured using lateral + frontal VL.

3.4.2. Subjects' Iris Reconstruction Results

The 3D iris surfaces of 10 subjects were reconstructed using the proposed methodology. Figure 3.10 shows, for each subject in one row, the frontal iris image followed by one image taken from the 3D model at a speciphic angle. It can be observed that the final 3D model captures the depth information, as well as detail information from the iris. In general, the regions of the iris with greater texture produced more intricate shapes in the final 3D iris reconstruction. Figure 3.11c shows a close up of model S01 from a lateral perspective. It illustrates the relief of the iris, which is captured by our method. Table 3.3 shows information about the reconstructed 3D models. The

average peak-to-peak depth value among the models is 1.38 mm, considering only the iris region. The variability of the depth value is characterized by the standard deviation. Its average value is 270 µm including all the models. Finally, the average number of reconstructed 3D points is 11,105, and the average number of polygons in the 3D mesh is 54,605. The results of the cross section comparison between the 3D model and the OCT are presented in Figure 3.11. The result of the OCT scan in Figure 3.11b shows, as expected, the presence of a non-planar relief in the surface of the iris of subject 1. Figure 3.11d shows the slice of the 3D model.

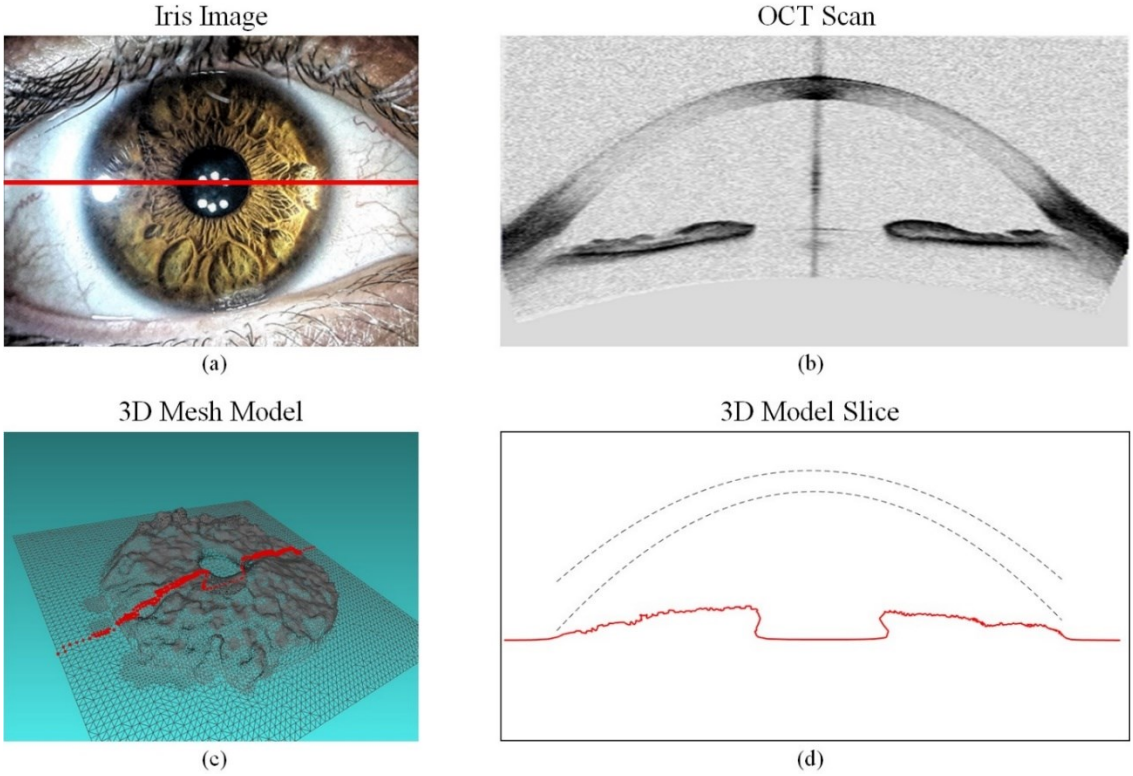


Figure 3.11: Cross-section comparison between the OCT scan and the 3D mesh model of subject 1. Red lines indicate the cutting plane. Dashed lines in (d) represent the cornea in (b). Lineal dilation was applied in (d) so that the 3D model has the same pupil dilation level as the OCT scan.

TABLE 3.3
RECONSTRUCTED 3D-MODEL SUMMARY

Subject	3D Points	Polygons	PP [mm]	STD[mm]
S01	12,006	66,273	1.53	0.31
S02	9,430	53,261	1.27	0.25
S03	13,332	76,902	1.09	0.22
S04	9,683	51,146	1.21	0.24
S05	12,919	73,478	1.47	0.29
S06	6,840	31,938	1.47	0.29
S07	14,030	42,558	1.43	0.24
S08	9,934	22,376	1.38	0.24
S09	12,726	71,506	1.53	0.27
S10	10,147	56,611	1.46	0.30
Average:	**11,105**	**54,605**	**1.38**	**0.27**

3.5. Conclusions

We presented a new method for building a 3D model of the iris surface from several 2D iris images. Our method uses 2D VL images of the iris from different perspectives, and a modified SfM algorithm to reconstruct a 3D mesh representation of the iris. The use of a combination of lateral and frontal VL illumination is a key element in the image acquisition process, as it enhances the visibility of the structures on the iris for both dark and light colored eyes [24]. Since several 2D images from different views are acquired to build the 3D model, illuminating both eyes with the same amount of light helps maintain stable dilation. However, small dilation variation between images still occurs. To further reduce dilation related problems, an image selection process was performed identifying a set of images with very similar dilation for generating a 3D model. As a result, the number of 3D points, the shape of the reconstructed irises, and the robustness of the method improved relative to our previous research [23].

The spatial resolution of the proposed method was measured using three 3D patterns of known dimensions (75 μm, 150 μm, and 375 μm). The level of the noise present in our method is consistent for the three models with a value of 11 μm ± 1 μm. This implies that the smallest detail we can detect is 11 μm, which means that we can measure details six times smaller than current 3D common-object scanners [87]. The resolution of our proposed method is similar to that of conventional OCT scanners, which operate in the neighborhood of 10 μm [88]. However, modern OCT scanners can reach resolutions as fine as 2.2 μm [89]. To put those numbers in context, the thickness of the iris tissue is about 400μm [90][91].

We reconstructed the 3D iris models for 10 subjects, and it was observed that the final 3D model captures the depth information, as well as detail information, from the iris. We also compared an OCT for one iris to the respective crosscut performed over the 3D model of the iris. Similar spatial features were identified with both methods. To increase the precision in depth measurement, camera resolution can be improved. This would increase the number of keypoints that are used to create the 3D model and enhance the resolution.

The proposed methodology has the potential of being applied in biometrics and in ophthalmology. With the additional depth information extracted by our method, a possible line of research is to develop methods that make use of 3D data in iris recognition. This may be of special interest when only segments of the iris are available because they would contain more information than 2D segments.

Additionally, our proposed method could be useful for detecting textured contact lenses, and therefore is potentially useful in the presentation attack detection (PAD) area of research in iris recognition [92][93][94]. The 3D model of a textured contact lens will have a spherical shape since the dominant texture is printed on the surface of the contact lens.

This method could be applied in the diagnosis of acute angle glaucoma [23][24]. A current tool to diagnose this disease is Optical Coherence Tomography (OCT) [26]. OCT forms an image of a transversal cut of the eye. Ophthalmologists measure various parameters such as: the

width of the iris, the angle between the iris and the cornea, and the distance between the anterior chamber and the top of the iris, for the diagnosis [26].

A 3D mesh representation of the iris can also be used as a tool for studying pupil dilation in greater detail. The deformation of the iris tissue due to pupil dilation is typically treated as linear although it is known that this is only an approximation [5][10]. This motivated research into more accurate iris normalization models that improve iris recognition performance under dilation differences [27][95][96][40]. For instance, Clark et al. generated a non-linear iris normalization model based on the biomechanical properties of the iris [27]. Studying this phenomenon in 3D with real irises could produce valuable insights, and it might lead to the formulation of a more accurate iris normalization model.

A limitation of our method is the specular reflection of the lateral LED. Due to the saturation of this light, a segment of the iris is lost in the reconstruction. A possible solution to this problem is to place an additional lateral LED on the other side of the eye, and capture two sets of images that then can be combined for the 3D models.

3.6. Acknowledgment

This research has been funded by CONICYT through project FONDECYT 1161034 and by the Department of Electrical Engineering, Universidad de Chile. The authors would also like to thank the students at the School of Engineering, Universidad de Chile, who participated enthusiastically as volunteers for iris image acquisition.

Chapter 4

A 3D Iris Scanner from a Single Image using Convolutional Neural Networks

A 3D model of the human iris provides an additional degree of freedom in iris recognition, which could help identify people in larger databases, even when only a piece of the iris is available. Previously, we reported developing a 3D iris scanner that uses 2D images of the iris from multiple perspectives to reconstruct a 3D model of the iris. This paper focuses on the development of a 3D iris scanner from a single image by means of a Convolutional Neural Network (CNN). The method is based on a depth-estimation CNN for the 3D iris model. A dataset of 26,520 real iris images from 120 subjects, and a dataset of 72,000 synthetic iris images with their aligned depthmaps were created. With these datasets, we trained and compared the depth estimation capabilities of available CNN architectures. We analyzed the performance of our method to estimate the iris depth in multiple ways: using real step pyramid printed 3D models, comparing the results to those of a test set of synthetic images, comparing the results to those of the OCT scans from both eyes of one subject, and generating the 3D rubber sheet from the 3D iris model proving the correspondence with the resulting 2D rubber sheet and binary codes. On a preliminary test the proposed 3D rubber sheet model increased iris recognition performance by 48% with respect to the standard 2D iris code. Other contributions include assessing the scanning resolution, reducing the acquisition and processing time to produce the 3D iris model, and reducing the complexity of the image acquisition system.

4.1. Introduction

The human iris is composed of two muscle systems and a sphincter to control the amount of light entering the retina [10]. These muscular fibers, as well as the pigmentation, provide a unique texture to each iris that can be used for identification [8]. Traditionally, the texture of the iris has been analyzed using 2D images to produce accurate iris recognition [3][8] [27][43] [47][56][64]. However, in recent years, a 3D iris scanning method that exploits the 3D relief of the iris has been proposed [23][25][35]. This method reconstructs a 3D model of the iris surface using images from several perspectives and Structure from Motion (SfM) algorithms [12][13]. The 3D iris model opens new frontiers for biometric applications, as well as in ophthalmology [25]. For example, the 3D iris model can potentially be used as a screening method for Closure Angle Glaucoma, a disease currently diagnosed with Optic Coherence Tomography (OCT) scans [23][25][26].

A method for reconstructing a 3D model of the iris surface from several images was introduced by Bastias et al. [23] and improved by Benalcazar et al. [25]. The improved method consists of the following steps: First, visible light (VL) images of the iris are captured from different perspectives. These images are acquired with a custom device that illuminates the iris with Lateral and Frontal Visible Light (LFVL) [24]. Then, a modified SfM algorithm estimates the camera pose of every image jointly with a sparse 3D model of the iris [23][25]. Then, a dense 3D point-cloud reconstruction is performed by extracting Shi-Tomasi keypoints from each image [25][85]. Finally, the point-cloud model is converted into a mesh surface by the Screened Poisson Surface Reconstruction technique [67]. This mesh helps interpolate the depth information in areas of the iris with low texture [25]. The result is a 3D model that incorporates both depth and color information of the iris surface. The additional dimension aims to increase iris recognition accuracy particularly when the iris is occluded by eyelids, eyelashes, and reflections [23][35]. The system recently developed by Cohen et al. [35] tracks fiducial points from two or more near-infrared (NIR) images of the eye to create the 3D model. They then calculate the geometric error between two 3D models as the Mean Square Error (MSE) of candidate matching points. They tested their method on a dataset of 20 irises, correctly classifying all of them.

As previously described, the 3D iris scanning method can produce a complete model of the human iris, but there are limitations to this technique. First, the SfM method requires a moving camera, which adds complexity to the system. Second, SfM was conceived to scan inanimate objects; however, the human iris can dilate from frame to frame, adding a source of distortion. This was solved by acquiring many images per position, and selecting those with a consistent dilation level [25]. This solution increases both acquisition and processing time. Third, because SfM relies on keypoints and descriptors, irises with richer texture generate more 3D points than those with fewer details. Finally, it is difficult to acquire 3D points from areas in the image that present no texture; thus the point-cloud 3D model has an uneven distribution of points in space. The mesh representation solves this issue at the expense of more processing time [25].

However, SfM is not the only method that can produce 3D scene reconstruction from 2D images. In recent years, Convolutional Neural Networks (CNN) have increased accuracy in depth prediction tasks [20][97][21]. Most of the CNNs rely on training an encoder-decoder architecture with the image of a scene as the input, and an aligned depthmap as the target [19][98][99]. As a result, the CNN learns to identify visual cues, such as perspective, that allow prediction of the depth of every object in the scene. The output depthmap captures the depth value of every pixel, even in low texture areas such as uniform color furniture or roads [19]. Therefore, the 3D model is always complete and evenly sampled regardless of the texture in the image.

The main contribution of this paper is to propose a new method to obtain a 3D model of the iris from a single image using CNNs. The method is based on a depth-estimation CNN for the 3D iris model. A dataset of real iris images from 120 subjects, and a dataset of synthetic iris images with their aligned depthmaps were created. Then, depth-estimation CNNs were trained using the real and synthetic irises [15][20][97][100], and two network architectures were combined to improve performance. We analyzed the performance of our method in predicting the iris depth by using real step pyramid printed 3D models, comparing the results to those of a test set of synthetic images, comparing the results to those of the OCT scans from both eyes of one subject and generating the 3D rubber sheet from the 3D iris model, and proving the correspondence with

the resulting 2D rubber sheet and binary codes. Other contributions of the proposed method include assessing the scanning resolution, reducing the acquisition and processing time for producing the 3D iris model, and reducing the complexity of the image acquisition system since the camera does not need to move to scan the iris.

4.2. Related Methods in Depth Estimation Using Convolutional Neural Networks

Depth estimation by a CNN can be formulated as a regression problem, in which the input is an image, and the target is the depth value of every pixel, also known as the depthmap. Eigen et al. [98] used a single image of an indoor scene as input, and the aligned depthmap of the same scene as the target. Such a depthmap had been acquired previously with an RGB-D camera. As a result, the CNN learned the depth of the walls and objects in indoor environments with great accuracy from their contexts [98]. Since then, several methods have been reported in the literature that have used similar training schemes and improved architectures with excellent depth estimation performance [15] [19][20] [21][97][99].

The architecture of some depth estimation CNNs has been improved to produce more robust solutions. Eigen and Fergus [99] expanded their previous work to also predicting surface normals and labels. Laina et al. [19] trained a ResNet50 [101] based auto-encoder to increase accuracy. Alhashim and Wonka [20] developed DenseDepth, a DenseNet-169 based encoder with upsampling layers in the decoder to obtain high resolution depthmaps of indoor and outdoor scenes. Xu et al. [102] integrated Convolutional Neural Fields and a structured attention model to generate pixel precision in depth estimation. Fu et al. [97] developed DORN, with a space-increasing discretization strategy to recast depth estimation as an ordinal regression problem. CNNs have been trained to produce more complex methods for map reconstruction and navigation. For example, the CNN SLAM not only estimates depth from a single frame, but also integrates successive predictions of a video feed into a larger and more complete map of the environment [103]. Another deep network, FastDepth, by Wofk et al. [21] focused on a real time implementation for robotic navigation.

One limitation of the previously described methods is the need for a large number of aligned depthmaps for training. That is why Godard et al. [14] developed Monodepth, an encoder-decoder CNN that is trained with stereo images. The input of that network is the left image and generating the right image is the target. In this sense, the network has to understand the 3D geometry of the scene implicitly to perform the task. Kuznietsov et al. [104] combined stereo image information with sparse depthmap ground truth to produce a semi-supervised implementation. Their approach uses a small number of aligned image-depthmap pairs as ground truth in a supervised manner, along with a greater number of stereo image pairs in an unsupervised manner [104]. The latter two methods [14][104], outperformed previous existing methods in depth estimation. However, the most recent methods, DenseDepth [20], and DORN [97] have already achieved better results.

Another solution for the limited availability of training data in depth estimation is the use of synthetic images. Tian et al. [22] trained detection and classification networks using a

combination of real and synthetic images. In their work, CNNs trained with real and synthetic data outperformed those trained with only real images [22]. Moreover, Zheng et al. [15] developed a depth-estimation CNN (Translation and Task Network, T^2Net) that incorporates the use of synthetic and real images in its architecture. The T^2Net is composed of a Generative Adversarial Network (GAN) that translates synthetic images to the domain of the real ones. The task component is an encoder-decoder that then predicts depth from the translated images [15]. T^2Net achieved state-of-the-art results in widely used datasets, such as NYU-DepthV2, and KITTY [15].

Zheng et al. [15] analyzed various strategies for incorporating synthetic data in depth estimation tasks. As a result, they propose that the best alternative is incorporating both the translation and the task in the same training loop. In this way the GAN will learn to modify synthetic images only in their appearance while keeping the main features aligned with their depthmaps. They call it the full approach, and it had the best results among the other strategies analyzed [15].

4.3. Methodology

Our methodology for developing a new method to obtain the 3D model of the iris from a single image using CNNs is based on a depth-estimation CNN. First we defined the requirements of the training images so that the CNNs could infer depth information from visual cues. Then, we acquired both real and synthetic iris datasets with the desired characteristics. After that, we used our datasets to train available depth-estimation CNNs for 3D iris scanning. We then analyzed the performance of our method in predicting iris depth, and using printed 3D step pyramid models, we compared the results to those of a test set of synthetic images, compared the results to those of the OCT scans from both eyes of one subject, and generated the 3D rubber sheet from the 3D iris model demonstrating the correspondence between the resulting 2D rubber sheet and binary codes.

4.3.1. Learning Depth Information

Several visual cues provide depth information to humans. Cutting and Vishton [105] identified nine distinct mechanisms from which humans perceive depth. Occlusions indicate whether an object is behind or in front of another. The relative size of an object also indicates depth. Due to perspective, an object that is closer to a camera appears bigger than another that is farther away [70]. Similarly, the texture density of a cobble road appears to be coarser close to the viewer than farther away [105]. Binocular disparity allows triangulation to compute the distance of an object from the camera depending on how its position changes from one view to the next [12][13][70]. These visual cues are exploited by most SfM and CNN systems to reconstruct the precise 3D model of an object or a scene [12][98].

Depth information of the human iris images has some particular issues that are different from those of general visual scenes. In iris images, the iris is the main object in the scene, and its size is normalized. Therefore, depth information cannot be inferred by occlusions or perspective. However, shadows cast by objects are another type of visual cue that provides depth information [105]. Elevations and craters can be identified by the shadows they cast. Similarly, in our method

it is desirable to learn the relationship between the shadows on the surface of the iris, and the depth of the features that produce them.

In order for the iris features to cast shadows, a lateral source of illumination is needed. For this purpose, we developed a device with lateral and frontal lighting [24]. The device has a black frame that blocks external light sources, and has six white LEDs in front of each iris and three white LEDs on the side of each iris (LFVL illumination), to illuminate both eyes. The lateral illumination creates shadows from the relief of the iris surface, increasing the texture in the image [24]. This texture improved results in iris recognition [24]. LFVL illumination has also been used in 3D iris scanning with good results [25]. It was shown in [25] that LFVL improved the iris texture by producing more keypoints for 3D iris reconstruction. In our work, however, the use of LFVL illumination is important because shadows from iris features carry depth information.

4.3.2. Real Iris Dataset

The real iris dataset contains iris images with a wide range of dilation levels from 120 subjects. The study was properly approved as states the resolution No.011, on May 9, 2019, by the Ethics and Biosafety Committee for Research, Faculty of Physical and Mathematical Sciences, Universidad de Chile. Each of the 120 subjects signed a letter of consent for participating in this study. Iris images were captured under LFVL illumination using the device described in both the previous section, and in [24]. Iris images were captured in 3-second videos of pupil reaction to light changes. The pupil reaction test consisted of dark adaptation for 10 seconds, so that pupils would dilate, followed by turning on the LFVL illumination for 3 seconds. This experiment is harmless to the human eye since the LEDs used in this study are catalogued Risk Group 0-1 [106]. The maximum admissible exposure time is 10,000 s for those risk groups, and our subjects were only exposed for 3 seconds [106].The video captures how the pupil contracts from a dilated state, frame by frame, at 30 f/s (frames per second). Figure 4.1 shows some frames of the pupil reaction experiment for 3 subjects, while Figure 4.2 illustrates the evolution of the dilation level over time for one subject. The dilation level is measured as the ratio between the radii of the pupil/iris boundary (Rp) and the iris/sclera boundary (Rs) [10][107]:

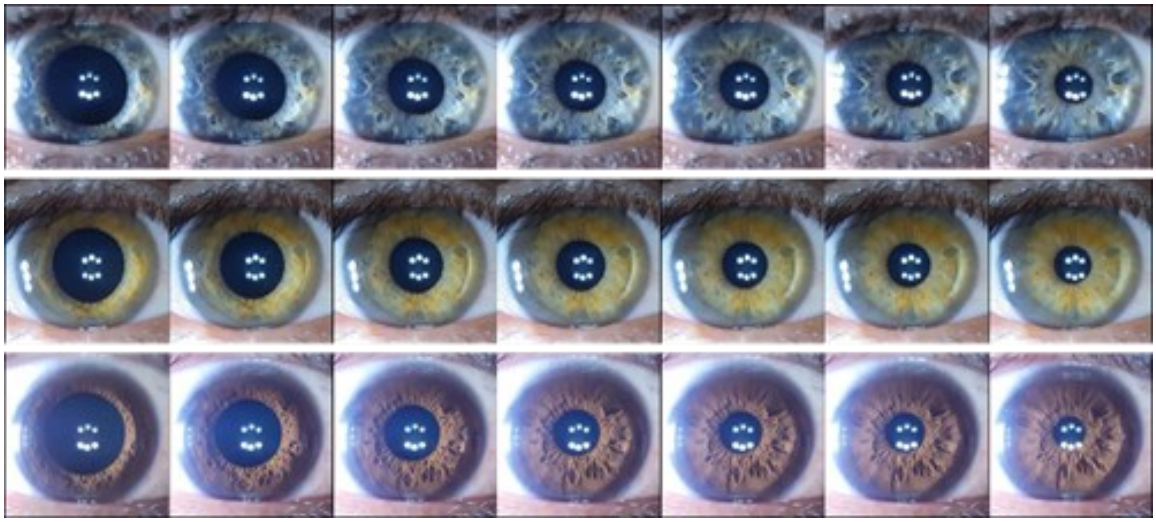


Figure 4.1: Seven frames from the pupil reaction from 3 different subjects with different pupil dilations.

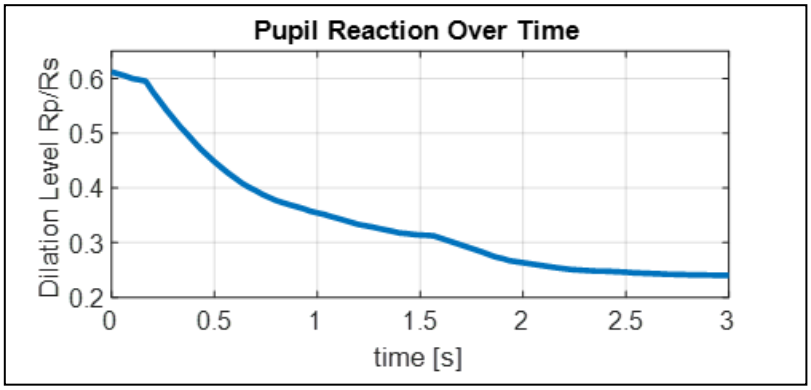


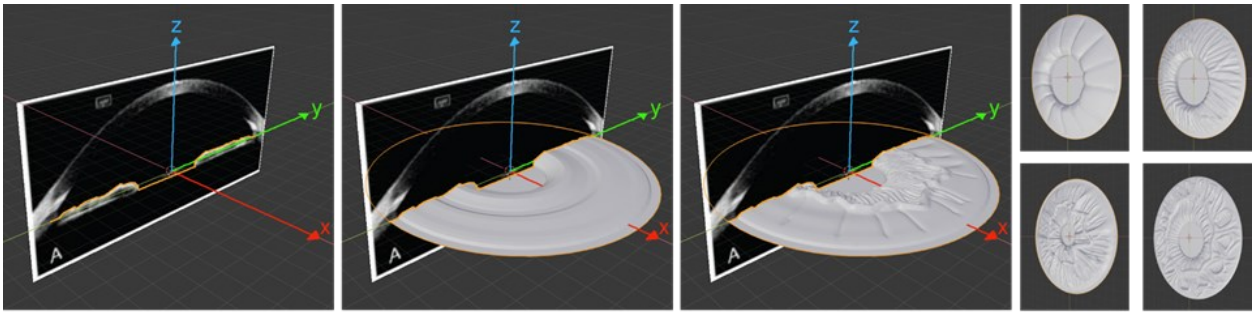
Figure 4.2: Pupil reaction experiment for the right eye of subject 003. Dilation level decreases over time in a non-linear manner due to transition from low light to brighter light.

$$\lambda = Rp/Rs. \tag{4.1}$$

In order to remove artifacts and normalize the number of images per subject, 60 valid frames were selected per video. At 30 f/s, each video has 90 available frames; however, some frames in the videos contained motion blur, occasioned by eye movements and blinking. Additionally, there were redundant frames with similar dilation levels, as can be seen in Figure 4.2 in the interval between 2.5 s and 3 s. Therefore, all images with motion blur or artifacts were removed manually, and 60 frames with different dilation levels were selected from the remaining images. The selection consisted of keeping the images with a steeper slope in the curve of Figure 4.2, and randomly sampling the images in the plateaus until 60 images were selected. Therefore, all the videos contain exactly 60 valid frames in the dataset. We captured two videos of pupil reaction from each eye of each subject. From the 480 videos of the 120 subjects, 38 were eliminated since the number of available frames without motion blur or artifacts was less than 60. Therefore, a total of 442 were available from the 120 subjects. The total number of iris images available was 26,520.

The dataset was acquired from 120 subjects with an average age of 23.2±5.0 years old. Of these subjects, 67% were male and 33% were female. Of the 120 subjects, their iris colors were 48 dark brown, 49 light brown, 19 green, 3 blue, and one gray iris. The average minimum and maximum dilation levels per iris among the subjects were 0.24 and 0.54 respectively in the dataset. However, the overall minimum and maximum dilation levels were 0.16 and 0.77 respectively.

The real iris dataset was partitioned in the following manner: 96 subjects were selected randomly for training, 12 for validation, and 12 for testing. There are, therefore, 20,940 training images, 2,700 validation images, and 2,880 testing images. It is worth mentioning that we have OCT scans available of both eyes of one subject in the dataset. This subject was placed in the test set in order to assess the generalization capacity of the 3D models in comparison to OCTs of that subject. Finally, each video was captured at a resolution of 8 Mpx, and the iris diameter is 800 pixels on average. However, due to GPU limitations, we resized the iris images to a resolution of 256×256. The resized images are similar in size to iris images in current commercial iris sensors.



(a) (b) (c) (d)

Figure 4.3: Virtual iris formation from OCT images in Blender [108]. (a) The OCT is aligned with the *yz* plane and the contour, in orange, is traced. (b) The contour of the iris is used to make a revolution surface. (c) 3D texture is sculpted in Blender. (d) Four different examples of virtual irises with texture and dilation levels to simulate the variability of those parameters found in the real iris dataset.

4.3.3. Synthetic Iris Dataset

In order to acquire a synthetic iris dataset we used Blender, an open-source 3D-design application [108]. Blender can produce 3D models, simulate light sources and materials, render 2D images, and produce aligned depthmaps [108]. These characteristics allowed us to simulate LFVL illumination in virtual irises. We sculpted 100 virtual irises by obtaining texture information from the real iris dataset, and depth information from 36 OCT scans gathered from the internet. Figure 4.3 illustrates the process of sculpting irises using Blender. In this study, we define the *xy* plane as the same plane used in 2D iris images, while the *z* axis represents depth. First, one slide from one OCT is aligned with the *yz* plane. Then, the iris contour is carefully traced, and a revolution surface is created by revolving the OCT slice around the *z* axis. The 3D texture is then added to the model so that it will resemble that of the real iris. Each of the 100 virtual irises has a different dilation level, depth profile, and texture. To illustrate, Figure 4.3d shows four virtual irises that come from different OCTs, and therefore have different textures and dilation levels.

We then rendered synthetic iris images from those 3D models simulating LFVL illumination [24]. Thus, all the images have illumination sources from the side, and from the front. We used the same resolution of the real iris dataset, which is 256×256. In the synthetic images a virtual iris of 12.1 mm in diameter was assigned 230 pixels in the image. This diameter corresponds to the average diameter of a human iris [91]. Figure 4.4 shows examples of synthetic images and their respective depthmaps. The shadows in a synthetic image (Figure 4.4) are simulated from the interactions of LFVL light with the 3D relief of virtual irises (Figure 4.3). Next, we used data augmentation on the 3D models rather than on the 2D images to avoid aliasing and distortions. For this purpose, we changed rotation, translation, scaling, mirroring, and color in the 3D models. We used 4 colors, 9 positions, 5 rotations, 2 scales, and mirroring, generating a total of 720 images per each virtual iris. The synthetic iris dataset therefore has 72,000 images. Since the 3D information of each model is known, the corresponding synthetic images are accompanied by their aligned depthmaps. However, since color swapping produces the same depthmap, there are only 18,000 depthmaps in the dataset. The depthmaps were encoded using 8 bits (0-255). The scale range of 255 is equivalent to 1.936 mm in Blender for our virtual irises.

We also added eyelids, eyelashes, and reflections to the synthetic images, emulating the real iris dataset. This step also helps the networks to learn to predict depth information even in the

Figure 4.4: Examples of synthetic iris images, without eyelids, and their corresponding depthmaps.

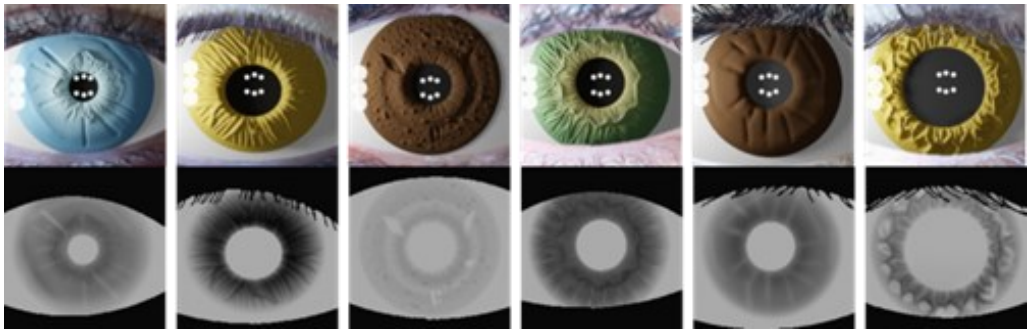


Figure 4.5: Examples of synthetic images with eyelids, eyelashes and reflections.

presence of specular highlights. This will also allow the network to learn how to segment eyelids and eyelashes from the iris. Figure 4.5 shows the synthetic images with the characteristics described. Eyelids were given a depth value of 10 on the scale of 0-255. This number was selected to avoid saturations during training using backpropagation.

We then partitioned the synthetic iris dataset randomly, using 80 virtual irises for training, 10 for validation, and 10 for testing. We thus have 57,600 synthetic images for training, 7,200 for validation, and 7,200 for testing. The synthetic iris dataset will be available on GitHub**:** https://github.com/dpbenalcazar/irisDepth.

4.3.4. Network Architecture and Training

In this work we trained several state-of-the-art CNNs to compare their performances in solving the iris depth estimation problem. We compared DenseDepth [20], DORN [97], and T^2Net [15], those that have demonstrated great depth estimation performance in outdoor and indoor settings. We then introduce irisDepth, which combines the GAN of T^2Net with the sophisticated depth prediction architecture of DenseDepth, to increase performance. Figure 4.6 shows the architectures of T^2Net and irisDepth. The yellow module $G_{S\rightarrow R}$ is a GAN that is shared in both networks. In order to use iris images with these networks, we added lateral illumination (LFVL) of the iris, which enhances shadows produced by iris features [24]. Thus, LFVL illumination allows the networks to relate shadows in RGB images to depth information. The networks were then trained to relate shadows in RGB images to depth information [105]. Both real and synthetic images were illuminated with LFVL in this work.

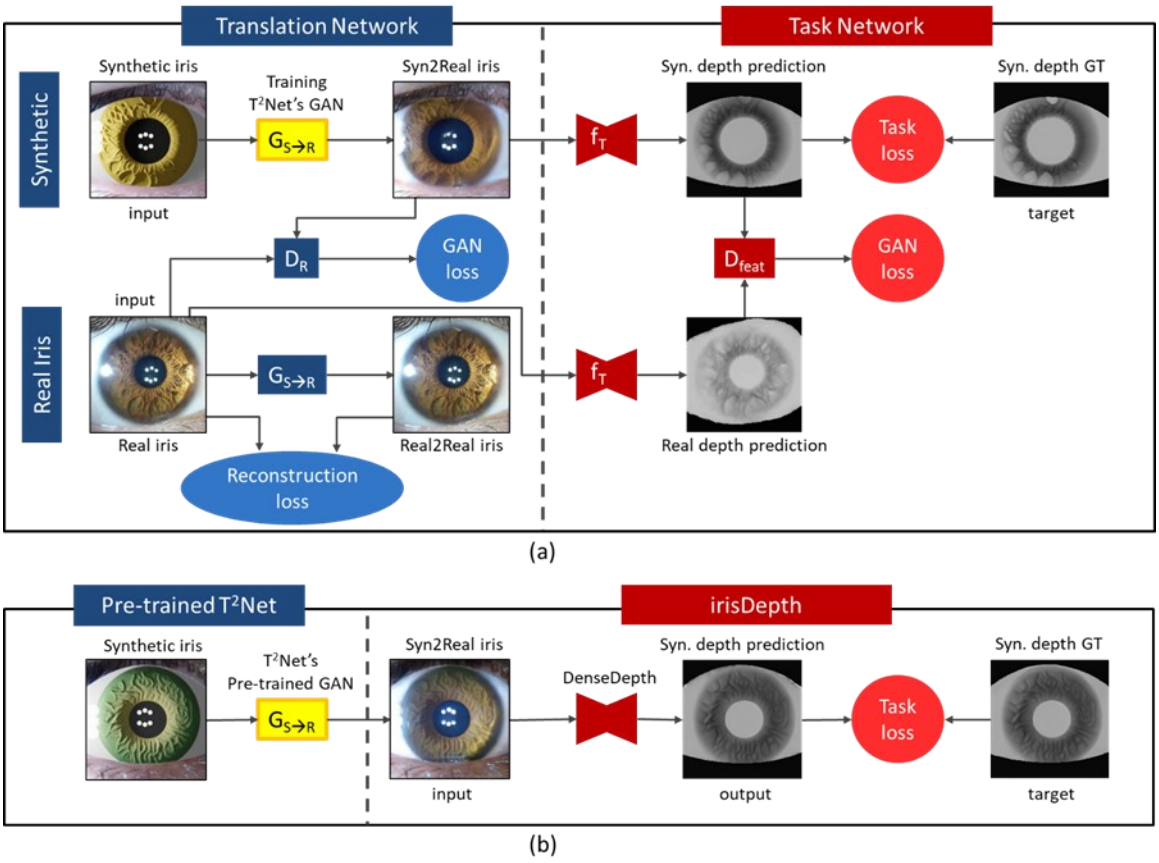


Figure 4.6: Architectures of T²Net [15] and irisDepth in the context of iris depth estimation. (a) T²Net consists of two parts translation, in blue, and task, in red. The translation network is comprised of a GAN that enhances the realism of synthetic images. The task part is comprised of an encoder-decoder architecture f_T, which makes depth predictions from real and translated images. (b) irisDepth uses the DenseDepth [20] architecture to improve depth prediction performance. A pre-trained T²Net GAN enhances the realism of synthetic images while leaving iris features aligned with the corresponding depth features. After training with realistic irises with aligned depthmaps, irisDepth can make depth predictions in real iris images. The yellow module $G_{S\rightarrow R}$ is first trained in (a), and then used in (b) to generate the inputs.

To make use of synthetic and real data in the training process, Zheng et al. described two training schemes, called vanilla and full [15]. In the vanilla approach, the translation component is trained first, and the task component is trained afterwards. In the full approach, both translation and task are trained simultaneously. In the context of iris depth estimation, the translation component performs domain adaptation to the synthetic iris images to look realistic, and the task component estimates the depth value of every pixel in the iris image. We used both vanilla and full approaches to train available state-of-the-art networks for 3D iris scanning with the datasets that were described in the Methodology, subsections 4.3.2 and 4.3.3.

For the vanilla approach, we trained CycleGAN [109][110] to perform domain adaptation on synthetic images. We used the synthetic iris images as the input, and the real iris images as the target. We trained the network using the train partition of both datasets, and the stop epoch was determined with the validation set. After that, we used CycleGAN to translate all 72,000 of the synthetic images, and thus formed a photo-realistic iris dataset. This dataset was partitioned identically to that of the synthetic iris dataset. Then, with the photo-realistic irises as the input, and the depth ground truth of the synthetic images as the target, we trained DenseDepth [20],

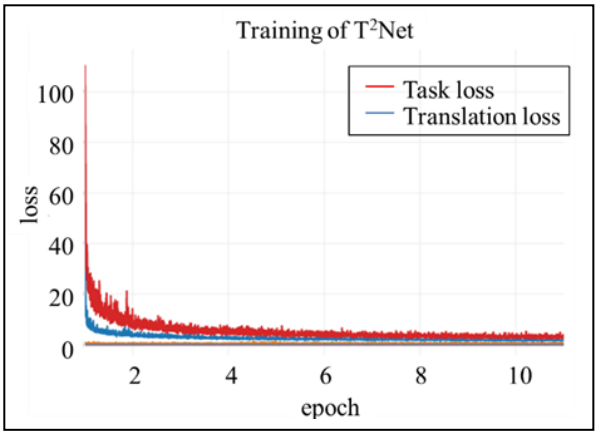


Figure 4.7: Evolution of the loss functions in the training process of T2Net, as an example of convergence.

DORN [97], pix2pix [100] and T^{2}Net [15]. In all these cases, we used the same networks available on the original code, with the exception of adjusting image sizes to 256×256. We used the The full version of T^{2}Net, shown in Figure 4.6a, was trained using a similar procedure. We also made no changes in the network architecture other than adjusting input and output image sizes. The GAN part of T^{2}Net ($G_{S\to R}$) is based on SimGAN in the generator and PatchGAN in the discriminator [15]. The task network (f_T) uses ResNet-50 in the encoder and up-sampling layers in the decoder [15]. Due to GPU constraints, we had to reduce image resolution to 192×192 only for this network. Then, we used the train partitions of both real iris and synthetic iris datasets as the input, and the depth ground truth of the synthetic images as the target. Using the validation partition, we determined the stop epoch.

We propose a method to increase performance by merging DenseDepth and T^{2}Net. As Zheng et al. described in their paper [15], the problem with the vanilla approach is that while the GAN could morph image features in favor of better appearance, those image features might no longer be aligned with depth features in the corresponding depthmap [15]. We experienced this phenomenon with CycleGAN. As a solution to this problem, we propose using the GAN prediction of a pre-trained T^{2}Net along with the auto-encoder of DenseDepth, instead of using a GAN that is blind to depth information. We call this approach irisDepth, and it makes use of the precision of DenseDepth while solving the main problem of the vanilla approach. Figure 4.6b illustrates irisDepth's architecture.

The following steps were performed for the purpose of using irisDepth in our problem: First, we changed the configuration of T^{2}Net to handle images with a resolution of 256×256. We used 6 down-sample layers in the transform network, 3 down-sample layers in the task network, 3 down-sample layers in the discriminator, and kept the rest of the parameters of the original configuration of T^{2}Net. Then, we trained T^{2}Net (Figure 4.6a) with our datasets, and used the validation set to find the stop point. Figure 4.7 shows the evolution of the translation and task loss functions. This illustrates an example of convergence with the proposed method. We then discarded the task part of this T^{2}Net, and used only its GAN at the best epoch for the next steps. This is the yellow $G_{S\to R}$ module in Figure 4.6. After that, we translated all the images in the synthetic iris dataset to obtain a realistic dataset. We partitioned this dataset to be identical to the synthetic iris dataset. Finally, we trained the standard version of DenseDepth using the train partition of the realistic dataset as the input, and the corresponding depthmaps of the original

synthetic images as targets, as illustrated in Figure 4.6b. In this way, our irisDepth uses a GAN with information about depth data and a robust auto-encoder for the task part.

4.3.5. Depth Evaluation with Synthetic Images

As one performance evaluation, we compared each network depth estimation capacity using the test set of 7,200 synthetic images. The goal of this test is to evaluate the depth estimation part of each network rather than the photo-realism of the translated images. The results of this test do not generalize to the performance on a real iris, but give a good indication of the precision of each network in the depth estimation task. First, the synthetic images were translated to the realistic domain using CycleGAN for the vanilla networks, as well as their respective GAN for the full networks. Both T^2Net and irisDepth have loss functions for the translation, as well as for the task part. Therefore, the networks perform domain adaptation instead of leaving synthetic images unchanged. Depthmaps were then predicted from the translated images using each network. Finally, we evaluated how similar the depthmaps that were predicted from the translated images were to the ground-truth depthmaps of the synthetic images. For this purpose, we used the standard metrics: Absolute Relative Difference (abs_rel), Squared Relative Difference (sq_rel), Root Mean Square Error (rmse), Logaritmic Root Mean Square Error (rmse_log), and the Accuracy Metrics (a_1, a_2 and a_3) [15][19][20][21][97][98][99]. The accuracy metrics a_1, a_2 and a_3 are computed using:

$$\mathrm{th}(u,v) = \max\left(\frac{\mathrm{depth}(u,v)}{\mathrm{GT}(u,v)}, \frac{\mathrm{GT}(u,v)}{\mathrm{depth}(u,v)}\right) \text{ [15]}, \tag{4.2}$$

$$a_n = (W \cdot H)^{-1} \cdot \sum_{u,v}(\mathrm{th}(u,v) < 1.25^n) \text{ [15]}, \tag{4.3}$$

where u and v are the coordinates of a pixel, depth(u,v) is the intensity of the predicted depthmap at the (u,v) coordinate, GT(u,v) is the intensity of the ground truth depthmap at the same coordinate, and n = {1,2,3}.

4.3.6. 3D Reconstruction of Human Irises

After all the networks are trained and tested, they can be used to generate depth estimates on human iris images. With an iris image and the predicted depthmap we can construct a 3D model of the iris. The 3D point-cloud model consists of a list of (x,y,z) coordinates of each 3D point. The x and y coordinates come directly from scaling the position of the pixels in the image, while the z coordinate is related to the depth value. If we use u and v to describe the horizontal and vertical position of a pixel in the image, and x, y and z to describe the 3D position of a point in the point-cloud model, the coordinates of such a point in millimeters are obtained by:

$$x(u,v) = \frac{13.47}{W}\left(u - \frac{W}{2}\right), \tag{4.4}$$

$$y(u,v) = -\frac{13.47}{H}\left(v - \frac{H}{2}\right), \tag{4.5}$$

$$z(u,v) = \frac{1{,}936}{255}(255 - \text{depth}(u,v) + \min(\text{depth})), \quad (4.6)$$

where W is the image width, and depth(u,v) is the intensity value of the predicted depthmap at the (u,v) coordinate. The constants in (4.4)-(4.6) depend on the size of the virtual iris and the distance to the camera. The constant 13.47 in the xy plane is computed assuming a design criterion where a virtual iris of 12.1 mm in diameter uses 230 pixels in the rendered image. Therefore, 256 pixels are 13.47 mm. The constant 1.936 mm is the maximum depth size equivalent of a variation of 255 levels in the depth map. Then, a 3D mesh model is formed by connecting neighboring points in the point-cloud. As a result, two 3D model representations are formed, and they are compatible with our previous SfM approach [25]. These models can easily be sliced and compared with OCT scans.

4.3.7. Depth Evaluation with OCT Scans

For one subject in the test dataset, we acquired four Anterior-Segment OCT slices of each eye, using the Visante™ OCT system [111]. These 8 OCT slices provide a ground truth for the evaluation of depth estimation from real iris images. First, we normalized the scale of the OCTs and rotated them so that the intersections of the cornea and the iris lay in a horizontal line. Then we used Canny edge detection to obtain the positions of the points on the iris surface. Figure 4.8 shows an example of the OCT with its corresponding iris surface in red. After that, one 3D model was estimated for each iris using real images of the same subject, and using the trained CNNs. We also produced one 3D model for each iris using the SfM 3D-iris-scanning method described in [25]. Then we sliced each 3D model using the same angles as in the available OCTs: 0°, 45°, 90° and 145°. To compensate for dilation differences between the OCTs and the iris images, we transformed the 3D model slices linearly to match the beginning and ending points of the irises. Finally, we compared each 3D model slice with the corresponding OCT slice, and measured the mean absolute error (MAE). The scale information on the OCT scans allowed us to calculate MAE in micrometers.

4.3.8. Resolution Assessment

We assessed the minimum depth that we could detect with our method, as well as the amount of error on all three axes. For the analysis, we manufactured and scanned 3D patterns of known dimensions. We printed them in 3D real truncated pyramids of various heights, as shown in

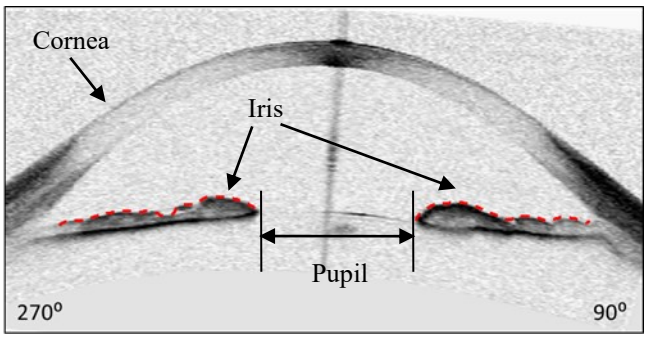
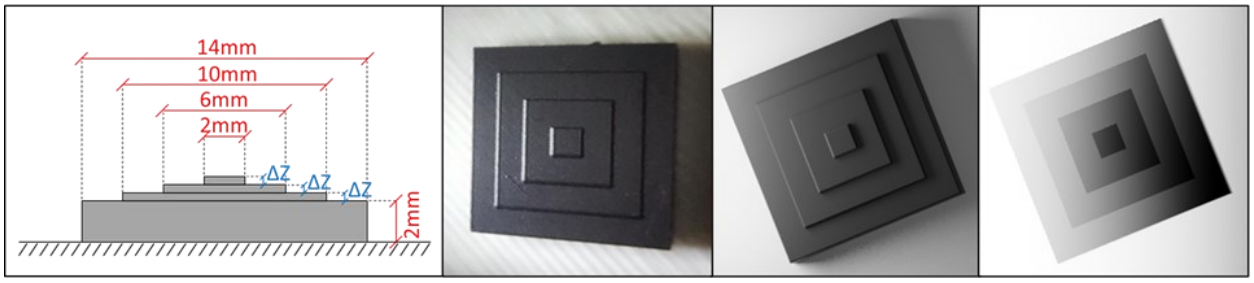
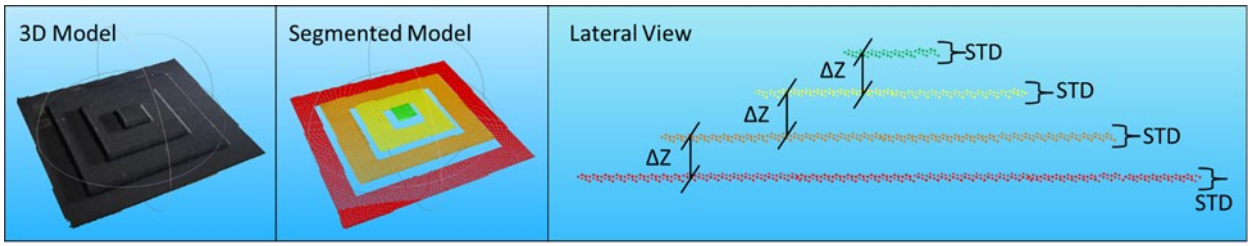


Figure 4.8: OCT edge detection of one subject in the test dataset (red dashed line).



(a) (b) (c) (d)

Figure 4.9: Truncated 3D pyramids for resolution assessment. (a) General shape of all 3D patterns. Red dimensions are fixed for every pyramid, while ΔZ, blue, change from model to model. Within a 3D model, all 3 steps have the same ΔZ value. (b) Example of a real image of the 3D patterns captured under LFVL illumination. (c) Example of a synthetic image of the 3D pattern, produced in Blender. (d) Depthmap of the synthetic image in (c).



(a) (b) (c)

Figure 4.10: Analysis of the 3D reconstructed truncated pyramid model. (a) Example of one 3D truncated pyramid model reconstructed by irisDepth. (b) Segmentation of each step of the truncated pyramid model. (c) Description of the height estimation of each step (ΔZ) and the standard deviation of the 3D points (STD) on the truncated pyramid model.

Figure 4.9a. The *x* and *y* dimensions of every step are fixed, and the step height ΔZ varies from 25 µm to 500 µm in increments of 25 µm. In total, we manufactured 20 pyramids for training and 5 for testing, using the FORMLABS FORM-2 stereolithography 3D printer. We set the 3D printer for the best resolution, which is 25µm per layer. Then we trained our irisDepth network with images of the 3D patterns. We used the same architecture and the same training scheme described in the Methodology, in subsection D. In this way, we used real images, as well as synthetic images with aligned depthmaps.

For the real pyramid image dataset, we used the same device and setup that was used for the iris images to assess the depth performance of our method. Figure 4.9b shows one image captured under these conditions as an example. We captured 360 images of the 20 real step pyramid printed 3D models, which included 6 different angles on the *z* axis and 3 angles on the *y* axis. We augmented the data using translation and scaling to produce a total of 7,200 images.

For the synthetic dataset, we used Blender to create 20 virtual pyramids with similar characteristics to those of the 3D printed ones. Then, we simulated the same LFVL illumination as was used in the synthetic iris dataset. Figure 4.9c shows an example of a synthetic image, and Figure 4.9d shows its corresponding depthmap. Using 3D data augmentation, we included 45 different angles from the *z* axis, and 6 angles from the *y* axis, rendering 5,400 synthetic images with aligned depthmaps. Then, using 2D data augmentation of 6 random translations and scales, we obtained 32,400 synthetic images. Finally, we partitioned the image dataset into 80% (25,920) for training, 10% (3,240) for validation, and 10% (3,240) for testing.

We followed the same procedure for training our irisDepth network with the real and synthetic pyramid dataset as was used for real and synthetic irises. Using the trained irisDepth network, we reconstructed five 3D models from images of the real truncated pyramids, one for each of the five different heights (from 25 μm to 500 μm in increments of 25 μm). We then measured the height of each step in the reconstructed pyramids along the *x* and *y* axes. Figure 4.10a shows a reconstructed 3D pattern, and Figure 4.10b and Figure 4.10c show the segmented version of the 3D pattern in Figure 4.10a. After that, we measured the average *z* value, as well as the standard deviation (STD), of the 3D points that form each step. Figure 4.10c shows the height of each step, and the mean step size ΔZ of the 3D model. We determined the measurement errors on each axis, using the absolute difference between the measured step on the 3D model and the measured step on the ground truth. The ground truth values (ΔZ_{GT}) were measured using a Mitutoyo 293-330 micrometer on the real truncated pyramids. The precision of the ground truth measurements is given by the micrometer precision, which is ±1 μm.

4.3.9. 3D Rubber Sheet Model Proof of Concept

As indicated in the Introduction, a 3D model of the human iris could be used in the future to improve accuracy in iris recognition. In this paper we explore a proof of concept of constructing a 3D rubber sheet from the 3D iris. Additionally, we evaluate iris recognition performance in the test set of 12 subjects.

With the purpose of building the 3D rubber sheet model, we applied a slicing procedure at regular intervals as described in the Methodology section, in sub-section 4.3.6. Each slice is a 2D curve that represents the relief of the iris in a radial manner. If the radial axis of the slices is normalized between 0 and 1, the 3D rubber sheet is resilient to dilation within certain ranges, as is the case with 2D rubber sheet models. The slices, then, obtained at different angles, are concatenated linearly to form a 3D structure. We built the 3D rubber sheet of the same subject used in the OCT test. We tested the similarity of a regular rubber sheet obtained from a 2D image [56] with the flattened version of the 3D rubber sheet. We tested separately the similarity using MAE, the zero-crossing normalized cross correlation (ZNCC) [112], as well as with the Hamming Distance (HD) [8] of the iris codes from both rubber sheets. A close similarity would indicate that our 3D models contain the same information on the *xy* plane as a 2D iris image; but we would have additional information available on the *z* axis to be exploited.

A preliminary 3D iris recognition method was implemented using a 3D rubber sheet model to extract 3D keypoints and descriptors, and to compare their distances. For this purpose, we constructed 480 3D rubber sheet models using 20 images per eye of the 12 subjects in the test set. We enrolled the 20 images with the dilation level closest to the median value of the subject, as recommended by Ortiz et al. [113]. We constructed the 3D rubber sheet models using 75 samples on the radial axis, and 360 slices on the angular axis. Our 3D rubber sheets, therefore, contain 75×360 = 27,000 3D points. Our proposed method for iris recognition in 3D has the following steps: First, we sample the 3D rubber sheet model with a 4×15 grid to find 60 keypoints. Then, we obtain the Spin Image descriptor [33][34] for each keypoint. Finally, we assess the similarity of two 3D rubber sheet models as the average ZNCC [112] between corresponding Spin Images on the sampling grid. As with the 2D iris code, we account for small angular displacements by translating the 3D rubber sheet ±5° and storing the best result [8].

We compared the iris recognition performance of our 3D proposed method with that of the 2D iris code. For this purpose, we obtained the 2D rubber sheets and iris codes of the 480 images in the test set using Osiris V4.1 [114]. We then used the *d'* index to score iris recognition performance [8]. This index shows how well we can separate intra-class from inter-class comparisons, and it is computed using:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{0.5*(\sigma_1{}^2 + \sigma_2{}^2)}} \text{ [8]}, \tag{4.7}$$

where μ_1 and μ_2 are the mean values of the intra-class and inter-class distributions, respectively, and σ_1 and σ_2 are the standard deviations (STD) of both distributions. The higher the d' value, the easier it is to separate intra-class from inter-class distributions.

4.4. Results and Analysis

4.4.1. Resolution Assessment

The results on the 3D real truncated pyramids of different step sizes are as follows: Table 4.1 shows the five sizes for the 3D printed pyramids of the test set with a designed step size (ΔZ_{Design}) of 25 µm, 50 µm, 100 µm, 200 µm, and 400 µm. The values of the step size measured with the micrometer are the ground truth for our depth measurements (ΔZ_{GT}); the mean step sizes measured in the 3D reconstructions (ΔZ); the standard deviation of the 3D points that form each step (STD); as well as the absolute errors measured along each *z* (Z_{err}), *x* (X_{err}), and *y* (Y_{err}) axis.

TABLE 4.1:
DESIGNED AND MEASURED STEP SIZES AS WELL AS ESTIMATED ERRORS ON THE *X*, *Y* AND *Z* AXES, FOR THE 3D TRUNCATED PYRAMIDS

ΔZ_{Design} [µm]	ΔZ_{GT} [µm]	ΔZ [µm]	STD [µm]	Z_{err} [µm]	X_{err} [µm]	Y_{err} [µm]
25	29.6	31.9	18.4	2.3	36.7	33.3
50	55.9	55.9	16.6	3.0	24.3	46.8
100	101.9	98.0	17.6	3.9	59.9	49.8
200	206.2	202.2	17.3	4.0	46.3	47.9
400	402.4	395.3	18.6	7.1	40.4	42.8
		Average:	**17.7**	**4.1**	**42.0**	**43.6**

The results of Table 4.1 show that the measured step size ΔZ is close to the ground truth value (ΔZ_{GT}) for all five 3D patterns. The average absolute error on the *z* axis is 4.1 µm. The standard deviation represents how much the 3D points deviate from a perfect plane [25]. Its average value is 17.7 µm. This means that a feature on the *z* axis that is smaller than 17.7 µm is within the noise level of the 3D points. Features larger than 17.7 µm, however, can be detected by our system. Therefore, the resolution limit of our method is 17.7 µm. This figure is about 1/30th of the iris thickness [91]. Additionally, the resolution limit of 17.7 µm is almost twice as high as the 10µm of conventional OCT scans, as well as the 11 µm reported in [25] for SfM. Our results show a reasonable level of precision from a single 256×256 image.

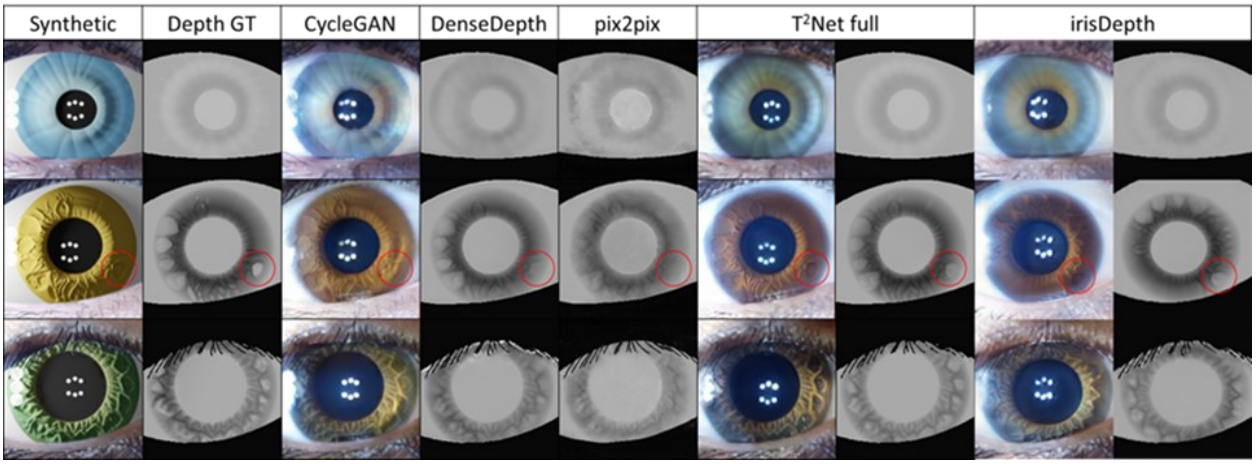


Figure 4.11: Examples of depth estimation using synthetic images. Each row is a different example. The first two columns are synthetic images and their corresponding ground truth depthmap (Depth GT). The succeeding columns show the outputs of each network. DenseDepth and pix2pix make depth predictions from the synthetic images translated by CycleGAN. T^2Net and irisDepth make depth predictions from the results of their own GANs. The second row shows a red circle highlighting an iris feature that can be followed into the corresponding depthmaps.

The scale values on the OCT scans, as well as equations (4.2), (4.3) and (4.4) allow estimating the theoretical resolution of our method. According to (4.2) and (4.3), a variation of 1 mm on the *x* or *y* axis produces a variation of 19 pixels for the 256×256 images. Therefore, the resolution of the 3D model on the *xy* plane is 52.6 μm/px. This figure is around $1/230^{th}$ of the iris diameter [91] and can be improved by increasing image resolution. For instance, if we used 800×800 images, equations (4.2) and (4.3) yield a resolution of 16.8 μm/px. A variation of 1mm on the *z* axis produces a depth change of 132 on the depth scale between 0 and 255. Therefore, the resolution on the *z* axis is 7.56 μm. Measurements are therefore 7 times more precise along the *z* axis than on the *xy* plane. These figures roughly match those shown in the experimental results of Table 4.1, where there is almost 10 times more error along the *xy* plane than on the *z* axis.

4.4.2. Depth Evaluation with Synthetic Images

This test illustrates the precision of each network in the depth estimation task. The ground truth in this experiment comes from the depthmaps in the synthetic iris dataset, while the inputs are translated images. Figure 4.11 shows examples of ground truth synthetic images in the test set, results of translated images, and the network predicted depthmaps. The vanilla networks, such as DenseDepth, DORN, and pix2pix, make up a depth estimation from the photorealistic images produced by CycleGAN. T^2Net and irisDepth make depth estimations from the output of their own GANs. Figure 4.11 also illustrates the problem of training a GAN blindly from depth estimation. The ground truth example in the second row has a concave feature highlighted with a red circle. Since this feature is not reproduced by CycleGAN, neither DenseDepth nor pix2pix can estimate its depth. However, the GANs trained in the full approach learn to reproduce this feature. Both T^2Net and irisDepth were able to estimate the depth of this concave feature correctly.

The results of the depth evaluation with the 7,200 synthetic images in the test set are presented in Table 4.2. For abs_rel, sq_rel, rmse, and rmse_log metrics, a lower value means a better result,

while for a_1, a_2 and a_3, a higher value is better [15][19][20]. The accuracy metrics a_n are computed using (4.2)-(4.3). The best result of each column was highlighted in bold. Table 4.2 shows that irisDepth produced the best results on almost all the tests. DenseDepth and DORN also produced good results due to their specialized architectures in depth prediction tasks. IrisDepth produced the best overall results since it combines a GAN that has information on depth data, and the powerful depth prediction architecture of DenseDepth.

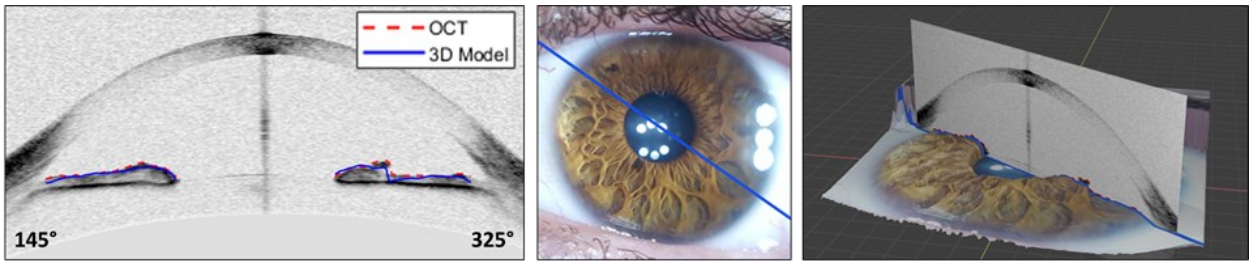
TABLE 4.2:
SIMILARITY USING STANDARD METRICS BETWEEN DEPTHMAPS PREDICTED FROM THE TRANSLATED IMAGES AND DEPTHMAPS OF THE SYNTHETIC IMAGES IN THE TEST DATASET OF 7,200 IMAGES.

Method	abs_rel	sq_rel	rmse	rmse_log	a_1	a_2	a_3
DenseDepth [20]	0.0520	1.468	9.876	0.1150	0.9651	0.9952	0.9976
DORN [97]	0.0525	1.775	11.893	0.1203	0.9714	0.9895	0.9931
pix2pix [100]	0.2219	9.498	16.039	0.4360	0.7302	0.8630	0.9174
T^2Net_vanilla [15]	0.0715	2.972	8.649	0.1849	0.9567	0.9847	0.9881
T^2Net_full [15]	0.0576	1.863	**7.350**	0.1440	0.9619	0.9903	0.9931
irisDepth (ours)	**0.0475**	**1.161**	8.878	**0.1105**	**0.9728**	**0.9958**	**0.9978**
	Lower is better				Higher is better		

4.4.3. Depth Evaluation with OCT Scans

We also assessed the performance of our method by comparing the generated 3D models against the depth ground truth provided by iris OCT slices. Figure 4.12 shows the comparison between an iris 3D model slice and the corresponding OCT. Figure 4.12a shows the OCT image with markings of the ground truth iris surface, and the slice of the iris 3D model. Figure 4.12b illustrates the angle of the slice and the iris features that are present along this line. Figure 4.12c shows a spatial comparison of the 3D model with the OCT. This visual comparison illustrates the changes in the 3D model across the profile, and shows how they closely match the OCT.

We then compared the difference quantitatively between the ground-truth iris surface in the OCT slices and the corresponding slices of the 3D models produced by both SfM, and the different CNNs trained in this work. Figure 4.13 shows close-up comparisons between OCT slices and all the various 3D models produced by the different methods. Figure 4.13 shows that the models produced by DenseDepth, T^2Net_full, and irisDepth follow the depth ground truth of the OCT closely. The model produced by SfM has a great resemblance on the left side, but a



(a) (b) (c)

Figure 4.12: Example for the comparison between the OCT slice with the corresponding iris 3D model slice. (a) OCT slice with the ground-truth iris surface in red, and the 3D model slice in blue. (b) An iris image showing the slice angle in blue. (c) The 3D iris model with the OCT superimposed at the same angle.

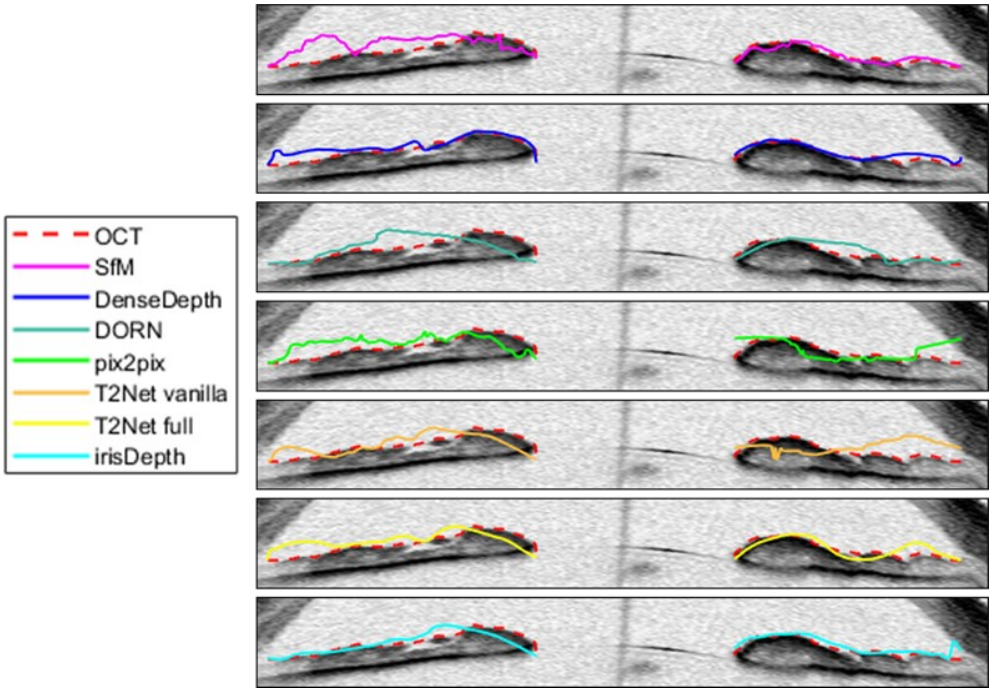


Figure 4.13: Comparison of one OCT slice of the iris with all the 3D models produced by SfM, and all the trained CNNs.

significant difference on the right side of the iris. For each method, we have the curve of the OCT ground truth, and that of the 3D model slice. We computed the mean absolute error to quantify the error between both curves. We compared the 3D iris models that are produced by each method for the left and right eyes of the subject to the total of 8 available OCT slices for the right and left eyes (4 for each eye). Table 4.3 shows the results of the mean absolute error in micrometers when comparing each 3D iris model to the ground truth (OCT). The minimum average error of 77 µm was obtained with our model irisDepth. The typical thickness of the iris is around 500 µm [91], and therefore, the error achieved with the irisDepth method is within 15% of the thickness. Figure 4.13 also shows that irisDepth is the method that follows the ground truth the most closely. SfM produced the second to last good performance, and the error of SfM is 60% greater than the best CNN method (irisDepth). This indicates that the CNN irisDepth produces a more accurate 3D model from a single image than was achieved with SfM from multiple images.

TABLE 4.3:
DEPTH ESTIMATION ERRORS IN µM BETWEEN 3D MODEL SLICES AND OCT SCANS.

Method	Left Eye MAE [µm]	Right Eye MAE [µm]	Average [µm]
SfM [25]	111	134	123
DenseDepth [20]	74	98	86
DORN [97]	71	115	93
pix2pix [100]	92	129	111
T^2Net_vanilla [15]	119	160	140
T^2Net_full [15]	92	92	92
irisDepth (ours)	**69**	**86**	**77**

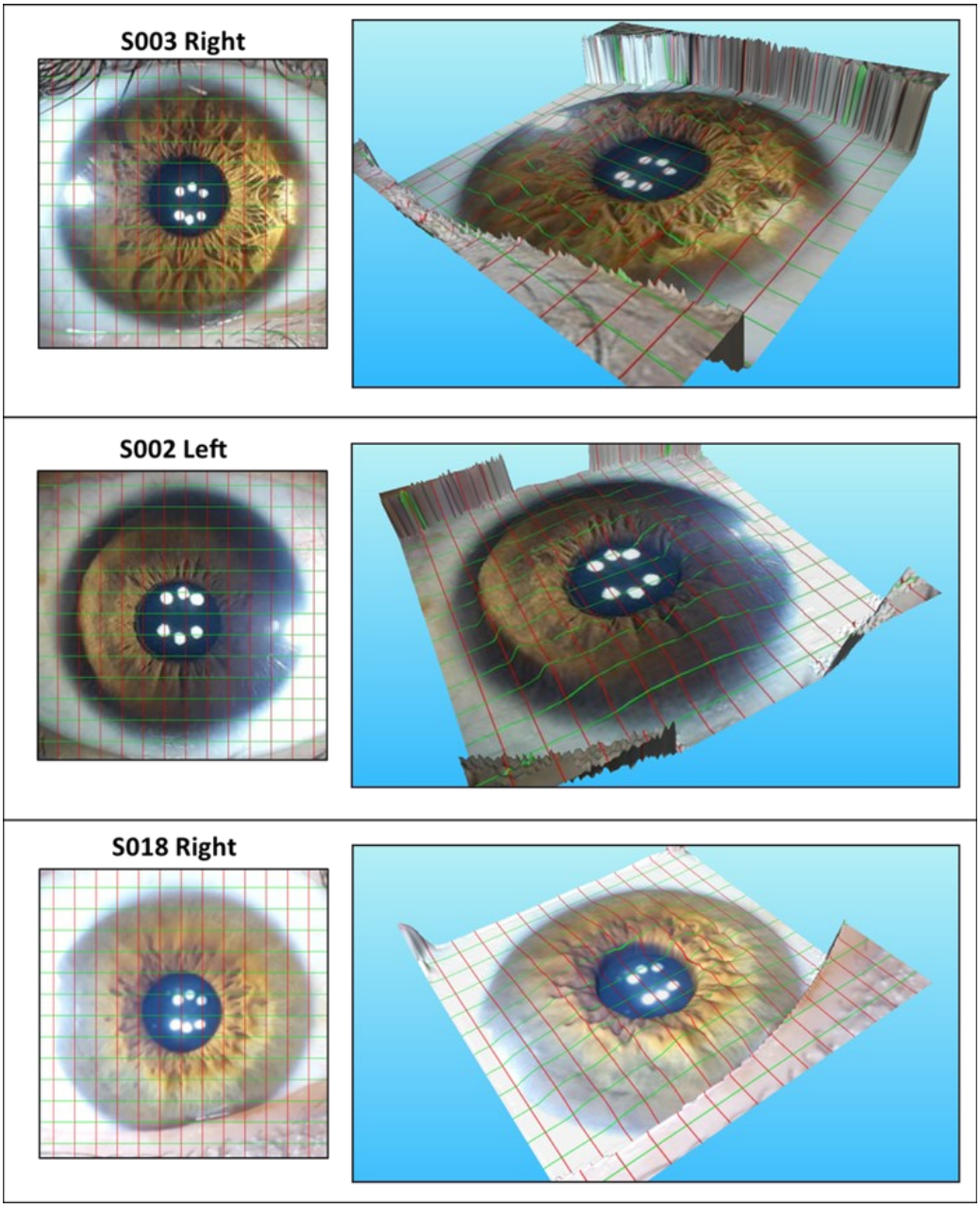


Figure 4.14: Example of human 3D iris models. A red-green grid was drawn on the surface of the models for a better visualization of 3D features. Each row shows the iris image of the subject on the left, and the iris 3D model on the right.

4.4.4. 3D Reconstruction of Human Irises

We produced point-cloud and mesh 3D models of the subjects in the test set using irisDepth. Figure 4.14 shows examples of 3D mesh models for five different subjects. For the purpose of appreciating the 3D information in a 2D image, a red-green grid was drawn on the surface of the 3D model. In this way, deformations in the grid illustrate depth variations across the iris surface. This figure also shows the estimation of the 3D information performed by irisDepth from a single image of the human iris. The point-cloud models produce depth predictions from every pixel in the image. At a resolution of 192×192, the models have 36,864 3D points, and at 256×256pixels,

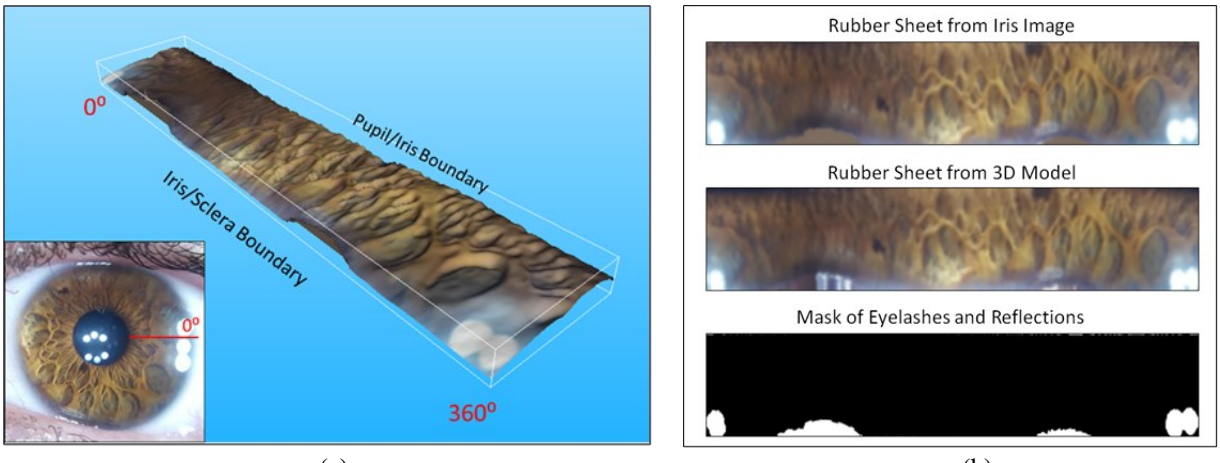
there are 65,536 3D points. In contrast, the SfM method reported an average production of 11,105 3D points [10]. Therefore, our CNN approach has more information available for producing the 3D model of the iris compared to that of the SfM approach.

Our results show that there are advantages to using CNNs over SfM for 3D iris model generation. Besides using multiple images at a greater resolution, SfM has problems producing 3D points in areas of the iris that have no significant texture. In contrast, the CNN models produce a uniform distribution of points regardless of iris texture. The number of 3D points obtained by CNNs is always constant, and it can be 6 times greater than those of SfM. Additionally, artifacts such as lateral reflections produced noisy points in the SfM model. One of the main advantages of our proposed method is that CNNs require only a single image for the 3D model estimation. This saves acquisition and processing time, as well as storage space. The acquisition time is relevant for subjects in the use of biometric applications. The SfM approach [10] requires capturing a burst of about 10 images per camera position for the 3D model construction. A set of one-hundred 16Mpx images, therefore, is typically used to reconstruct a single 3D model. Consequently, obtaining a 3D model from a single image is a significant improvement.

4.4.5. Rubber Sheet Model and 3D Iris Recognition Proof of Concept

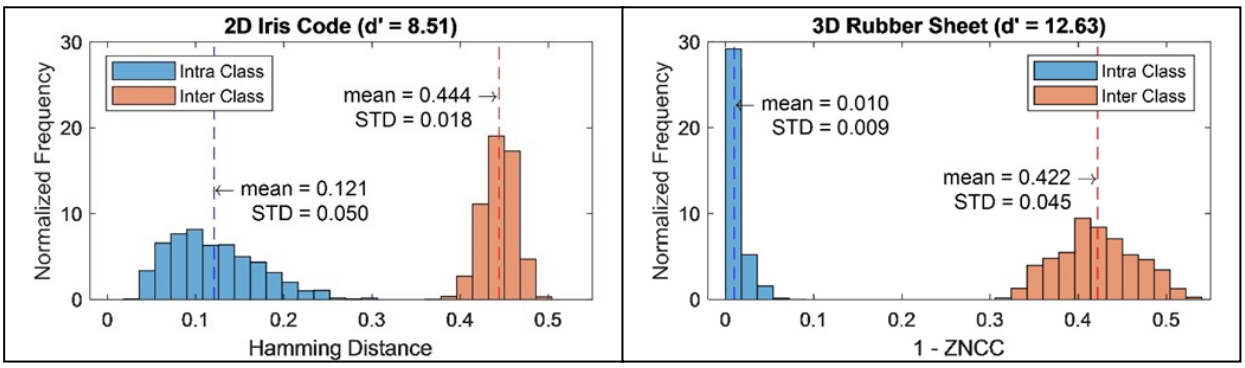
We reconstructed the 3D rubber sheet from the 3D model in Figure 4.12c by obtaining one 2D slice every 1°. The 3D rubber sheet is shown in Figure 4.15a. The 3D rubber sheet captures the color information of the 2D image, as well as the depth of the iris. Just like a 2D rubber sheet, this is a representation of the human iris that normalizes dilation changes in a linear manner [2].

We then compared the rubber sheet from the iris image of Figure 4.12b with the projection of the 3D rubber sheet of Figure 4.15a onto the *xy* plane. Figure 4.15b shows the resulting rubber sheets, as well as the mask of eyelids, eyelashes, and reflections. This mask was used to ensure



(a) (b)

Figure 4.15: (a) 3D Rubber Sheet obtained from 360 slices of the 3D model in Figure 4.12c. The iris image of the subject is shown on the bottom left corner along with the 0º line of the slicing process. (b) Comparison between the rubber sheet of an iris image in Figure 4.12b and the flattened version of the 3D rubber sheet in (a). A mask was used in the comparison to avoid the effects of eyelids, eyelashes, and reflections **[2]**.



(a) (b)

Figure 4.16: Iris Recognition performance in our test set of 12 subjects and 480 images. (a) Using the 2D Iris Code in Osiris V4.1 [8][114]. (b) Using the proposed 3D rubber sheet model, with Spin Image Descriptors [33][34] and ZNCC [112].

that those artifacts would not affect the comparison. The resulting MAE value for the comparison is 0.0313; ZNCC is 0.9385; and HD is 0.226. These values indicate a small error and a large correlation between the two images. This means that the reconstructions of the 3D model and the 3D rubber sheet preserve the information along the *xy* plane with a small error. Additionally, the low HD ensures a true positive in biometric tests. For context, in a previous work, we analyzed that the mean intraclass HD of LFVL images is 0.243, while that of the interclass distribution is 0.48 [15]. Therefore, the HD value of 0.226 falls in the range of two different images from the same individual.

The results of the 3D iris recognition are presented in Figure 4.16 which shows the iris recognition performance of the 3D rubber sheet compared to that of the 2D iris code in our test set of 12 subjects and 480 images. The distributions in Figure 4.16 are normalized so that they have an area of 1. The results with the 2D iris code yielded a *d'* of 8.51, using Osiris V4.1. The 3D rubber sheet achieved a *d'* of 12.63, which is 48% higher. The mean value of the intra-class distribution is similar for both methods, with a value of approximately 0.45. However, the mean value of the intra-class distribution is 0.111 units less for the proposed 3D method. The results of this preliminary test show that the 3D characteristics extracted from the human iris are more discriminative than the 2D iris code.

The preliminary results of iris recognition in the test set of 12 subjects, along with the proof of concept of the Rubber Sheet model, and the depth evaluation tests with 8 ground-truth OCT slices of one subject illustrate the capabilities of the proposed method to reconstruct the surface of the human iris, and its applications in iris recognition. The tests with stepped pyramids of known dimensions demonstrate the smallest resolution our method can measure. All these evaluations show that our method can reconstruct a 3D model of the human iris with good performance.

4.5. Conclusions

Our proposed method for 3D iris model estimation from a single image produced complete 3D representations of the human iris using CNNs. Our method, irisDepth, uses the GAN part of a pre-trained T^2Net with the depth prediction of DenseDepth. Therefore, the GAN is not blind to

depth information during training, and the depth prediction is more powerful than T^2Net alone. IrisDepth produced the best performance among the trained networks in both the synthetic and real iris tests. We used a dataset of 96 subjects randomly selected for training, 12 for validation and 12 for testing. There are 20,940 training images, 2,700 validation images and 2,880 testing images. We also used synthetic irises with 72,000 images. Both datasets used lateral illumination of the iris (LFVL) to enhance the shadows produced by iris features [15]. Thus, lateral illumination allowed the networks to relate shadows in RGB images to depth information.

We validated the results of our method for modeling the human iris by comparing slices of the 3D models with corresponding OCT slices of both eyes of one subject. The overall shape of the 3D models matches that of the OCT. Our method produced 65,536 3D points, with an absolute error of 77 μm on average. These numbers represent 6 times more 3D points and a 60% increase in accuracy with respect to previous 3D iris models based on SfM [25]. We proposed a 3D rubber sheet model proof of concept, which had a 0.9385 correlation with a 2D rubber sheet on the *xy* plane, and additional information on the *z* axis to be exploited. On a preliminary test with 480 images, the proposed 3D rubber sheet model increased iris recognition performance by 48% with respect to the standard 2D iris code [8]. Finally, the resolution of our method is 17.7 μm, as was measured by scanning 3D pyramids of known dimensions. This is roughly $1/30^{th}$ of the iris thickness.

A 3D model of the iris may open research lines in iris recognition and ophthalmology. In addition to increasing accuracy in iris recognition [35], obtaining 3D information of the iris could help in extreme pose detection [11][115][116][117][118]. Additionally, a 3D model of the iris could produce information similar to that of an OCT, which could help ophthalmologists in the detection of closure angle glaucoma [25][26].

Future improvements could increase the precision of our method. First, modifying the architecture to train with OCT slices or OCT based 3D models would produce 3D iris models that correlate more closely with actual OCT scans. Also, although CNN and SfM are traditionally used separately, a combination of them could yield a more robust method [103]. The CNN prediction could be the starting point for SfM, which could output more 3D points from several views at a higher resolution, thus improving the 3D model [12][13].

4.6. Acknowledgements

This research has been funded by CONICYT through project FONDECYT 1191610, by the Department of Electrical Engineering, and Advanced Mining Technology Center (CONICYT Project AFB180004), Universidad de Chile. We would like to thank Prof. Javier Ruiz-del-Solar and Dr. Patricio Loncomilla from AMTC for their insights on training CNNs. The acquisition of the OCTs for this research was possible thanks to the help of Dr. Claudio I. Perez from Fundación Oftalmológica Los Andes. We would also like to thank the students at the School of Engineering, Universidad de Chile, who participated enthusiastically as volunteers for iris image acquisition.

Chapter 5

3D Iris Recognition using Spin Images

The high demand for ever more accurate biometric systems has driven the search for methods that reconstruct the iris surface in a 3D model. The intent in adding the depth dimension is to improve accuracy even in large databases. Here, we present a novel approach to iris recognition from 3D models. First, the iris 3D model is reconstructed from a single image using irisDepth, a CNN based method. Then, a 3D descriptor called Spin Image is obtained for keypoints of the 3D model. After that, matches are found between keypoints in the query and the reference 3D models using k-dimensional trees. Finally, those keypoint matches are used to determine the spatial transformation that best aligns the 3D models. A combination of the transformation error and the inlier ratio is used as the metric to assess the similarity of two iris 3D models. We applied this method in a dataset of 100 eyes and 2,000 iris 3D models. Our results indicate that using the proposed method is more effective than alternative methods, such as Daugman's iris code, point-to-point distance between the 3D models, and the 3D rubber sheet model model, and CNN-based methods.

5.1. Introduction

Iris recognition is one of the most successful biometric techniques because the human iris has great variability among individuals [1][8]. For instance, Daugman's iris code can achieve a false match rate of only 1 in 4 million under a fractional Hamming Distance of 0.33 [2]. This is why, from small to large scale, iris recognition applications are in use, such as India's Unique ID program [6], the United Arab Emirates' border-crossing control [5], and the voter registration list in Somaliland [7].

Although iris recognition is a mature technology, it still has some limitations. For instance, a false match rate of 1 in 4 million in a dataset of one billion individuals would mean that 250 people have similar iris codes [8][9][25]. In addition, the area of the iris in the image is usually occluded by eyelids, eyelashes, and reflections [25]. This limits the amount of information that can be extracted from an iris image. To improve iris recognition performance under these limitations, systems that produce 3D representation of the iris have been proposed [25][23] [31][35][36]. It is theorized that adding depth information can improve accuracy when only a portion of the iris is available for analysis [25][31]. A 3D representation could also improve iris recognition performance under extreme poses [11][31], just as it does in face recognition [116][118][117].

One of the most accurate methods for 3D iris reconstruction is irisDepth [31]. It produces an iris 3D model from a single image under lateral and frontal visible light (LFVL) [24], using Convolutional Neural Networks (CNN). This architecture outperformed existing depth estimation methods [15][20][97], as well as a Structure from Motion (SfM) method [12][13][25], in iris 3D reconstruction against 8 Optic Coherence Tomographies (OCT) as the ground truth. In this work, we retrain irisDepth using a different partition of the dataset used in [31], which allowed us to test in 50 subjects. We use both 5, and 20 images, per eye to obtain test sets of 500 and 2,000 iris images.

Our method consists of computing 3D keypoints and descriptors to assess the similarity of two iris 3D models [70]. We propose using Spin Images for the 3D descriptors since that technique has excellent results in 3D object and face recognition [33][34][119][120][121][122][123][124][125]. We compare our results with Daugman's iris code, the mean point-to-point distance between 3D iris models [35][36], the 3D rubber sheet model [31], as well as CNN based iris recognition methods [126][127].

Our contributions presented in this paper are the following: (1) developing a novel approach to iris recognition by means of iris 3D models and Spin Images, (2) testing the iris recognition performance of our method on a dataset of 2,000 iris 3D models, and (3) comparing the performance of our method with that achieved by state-of-the-art 3D and CNN methods.

5.2. Related Methods

A method to reconstruct a 3D model of the iris surface was introduced by Bastias et al. [23]. This method utilized near-infrared (NIR) iris images acquired from several perspectives, and SfM [12][13] to reconstruct a point cloud iris 3D model [23]. Benalcazar et al. [25] improved this method by utilizing LFVL images, which enhance iris texture [24]. The improved method used SfM [12] [13] to create a dense point cloud model, and a mesh 3D model of the iris. The resulting 3D model incorporates both color and depth information of the iris surface. This method produced an average of 11,105 points and 56,605 polygons with a precision of 11μm [25].

Cohen et al. [35][36] used a different approach to reconstruct an iris 3D model, employing triangulation to produce 3D salient points from fiducial points in a set of two or more NIR iris images [35][36]. Then, loop subdivision was used to create a denser set of feature points and a high-resolution iris 3D mesh model [35][36]. They then used their iris 3D models for iris recognition in 3D. They used a dataset of 20 base models and 20 test models. The similarity between two models was evaluated using the mean square error (MSE) between each fiducial point in the test model and the closest fiducial point in the base model [35][36]. They were able to classify the iris 3D models in their dataset [35][36] with only a few false positives.

Benalcazar et al. [31] developed irisDepth, a CNN method that estimates the 3D model from a single image. This method estimates the depth value of every pixel in the iris image. At a resolution of 256×256, the iris 3D models have 65,536 evenly distributed 3D points and 130,050 polygons regardless of iris texture [31]. The architecture of irisDepth incorporates the Generative Adversarial Network (GAN) of T2Net [15] with the depth prediction network of denseDepth

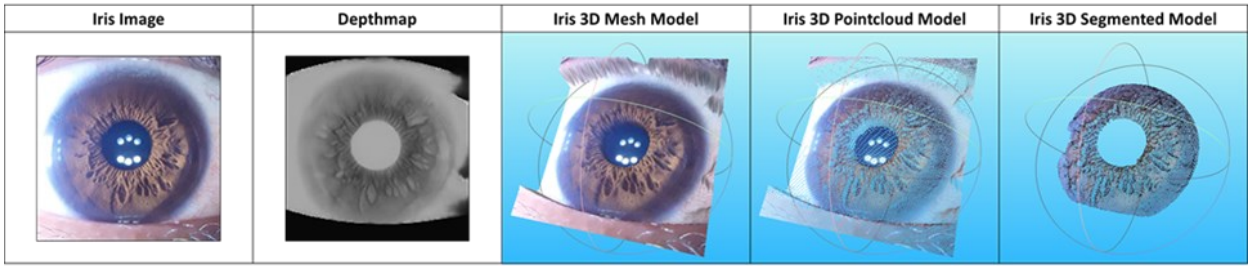
[20]. Performance of several methods was evaluated by computing the mean absolute error between the iris 3D models produced against 8 OCT slices. IrisDepth outperformed DenseDepth [20], DORN [97], T2Net [15], and SfM [25] in this test. Also, [31] proposed a 3D rubber sheet model for iris recognition. The method consists of changing the iris 3D model from cartesian to polar coordinates, as in the 2D rubber sheet model. Then, 60 fixed keypoints in a square grid are sampled, and Spin Image descriptors are obtained [33]. The mean correlation of corresponding keypoints was used to score the similarity between two 3D models. Results showed a 48% increase in iris recognition performance with respect to Daugman's iris code in a test set of 12 subjects and 480 images.

5.3. Methodology

Our work is a novel approach to iris recognition in 3D using keypoints and descriptors. The iris 3D models in this work come from depthmap predictions of irisDepth using a single image. We improved the method proposed in [31] by extracting 15 times more keypoints to find the best transformation between two models, and also by eliminating a rigid alignment between models assumed in [31]. This section describes the iris-3D-model dataset, the computation method for keypoints and descriptors, the comparison method between two iris 3D models, as well as the performance evaluation tests using our proposed method, Cohen's MSE distance [35][36], Daugman's iris code [8], and the 3D rubber sheet [31].

5.3.1. Iris-3D-Model Dataset

IrisDepth was trained using 72,000 synthetic images, as well as 65,520 real images from 120 subjects [31]. In this work, we retrained irisDepth using the same datasets, but we split 65 subjects for training, 5 subjects for validation and 50 subjects for iris recognition tests. Using the 50 test subjects, and capturing 20 images per eye, we formed a test set of 2,000 iris images. We call this test set: Set_{2000}. For each subject, the 20 images with the dilation level closest to the median value were enrolled, as recommended by Ortiz et al. [113]. The dilation level is dimensionless since it is the quotient between the pupil radius and the iris radius. Among the 50 subjects, the median dilation level varies between 0.23 and 0.52. In the literature, it is common to use five images per eye in iris recognition tests [8][9][24][45][60][65][95] which is why we also obtained a subset of 5 images per eye from those 50 test subjects for a total of 500 images. We call this test set: Set_{500}. As expected, there is less dilation variation in the Set_{500} than that of the Set_{2000}. In average, the dilation level changed in ±0.012 and ±0.028 for a subject in the Set_{500} and in the Set_{2000}, respectively. All iris images were captured under the same illumination conditions using LFVL [24] at a resolution of 256×256. All the iris images in the dataset were normalized so that the iris has a diameter of 230 pixels, and is placed in the center of the image.



(a) (b) (c) (d) (e)

Figure 5.1: Iris 3D model formation and segmentation. (a) Input iris image. (b) Depthmap predicted by irisDepth [31]. (c) Resulting iris 3D model in mesh format with 130,050 polygons. (d) Resulting iris 3D model in point-cloud format with 65,536 3D points. (e) Segmented model with 29,799 3D points after removing eyelids, eyelashes, the pupil, the sclera, and reflections from (d).

Then, we used irisDepth [31] to produce a depthmap prediction for each one of the 2,000 iris images. Figure 5.1a shows an iris image and Figure 5.1b shows the predicted depthmap. To make our method robust to dilation changes, we linearly normalized the dilation level of the iris images and depthmaps to the median dilation level of each subject. Then, the iris 3D models were constructed in the point cloud format using the following equations:

$$x(u,v) = \frac{13.47}{W}\left(u - \frac{W}{2}\right) \text{ [31]}, \tag{5.1}$$

$$y(u,v) = -\frac{13.47}{H}\left(v - \frac{H}{2}\right) \text{ [31]}, \tag{5.2}$$

$$z(u,v) = \frac{1.936}{255}(255 - D(u,v) + \min(D)) \text{ [31]}, \tag{5.3}$$

where x, y, and z are the coordinates in mm of a 3D point in the point cloud model; u and v are the coordinates of a pixel in the iris image; H and W are the image's height and width; and finally, $D(u,v)$ is the intensity value of the predicted depthmap at the (u,v) coordinate [31]. The constants 13.47 and 1.936 are scale factors that preserve the aspect ratio [31]. Figures 5.1c and 5.1d show the resulting iris 3D model created by applying (5.1), (5.2), and (5.3) on the image of Figure 5.1a, and the depthmap of Figure 5.1b.

5.3.2. Post-processing Stage

The depthmaps produced by irisDepth encode eyelid segmentation as a small depth value [31]. For example, the eyelids in Figure 5.1a are segmented in black in Figure 5.1b. Using this information, eyelids and eyelashes can be removed from the 3D model. The pupil and the sclera can be removed using circular segmentation. We segmented the 2,000 images in the dataset using an improved version of Daugman's method [8]. The main difference in our improved version is that we search for the pupil in the saturation channel of the HSV color space [128]. This increases accuracy in LFVL images [24]. After that, reflections can be segmented as regions of high-intensity value in the iris image. Figure 5.1e shows a segmented iris 3D model.

Finally, we obtained the normal vector of every point in the point cloud model. The normal vector of a 3D point was calculated as the unit vector perpendicular to the best fitting the plane of the 6 neighboring 3D points, as proposed by Hoppe et al. [129].

We obtained the iris 3D models and applied the post-processing described above to all 2,000 images in the dataset. The average number of 3D points in the segmented iris 3D models is 27,553±2,880. The standard deviation of 2,880 3D points is due to differences in dilation levels and eyelash occlusions between subjects.

5.3.3. Keypoints and Descriptors

Keypoints can be extracted from corners or fiducial points [35][70]. They are reliable since they can be found in the same place for different instances of the same object [70]. However, iris images with low texture can produce an insufficient number of keypoints for the analysis. A common alternative to deal with this issue is to obtain evenly-spaced keypoints on a grid [130]. That is why in this paper we used a 3D grid of 40×40×20, obtaining 1,039±287 keypoints on average with this method. The number of keypoints varies from image to image since they all have different levels of dilation and eyelid occlusion.

For the descriptors, we used Spin Images [33] since they were conceived for 3D object recognition. A Spin Image describes one keypoint $\boldsymbol{p}$, with normal $\boldsymbol{n}$, as a function of the neighboring 3D points $\boldsymbol{x}_k$ in a unique manner. All $\boldsymbol{x}_k$ points are projected onto two axes α and β using:

$$\alpha_k = \sqrt{\|\boldsymbol{x}_k - \boldsymbol{p}\|^2 - (\boldsymbol{n} \cdot (\boldsymbol{x}_k - \boldsymbol{p}))^2} \quad [33], \tag{5.4}$$

$$\beta_k = \boldsymbol{n} \cdot (\boldsymbol{x}_k - \boldsymbol{p}) \quad [33]. \tag{5.5}$$

The values of α_k and β_k are continuous. However, for a more efficient comparison between Spin Images, they are discretized [33][119]. The (α, β) space is divided into square cells with height dxBin, and the number of cells is Nbins^2 [33]. The coordinates of an individual cell or bin are given by (i_k, j_k), and are computed by:

$$i_k = \text{floor}(\text{Nbins}/2 - \beta_k/d\text{xBin}) \quad [33], \tag{5.6}$$

$$j_k = \text{floor}(\alpha_k/\text{dxBin}) \quad [33]. \tag{5.7}$$

The Spin Image is finally formed in the following manner: First, it is initialized as a matrix full of zeros of size Nbins by Nbins. Then, α_k, β_k, i_k, and j_k are computed for each 3D point x_k. Each 3D point increases the (i_k, j_k) bin in the Spin Image by 1 [33]. When the value of i_k or j_k is greater than Nbins, no bin is increased in value. Therefore, the number of 3D points that participate in a Spin Image is determined by Nbins and dxBins [33]. The resulting Spin Image is a matrix that depends on the position, or shape, of the neighboring points $\boldsymbol{x}_k$ in a certain region around the keypoint $\boldsymbol{p}$. Figure 5.2 shows examples of Spin Images for variations in Nbins and dxBin. This figure illustrates the abstract nature of the Spin Images as local descriptors [33][124]. It can also be seen that small values of Nbins and dxBin emphasize local features, whereas larger values represent global ones. Local features are preferred because they help to discriminate a specific iris 3D point [70]. Additionally, a large Nbins value uses more computational resources. In this study, we used the value of 0.15 mm for dxBin, and 20 for Nbins.

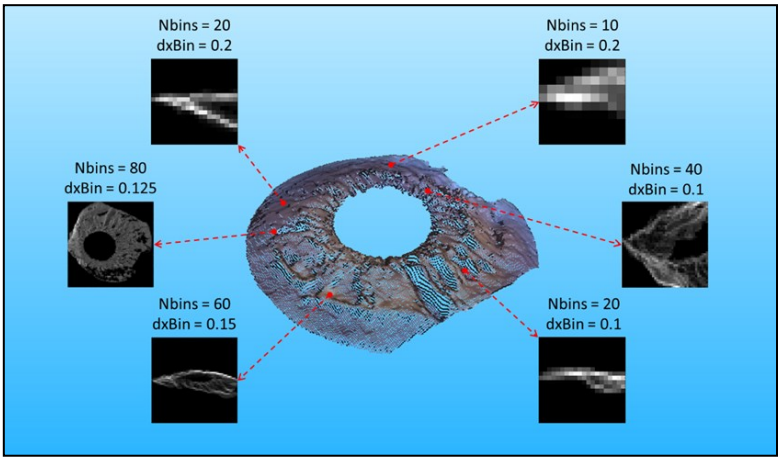
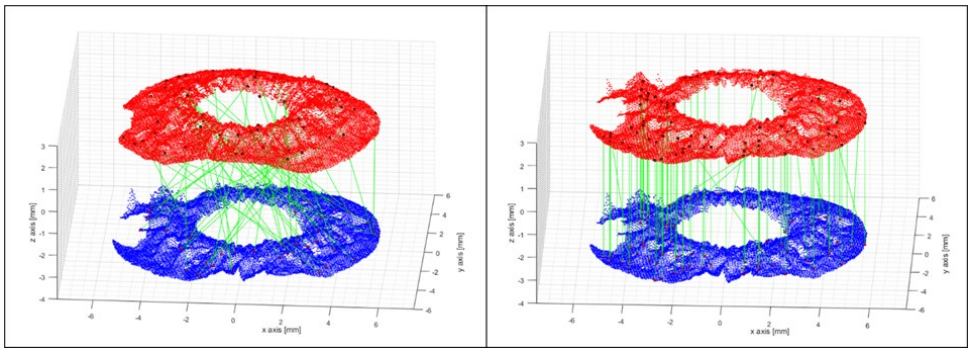


Figure 5.2: Examples of Spin Images obtained from different keypoints in the same iris 3D model. All the spin images were calculated using different combinations of Nbins and dxBin.



(a) (b)

Figure 5.3: Matches between two iris 3D models using Spin Images and KD trees. Only the 60 matches with the lowest SAD values were plotted. (a) Matching iris 3D models from different subjects. (b) Matching two iris 3D models from the same subject.

5.3.4. Finding Descriptor Matches

In comparing two iris 3D models, all the descriptors in one model have to be compared with all the descriptors in the other to determine corresponding points. Since we have about 936 keypoints and descriptors per iris 3D model, this is a costly operation. That is why we use the k-dimensional trees (KD trees) technique to guide the search [131].

We rank the quality of a match between two spin images using the Sum of Absolute Differences (SAD) [70]. A low value indicates that the descriptors match. Figure 5.3 show examples of finding matching points with the proposed method; only the 60 matches with the smallest SAD value were plotted. When different subjects are compared, matched points appear at random places of the two models, as seen in Figure 5.3a. However, when two models of the same subject are compared, there are more matched points in corresponding places of the two models, as seen in Figure 5.3b. Best performance was obtained keeping only the 500 matches with smallest SAD.

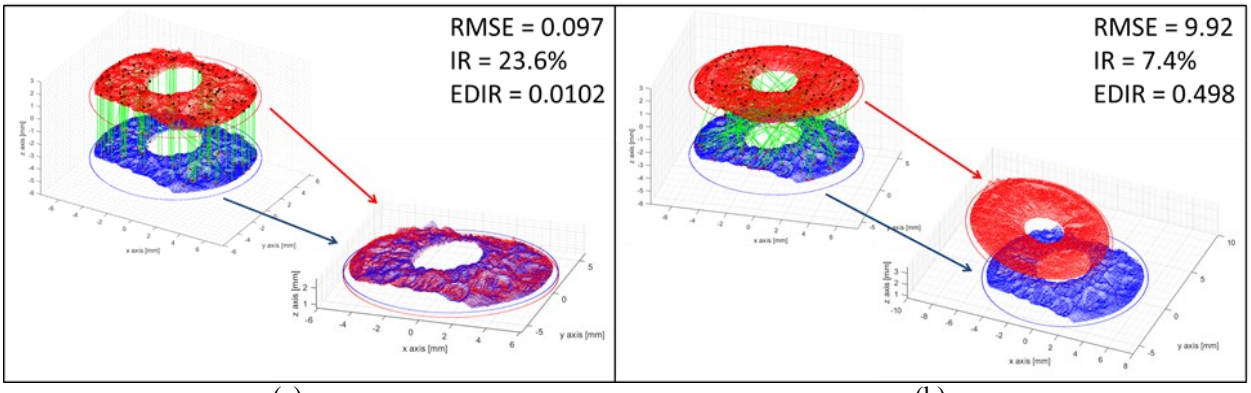
5.3.5. Iris-3D-Model Similarity Assessment

In order to assess the similarity between two iris 3D models, we suggest using the transformation that best suits the descriptor matches found above. If two iris 3D models have similar descriptors in the same places, the transformation would be close to the identity. However, if the two models have similar descriptors in different places, the resulting transformation would have to deform one 3D model to make them match.

For the transformation we consider variations in rotation, translation, and scale. Therefore, our method is robust even when the two iris 3D models have different scales, and when they are rotated and shifted with respect to each other. We use the Random Sampling and Consensus (RANSAC) technique [84] to find inliers in the matched points. In our method, only the 80 matches with smaller SADs can produce an appropriate transformation proposal, however, all the matches can vote to find the consensus. We find the transformation that best suits the inliers, and store the inlier ratio, which is the number of inliers divided by the total number of matches.

Then, we score the deformation suffered by the query 3D model as the distance, or Root Mean Square Error (RMSE), between the original 3D points and the transformed 3D points. As a means to accelerate the computation, we compute the distance using only a circle of 360 3D points in the iris perimeter instead of the 27,553 3D points in the whole iris 3D model. Figure 5.4 shows examples of the resulting transformations. In Figure 5.4a, a transformation between models of the same subject makes them align almost perfectly, and the error between the circles is small. In Figure 5.4b, a transformation between inter-class models satisfies only a few inliers and the 3D models are completely misaligned. This increases the Root Mean Square Error (RMSE) between the two circles.

Finally, we combine the RMSE distance with the inlier ratio (IR) of the transformation to produce a robust score of similarity between the two models. We call it Exponential Distance to Inlier Ratio (EDIR) and it is calculated using:



(a) (b)

Figure 5.4: Examples of transformations using the RANSAC consensus of keypoint matches. (a) Comparing two iris 3D models from the same subject. (b) Comparing two iris 3D models from different subjects. In both (a) and (b), the reference model is shown in blue and the query model is shown in red. The circular perimeters of both iris 3D models are also drawn in the same colors. The transformation was applied to the query model. In (a), both models being compared match almost perfectly, and the RMSE between the two circles is low. In (b), there is a random transformation that satisfies a few inliers, and the RMSE between the two circles is high.

$$\text{EDIR} = 0.5\left(1 - \exp\left(-\frac{\text{RMSE}}{20*\text{IR}}\right)\right). \quad (5.8)$$

We use an exponential function for easier comparison with the Hamming Distance (HD). The HD of intra-class matches is close to zero while the HD of inter-class matches is close to 0.5 [8]. In our similarity score, genuine matches are also close to zero; however, without the exponential function, the distance just increases for inter-class matches. That is why the exponential function helps to saturate the score of inter-class matches to 0.5 just like HD.

5.3.6. Performance Evaluation

We compared the recognition accuracy of our method against Daugman's Iris Code [8], Cohen's MSE [35][36], the 3D rubber sheet [31], and CNN based methods [126][127]. For this purpose, we compute both intra-class and inter-class distributions. In our dataset of 2,000 images, the distributions are comprised of 19,000 intra-class comparisons and 1,980,000 inter-class comparisons. We evaluate of each method by means of the decidability index, d':

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{0.5*(\sigma_1{}^2 + \sigma_2{}^2)}} \text{ [8]}, \quad (5.9)$$

where μ_1 and μ_2 are the mean values of the intra-class and inter-class distributions respectively, and σ_1 and σ_2 are the standard deviations (STD). The higher the d' value, the easier it is to separate intra-class from inter-class distributions [8].

We obtain the iris code of the 2,000 iris images in our dataset using Osiris V4.1 implementation [39]. We also evaluate the iris recognition performance of the iris code obtained from the depthmap image using Osiris V4.1. This test indicates whether or not the depth of the iris alone is sufficiently discriminative.

Finally, we compare our performance to that of state-of-the-art CNN-based methods. Nguyen et al. [127] proposed extracting deep features from rubber sheets using pretrained CNNs. Based on the best results of that method, we extracted deep features from the 9th layer of VGG-19 [66], the 11th layer of ResNet-18 [101] and the 5th layer of DenseNet [132]. Then, we used the deep-feature binarization method proposed by Zhao et al. [126] so that we can compare the binary feature vector of two irises using the Hamming Distance. This method consists of binarizing the feature vector using its mean value as the threshold [126].

5.4. Experimental Results and Analysis

Examples of the intra-class and inter-class distributions can be seen in Figure 5.5. Iris recognition performance evaluation of the different methods is summarized in Table 5.1. All the tests were conducted on both sets of 500 and 2,000 images (Set_{500} and Set_{2000}). The mean and the STD of intra-class and inter-class distributions, as well as the decidability index d' are shown in each Figure. Due to the great difference between the number of intra and inter-class comparisons, the distributions were normalized to have an area of one. All the methods produced perfectly separable distributions, except for the MSE in Figure 5.5b.

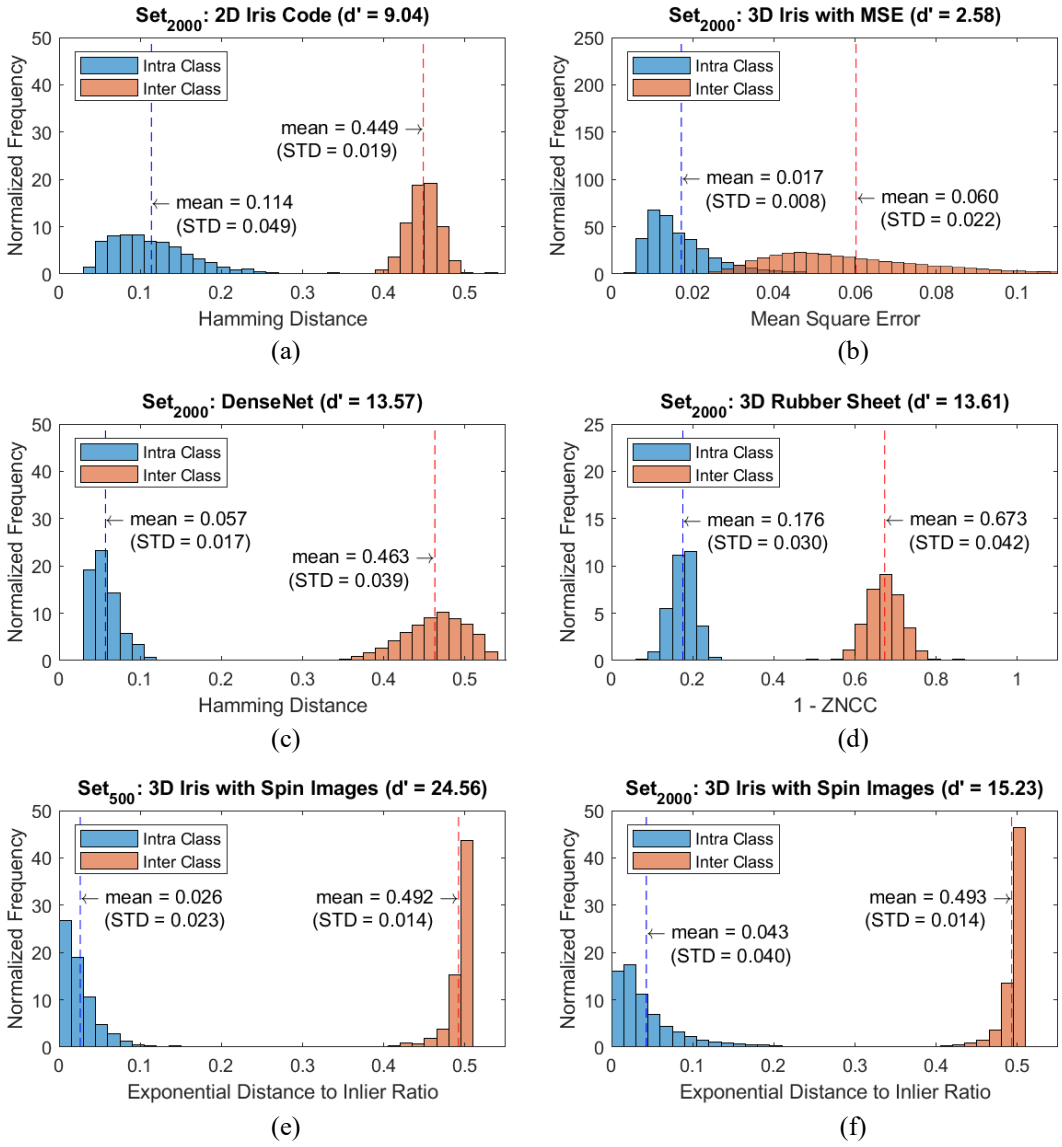


Figure 5.5: Intra and inter class distributions of the tested methods. (a) Using the 2D iris code [8] in Set_{2000}. (b) Using Cohen's MSE method [35][36] in Set2000. (c) Using binarized deep features from the 5th layer of DenseNet [126][127][132] in Set_{2000}. (d) Using the 3D rubber sheet model [31] in Set_{2000}. (e) Using the proposed method in Set_{500}. (f) Using the proposed method in Set_{2000}.

Table 5.1. Decidability index (*d'*) obtained by the different methods in the test sets of 500 and 2,000 images

Method	d'_{500}	$d'_{2,000}$
2D Iris Code using Osiris [8][114]	15.66	9.04
Depthmap Iris Code using Osiris [114]	10.43	7.34
VGG-19 [126][127][66]	15.82	12.74
ResNet-18 [126][127][101]	16.33	13.46
DenseNet [126][127][132]	16.92	13.57
3D Rubber Sheet [31]	16.03	13.61
Iris 3D Model with MSE [35][36]	2.92	2.58
Iris 3D Model with Spin Images (ours)	24.56	15.23

The shape of the distributions is consistent in both test sets. For example, the proposed method produced skewed distributions of similar shapes in Figure 5.5e and Figure 5.5f. The shape of the distributions that use Daugman's iris code are similar to the one presented in Figure 5.5a. This is because Daugman's method makes the codes of different irises uncorrelated, with a Bernoulli distribution of mean 0.5 [8]. Similarly, all the distributions of CNN-based methods resembled the one shown in Figure 5c. Even though the shape of the distributions of all methods are similar in Set_{500} and Set_{2000}, the standard deviation of intra-class comparisons are higher in the Set_{2000} because this set has iris images with greater dilation changes. Consequently, Table 5.1 shows that performance drops 30% in average from Set_{500} to Set_{2000} in all the compared methods.

Osiris implementation of Daugman's iris code [8], shown in Figure 5a, is the baseline for the analysis. Table 5.1 shows *d'* of 15.66 and 9.04 for Set_{500} and Set_{2000} respectively. These values are 1.56 and 1.74 greater than that achieved by Daugman under ideal and realistic image capture conditions respectively [8]. This correlates with our experiment since, as mentioned in section 3.1, Set_{500} has small dilation variation, and Set_{2000} has more dilation variation.

The proposed method shown in Figures 5.5e and 5.5f has the highest *d'* values among all the tested methods in Table 5.1. In Set_{2000}, its *d'* value of 15.23 is 68% higher than that of the 2D iris code. The intra-class distribution in Figure 9b shows that there is a small EDIR of 0.043 when comparing iris 3D models of the same individual. On the other hand, the inter-class distribution has a mean EDIR of 0.493. Therefore, the 3D information, along with the EDIR score, helps separating genuine from impostor matches with great precision. The proposed method is also more robust than the 3D rubber sheet model [31]. Both methods are resilient to dilation changes, however, our proposed method finds the transformation that best aligns the two models with RANSAC. This helps overcome slight misalignments that might occur in the segmentation stage, as well as severe pose and scale changes that are common in SfM based iris 3D models [23][25]. As a result, our proposed method increased d' in 12% with respect to the 3D rubber sheet in Set_{2000}.

Extracting the iris code from the depthmap image with Osiris V4.1 produced separable distributions with *d'* values of 10.43 and 7.34, as can be seen in Table 1. This indicates that the depthmaps produced by irisDepth were sufficiently discriminative for iris recognition. However, there is a benefit in using 3D features since in Set_{2000}, *d'* increased 1.85 times using the 3D rubber sheet and 2.07 times using the 3D model, compared to the depthmap alone. The 3D rubber sheet produced a similar level of performance as using CNN features as shown in Figure 5.5b, Figure 5.5c, and Table 5.1. Just like in Nguyen's paper [127], DenseNet and ResNet produced deep features that classified irises better than VGG and the iris code. However, the proposed 3D method obtained a higher *d'* index with respect to the other methods in Table 5.1.

Comparing the iris 3D models point by point with MSE was not effective; Figure 5b and Table 5.1 show that the *d'* values are only 2.96 and 2.58 in Set_{500} and Set_{2000}, respectively. Additionally, there is overlapping in the distributions. The overlapping area produces false positives and false negatives, which are not present in the other methods (Figures 5.5a, and 5.5c through 5.5f). The overlap is also observable in the Table 5.1 of Cohen's paper [35], where the maximum distance between corresponding irises is 0.09, and there are other matches between different irises that have less or equal distance.

Finally, our method uses only the shape information of the iris 3D models for iris recognition. Performance could be increased if we combined shape and color information, which is also available in the iris 3D models produced by irisDepth. This could be explored in the future. Moreover, using more than 360 3D points in the iris perimeter could also increase the performance at the expense of computation time.

5.5. Conclusions

We have described a novel approach for iris recognition in 3D using keypoints and descriptors to find the best fitting transformation between two iris 3D models. We assessed the similarity between two iris 3D models using the EDIR score, which is a function of the transformation error and the inlier ratio. The results show a 68% increase in *d'* with respect to Daugman's iris code and 12% increase with respect to the 3D rubber sheet model, indicating that our method better separates the intra-class and inter-class distributions. Furthermore, we obtained over 12% increase in iris recognition performance concerning state-of-the-art CNN-based methods. Therefore, the recognition with our method would be more robust and less prone to errors on larger datasets and in real applications.

The high performance achieved with our method is due to the depth information in the iris 3D models, which is absent in 2D images. The Spin Image descriptor allowed us to use the 3D information effectively to identify keypoints in the 3D model, and to discriminate among 3D iris models. Future research should consider evaluating the proposed system on larger datasets, and studying ways to reduce the computational cost of the complete pipeline.

5.6. Acknowledgement

This research has been funded by ANID through project FONDECYT 1191610, ANID Project AFB180004, and by the Department of Electrical Engineering, Universidad de Chile.

Chapter 6

Conclusions

6.1. General Conclusions

We developed a device that captures close range images of the human iris from several perspectives in a controlled VL environment [24]. The device uses a Samsung S6 cellphone-camera along with a macro lens to acquire 16Mpx images of the iris. Both eyes are illuminated using white-light LEDs, which produce LFVL. This kind of illumination enhances iris features, which both increases the number of keypoints in SfM [25] and allows CNNs to relate iris features with depth information [31].

We obtained a slight improvement in iris recognition performance by using LFVL images with respect to NIR in a dataset of 96 subjects and 1,920 images [24]. We used Daugman´s iris code as the benchmark for both illumination methods. This result indicates that the proposed device and illumination method can obtain iris images with similar amount of information to traditional NIR images, but the proposed device captures them at much higher resolution and frame rate than conventional NIR devices [24].

We developed a 3D iris scanning algorithm using SfM, which accounts for pupil dilation and lens distortion [25]. First, LFVL images of the iris are captured from different perspectives using the proposed device [24]. Lens distortion is corrected in each image using camera calibration process [28][29]. Pupil dilation is mitigated by selecting iris images with the dilation level that is present in the greatest number of camera positions [25]. Then, a customized SfM algorithm estimates the camera pose of every image jointly with a sparse 3D model of the iris [23][25]. Then, a dense 3D point-cloud reconstruction is performed by extracting Shi-Tomasi keypoints from each image [25][85]. Finally, the point-cloud model is converted into a mesh surface by the Screened Poisson Surface Reconstruction technique [67]. This mesh helps interpolate the depth information in areas of the iris with low texture [25]. The result is a 3D model that incorporates both depth and color information of the iris surface.

We developed a 3D iris scanning methodology using CNNs, which accounts lens distortion [31]. The proposed method produces a 3D model of the human iris from a single image, at any dilation level. In order to train the networks, we acquired a dataset of real irises with 26,520 images and another dataset of synthetic irises with 72,000 images. Lens distortion was corrected in each real image using camera calibration process [28][29]. Both datasets were captured under LFVL illumination to enhance the features of the iris surface [25][24]. Thus, lateral illumination allowed the networks to relate shadows in RGB images to depth information [31]. The proposed architecture, irisDepth, combines the Generative Adversarial Network (GAN) of T^2Net [15] with the encoder-decoder architecture of DenseDepth [20]. T^2Net allows our method to train with real and synthetic iris images, while DenseDepth allows predicting a depthmap from an input image

with great accuracy [31]. After training with both real and synthetic datasets, the method was able to estimate the depth value of every pixel in an iris image.

Using 3D patterns of known dimensions, as well as OCT scans, we evaluated the performance of the proposed SfM and CNN methods in terms of number of produced 3D points, precision, and accuracy. Despite using 16Mpx images, the SfM method produced only11,105 3D points in average and 57,937 polygons [25]. On the other hand, at a resolution of 256×256, the CNN method always produces 65,536 evenly distributed 3D points and 130,050 polygons regardless of iris texture [31]. The precision of each method is measured as the resolution of the 3D models they produce. This feature was evaluated scanning physical 3D patterns of known dimensions. The SfM method has a higher precision than the CNN method since they produced resolutions of 11μm and 17.7μm respectively. For context, the resolution of conventional OCT scans is 10μm [111]. Finally, we evaluated the accuracy of each method using 4 OCT slices per eye of a subject. We scored the Mean Absolute Error (MAE) between an OCT slice and the corresponding slice of the 3D model reconstructed with the proposed methods. Averaging for the 8 available OCT slices, SfM produced a MAE of 123μm, while MAE was 77μm using CNN. Thus, the CNN method produced a 60% increase in accuracy with respect to SfM [31].

We developed two methods for extracting information from an iris 3D model and evaluated their performances against the standard 2D iris Code and alternative 3D iris recognition methods [35][36]. Using the proposed 3D rubber sheet model, iris recognition performance increased in 48% with respect to the 2D iris code in a dataset of 12 subjects and 480 images [31]. However, extracting Spin Image descriptors [33][124] from the iris 3D model and using Random Sampling and Consensus [84] to find the best spatial transformation between two 3D models resulted in a more successful approach. The latter method obtained a decidability index of 15.23 when that of the 3D rubber sheet and the 2D iris code were 13.61 and 9.04 respectively, in a dataset of 50 subjects and 2,000 images [32]. All these results illustrate the increase in discriminative information that is gained by using iris 3D models instead of 2D images.

6.2. Discussion and Future Work

In this thesis we developed two methods that reconstruct a 3D model of the iris surface [25][31], as well as two methods to use the 3D information for iris recognition [31][32]. Our results show an important increment in decidability index *d'* using the 3D model with respect to 2D methods [32]. Even though in small datasets both 2D and 3D methods can show perfect separability of the inter and intra-class distributions, this increment in *d'* indicate the superior performance of the proposed 3D methods. An increasing level of performance is always desirable in the biometrics field since datasets are growing every day. For instance India migrated from a fingerprint-based ID system to an iris based one because the false match rate of fingerprint is inadmissible for a dataset of the magnitude of India's population [6]. The same could become true for 2D iris recognition in the future. For instance, Daugman demonstrated that a false match rate of 1 in 4 million could be achieved when using a fractional Hamming Distance of 0.33 [8]. In a database of one billion individuals, a false match rate of 1 in 4 million indicates that 250 people have similar iris codes. In this scenario, 3D iris recognition methods could be an ideal solution. Additionally, most of the times iris images present occlusions due to eyelids, eyelashes and

reflections [4]. Extracting additional depth information from iris images can help identify people even when only a portion of the iris is available for the analysis [23].

Additionally, obtaining a 3D model of the human iris opens new research lines. For instance, new research could be made to improve the computational time and complexity of the 3D models towards a more compact representation such as a 3D iris code. This not only improve the technical aspects of the method, but also make it more accessible and commercially feasible. Additionally, a 3D iris model could be used to improve iris recognition in extreme poses [11], just like 3D information has for face recognition [115][116][117]. Furthermore, 3D iris models could enable studying the phenomenon of pupil dilation in greater detail. Pupil dilation creates a non-affine deformation of the iris [8][133][134]. Studies have shown that the performance of iris recognition systems falls when images with different dilation levels are compared [10] [135][136][137][138][139]. In those research works several iris normalization methods have been proposed to increase iris recognition performance. However, those methods only assume 2D information and encounter some artifacts in the motion of the tissue [136]. Perhaps those artifacts could be explained due to tissue motion in the depth axis. Therefore, 3D iris models could provide a better insight on this phenomenon, as well as better iris normalization mechanisms [139].

On the other hand, it is also worth analyzing the disadvantages of the 3D methods with respect to 2D. The 2D iris code is much faster to obtain and to compare than the proposed 3D methods. Additionally, most 2D iris imaging devices do not require the subject's face to be in contact with the device, unlike the proposed LFVL device [4]. The disadvantages of the SfM method over 2D are the need for a moving camera, the additional acquisition and processing time that is required for capturing bursts of 10 images for each camera position, and the storage needed for all those images [25]. The CNN method possesses less limitations since it uses a single image. The main disadvantage over the 2D iris code is the need for extra computational resources such as GPUs to run the networks [31]. Finally, a camera calibration step is required for both SfM and CNN methods; however, this step is only executed once in the lifetime of the device and does not represent a major drawback with respect to 2D. As with any technology there is a tradeoff between performance and simplicity. All the drawbacks of the 3D methods can be easily compensated by the superior performance they achieved with respect to 2D, especially if they are planned to be deployed in large datasets.

Finally, the proposed 3D iris scanning methods could be applied in the detection of Closure Angle Glaucoma (CAG). An Anterior-Segment Optical Coherence Tomography (AS-OCT) is the gold standard in ophthalmology for the detection of most of the types of glaucoma [26]. CAG in particular is produced when the Anterior Chamber Angle (ACA), which is the angle between the iris and the cornea, is too narrow [140]. This blocks the natural drain channels of aqueous humor; thus, increasing intra-ocular pressure and damaging the optic nerve [140]. The AS-OCT allows measuring the ACA for diagnosis [140]. In recent years automatic methods to assess the ACA have been proposed; some of them identify markers in the AS-OCT [141], while others use CNNs to learn handmade annotations from visual ques [142]. A 3D iris scanner could also be used to detect CAG. Eyes with a narrow ACA present a characteristic curvature in the iris surface [141]. This curvature could be easily detected in a 3D model since depth information is known. However, further studies are needed with more training and test subjects with and without CAG.

Those studies would require both the AS-OCT of each subject, as well as the 3D models, for a proper validation of the method.

Bibliography

[1] J. Daugman, “The importance of being random: Statistical principles of iris recognition,” Pattern Recognit., vol. 36, no. 2, pp. 279-291, 2003.

[2] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 4-20, Jan. 2004.

[3] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, “Image understanding for iris biometrics: A survey,” Comput. Vis. Image Understand., vol. 110, no. 2, pp. 281-307, 2008.

[4] K. W. Bowyer and M. J. Burge, “Introduction to the handbook of iris recognition,” in Handbook of Iris Recognition. London, U.K.: Springer, 2016, pp. 1-14.

[5] J. Daugman and I. Malhas, “Iris recognition border-crossing system in the UAE,” Int. Airport Rev., vol. 8, no. 2, pp. 49-53, 2004.

[6] J. Daugman, “600 million citizens of India are nowenrolled with biometric ID,” Proc. SPIE Newsroom, vol. 7, pp. 1-4, May 2014.

[7] K. W. Bowyer, E. Ortiz, and A. Sgroi, “Trial Somaliland voting register de-duplication using iris recognition,” in Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit., vol. 2, May 2015, pp. 1-8.

[8] J. Daugman, “How iris recognition works,” IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 21-30, Jan. 2004.

[9] J. Daugman and C. Downing, “Epigenetic randomness, complexity and singularity of human iris patterns,” Proc. Roy. Soc. B, Biol. Sci., vol. 268, no. 1477, pp. 1737-1740, 2001.

[10] K. Hollingsworth, K.W. Bowyer, and P. J. Flynn, “Pupil dilation degrades iris biometric performance,” Comput. Vis. Image Understand., vol. 113, no. 1, pp. 150-157, 2009.

[11] E. Celik and M. Karakaya, “Pupil dilation at synthetic off-angle iris images,” in Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG), Sep. 2016, pp. 1-5.

[12] Y. Furukawa and C. Hernandez, “Multi-view stereo: A tutorial,” Found. Trends Comput. Graph. Vis., vol. 9, nos. 1-2, pp. 1-148, 2013.

[13] C.Wu, “Towards linear-time incremental structure from motion,” in Proc Int. Conf. 3D Vis., Jun./Jul. 2013, pp. 127-134.

[14] C. Godard, O. M. Aodha, and G. J. Brostow, “Unsupervised monocular depth estimation with left-right consistency,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6602_6611.

[15] C. Zheng, T. J. Cham, and J. Cai, “T2Net: Synthetic-to-realistic translation for solving single-image depth estimation tasks,” in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 767_783.

[16] Y. Furukawa and J. Ponce, “Accurate camera calibration from multi-view stereo and bundle adjustment,” Int. J. Comput. Vis., vol. 84, no. 3, pp. 257-268, Apr. 2009.

[17] S. Agarwal, Y. Furukawa, and N. Snavely, “Building rome in a day,” in 2009 IEEE 12th international conference on computer vision, 2009, pp. 72–79, doi: 10.1145/2001269.

[18] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 8, pp. 1362-1376, Aug. 2008.

[19] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, “Deeper depth prediction with fully convolutional residual networks,” in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 239_248.
[20] I. Alhashim and P. Wonka, “High Quality Monocular Depth Estimation via Transfer Learning,” arXiv e-prints, 2018, [Online]. Available: http://arxiv.org/abs/1812.11941.
[21] D. Wofk, F. Ma, T.-J. Yang, S. Karaman, and V. Sze, “FastDepth: Fast monocular depth estimation on embedded systems,” in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 6101_6108.
[22] Y. Tian, X. Li, K. Wang, and F.-Y. Wang, “Training and testing object detectors with virtual images,” *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 539_546, Mar. 2018.
[23] D. Bastias, C. A. Perez, D. P. Benalcazar, and K. W. Bowyer, “A Method for 3D Iris Reconstruction from Multiple 2D Near-Infrared Images,” IJCB 2017 Int. Jt. Conf. Biometrics, pp. 503–509, 2017.
[24] D. P. Benalcazar, C. A. Perez, D. Bastias, and K. W. Bowyer, “Iris Recognition: comparing Visible-Light Lateral and Frontal Illumination to NIR Frontal Illumination,” IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2019.
[25] D. P. Benalcazar, D. Bastias, C. A. Perez, and K. W. Bowyer, “A 3D iris scanner from multiple 2D visible light images,” *IEEE Access*, vol. 7, pp. 61461_61472, 2019.
[26] I. I. Bussel, G. Wollstein, and J. S. Schuman, “OCT for glaucoma diagnosis, screening and detection of glaucoma progression,” *Brit. J. Ophthalmology*, vol. 98, no. 2, Jul. 2014, pp. ii15_ii19.
[27] A. Clark, S. Kulp, I. Herron, and A. A. Ross, A Theoretical Model For Describing Iris Dynamics. London, U.K.: Springer, 2016, pp. 417-438.
[28] Z. Zhang, “A Flexible New Technique for Camera Calibration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, Dec. 2000.
[29] D. Scaramuzza, A. Martinelli, and R. Siegwart, “A Toolbox for Easily Calibrating Omnidirectional Cameras,” in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 5695–5701.
[30] S. Paris, S.W. Hasinoff, and J. Kautz, “Local Laplacian filters: Edge-aware image processing with a laplacian pyramid,” ACM Trans. Graph., vol. 58, no. 3, pp. 81-91, 2011.
[31] D. P. Benalcazar, J. Zambrano, D. Bastias, C. A. Perez, and K. W. Bowyer, “A 3D Iris Scanner from a Single Image using Convolutional Neural Networks,” IEEE Access, vol. 8, no. 1, pp. 98584-98599, 2020.
[32] D. P. Benalcazar, D. Montecino, J. Zambrano, C. A. Perez, and K. W. Bowyer, “3D Iris Recognition using Spin Images,” Accepted in the 2020 IEEE International Joint Conference on Biometrics (IJCB), 2020, pp. 1–8.
[33] A. E. Johnson and M. Hebert, “Using Spin-Images for Efficient Object Recognition in Cluttered 3-D Scenes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. July, pp. 433–449, 1998.
[34] A. Eleliemy, A. Mohammed, and F. M. Ciorba, “Efficient Generation of Parallel Spin-images Using Dynamic Loop Scheduling,” in 2017 IEEE 19th International Conference on High Performance Computing and Communications Workshops (HPCCWS), Dec. 2017, pp. 34–41.
[35] F. Cohen, S. Sowmithran, and C. Li, “Iris Identification in 3D,” Scand. Conf. Image Anal., pp. 324–335, Jun. 2019.

[36] F. Cohen, S. Sowmithran, and C. Li, “3D iris model and reader for iris identification,” Concurr. Comput. Pract. Exp., no. August 2018, pp. 1–15, Jan. 2020.
[37] R. P. Wildes, “Iris recognition: an emerging biometric technology,” Proc. IEEE, vol. 85, no. 9, pp. 1348–1363, 1997.
[38] J. Daugman, “Information Theory and the IrisCode,” IEEE Trans. Inf. Forensics Secur., vol. 11, no. 2, pp. 400–409, 2016.
[39] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, and A. Ross, “Long range iris recognition: A survey,” Pattern Recognit., vol. 72, pp. 123–143, Dec. 2017.
[40] M. Vatsa, R. Singh, and A. Noore, “Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing,” IEEE Trans. Syst. Man, Cybern. Part B Cybern., vol. 38, no. 4, pp. 1021–1035, 2008.
[41] T. T. Li Ma, Y. W., and D. Z., “Efficient iris recognition by Charactering Key Local Variation,” IEEE Trans. Image Process., vol. 13, no. 6, pp. 739–750, 2004.
[42] J. E. Tapia, C. A. Perez, and K. W. Bowyer, “Gender Classification from Iris Images using Fusion of Uniform Local Binary Patterns,” Comput. Vis. - ECCV 2014, vol. 8927, no. September, pp. 751–763, 2014.
[43] J. E. Tapia, C. A. Perez, and K. W. Bowyer, “Gender Classification From the Same Iris Code Used for Recognition,” IEEE Trans. Inf. Forensics Secur., vol. 11, no. 8, pp. 1760–1770, Aug. 2016.
[44] H. Dehghani, B. W. Pogue, J. Shudong, B. Brooksby, and K. D. Paulsen, “Three-dimensional optical tomography: resolution in small-object imaging,” Appl. Opt., vol. 42, no. 16, p. 3117, Jun.
[45] M. S. Hosseini, B. N. Araabi, and H. Soltanian-Zadeh, “Pigment Melanin: Pattern for Iris Recognition,” IEEE Trans. Instrum. Meas., vol. 59, no. 4, pp. 792–804, 2010.
[46] A. Czajka and K. W. Bowyer, “Presentation Attack Detection for Iris Recognition: An Assessment of the State of the Art,” ACM Comput. Surv., p. 35, 2018.
[47] H. Proenca, “Iris Recognition in the Visible Wavelength,” in Handbook of Iris Recognition, 2013, pp. 151–169.
[48] M. A. M. Abdullah, F. H. A. Al-Dulaimi, W. Al-Nuaimy, and A. Al-ataby, “Efficient Small Template Iris Recognition System Using Wavelet Transform,” Int. J. Biometrics Bioinforma., vol. 5, no. 1, pp. 16–27, 2011.
[49] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, “The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-the-Move and At-a-Distance,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 8, pp. 1529–1535, Aug. 2010.
[50] K. W. Bowyer, “The results of the NICE.II Iris biometrics competition,” Pattern Recognit. Lett., vol. 33, no. 8, pp. 965–969, Jun. 2012.
[51] M. De Marsico, D. Riccio, and H. Wechsler, “Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols,” Pattern Recognit. Lett., vol. 57, pp. 17–23, May 2015.
[52] Google. Google Cardboard Google VR. Accessed: Sep, 13, 2018. [Online]. Available: https://vr.google.com/cardboard/
[53] A. Mariakakis et al., “BiliScreen: Smartphone-Based Scleral Jaundice Monitoring for Liver and Pancreatic Disorders,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol, vol. 1, no. 26, 2017.

[54] The MathWorks, Single Camera Calibrator App MATLAB Simulink. Accessed: Sep, 12, 2018. [Online]. Available: https://la.mathworks.com/help/vision/ug/single-camera-calibrator-app.html
[55] G. W. Quinn, J. Matey, E. Tabassi, and P. Grother, "IREX V Guidance for Iris Image Collection NIST Interagency Report 8013 Information Access Division National Institute of Standards and Technology," 2014.
[56] L. Masek, "Recognition of human iris patterns for biometric identification," M.S. thesis, School Comput. Sci. Softw. Eng., Univ. Western Australia, Perth, WA, Australia, 2003.
[57] D. J. Ketcham, "Real-Time Image Enhancement Techniques," in Image processing; Proceedings of the Seminar, 1976, vol. 74, pp. 120–125.
[58] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Image quality assessment for iris biometric," 2006, vol. 6202, p. 62020D.
[59] J. Canny, "A Computational Approach to Edge Detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, 1986.
[60] M. A. M. Abdullah, J. A. Chambers, W. L. Woo, and S. S. Dlay, "Iris biometrie: Is the near-infrared spectrum always the best?," 3rd IAPR Asian Conf. Pattern Recognition, ACPR 2015, pp. 816–819, 2016.
[61] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," Proc. IEEE, vol. 69, no. 5, pp. 529–541, 1981.
[62] M. Arsalan et al., "Deep learning-based iris segmentation for iris recognition in visible light environment," in 2016 IEEE Int'l Conference on Image Processing (ICIP), 2017
[63] N. Liu, H. Li, M. Zhang, Jing Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in 2016 International Conference on Biometrics (ICB), Jun. 2016, pp. 1–8.
[64] A. Gangwar and A. Joshi, "DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in 2016 IEEE International Conference on Image Processing (ICIP), Sep. 2016, pp. 2301–2305.
[65] S. Minaee, A. Abdolrashidiy, and Y. Wang, "An experimental study of deep convolutional features for iris recognition," in 2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Dec. 2016, pp. 1–6.
[66] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014, arXiv:1409.1556.
[67] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," ACM Trans. Graph., vol. 32, no. 3, pp. 1–13, Jun. 2013.
[68] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM.," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 1052–67, 2007.
[69] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," IEEE Trans. Robot., vol. 31, no. 5, pp. 1147–1163, 2015.
[70] P. Corke, "Vision," in *Robotics, Vision and Control*, 2nd ed. Brisbane, Australia: Springer, 2013, pp. 251-283.
[71] G. Silveira, E. Malis, and P. Rives, "An Efficient Direct Approach to Visual SLAM," IEEE Trans. Robot., vol. 24, no. 5, pp. 969–979, Oct. 2008.

[72] W. Fang, Y. Zhang, B. Yu, and S. Liu, “FPGA-based ORB feature extraction for real-time visual SLAM,” in Proc. Int. Conf. Field Program. Technol. (ICFPT), Dec. 2017, pp. 275-278.
[73] A. Pumarola, A. Vakhitov, A. Agudo, and A. Sanfeliu, “PL-SLAM : Real-Time Monocular Visual SLAM with Points and Lines,” pp. 4503–4508, 2017.
[74] R. Egodagamage and M. Tuceryan, “A Collaborative Augmented Reality Framework Based on Distributed Visual Slam,” in 2017 International Conference on Cyberworlds (CW), Sep. 2017.
[75] B. Triggs et al., “Bundle Adjustment – A Modern Synthesis,” Vis. Algorithms, pp. 298–372, 1999.
[76] N. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla, “Using Multiple Hypotheses to Improve Depth-Maps for Multi-View Stereo,” Springer, Berlin, Heidelberg, 2008, pp. 766–779.
[77] C. Hernandez, G. Vogiatzis, and R. Cipolla, “Probabilistic visibility for multi-view stereo,” in 2007 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2007, pp. 1–8.
[78] C. Hernández Esteban and F. Schmitt, “Silhouette and stereo fusion for 3D object modeling,” Comput. Vis. Image Underst., vol. 96, no. 3, pp. 367–392, Dec. 2004.
[79] J. Mélou, Y. Quéau, J.-D. Durou, F. Castan, and D. Cremers, “Beyond Multi-view Stereo: Shading-Reflectance Decomposition,” Springer, Cham, 2017, pp. 694–705.
[80] J. Z. Lim, H. Suzuki, S. Utsugi, and H. Katai, “Experimental Development of a Multi-view Stereo Endoscope System,” pp. 14–18, 2017.
[81] P. Loncomilla, J. Ruiz-del-Solar, and L. Martínez, “Object recognition using local invariant features for robotic applications: A survey,” Pattern Recognit., vol. 60, pp. 499–514, 2016, doi: 10.1016/j.patcog.2016.05.021.
[82] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features (SURF),” Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346–359, Jun. 2008.
[83] B. Horn. (Jun. 4, 2014). Recovering Baseline and Orientation from Essential Matrix. [Online]. Available: http://ocw.mit.edu/courses/
[84] M. a Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with,” Commun. ACM, vol. 24, pp. 381–395, 1981.
[85] J. Shi and C. Tomasi, Good Features to Track. Ithaca, NY, USA: Cornell Univ. 1993.
[86] P. Cignoni et al., “MeshLab: an open-source mesh processing tool,” Eurographics Ital. Chapter Conf., pp. 129–136, 2008.
[87] M. Lansard. (2018). The 10 Best 3D Scanners in 2018 Reviews and Buying Guide. ANIWAA. Accessed: Jan. 22, 2019. [Online]. Available: https://www.aniwaa.com/best-3d-scanner/
[88] A. Fouad et al., “Variations in optical coherence tomography resolution and uniformity: A multi-system performance comparison,” Biomed. Opt. Express, vol. 5, no. 7, pp. 2066-2081, Jul. 2014.
[89] Thorlabs. (2019). OCT Selection Guide. Thorlabs, Inc. Accessed: Jan. 22, 2019. [Online]. Available: https://www.thorlabs.com.
[90] A. Invernizzi, M. Cigada, L. Savoldi, S. Cavuto, L. Fontana, and L. Cimino, “In Vivo analysis of the iris thickness by spectral domain optical coherence tomography,” Brit. J. Ophthalmol., vol. 98, no. 9, pp. 1245-1249, Sep. 2014.

[91] M. He, D. Wang, J. W. Console, J. Zhang, Y. Zheng, and W. Huang, “Distribution and heritability of iris thickness and pupil size in chinese: The guangzhou twin eye study,” Investigative Ophthalmol. Vis. Sci., vol. 50, no. 4, pp. 1593-1597, Apr. 2009.
[92] Z. He, Z. Sun, T. Tan, and Z.Wei, Efcient Iris Spoof Detection Via Boosted Local Binary Patterns. Berlin, Germany: Springer, 2009, pp. 1080-1090.
[93] K. Hughes and K. W. Bowyer, “Detection of contact-lens-based iris biometric spoofs using stereo imaging,” in Proc. 46th Hawaii Int. Conf. Syst. Sci., Jan. 2013, pp. 17631772.
[94] A. Czajka, Z. Fang, and K. Bowyer, “Iris presentation attack detection based on photometric stereo features,” in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Jan. 2019, pp. 877-885.
[95] I. Tomeo-Reyes, A. Ross, A. D. Clark, and V. Chandran, “A biomechanical approach to iris normalization,” in 2015 International Conference on Biometrics (ICB), May 2015, pp. 9–16.
[96] S. Shah and A. Ross, “Iris Segmentation Using Geodesic Active Contours,” IEEE Trans. Inf. Forensics Secur., vol. 4, no. 4, pp. 824–836, Dec. 2009.
[97] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, “Deep Ordinal Regression Network for Monocular Depth Estimation,” IEEE Int’l Conf. Comput. Vis. Pattern Recognit., pp. 2002–2011, 2018.
[98] D. Eigen, C. Puhrsch, and R. Fergus, “Depth map prediction from a single image using a multi-scale deep network,” Adv. Neural Inf. Process. Syst., vol. 3, no. January, pp. 2366–2374, 2014.
[99] D. Eigen and R. Fergus, “Predicting depth , surface normals and semantic labels with a common multi-scale convolutional architecture,” Int. Conf. Comput. Vis., pp. 2650–2658, 2015.
[100] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” IEEE Int’l Conf. Comput. Vis. Pattern Recognit., pp. 1125–1134, 2017.
[101] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770_778.
[102] D. Xu, W. Wang, H. Tang, H. Liu, N. Sebe, and E. Ricci, “Structured Attention Guided Convolutional Neural Fields for Monocular Depth Estimation,” IEEE Int’l Conf. Comput. Vis. Pattern Recognit., pp. 3917–3925, 201.
[103] K. Tateno, F. Tombari, I. Laina, and N. Navab, “CNN-SLAM : Real-time dense monocular SLAM with learned depth prediction,” IEEE Conf. Comput. Vis. Pattern Recognit., pp. 6565–6574, 2017.
[104] Y. Kuznietsov, J. Stückler, and B. Leibe, “Semi-supervised deep learning for monocular depth map prediction,” IIEEE Int’l Conf. Comput. Vis. Pattern Recognit., pp. 6647–6655, 2017.
[105] J. E. Cutting and P. M. Vishton, “Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth,” in Perception of Space and Motion, Academic Press, 1995, pp. 69–117.
[106] F. Behar-Cohen et al., “Light-emitting diodes (LED) for domestic lighting: Any risks for the eye?,” Prog. Retin. Eye Res., vol. 30, no. 4, pp. 239–257, 2011.
[107] X. Yuan and P. Shi, “A Non-linear Normalization Model for Iris Recognition,” Int. Work. Biometric Pers. Authentication, pp. 135–141, 2005.

[107] Blender 2.82. (2019). *Stichting Blender Foundation*. [Online]. Available: https://www.blender.org/

[109] J. Zhu, T. Park, A. Iso, a, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," 2017 IEEE Int. Conf. Comput. Vis., pp. 2242–2251, 2017.

[110] C. Chu, A. Zhmoginov, and M. Sandler, "CycleGAN, a master of steganography," 2017, *arXiv:1712.02950*. [Online]. Available: http://arxiv.org/abs/1712.02950

[111] L. Sorbara, J. Maram, D. Fonn, C. Woods, and T. Simpson, "Metrics of the normal cornea: anterior segment imaging with the Visante OCT," Clin. Exp. Optom., vol. 93, no. 3, pp. 150–156, Apr. 2010.

[112] S. Sadek, M. G. Iskander, and J. Liu, "Accuracy of Digital Image Correlation for Measuring Deformations in Transparent Media," J. Comput. Civ. Eng., vol. 17, no. 2, pp. 88–96, Apr. 2003.

[113] E. Ortiz, K. W. Bowyer, and P. J. Flynn, "Dilation-aware enrolment for iris recognition," IET Biometrics, vol. 5, no. 2, pp. 92–99, 2016.

[114] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "OSIRIS: An open source iris recognition software," Pattern Recognit. Lett., vol. 82, pp. 124–131, Oct. 2016.

[115] A. Jourabloo and X. Liu, "Large-Pose Face Alignment via CNN-Based Dense 3D Model Fitting," IEEE Int'l Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 4188–4196, 2016.

[116] P. Dou, S. K. Shah, and I. A. Kakadiaris, "End-to-end 3D face reconstruction with deep neural networks," IEEE Int'l Conf. Comput. Vis. Pattern Recognit., vol. 2017-Janua, pp. 1503–1512, 2017.

[117] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, "Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression," IEEE Int'l Conf. Comput. Vis. Pattern Recognit., vol. 2017-Octob, pp. 1031–1039, 2017.

[118] A. Tewari et al., "MoFA: Model-Based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction," IEEE Int'l Conf. Comput. Vis. Pattern Recognit., vol. 2017-Octob, pp. 3735–3744, 2017.

[119] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 5, pp. 433–449, May 1999.

[120] J. Assfalg, G. D'Amico, A. Del Bimbo, and P. Pala, "3D content-based retrieval with spin images," in 2004 IEEE International Conference on Multimedia and Expo (ICME), pp. 771–774, 2004.

[121] Y. Mei and Y. He, "A new spin-image based 3D Map registration algorithm using low-dimensional feature space," in 2013 IEEE International Conference on Information and Automation (ICIA), Aug. 2013, pp. 545–551.

[122] D.-H. Kim and K.-S. Choi, "Angular-partitioned spin image descriptor for robust 3D facial landmark detection," Electron. Lett., vol. 49, no. 23, pp. 1454–1455, Nov. 2013.

[123] A. Eleliemy, D. Hegazy, and W. S. Elkilani, "MPI parallel implementation of 3D object categorization using spin-images," in 2013 9th International Computer Engineering Conference (ICENCO), Dec. 2013, pp.

[124] B. Apostol, C. R. Mihalache, and V. Manta, “Using spin images for hand gesture recognition in 3D point clouds,” in 2014 18th International Conference on System Theory, Control and Computing (ICSTCC), Oct. 2014, pp. 544–549.
[125] A. Eleliemy, M. Fayze, R. Mehmood, I. Katib, and N. Aljohani, “Loadbalancing on Parallel Heterogeneous Architectures: Spin-image Algorithm on CPU and MIC,” Proc. 9th EUROSIM Congr. Model. Simulation, EUROSIM 2016, 57th SIMS Conf. Simul. Model. SIMS 2016, vol. 142, pp. 673–679, 2018.
[126] Z. Zhao and A. Kumar, “Towards More Accurate Iris Recognition Using Deeply Learned Spatially Corresponding Features,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob, pp. 3829–3838, 2017.
[127] C. G. Campbell, D. T. Laycak, W. Hoppes, N. T. Tran, and F. G. Shi, “High concentration suspended sediment measurements using a continuous fiber optic in-stream transmissometer,” J. Hydrol., vol. 311, no. 1–4, pp. 244–253, Sep. 2005.
[128] Computer Graphics staff, “Status report of the graphic standards planning committee,” ACM SIGGRAPH Comput. Graph., vol. 13, no. 3, pp. 1–10, Aug. 1979.
[129] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, “Surface reconstruction from unorganized points,” in Proceedings of the 19th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’92, 1992, pp. 71–78.
[130] K. Higa, K. Iwamoto, and T. Nomura, “Multiple object identification using grid voting of object center estimated from keypoint matches,” in 2013 IEEE International Conference on Image Processing, Sep. 2013, pp. 2973–2977.
[131] C. Silpa-Anan and R. Hartley, “Optimised KD-trees for fast image descriptor matching,” in 2008 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2008, pp. 1–8.
[132] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, pp. 2261–2269, 2017.
[133] J. Thompson, P. Flynn, K. Bowyer, and H. Santos-Villalobos, “Effects of iris surface curvature on iris recognition,” IEEE 6th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS), 2013.
[134] E. Ortiz, K. W. Bowyer, and P. J. Flynn, “A linear regression analysis of the effects of age related pupil dilation change in iris biometrics,” IEEE 6th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS), pp. 0–5, 2013.
[135] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, “Pupil dilation degrades iris biometric performance,” Comput. Vis. Image Underst., vol. 113, no. 1, pp. 150–157, Jan. 2009.
[136] I. Tomeo-Reyes, A. Ross, A. D. Clark, and V. Chandran, “A biomechanical approach to iris normalization,” in 2015 International Conference on Biometrics (ICB), May 2015, pp. 9–16.
[137] K. W. Bowyer, S. E. Baker, A. Hentz, K. Hollingsworth, T. Peters, and P. J. Flynn, “Factors that degrade the match distribution in iris biometrics,” Identity Inf. Soc., vol. 2, no. 3, pp. 327–343, 2009.
[138] I. Tomeo-Reyes and V. Chandran, “Effect of pupil dilation and constriction on the distribution of bit errors within the iris,” IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., no. June, pp. 40–47, 2014.
[139] J. E. Zambrano, D. P. Benalcazar, D. A. Montecino, C. A. Perez, and K. W. Bowyer, “Analysis of Iris Deformation Patterns due to Pupil Dilation Using Non-Rigid Image

Registration and Displacement Fields," Submitt. to 2021 IEEE Winter Conf. Appl. Comput. Vis., pp. 1–10, 2021.
[140] C. K. S. Leung and R. N. Weinreb, "Anterior chamber angle imaging with optical coherence tomography," Eye, vol. 25, no. 3, pp. 261–267, 2011.
[141] H. Fu et al., "Segmentation and quantification for angle-closure glaucoma assessment in anterior segment oct," IEEE Trans. Med. Imaging, vol. 36, no. 9, pp. 1930–1938, 2017.
[142] H. Fu et al., "Multi-context deep network for angle-closure glaucoma screening in anterior segment OCT," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, pp. 356–363.