Image-based Eye Pose and Reflection Analysis for Advanced Interaction Techniques and Scene Understanding

January 2011

Christian NITSCHKE

Image-based Eye Pose and Reflection Analysis for Advanced Interaction Techniques and Scene Understanding

A dissertation
submitted to the
Graduate School of Information Science and Technology
Osaka University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

January 2011

Christian NITSCHKE

Thesis Committee:

Prof. Haruo Takemura (Osaka University)
Prof. Yasushi Yagi (Osaka University)
Assoc. Prof. Katsuyoshi Miura (Osaka University)
Lect. Atsushi Nakazawa (Osaka University)

List of Publications

Journals

1. C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration from eye reflections. In Japanese. *IEICE T. Inf. Syst. (Special Section on Image Recognition and Understanding)*, J93-D(8):1450–1460, 2010.

2. C. Nitschke, A. Nakazawa, and H. Takemura. Real-time space carving using graphics hardware. *IEICE T. Inf. Syst. (Special Section on Image Recognition and Understanding)*, E90-D(8):1175–1184, 2007.

International Conferences

Peer-reviewed

1. C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration from eye reflections. *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1226–1232, 2009.

2. C. Nitschke, A. Nakazawa, and H. Takemura. Eye reflection analysis and application to display-camera calibration. *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 3449–3452, 2009.

3. O. Bimber, G. Wetzstein, A. Emmerling, and C. Nitschke. Enabling view-dependent stereoscopic projection in real environments. Best paper. *Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 14–23, 2005.

4. O. Bimber, G. Wetzstein, A. Emmerling, C. Nitschke, and A. Grundhöfer. Enabling view-dependent stereoscopic projection in real environments. Demonstration. *Proc. ACM SIGGRAPH Sketches/Emerging Technologies*, 2005.

5. G. Schatter, E. Züger, and C. Nitschke. A synaesthetic approach for a synthesizer interface based on genetic algorithms and fuzzy sets. *Proc. International Computer Music Conference (ICMC)*, pages 664–667, 2005.

Non-peer-reviewed

6. A. Nakazawa, C. Nitschke, K. Kiyokawa, and H. Takemura. Real-time space carving using graphics hardware and its applications. Invited talk. *Proc. Asiagraph*, pages 43–48, 2007.

7. Y. Ai, R. Gerling, M. Neumann, C. Nitschke, and P. Riehmann. TIRA – Text-based information retrieval architecture. *Proc. International Workshop on Text-Based Information Retrieval (TIR)*, pages 67–75, 2005.

Domestic Conferences

Peer-reviewed

1. C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration from eye reflections. *Proc. Meeting on Image Recognition and Understanding (MIRU)*, pages 1388–1395, 2010.

Non-peer-reviewed

2. C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration from eye reflections. *Technical report of IEICE. PRMU*, 109(470): 205–210 (PRMU2009-268), 2010.

3. T. Ohnishi, C. Nitschke, A. Nakazawa, and H. Takemura. A handcart-type 3D reconstruction system using a laser range finder and a single camera. *Proc. IEICE General Conference*, page 199 (D-12-88), 2010.

4. C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration from eye reflections. *Proc. Meeting on Image Recognition and Understanding (MIRU)*, pages 521–528, 2009.

5. C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration from eye reflections. *IPSJ SIG Notes. CVIM*, 2008(82):115–122 (2008-CVIM-164-(19)), 2008.

6. C. Nitschke, A. Nakazawa, and H. Takemura. Real-time space carving using graphics hardware. *Proc. Meeting on Image Recognition and Understanding (MIRU)*, pages 928–933, 2006.

7. G. Schatter, E. Züger, C. Nitschke, M. Pradella, and D. Linke. Intuitive graphical user interfaces for the electronic sound generation based on genetic algorithms and fuzzy sets. In German. *Proc. VDT Audio Convention*, 2004.

Theses

- C. Nitschke. A framework for real-time 3D reconstruction by space carving using graphics hardware. Diplom thesis. Faculty of Media, Bauhaus-Universität Weimar, Germany, 2006.

Books

- C. Nitschke. 3D Reconstruction – Real-time Volumetric Scene Reconstruction from Multiple Views. ISBN 3836410621. VDM Verlag Dr. Müller, Saarbrücken, Germany, 2007.

Patents

- A. Nakazawa, and C. Nitschke. Apparatus and method for calibration-free eye gaze tracking. Osaka University, Application 2010-199849, 2010.

Abstract

This work proposes a theory of the light transport at the corneal surface of the human eye including multiple eye poses. The theory is subsequently applied to solve two practical problems in scene reconstruction and interaction techniques. Related with these are the solutions to two general problems in scene reconstruction from multiple eye images.

As the eyes are the interface between a human body and the visual information of the physical world, their movements also convey rich details for interpreting a person's affective state, behavior, and relation with the environment. Despite having numerous applications in a variety of fields, current approaches to extract information from eyes are lacking, being intrusive, restricted to laboratory conditions, and not providing sophisticated ways to integrate the eye with its environment. Recently, the geometric relation between eye and camera, that can be obtained from a face image, has been formalized to analyze light reflections in the cornea of a single eye or in a pair of eyes to recover simple scene structure. Nevertheless, there exist no solutions for relating reflections among multiple eyes, probably imaged by different cameras, with the structure of the surrounding environment. This, however, is crucial to develop sophisticated strategies for geometric eye analysis in arbitrary environments. This study aims to provide a solution.

The first practical problem lies in calibration to obtain display pose in display-camera setups. Understanding the combination of display and camera as a controlled system enables a range of interesting vision applications in non-professional environments, including object/face reconstruction and human–computer interaction, but attempting to do this in average homes has been difficult as current approaches require special hardware and tedious user interaction. This work proposes a novel calibration approach that simplifies this by building on the observation that a user is commonly located in front of the setup and that screen reflections in the cornea of the eye can be extracted from face images. Despite the difficult constraints, results obtained are feasible and should be sufficient for many applications involving non-intrusive calibration-free dynamic setups.

The main question then becomes what accuracy can be expected for scene reconstruction from multiple eye images. For this discussion, significant factors that affect accuracy are identified among individual eye geometry, camera parameters and geometric relation in the setup. Comprehensive experimental evaluation shows that, due to common errors in eye image processing and an unknown shape for the individual eye, scene reconstruction results in a large error and cannot be applied directly. To compensate for this, an optimization strategy is developed that exploits geometry constraints within the system to jointly improve eye poses and scene structure. Results show that the method performs accurately and stably with respect to varying subjects, scene alignments, eye positions, and gaze directions.

The second practical problem relates to non-intrusive eye gaze tracking in arbitrary environments. Flexible techniques for tracking a person's point of regard enable human–computer interaction and diagnostic studies with a range of applications in different fields. While eye gaze tracking has been an active area of research for over five decades, state-of-the-art approaches share major limitations restricting applications to controlled laboratory conditions with experienced personnel and a high degree of intrusiveness. This work proposes a novel system architecture that overcomes this by building on the observation that projected invisible structured light assigns environment locations with information that can be uniquely identified from corneal reflections. The approach is the first to support arbitrary surfaces and not require geometric calibration. Combined with unobtrusiveness and robustness to practical conditions, it enables a wide range of applications for novel user groups and situations.

Applying invisible structured light projection to corneal reflection analysis provides a solution to the general problem of accurate and robust feature matching among multiple eye images. The existing approach based on the epipolar geometry between a pair of eyes suffers from several shortcomings related to dependency on pose, shape, and reflection properties of the eye. Beside eliminating these, the proposed strategy provides a dense matching, is purely image-based, and thus, naturally enables feature matching between eye and conventional images. This is crucial for combining eye-specific information such as point of regard, peripheral vision, and visual field with high quality image data or scene geometry.

The results of this work have implications on several fields. The findings provide general insight on the application of eye reflections for geometric reconstruction and are an important contribution. Linking eye and environment information can lead to novel insights and understanding.

Keywords— Computer vision, image analysis, calibration, reconstruction, interaction, visual system, eye model, eye pose estimation, eye gaze tracking, catadioptric imaging system, corneal reflection, light transport, display-camera system, projector-camera system, structured light

Acknowledgments

The writing of this dissertation has been an incredible journey and a monumental milestone in my academic life. I could not have embarked on and accomplished this endeavor without the passionate and continued support of advisors, colleagues, family, and friends. I offer my regards and blessings to all of those who supported me in so many ways during the completion of the project.

This work was done under the supervision of Prof. Haruo Takemura of the Graduate School of Information Science and Technology at Osaka University. I would like to express profound gratitude to his continuous encouragement, guidance, and invaluable advice, especially at key moments in my work.

I am especially grateful for the indispensable assistance of my supervisor, Lect. Atsushi Nakazawa, for his inspiration, guidance, and dedication at any time of the day or night. Throughout my research and thesis-writing, he provided encouragement, academic experience, sound advice, and discussions with many good ideas.

I would like to acknowledge Prof. Yasushi Yagi and Assoc. Prof. Katsuyoshi Miura, as members of the thesis committee, for their efforts and discussions with insightful comments on my research.

I express my sincere gratitude to the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) for providing me with the prestigious Monbukagakusho Scholarship and academic support throughout the doctoral program.

I have also benefited greatly through being able to meet and work with a number of people at Osaka University. I want to thank all the members—staff and students—of Takemura Laboratory for their help and support. I especially would like to mention and thank Assoc. Prof. Kiyoshi Kiyokawa for his encouragement and support, Anuraag Agrawal for being such a great source of help with proofreading publications; and the students of the modeling team, especially Alexander P. Radkov, Takayuki Ohnishi, and Masato Takami, who were related to my research. Furthermore, I am grateful to the staff at the Graduate School of Information Science and Technology and Cybermedia Center for their guidance and support with official procedures.

I appreciate the assistance of Dr. Ko Nishino for making available a copy of his VisualEyes software, Dr. Yannick Francken and Nathan Funk for sharing insights on their display-camera calibration methods, Dr. Micah K. Johnson for supplying example code on rendering synthetic eye images, and Ferry Häntsch for providing information on hardware-triggering for projector-camera synchronization. I am indebted to everyone who participated in the numerous exhausting experimental studies. Without their commitment and endurance, this work would not have been possible.

Special thanks are given to my former supervisor, Prof. Oliver Bimber, who gave me the opportunity to carry out my research for the Diplom thesis

at Osaka University, and with that, made this current study possible. I also would like to express my gratitude to Prof. Bernd Fröhlich, Prof. Benno Stein, Dr. Günther Schatter and Dr. Bernd Schalbe for their supervision, guidance and support during my time at the Bauhaus-Universität Weimar.

I am grateful to Prof. Hiroki Iwai at the Institute for Higher Education Research and Practice, Osaka University, for giving me the opportunity to work as a teaching assistant in undergraduate classes on German as a foreign language. Beside providing a good excuse to get away from my study, it let me gain a new perspective on my mother tongue and culture.

Many thanks to Kumi for always believing in me and for the patience and understanding of all the unexpected deadlines that may happen during doctoral study.

Lastly, and most importantly, I wish to express my sincere thanks to my parents, Petra Nitschke and Detlef Nitschke, for their understanding and support, and for giving me the freedom and opportunity to pursue my own interests. To them I dedicate this thesis.

Christian Nitschke
Osaka University
January 2011

Contents

List of Tables

List of Figures

List of Acronyms

ANOVA	Analysis of variance
BRDF	Bidirectional reflectance distribution function
CFF	Critical flicker frequency
CPU	Central processing unit
CRT	Cathode ray tube
CSL	Coded structured light
DCT	Discrete cosine transform
DLP	Digital Light Processing
DMD	Digital Micromirror Device
EGT	Eye gaze tracking
EM	Expectation maximization
EOG	Electro-oculography
FOV	Field of view
FPS	Frames per second
GPU	Graphics processing unit
GT	Ground truth
GUI	Graphical user interface
HCI	Human–computer interaction
HUD	Head-up display
IEEE	Institute of Electrical and Electronics Engineers
IR	Infrared
LCD	Liquid crystal display
LED	Light-emitting diode
LoG	Line of gaze

LoS	Line of sight
MRF	Markov random field
PC	Personal computer
PCCR	Pupil center corneal reflection
PoR	Point of regard
PTZ	Pan, tilt and zoom
RANSAC	Random sample consensus
RMSE	Root mean square error
ROI	Region of interest
SIFT	Scale-invariant feature transform
SLAM	Simultaneous localization and mapping
SNR	Signal-to-noise ratio
SVD	Singular value decomposition
TFT	Thin-film transistor
TV	Television
USB	Universal Serial Bus
VGA	Video Graphics Array
VOG	Video-oculography

Chapter 1
Introduction

Our eyes are one of the most important sense organs allowing vision and providing us with rich information content about our physical world. They are important to the exploration, analysis, perception of, and interaction with visual information. Thus, eye movements contribute a key part to the interpreting and understanding of a person's wishes, needs, tasks, cognitive processes, affective states, and interpersonal relations. As this information is relevant to a large number of applications in a variety of fields (Duchowski, 2002), eye gaze tracking is one of the most common problems with a long tradition in the image-based analysis of eye-related information (Duchowski, 2007; Young and Sheena, 1975; Hansen and Ji, 2010). At the present day, however, it has not emerged from the status of merely being applied as a research tool in laboratories, operated by professionals with technical knowledge and long-time experience. It is necessary to develop novel strategies enabling eye gaze tracking and eye context analysis to meet the requirements of the outside world.

Traditionally, interaction with computer systems is restricted to a small number of input modalities, such as typed text using a keyboard, or free-form drawing and gestures using a mouse and other mouse-like pointing devices. The lack of sophisticated and powerful multimodal input capabilities remains a bottleneck in human computer interaction (Sharma et al., 1998; Dumas et al., 2009). This may not concern the average user, but it is essential for a disabled user with limited motor function depending on alternative forms of input. Another issue relates to the rapid progress in computing technology and usage scenarios. While barely 30 years ago the personal computer just emerged, we are now surrounded by networked information infrastructures and ubiquitous devices. Furthermore, electronic machinery in our environment comes with basic computing and input/output capabilities, what can be interpreted as an indicator for upcoming ambient environments. However, while technology rapidly evolves, a lot of effort is required to keep algorithms and paradigms up with this pace.

Image-based Eye Analysis. Vision-based analysis techniques have the ability to facilitate remote non-intrusive interfaces or smart sensors in an ambient environment. There exists a variety of visual information that can be exploited in images and videos of the human body, relating to vision tasks

such as body, hand, and head tracking; recognition of postures, gestures, and activity; face detection, location, recognition, and expression analysis; and eye gaze tracking (Turk, 2004). Due to the range of potential applications, especially eye gaze tracking receives large research interest. It is, however, not the only task in image-based eye analysis. Since eyes are one of the most salient features of the human face, their unique geometric and photometric properties provide important visual cues for obtaining face-related information, for application in face detection, recognition, or expression analysis. The unique appearance of structures of the eye is exploited in biometrics for iris recognition or retinal scanning. The cornea is the protective and optical outer layer of the eye covering the iris. Due to its transparency, the cornea itself is not relevant to image analysis. What is often overlooked, however, are its mirror-like reflection characteristics that cause specular reflection of environmental light. Regarding this property, the combination of camera and corneal reflector can be seen as a catadioptric imaging system capturing a wide-angle view of the environment. Calibrating this system by retrieving the pose of the eye—for example from eye features that can be detected in an image, such as the contours of iris and pupil—and aligning a model of the corneal surface geometry, enables a variety of applications for omnidirectional vision (Yagi, 1999). Refer to Figure 1.1 for close-up views of an eye showing corneal reflections, iris texture, and corneal surface characteristics.

Relation between Eye and Environment. More important than traditional vision applications solely focusing on information about the environment, however, are probably applications relating this information to the context of the individual (eye). For example, by projecting an environment map from corneal reflections onto the retina to obtain an image of what a person is seeing and looking at, or by computing a spherical frontal-view panorama to determine the location and situation under which a person is photographed. Relating environment and gaze allows for a combined analysis of stimulus and response in higher-level tasks, such as the study of human behavior and affect (Nishino and Nayar, 2006).

Scene Reconstruction from Corneal Reflections. When capturing two eyes of a person with a static camera—for example in a single face image—the two corneal reflectors and the camera act as a catadioptric stereo system where each environment location is imaged from two different viewpoints. Finding correspondences between both cornae, the 3D structure of the scene can be obtained by triangulation. The generalization of this strategy would be a scenario where a moving camera captures a moving eye—for example in flexible non-intrusive tracking applications. Recovering eye poses and matching feature correspondences among video images, a 3D model of the scene can be obtained and aligned with camera and eye pose trajectories. Such simul-

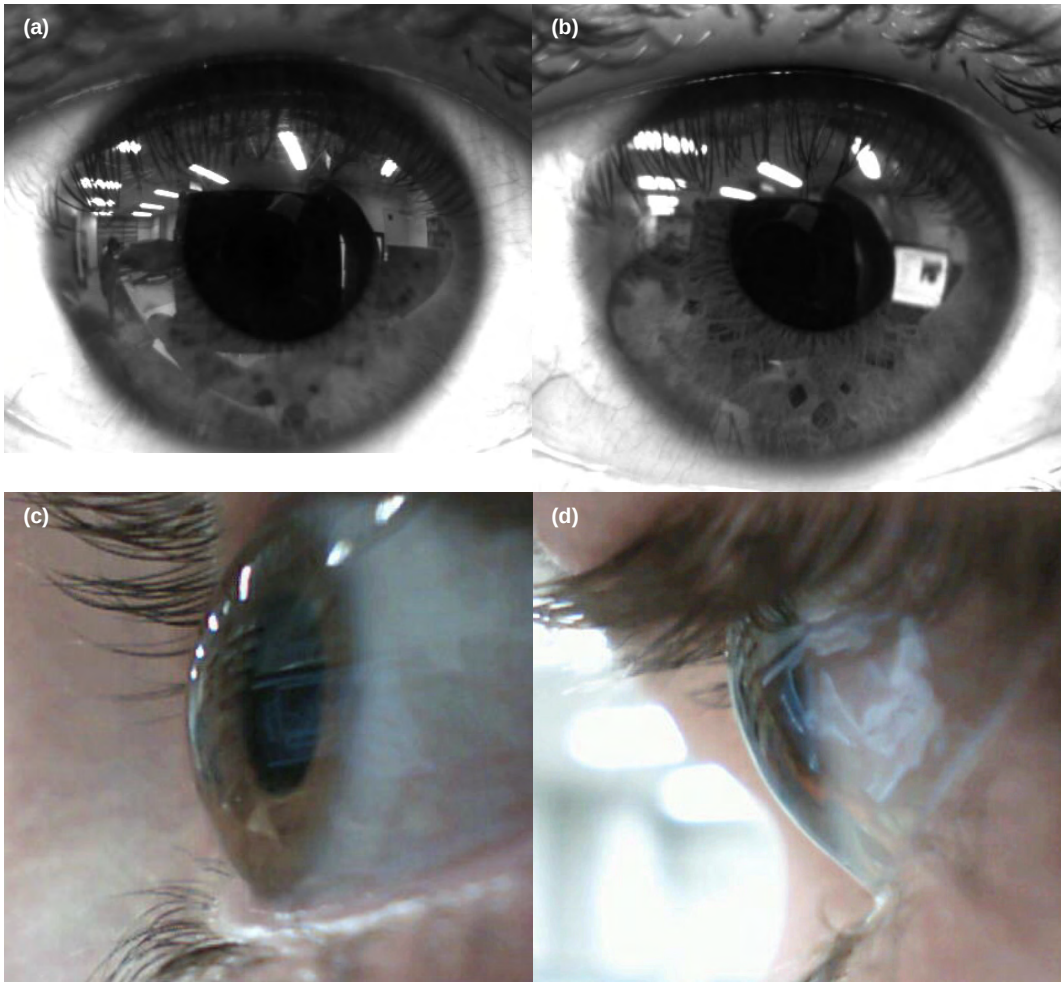


Figure 1.1: Corneal reflections. (a) The reflected office environment is clearly visible in the eye image. The superimposed diffuse reflections from the pattern of the iris tissue disturb the image of the specular reflections. Note that the camera is placed directly in front of the eye and, therefore, the central area remains dark. (b) A similar scene, but with focus on the pattern of the iris tissue. (c) A view from the side shows the transparent reflective surface of the cornea. (d) A close view reveals the corneal limbus which is the surface shape discontinuity where the transparent cornea dissolves into the white sclera with lower curvature.

taneous localization and mapping (SLAM) (Muhammad et al., 2009) for the corneal imaging system naturally integrates eye gaze tracking and environment information, enabling for the described applications as well as unobtrusive, uncalibrated future interfaces in ubiquitous and ambient scenarios.

Nevertheless, there is a long way from a first theoretical model of the corneal stereo system between two eyes (Nishino and Nayar, 2006) towards a practical strategy for corneal SLAM from multiple eyes, requiring comprehensive

knowledge about the geometry between camera, eyes, and scene, and solutions to a range of problems. With this work, we want to approach this goal and present the first study on corneal reflection and environment relation analysis under multiple eye and camera poses. In the next section, we identify particular problems and explain our proposed solutions.

1.1 Contribution

This work proposes a theory of the light transport at the corneal surface of the human eye including multiple eye poses (Chap. 3). The theory is subsequently applied to achieve a range of contributions for corneal reflection analysis from eye images, categorized into four topics. The first two topics are related to novel methods that are proposed in order to solve practical problems in display-camera calibration (Chap. 4) and eye gaze tracking (Chap. 5). The remaining two topics cover the general problems of accuracy in scene reconstruction and matching of feature correspondences. The respective findings and derived strategies are integrated with the novel methods and can be of general relevance to other work in corneal reflection analysis.

Beside this, an overview of the anatomic structures of the eye related to the model-based estimation of its pose from an image is given (Sec. 2.1). Reviewing studies on anthropometric variation, schematic eye models, and eye models applied within related work, a geometric eye model with spherical curvature and constant parameters is developed. The results of this work are based on that model.

In the context of eye pose estimation (Sec. 2.2), a detailed derivation is given for the projection of a circle with arbitrary 3D position and orientation into an image. Based on the imaged contour, two methods for either perspective and weak-perspective projection are explained to reconstruct the 3D pose of the original circle. It is shown how the methods can be applied to eye pose estimation from circular eye features, such as pupil and iris contour.

Display-Camera Calibration from Eye Reflections. With advances in vision algorithms, the webcam emerges from its status of solely being a tool for videoconferencing. Relating the camera to the physical context of the PC setup, camera and CRT/LCD monitor form a controlled system. In the past, there have been two major areas of application for such display-camera systems. One is the acquisition of object shape and reflectance using the display as a controlled illumination device, the other is human–computer interaction (HCI) where the result from image processing produces a feed-back on the screen. Applications typically require a calibration of the geometric relation between display and camera. Since display and camera face the direction of the user, calibration is achieved interactively where the user moves a planar or spherical mirror to make the display visible from the view of the camera.

A novel calibration technique for display-camera setups is described, building on the observation that the cornea of the human eye acts as a partial mirror. For a user moving in front of the setup, it exploits the fact that corneal reflections of display content can be extracted from face images to reconstruct the position of the display itself. The method has several benefits compared with other approaches:

- It requires only off-the-shelf hardware that commonly exists in PC setups.
- It does not need user interaction or awareness.
- It supports dynamic setups since at least a single face image is required.
- Additionally, it estimates eye poses which is beneficial when eye gaze tracking applications are involved.

The method is motivated by Nishino and Nayar's method for image-based eye pose estimation without using active controlled illumination (Nishino and Nayar, 2006) and Francken et al.'s method for display-camera calibration using a spherical mirror (Francken et al., 2007). Thorough experimental evaluation shows that the straightforward application of both methods results in a large error. To compensate for this, an optimization strategy is proposed that jointly improves display-camera calibration and eye pose estimation, subject to geometry constraints in the scene.

A closed-form linear least-squares solution is developed for the triangulation of multiple inverse reflection rays. The method reconstructs a 3D scene location as the point with minimal distance to the set of corresponding inverse reflection rays obtained under multiple eye poses. This can be relevant to scene reconstruction from eye images in general.

Calibration-free Non-intrusive Eye Gaze Tracking in Arbitrary Environments. While recent developments in the field of remote eye gaze tracking are promising, state-of-the-art techniques are still far from being unobtrusive and usable for practical applications. There are different characteristics that restrict their application to work-intensive controlled laboratory conditions with experienced instructors and trained users. Moreover, techniques still feature a high degree of intrusiveness ranging from setup requirements to operation restrictions due to their technical approaches and hardware limitations.

A novel system architecture is proposed to overcome several limitations of existing eye gaze tracking techniques. Specifically, it removes the need for geometric calibration and enables application with arbitrary dynamic scenes. The architecture combines a number of benefits compared with other approaches:

- no requirement for geometric calibration,

- no requirement for body-attachments,
- support of arbitrary environments,
- support of free head-movement,
- support of challenging conditions,
- improved accuracy.

The proposed system architecture allows for increased applicability, not possible with existing techniques. Due to easy setup, tolerance to environmental conditions, and the same time increased accuracy, it has the potential to make eye tracking available for non-professional users in everyday environments. Furthermore, due to absence of calibration, body-attachments, and tolerance to operation conditions, it enables practical applications generally requiring unobtrusiveness, either to achieve natural and unbiased conditions in diagnostic scenarios or to develop interactive interfaces for ubiquitous and ambient environments.

The system uses at least two cameras, a non-attached eye camera with either high-resolution or a pan-tilt-zoom (PTZ) tracking architecture to capture a close-up view of the eye, and one or multiple environment cameras capturing the gazed scene. A method is developed to estimate the corneal reflection of the gazed point of regard (PoR) in the eye image and map it into the environment images based on scene feature correspondences. In order to robustly obtain a large number of accurate feature matches under severe conditions in arbitrary environments, non-intrusive coded structured light projection is applied, which is not perceived by the user but recovered from the camera images. In case of a known geometric relation between projector and environment cameras, a 3D model of the scene can be reconstructed.

There are several contributions involved with the proposed method that are relevant to corneal reflection and environment relation analysis in general:

- A closed-form solution is developed to calculate the forward projection for an imaged sphere as the surface location where light from a source reflects into the direction of the camera. Five methods are proposed, regarding the available knowledge about the distance between sphere and light source (scene). The methods are applied to calculate the location where the PoR reflects on the corneal surface.
- A closed-form solution is developed to calculate the distance between back-projection rays after reflection at the surface of a sphere. Three methods are proposed, regarding the available knowledge about the distance between sphere and lights source (scene). The methods are applied to calculate interpolation weights for neighboring corneal reflection rays at image locations where no correspondence information is available.

Accuracy of Scene Reconstruction from Eye Images. There are a lot of parameters involved in scene reconstruction from corneal reflections, regarding individual eye geometry, camera settings, and image quality; and regarding the geometric relation between camera, eyes, and scene. Careful evaluation is important to understand their effect on the overall accuracy, where the insight is helpful to develop compensation strategies.

Applying the proposed display-camera calibration framework, a large number of comprehensive experimental studies is conducted with real and synthetic data to evaluate the accuracy of scene reconstruction from corneal reflections, comparing the results to ground-truth data and results obtained with spherical mirrors of known size. The findings provide a tool to assess the quality that is expected for a particular setup, and an aid for the decision where compensation strategies are best applied.

Straightforward triangulation of inverse reflection rays is found to generally result in a relative high error. To compensate for this, an optimization framework is proposed that performs joint refinement of eye poses, reflection rays, and scene points subject to known geometry constraints from the scene. The performance of this framework is demonstrated in the context of the display-camera calibration algorithm with a large number of comprehensive experimental studies. The proposed framework can be generally relevant to scene reconstruction from eye images when geometry constraints are available.

To obtain synthetic data, a framework is developed for physically based rendering of eye images with corneal reflections from environmental illumination. It uses an extended eye model with aspherics where eye structures are modeled as ellipsoids and cross sections. The framework provides a general tool to analyze the impact of different parameters on scene reconstruction from, especially where ground-truth measurements are difficult to obtain as with parameters related to the individual eye.

Accurate and Robust Correspondence Matching among Multiple Eye and Scene Images. While in theory the epipolar geometry of the corneal stereo system (Nishino and Nayar, 2004b, 2006) is a helpful tool to simplify the correspondence problem—providing a reduced search-space and, therefore, inherently increased accuracy—there are several problems related with this approach in practice:

Accuracy The epipolar geometry depends on the result of eye pose estimation. Small estimation errors lead to different epipolar curves resulting in false matches. Performing this strategy with multiple eye images causes error accumulation.

Robustness While the correspondence problem among views of perspective cameras is well studied (Tuytelaars and Mikolajczyk, 2008), the

available techniques cannot be directly applied to corneal reflection images. Specific problems are the overlap of iris texture and corneal reflections (Wang et al., 2008), the low reflectivity of the cornea of less than 1% (Kaufman and Alm, 2003), and the non-planar surface geometry of the cornea (Hansen et al., 2007; Scaramuzza et al., 2008). Furthermore, it is necessary to match correspondences not only among eye images, but also between eye and scene images to integrate eye and scene related information. This is an important requirement, since the result enables the computation of eye related information from conventional images with much higher quality. Solving the problem of accurate and robust correspondence matching is not only a key contribution to scene reconstruction, but can also be relevant to improve eye pose estimation itself.

To solve the described problems, a novel strategy is proposed for the matching of feature correspondences among corneal reflection and scene images, based on non-intrusive coded structured light. Up to the knowledge of the author, this is the first approach to apply coded structured light projection to eye image and corneal reflection analysis. The proposed strategy includes a range of benefits described in the following.

- The projected feature points define a calibration-free relation between eye and environment camera views.
- Projecting feature points into the environment is a flexible way to obtain high spatial resolution and wider area-coverage on the corneal surface than achieved with a small number of point light sources in front of the user, a common approach in active light methods for eye pose estimation.
- Applying coded structured light increases spatial resolution and allows for robust detection. Experimental results verify the robustness under challenging conditions, such as short exposure, image noise, and environmental light.
- Applying imperceptible or invisible structured light, the dynamic code projection is not perceived by human observers. Additionally, imperceptible codes can be removed from camera images to recover the texture of the scene.

1.2 Relation

1.2.1 Image-based Eye Analysis

Eyes are one of the most salient features of the human face. Thus, their unique geometric and photometric properties provide important visual cues for obtaining face-related information, for example in face detection (Hsu et al.,

2002), face recognition (Zhao et al., 2003), and facial expression understanding (Fasel and Luettin, 2003). A detailed overview of the interests and applications in image-based eye analysis follows. It focuses on the two major areas of corneal reflections and iris appearance.

1.2.1.1 Corneal Reflections

Corneal Shape and Position. Since the cornea of the human eye exhibits mirror-like properties, specular reflections from eye images have been exploited by several works in different areas. The main application is to obtain information about the eye. In biomedicine, a detailed knowledge of corneal surface geometry is required for various ophthalmologic applications, such as refractive surgery, change monitoring, disease diagnosis, or contact lens development. The corneal topography is usually measured by a non-intrusive optical technique known as videokeratography (Mandell, 1996; Bogan et al., 1990). Halstead et al. (1996) describe a popular algorithm that reconstructs a 3D surface model of the human cornea from an image. It analyzes specular reflections from a pattern of concentric rings generated by a special illumination device. A model is then fit to the surface normals computed from the imaged reflection features and the geometry of the videokeratograph device. Another application for specular highlights from known light sources is eye position and gaze direction estimation in eye gaze tracking.

Environmental Light. More recently, the cornea is exploited as a light probe to obtain information about the light distribution in the environment of a person. Without involving explicit eye modeling, Backes et al. (2008) present an eavesdropping technique to recover screen content from reflections in the user's eyes at faraway locations using a telescope mounted on a camera. As the resulting quality is largely affected by system parameters, the particular causes for blur from motion, defocus, and diffraction are analyzed. While it is appealing to exploit corneal reflections, drawbacks include the small radius and low reflectance of the cornea. This requires long exposure times and large apertures which leads to motion and defocus blur. To account for that, Backes et al. (2009) apply non-blind image deconvolution techniques. The corresponding point spread functions are determined by either an offline or online approach. Previous results in Backes et al. (2008) could be noticeably improved, allowing for larger distances and smaller content sizes. While this work is interesting in terms of improving the detail of environmental reflections in eye photography, it does not exploit information about the geometric and photometric properties of the eye.

The majority of applications, analyzing eye reflections by relying on 3D modeling of eye pose and geometry, do not require detailed knowledge of the individual corneal shape. In fact, since deviation among different persons is relatively small (Snell and Lemp, 1997; Kaufman and Alm, 2003), the

cornea can often be approximated as an ellipsoid with average parameters. In computer graphics and vision, corneal reflections are exploited because they capture the illumination distribution surrounding the person. Tsumura et al. (2003) are the first to recover information about lighting from specular highlights in the eye. They place three point light sources at known positions and extract the corresponding highlights in an eye image. The inverse reflection directions towards the light sources are then estimated to reconstruct a 3D face model by photometric stereo (Woodham, 1980) which is applied for face relighting. Johnson and Farid (2007) analyze corneal reflections to identify digital forgeries where an image is composed from individuals photographed under inconsistent lighting conditions. They perform eye pose estimation, specular highlight extraction, and inverse raytracing to estimate the direction of light sources from eye images. Internal camera parameters are automatically obtained from the perspective distortion of the iris contour.

Corneal Imaging System. Looking at a photograph of an eye, it can be observed that many details of environmental structure and illumination are captured over a wide angle. Thus, beyond the simple task of obtaining light directions from specular highlights, it is eligible to ask how to recover the entire visual information of the environment captured in an eye image. By formally describing the imaging characteristics of the eye-camera geometry, it becomes possible to apply standard vision theory and algorithms to flexibly analyze the system and process the information content.

Nishino and Nayar (2004b, 2006) provide the first comprehensive analysis of the visual information that is embedded within an image of the human eye. They find that the cornea and a camera viewing the eye form a catadioptric imaging system which they refer to as the corneal imaging system. In contrast to common catadioptric configurations, the relation between the corneal reflector and the camera does not remain fixed, and thus, calibration is required for each frame in the form of eye pose estimation. Due to the flexible relation and the individual shape of the cornea, the catadioptric system generally does not have a single viewpoint (Baker and Nayar, 1999; Geyer and Daniilidis, 2001) but rather a caustic of viewpoint locations (Kuthirummal and Nayar, 2006; Swaminathan et al., 2006). Nishino and Nayar further introduce a geometric model of the cornea based on anthropometric studies and describe how to determine its pose from an eye image. Several properties of the corneal imaging system, such as field of view, resolution, and locus of viewpoints, can be analyzed from this model. The extracted environment map can be transformed into a spherical panorama to obtain a frontal view of the environment or projected onto the retina to obtain an image of what a person is seeing. The detailed irradiance map allows for face reconstruction and relighting (Nishino and Nayar, 2004a).

Another application of eye images is security systems using biometrics to automatically identify a person. A method enabling non-intrusive and large-

scale crowd surveillance is face recognition (Zhao et al., 2003). Practical solutions are required to perform reliably under the large variation of facial poses and illumination found in real scenarios. Regarding illumination, Nishino et al. (2005) use the described corneal imaging system to estimate the environment map and propose an appearance-based approach for face recognition that exploits lighting conditions estimated from corneal reflections. Advantages are an increased recognition rate and the ability to use only a single database image per person, instead of multiple images taken under varying illumination. They analyze face appearance variation across different persons under target lighting conditions to synthesize database images for recognition. While the face information is purely image-based, 3D model-based face recognition may be realized with photometric stereo from corneal reflections (Nishino and Nayar, 2006; Tsumura et al., 2003).

Corneal Stereo System. The combination of two eyes imaged by a static camera, for example the two eyes in a face image, can be seen as a catadioptric stereo system (Nayar, 1988; Nene and Nayar, 1998). The system is calibrated by estimating the poses of both eyes. Then, not only the direction towards a scene location, but also its position can be reconstructed as the intersection of the inverse reflection rays from both eyes. As for a conventional stereo system consisting of two perspective cameras, the epipolar geometry can be also formulated for a catadioptric stereo system consisting of a single camera and two reflectors (Pajdla et al., 2001; Svoboda and Pajdla, 2002). Regarding the corneal stereo system (Nishino and Nayar, 2004b, 2006), the epipolar plane formed by the two viewpoints[1] and the reflection of a scene location on the surface of one cornea, intersects the surface of the second cornea in an epipolar curve. The search for the corresponding scene reflection, thus, reduces to a search along the epipolar curve.

Separation of Corneal Reflections and Iris Texture. Techniques for image-based eye analysis either exploit specular corneal reflections or diffuse reflections of iris texture. Often both are present in a single eye image and act as mutual noise. Iris recognition systems, for example, apply active infrared (IR) illumination to reduce specular corneal reflections. On the other hand, diffuse iris reflection can cause substantial distortion when analyzing corneal reflections, especially in blue or green eyes. Separating both kinds of reflections is an ill-posed task if no other constraints are available. Wang et al. (2005a, 2008) introduce a method that exploits reflections and physical characteristics of a pair of irides to separate corneal reflections of complex environments from diffuse iris texture.

[1] This is a feasible approximation as the viewpoint locus is small compared to the baseline between two eyes.

1.2.1.2 Iris Appearance

Light arriving at the eye which is not specularly reflected at the cornea is refracted and enters the eye. A large part of that light illuminates the iris and reflects back in a diffuse manner (Wang et al., 2008). The resulting iris images have been exploited for various applications.

Iris Analysis. A well known application is biometrics to identify persons based on unique bodily features. Iris recognition is a more intrusive, but highly reliable method that uniquely identifies individual patterns from the detail-rich structure and intricate texture of the iris in high-resolution images (Daugman, 1993, 2004; Wildes et al., 1996; Wildes, 1997; Bowyer et al., 2008). Advantages of the technique include robustness to glasses and contact lenses, validity of the database as the iris texture does not change with age, and comparison speed. The corresponding approaches for iris detection and segmentation can be relevant to other fields such as medical imaging or eye gaze tracking. Note that there exists another less common image-based biometrics method, retinal scanning. It requires a special camera to identify the unique structure of blood vessels (vascular pattern) in the retina at the backside of the inner eye (Simon and Goldstein, 1935; Hill, 2002).

Iris Synthesis. Eye information is also explored for iris synthesis in photorealistic rendering. Lefohn et al. (2003) adopt the layered approach that ocularists take when creating physical models of the iris. They give an in-depth explanation of their applied geometric eye model. Lam and Baranoski (2006) introduce the first biophysically-based light transport model for the iris that simulates the light scattering and absorption processes within the iradial tissues. Pamplona et al. (2009) extend iris synthesis to animation and derive a biophysically-based model for the pupil light reflex with iradial pattern deformation under varying illumination.

1.2.2 Eye Gaze Tracking

As one of the most prominent features of the human face, eyes and their movements contribute a key part to the interpreting and understanding of a person's wishes, needs, tasks, cognitive processes, emotional states, and interpersonal relations. Eye movements are important to the navigation, analysis, perception of, and interaction with visual information. As this information is relevant to a large number of applications in a variety of fields (Duchowski, 2002), eye gaze tracking is and has been an active research area for over five decades with foundations dating back further in time.

1.2.2.1 Methods

An eye tracker is a device for measuring eye position, orientation, and movement over time. Depending on physical type and implemented method, eye trackers either measure the position of the eye relative to the head or the pose of the eye in space. Depending on the application, the result is either the location of the eye or the point of regard (PoR) where the eye is looking. Mainly three types of methods for eye tracking systems are distinguished: the contact lens method, electro-oculography (EOG), and video-oculography (VOG).

Contact Lens Method. The most precise measurements of eye movement are achieved with techniques where a special contact lens is attached to the eye (Ditchburn and Ginsborg, 1953; Riggs et al., 1953; Yarbus, 1967). The tracking is done with an embedded radiant spot, mirror, or magnetic search coil. Since conventional contact lenses would slightly slip when the eye rotates, a tight fit is achieved by special designs and methods to increase pressure between contact lens and eye. This causes discomfort in all of the systems, up to requiring the application of a topical anesthetic.

Electro-Oculography. Another type of system tracks the position of the eye by measuring differences in electric potential with skin electrodes attached around the eye (Marg, 1951; Kris, 1960; Shackel, 1967). The source of the electrical energy is a potential field with the positive pole at the cornea and the negative pole at the retina. As the eye rotates, the change in the orientation of the dipole generates a variation in the EOG signal measured from the electrodes. The technique has several advantages: Since it does not require access to the eye, it can be applied even when the eye is closed. It is independent of illumination conditions and, thus, can be used in total darkness or bright outside conditions. Since the EOG signal directly describes the eye position, no further processing is required resulting in a low computational cost.

The contact lens method and EOG are highly intrusive and lack a handling of head movements. Both are early eye tracking techniques where the output has the form of electrical or simple image data that does not require complex post-processing. With advancements in computational power, these methods are nowadays only used for special applications when their unique characteristics are required. A good review is given by Young and Sheena (1975).

Video-Oculography. In this work, we focus on image-based eye tracking known as video-oculography where the information is obtained from a single or multiple cameras with possible use of external light sources emitting invisible light (IR). Tracking the eye consists of two subtasks: eye localization and gaze estimation. Eye localization includes different aspects, such as the detection of

the existence of eyes in an image, the detailed localization and representation of eye features, and the tracking of this information between subsequent frames in video data. Common features for eye localization are the contours and centers points of pupil and visible iris, less common ones are the eye lids and corners. Gaze estimation uses the information from eye localization and corneal reflection analysis to estimate and track the 3D pose of the eye or the PoR on the destination surface.

Types of Calibration. Depending on the particular eye gaze tracking technique, unknown parameters may be determined by a calibration procedure. There exist several classes of parameters that relate to different system properties and can, therefore, change independently. Often parameters from different categories are determined by a joint calibration. According to Hansen and Ji (2010), the following categories are commonly distinguished:

Camera calibration to determine intrinsic camera parameters. For static cameras, this can be done once per camera using standard methods (Hartley and Zisserman, 2003; Zhang, 2000; Bouguet, 2010). For dynamic zoom cameras, the changing focal length is either interpolated from calibration data obtained in advance, or estimated from scene features such as the shape of the visible iris (Wu et al., 2005b). If an explicit calibration cannot be performed, the intrinsic parameters may be completely estimated from the shape of the visible iris (Johnson and Farid, 2007).

Geometric calibration to determine the relative position and orientation among the components of the setup, such as camera, light sources, and scene model (typically a planar surface). This is required once per setup.

Personal calibration to determine parameters of the individual eye geometry, such as corneal curvature, and angular offset between visual and optical axes. This is required once per user and can be subsequently applied to different setups.

Gaze-mapping calibration to determine an implicit regression-based mapping function between eye image measurements (features or appearance) and corresponding gaze locations (PoR) on a planar surface under fixed head pose. This is required once per setup, user, and user pose; and typically performed by asking the user to look at known markers in the scene (Merchant et al., 1974; Stampe, 1993; Morimoto and Mimica, 2005).

Intrusive Stationary Systems. Early image-based systems measure eye movement under a fixed head position where a gaze-mapping to the viewed surface is calibrated in advance. Stationary table- or head-mounted systems track the PoR on a static computer screen, projection canvas, or other planar

surface in front of the user. Bite bars, chin rests, head-mounts, or other forms of fixation are used to ensure that the head remains in its initial pose. Several methods are proposed for compensating head-pose changes with a once calibrated gaze-mapping (Kolakowski and Pelz, 2006; Karmali and Shelhamer, 2006; Zhu and Ji, 2007; Li et al., 2008).

Intrusive Wearable Systems. With advancements in computational power and devices, wearable head-mounted eye gaze tracking systems are introduced more recently to combine mobile tracking in arbitrary environments with head-pose invariance between head and scene. There exist research (Babcock and Pelz, 2004; Li et al., 2006; Wagner et al., 2006) and commercial systems (Tobii Technology AB, 2011b; SR Research Ltd., 2011b; SensoMotoric Instruments GmbH (SMI), 2011a; Mangold International GmbH, 2011a). A tight head-mount ensures that the relation between head and camera remains fixed. The functional principle is similar to that of stationary systems, with the difference that the PoR is tracked in the image of an environment camera recording the scene. The corresponding 3D location of the PoR can be obtained via triangulation using either multiple cameras or tracking scene features in a single moving camera (Munn and Pelz, 2008; Takemura et al., 2010). Gaze-mapping calibration is performed by asking the user to look at a planar surface with a number of projected markers or other feature points that can be identified in the image of the environment camera. After calibration, head pose is required to remain fixed with respect to camera and surface. Therefore, while being wearable, the systems can basically not be applied to moving users and arbitrary environments. The discussed stationary and wearable systems achieve good accuracy, however, are intrusive, stationary, and require a tedious calibration, limiting their application to controlled laboratory environments with experienced personnel.

Requirements for Ideal Systems. An ideal gaze tracking system for practical application needs to follow two main requirements: non-intrusiveness and high accuracy. These, however, are typically mutually exclusive.

Non-intrusiveness There are three requirements related to this point:

Absence of attachments The system should not require attachments to head or body since these need a dedicated setup with time and technical understanding, lead to increasing fatigue even from lightweight attachments, and have an impact on natural behavior due to direct effects on motor activities and perception or due to indirect effects when interacting in society.

Absence of calibration The system should not require an interactive individual calibration since the task needs time, technical under-

standing, and training. That means, parameters are either not required or measured by an automatic calibration approach.

Free head-movement The system should allow relatively free head-movement without the assumption of a fixed relation between head, camera, and scene as in stationary systems. Similar to the use of head and body attachments, non-allowed head movement prevents the user to become comfortable with the system. It causes increasing fatigue from intrusive bite bars and chin rests or from the user trying to keep the head fixed, and error accumulation from possible movements (drift). Further, the configuration requires setup time and user training.

High accuracy The system should maintain a high accuracy. This typically contradicts non-intrusiveness which introduces simplification assumptions to increase usability. Eliminating these assumptions requires compensation with sophisticated, and probably complex, hardware and software architectures yet to be developed.

Non-intrusive Remote Systems. In the last few years a new type of eye tracking systems emerges, having the inherent potential to fulfill the described requirements. So-called remote eye gaze trackers are less intrusive by not requiring body attachments. They use stationary high-resolution cameras or dynamic camera systems to track close-up images of the eye. Dynamic systems have been proposed using either a wide-field-of-view camera for face tracking in conjunction with a moveable PTZ camera for eye tracking (Oike et al., 2004; Yoo and Chung, 2005; Reale et al., 2010), or a static camera with movable mirrors (Kim et al., 2004). Remote eye gaze tracking is usually model-based where imaged eye features are located to track the 3D pose of the eye. Thus, systems do not perform gaze-mapping calibration and instead rely on a combination of camera, geometric, and personal calibration. The gaze direction is obtained by recovering at least two points on the optical axis of the eye such as the centers of pupil, iris/limbus, cornea, or eyeball.

There exist several passive remote eye gaze tracking methods that estimate the gaze direction without using active illumination. These methods are based on locating the center of the limbus by tracking the contour of the visible iris, and either directly estimate the gaze direction like our proposed method (Sec. 2.2) (Wang and Sung, 2001, 2002; Wu et al., 2005b; Nishino and Nayar, 2006; Wu et al., 2007; Schnieders et al., 2010) or in conjunction with head-pose estimation (Chen and Ji, 2008; Yamazoe et al., 2008; Reale et al., 2010). The majority of remote eye gaze tracking methods applies active illumination commonly in form of IR LEDs, and is based on the pupil-center–corneal-reflections (PCCR) technique (Shih et al., 2000; Ohno et al., 2002; Guestrin and Eizenman, 2006; Villanueva and Cabeza, 2007; Zhu and Ji, 2007; Villanueva et al., 2009). The center of the cornea is not directly

visible and usually estimated from corneal reflections of two light sources and a single camera in case of average eye parameters, or two cameras in case of individual eye parameters (Shih et al., 2000; Guestrin and Eizenman, 2006). If required, at least a single-point calibration is necessary to calculate the offset to the visual axis (Shih and Liu, 2004; Villanueva and Cabeza, 2008). Since active illumination usually increases accuracy and robustness, this is the strategy commonly applied in commercial systems (Tobii Technology AB, 2011a,c; SR Research Ltd., 2011a; Smart Eye AB, 2011a,b; Seeing Machines Inc., 2011; SensoMotoric Instruments GmbH (SMI), 2011b; Mangold International GmbH, 2011b).

After computing the gaze direction, the PoR is obtained through geometric modeling as the first intersection of either optical or visual axis with the 3D model of the destination surface. Such a model can be obtained by triangulation using multiple calibrated environment cameras or a single moving camera in conjunction with feature tracking (Smart Eye AB, 2011b; SR Research Ltd., 2011a; Seeing Machines Inc., 2011). Model-based approaches usually assume spherical curvature for cornea and eyeball. Only a few allow generalization to spheroid or ellipsoid models that better describe the eye geometry in the periphery of the cornea and may lead to better accuracy at large gaze angles (Beymer and Flickner, 2003; Nishino and Nayar, 2006; Nagamatsu et al., 2010)

1.2.2.2 Applications

Robust non-intrusive eye detection and tracking is, essential for both, diagnostic offline information retrieval to understand human behavior, cognition, and affective states, and active online analysis with feedback generation to develop interactive and attentive user-interfaces.

There exists a wide range of applications for eye gaze estimation and tracking (Duchowski, 2002). As a tool for passive information retrieval, eye tracking provides an objective and quantitative representation of a person's visual processes. Eye movements are usually recorded to study attentional and behavioral patterns over a provided stimulus. In this context, the observation is typically performed unobtrusively, and the stimulus is not affected by the measurements. The obtained data is then evaluated offline. Experimental assessment is currently the main application of eye tracking, used in a wide variety of disciplines, including cognitive science, psychology, medicine, industrial engineering and human factors, marketing research and advertising. Specific applications are the analysis of human behavior, attention, cognition, and communication; the analysis and diagnosis of anomalies and disease patterns in the visual system or other systems affecting vision; studies on human factors and usability; or studies on the perception of advertising.

Advances in computer systems, computational power, display hardware, and interaction paradigms foster eye tracking to become a powerful input

method for graphical user interfaces and visual applications. While information retrieval is typically an offline process, eye movement as an input modality requires an interactive system that responds to and interacts with the user in real-time. There are mainly two types of paradigms: selective and gaze-contingent. Selective systems apply the PoR as a pointing device such as the mouse cursor to control an application. This is an alternative approach in scenarios where hands can hardly be used, or where eye tracking is found to be faster and less fatiguing than other means of communication. It is, however, a key approach to enable interaction for motor disabled persons who still maintain eye movement control. Gaze-contingent systems exploit knowledge of the user's gaze to adjust the behavior of an application. The paradigm is mainly applied to increase the performance of rendering and data transmission in complex graphical environments (Duchowski et al., 2004). In the future, an improvement of online information retrieval techniques for eye tracking may create novel high-level sensors to establish a broad foundation for gaze-contingent systems.

1.3 Dissertation Overview

The remainder of this dissertation is organized as follows:

Chapter 2 explains the structures of the eye related to this work and surveys studies on eye anatomy and anthropometric variation to introduce a geometric model of the eye. It then surveys approaches to detect and track the eye in an image, followed by introducing methods for estimating the pose of the eye from imaged features.

Chapter 3 builds on eye pose estimation to develop a theory of the light transport at the corneal surface. It introduces a corneal reflection model to enable inverse light path construction for light source direction estimation, and position estimation under multiple eye poses. It further explains the inverse case that searches the unknown light path corresponding to a (partially) known light source and provides a distance measure between light paths and gaze direction.

Chapter 4 applies the developed theory to introduce a novel method for the calibration of the geometric relation in display-camera setups. After explaining the algorithm, a comprehensive set of experiments is described to give insights on the accuracy of scene reconstruction from corneal reflections under multiple eye poses. A summary discusses the method and its implications.

Chapter 5 applies the developed theory to introduce a novel method for calibration-free eye gaze tracking in arbitrary environments and discusses experimental results. With this, a solution to the general problem of accurate and robust correspondence matching among multiple eye and scene images is developed and verified. A summary discusses the method and its implications.

Finally, Chapter 6 summarizes and concludes this dissertation with a detailed overview and discussion of the contributions and findings of this work.

Chapter 2

Eye Geometry

This chapter introduces a geometric eye model and explains how to obtain the pose of the model from an image of an eye.

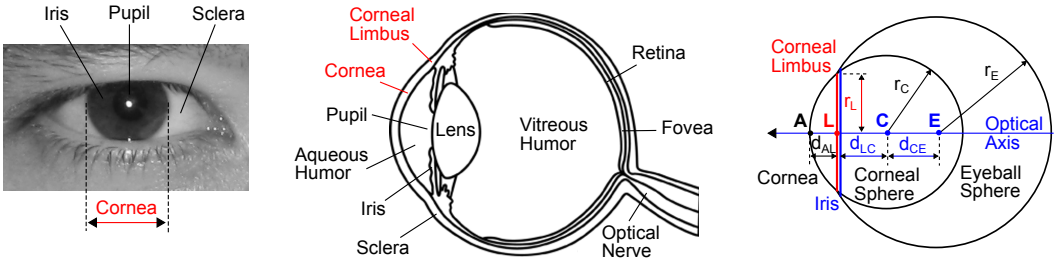
For this purpose, Section 2.1 reviews the anatomical background for structures of the eye, related to its shape and features visible on the outside. Variation in their individual manifestation is studied, surveying information in anatomical literature, and parameters of eye models describing either the imaging characteristics of the visual system or the geometric and photometric properties of the eye. A spherical-curvature model with average parameter values is developed and subsequently applied to the methods and results in the remainder of this work.

Section 2.2 then deals with estimating the pose of this model from an image, consisting of two tasks: image-based eye detection and tracking to identify and describe the eye and its features, and 3D eye pose estimation to calculate the 3D location and orientation of the model using this information. Eye detection and tracking methods are surveyed, leading over to passive methods based on circular eye features that are relevant to this work. After surveying directly related work in eye pose estimation, the perspective projection of arbitrary circles in 3D is explained. This forms the basis of two methods for circular pose estimation from elliptical feature contours, assuming perspective and weak-perspective projection models respectively.

2.1 Eye Model

2.1.1 The Human Eye

The human eye is the organ that provides the optics and photo-reception for the visual system. Received light energy is converted into nerve action potentials and sent through the optic nerve to the brain, where the information is further processed. The anatomy of the eye follows its function in this physiological process. Figure 2.1(a) shows an outer view of the human eye. The most distinctive components are the color-textured iris and the pupil in its center. The iris is surrounded by the white sclera, a dense and opaque fibrous tissue, mainly having protective function.



(a) Outer view (b) Cross-section (c) Geometric model

Figure 2.1: Geometric eye model. (a) Outer view and (b) cross-section of the human eye with important components marked red. (c) Geometric eye model with components involved in the eye pose estimation.

2.1.1.1 Eyeball

A cross-section of the eyeball in Figure 2.1(b) reveals that its main part is located behind skin and components visible from the outside. Geometrically, the eyeball is not a plain sphere; its outer layer can be subdivided into two approximately spherical segments with different radii and separated centers of curvature: the anterior corneal and the posterior scleral segment. The smaller anterior segment covers about one-sixth of the eye and contains the components in front of the vitreous humor, including the cornea, aqueous humor, iris, pupil, and lens. It has a radius of curvature r_{C} of ∼8 mm. The posterior segment covers the remaining five-sixths with a radius of curvature r_{E} of ∼12 mm. Both centers of curvature are separated by a distance $d_{\mathbf{CE}}$ of ∼5 mm. The eyeball is not symmetric; its diameters are approximately 23.5 mm horizontal (d_{H}), 23 mm vertical (d_{V}), and 24 mm anteroposterior (d_{AP}) (distance between anterior pole at the apex of the cornea and the posterior pole at the retina) (Remington, 2004). See Table 2.1 for an overview of parameter values from different sources.

Axes of the Eye. The eye has several axes. The two major ones are the optical axis and the visual axis. The optical axis (also axis of the eyeball) is usually defined as the line joining the centers of curvatures of the refractive surfaces. It is the line connecting the corneal apex **A**, the center of the limbus circle **L**, and the centers of corneal and eyeball sphere, **C** and **E**. The visual axis describes the gaze direction of the eye. It is defined as the line joining the fovea on the retina and the object being viewed, which slightly differs from the optical axis. Both axes intersect at the nodal point of the eye where the image of the object becomes reversed and inverted. The nodal point is located directly behind the back surface of the lens and remains within a distance of 1 mm from the center of corneal curvature for varying eye orientations (Young and Sheena, 1975). For a typical adult, the deviation of the visual axis is 4°–5° nasal and 1.5° superior to the optical axis with a standard deviation of

Table 2.1: Eye parameter variation [mm].

	Eyeball						Posterior segment	Anterior segment				
	d_{H}	d_{V}	d_{AP}	$d_{\mathbf{AL}}$	$d_{\mathbf{LC}}$	$d_{\mathbf{CE}}$	r_{E}	r_{C}	r_{L}	r_{LH}	r_{LV}	r_{I}
(a) Books on eye anatomy												
Snell and Lemp (1997)	23.50	23.00	24.00	—	—	—	12.00	7.70	5.575*	5.85	5.30	6.00
Crick and Khaw (2003)	—	—	—	—	—	—	—	—	5.75*	6.00	5.50	—
Kaufman and Alm (2003)	—	—	—	—	—	—	—	7.80	6.075*	6.30	5.85	—
Remington (2004)	23.50	23.00	24.00	—	—	5.70	12.00	7.80	5.75*	6.00	5.50	6.00
Khurana (2007)	—	—	—	—	—	—	12.00	7.80	5.675*	5.85	5.50	—
(b) Schematic eye models												
Gullstrand (1909) No. 1	—	—	24.385	3.60	—	—	—	7.80	—	—	—	—
Gullstrand (1909) No. 2	—	—	23.90	3.70	—	—	—	7.70	—	—	—	—
Le Grand and El Hage (1980) (1945)	—	—	24.197	3.60	—	—	12.30	7.80	—	—	—	—
Lotmar (1971)	—	—	24.197	3.60	—	—	12.30	7.80	—	—	—	—
Kooijman (1983)	—	—	24.147	3.55	—	—	—	7.80	—	—	—	—
Liou and Brennan (1997)	—	—	23.95	3.66	—	—	—	7.77	—	—	—	—
Escudero-Sanz and Navarro (1999)	—	—	23.92	3.60	—	—	12.00	7.72	—	—	—	—
(c) Work on eye modeling and applications												
Lefohn et al. (2003)	—	—	—	2.50	5.25	4.70	11.50	7.80	5.80	—	—	—
Johnson and Farid (2007)												
Morimoto and Mimica (2005)	—	—	—	3.53*	4.17	—	—	7.70	6.47*	—	—	—
Hua et al. (2006)	—	—	—	—	—	—	12.50	7.80	5.50	—	—	—
Nishino and Nayar (2006)	—	—	—	2.18	—	—	—	7.80	5.50	—	—	—
Li et al. (2007)	—	—	—	3.05	4.75*	5.70*	12.50*	7.80	6.19*	—	—	—
This work	—	—	—	2.27*	5.53*	5.70	—	7.80	5.50	—	—	6.00

*Calculated from given values.

3° (Hansen and Ji, 2010).

2.1.1.2 Cornea

The transparent cornea is the outer layer of the eye that covers the iris and dissolves into the sclera at the corneal limbus. Beside having protective function the cornea plays the main role for the eye as an optical system in focusing images on the retina. Its transparency and optical clarity stem from three factors (Kaufman and Alm, 2003; Crick and Khaw, 2003):

- the uniform size and arrangement of submicroscopic collagen fibrils in a special lamellar structure,
- the absence of blood vessels (avascularity), and
- the relative state of dehydration where the water content remains constant.

The internal pressure of the eye is higher than that of the atmosphere. This maintains the corneal shape and produces a smooth external surface. In addition, the surface is coated with a thin film of tear fluid which ensures that it remains smooth and helps to nourish the cornea. As a result, its surface shows mirror-like reflection characteristics.

Shape. Although the corneal surface approximates to a sphere, it has only spherical curvature near the apex and generally flattens towards the periphery. The cornea is subdivided into four anatomical zones with increasing radius from the optical axis (Snell and Lemp, 1997, Tab. 6-2): The central optical zone ($\leq$2.0 mm) is the most spherical and symmetric area which overlies the pupil. The paracentral/mid zone (2.0–3.5 mm) is mainly spherical but flatter. In the peripheral zone (3.5–5.5 mm) the cornea flattens the most, and finally transitions into the sclera at the limbal zone (5.5–6.0 mm).

Details of the corneal shape are examined by several studies since these are important to the fit of contact lenses and the modeling of the eye as an optical system, e.g., to predict aberrations in retinal image formation. Refer to Table 2.2 for an overview of the population distributions of corneal shape parameters. The general finding is that the surface curvature is steepest at the apex and progressively flattens towards the periphery. To model this asphericity, the corneal surface is often described by a three-dimensional conicoid expressed in the form

$$x^2 + y^2 + (1 + Q)\, z^2 - 2z\, r_{\mathrm{C}} = 0, \tag{2.1}$$

where z denotes the optical axis of the eye and r_{C} the radius of curvature at the corneal apex. The asphericity parameter Q specifies the form of the conicoid, where $Q < -1$ is a hyperboloid, $Q = -1$ is a paraboloid, $-1 < Q < 0$

Table 2.2: Population distributions for corneal vertex radii of curvature r_C and anterior surface asphericity coefficients Q (Atchison and Smith, 2000, Tabs. 2.2, 2.3).

	No. of subjects/eyes	Radius r_C [mm]	Asphericity Q
Donders (1864)			
females	38/—	7.80	—
males	79/—	7.86	—
Stenstrom (1948)	—/1000	7.86 ± 0.26	—
Sorsby et al. (1957)	—/194	7.82 ± 0.29	—
Lotmar (1971)	—	—	-0.286
Mandell and St Helen (1971)	8/8	—	-0.23 [-0.72, -0.04]
El Hage and Berny (1973)	1/1	—	$+0.16$
Lowe and Clark (1973)	46/92	7.65 ± 0.27	—
Kiely et al. (1982)	88/176	7.72 ± 0.27	-0.26 ± 0.18
Edmund and Sjøntoft (1985)	40/80	7.76 ± 0.25	-0.28 ± 0.13
Guillon et al. (1986)	110/220	7.78 ± 0.25	-0.18 ± 0.15
Koretz et al. (1989)			
females	68/—	7.69 ± 0.23	—
males	32/—	7.78 ± 0.24	—
Dunne et al. (1992)			
females	40/40	7.93 ± 0.20	—
males	40/40	8.08 ± 0.16	—
Patel et al. (1993)	20/20	7.68 ± 0.40	-0.01 ± 0.25
Lam and Douthwaite (1997)	60/60	—	-0.30 ± 0.13

Note: Values show the mean with standard deviation or range.

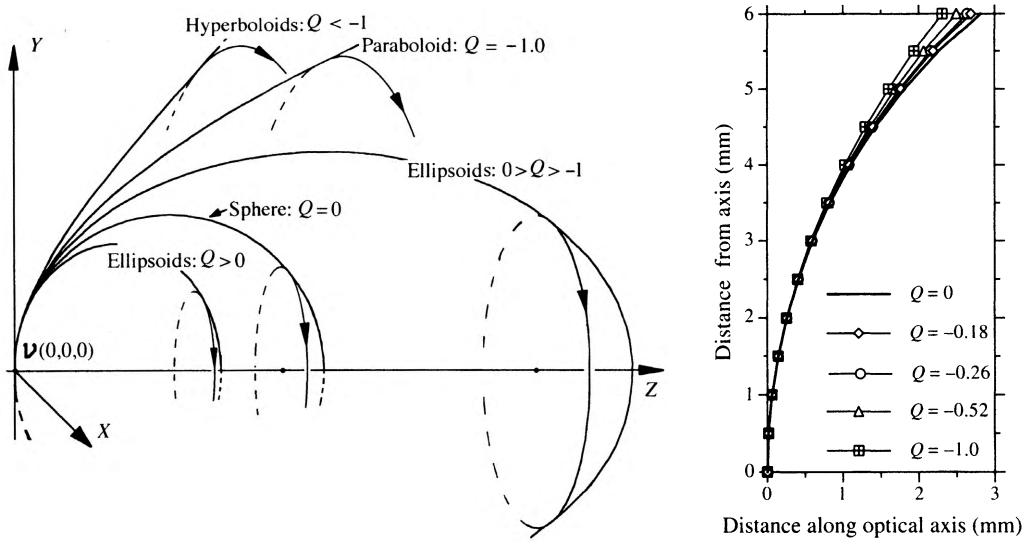


Figure 2.2: Asphericity of the cornea. (left) Effect of asphericity on the shape of a conicoid. All curves have the same apex radius of curvature. (right) Anterior surface of a cornea with a radius of curvature r_C of 7.8 mm and varying values of asphericity Q. (With permission from Atchison and Smith (2000), Atchison, D and Smith, G: Optics of the Human Eye, Butterworth-Heinemann, pp 14,15. Copyright © 2002 Elsevier Science Limited. All rights reserved.)

is a prolate ellipsoid with its major axis in the z-direction, $Q = 0$ is a sphere, and $Q > 0$ is an oblate ellipsoid with its major axis in the xy-plane (Fig. 2.2).

The eyeball is usually not rotationally symmetric around the optical axis but slightly flat in the vertical direction. This leads to a toricity in the corneal surface with the curvature being higher in the vertical direction.

Results from different studies on the shape of the eye show considerable individual variation in surface curvature, component separation, and axial length. The mean apex radius of curvature for the anterior surface of the cornea r_C is approximately 7.8 mm. The typical surface approximates to an ellipsoid with $-1 < Q < 0$. Asphericity values for individual eyes are widely distributed and can include some cases where the cornea steepens rather than flattens towards the periphery.

Recent advances in measurement techniques make it possible to acquire a detailed map of the corneal topography. Bogan et al. (1990) find that the variation of topographies in normal eyes can be classified into five qualitative patterns (Fig. 2.3). The distribution of patterns coincides with the beforementioned asphericity and toricity characteristics of the corneal contour. Another study by Liu et al. (1999) leads to broadly similar conclusions.

2.1.1.3 Limbus

The area where the transparent cornea dissolves into the opaque sclera is called the corneal or corneoscleral limbus. It is a band, approximately 1.5–2.0 mm wide, that surrounds the periphery of the cornea (Snell and Lemp, 1997; Remington, 2004). The radius of curvature immediately changes at this intersection, creating a shallow groove with a shape discontinuity on the outer surface of the eye. Refer to Table 2.1 for an overview of common values for horizontal radius r_{LH}, vertical radius r_{LV}, and average radius of the limbus r_L.

Histological, the limbus contains the transition from the regular lamellar structure of collagen fibrils of the cornea to the irregular and random organization of collagen bundles in the sclera. The layers of corneal tissue either merge into scleral tissue or terminate at different landmarks. The limbal area further contains blood vessels and lymphatic channels. This leads to a smooth and non-uniform transition (Fig. 2.4(1)).

2.1.1.4 Iris

The iris is a thin, pigmented, circular structure located directly in front of the lens. Its average radius r_I is 6 mm. The outer structures of the iris extend behind the limbus and the beginnings of the sclera. The area visible on the outside is delimited by the transparent corneal tissue that inhomogeneously dissolves at the limbus.

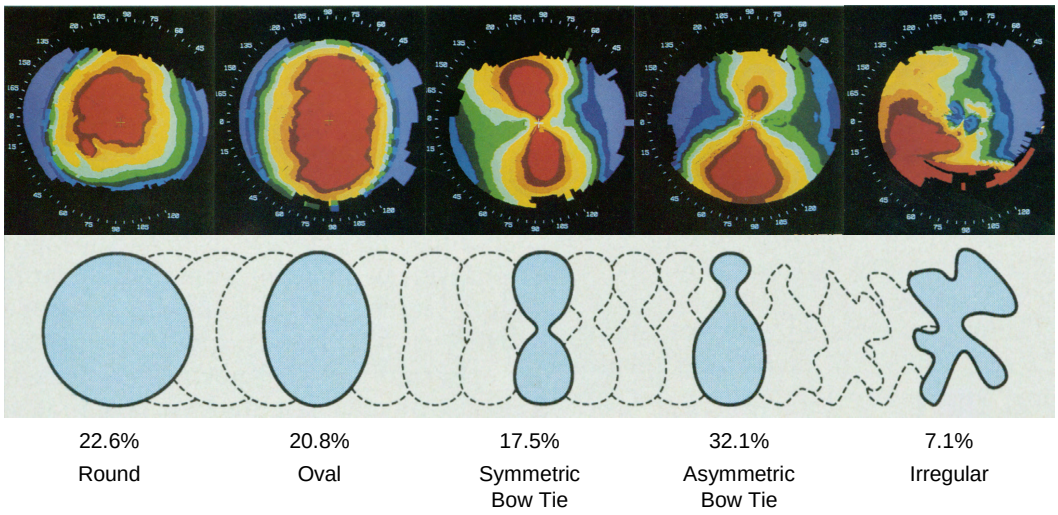
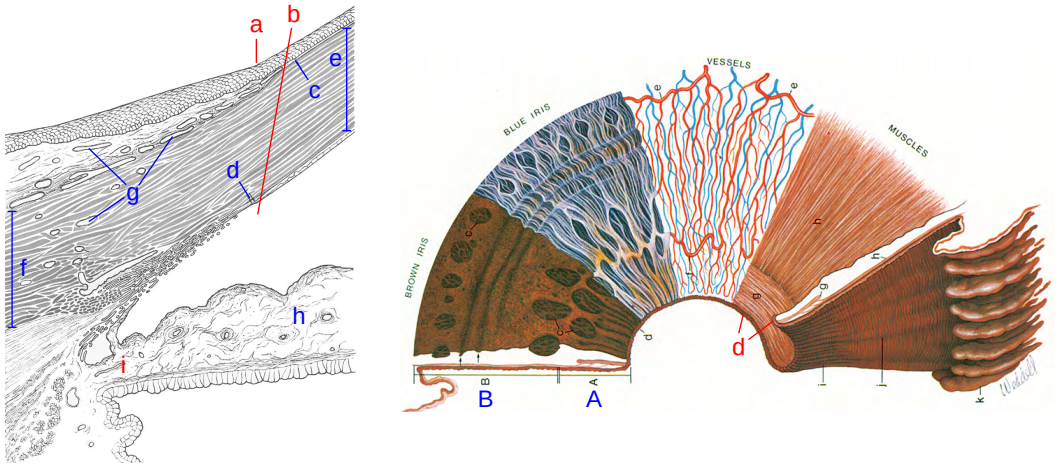


Figure 2.3: Topography of the cornea. Five quantitative patterns for corneal topography with their distribution in 216 normal eyes. (top) The degree of curvature is represented by 11 discrete color values from red (steep) to blue (flat). The range of curvature represented by each color varies among eyes since the scale is normalized to the particular degree of corneal asphericity. (bottom) The five patterns are found to probably form a continuum. (With permission from Bogan et al. (1990), Bogan, SJ et al.: Classification of Normal Corneal Topography Based on Computer-Assisted Videokeratography, Arch Ophthalmol. 1990;108(7):945–949, p 946. Copyright © 1990 American Medical Association. All rights reserved.)

Color. Iris colors for normal eyes range from light blue to dark brown, depending on the arrangement and density of connective tissue components, the density of pigment-producing cells, and the pigment-density within these cells. The color may vary between both eyes of the same person and different parts of the same iris (Snell and Lemp, 1997). The blue iris color results from light scatter and absorption of light with long wavelength in the iris tissue—the same effect that makes the sky appear blue. Darker iris colors are caused by the general amount of light absorption, which depends on the pigment density. The surface of a heavily pigmented brown iris appears smooth and velvety, whereas the surface of a lightly colored gray, blue, or green iris appears rough and uneven, with collagen fibrils in the iris tissue visible as white fibers arranged in radial columns (Fig. 2.4(2)).

Pupil. The iris forms the diaphragm of the optical system with a central circular aperture, the pupil. The size of the pupil controls retinal illumination with a diameter varying between 1 and 8 mm depending on lighting conditions. In about 25% of individuals it slightly differs in size (Snell and Lemp, 1997). The image of the pupil visible on the outside of the eye is the entrance pupil, which is magnified by the cornea and does not correspond to the location and size of the physical pupil. Compared to the smooth appearance of the



(1) Limbus (2) Surfaces and layers of the iris

Figure 2.4: Physiology of the visible iris. (1) The limbus where the transparent cornea dissolves into the opaque sclera marks the boundary of the visible iris. (a) Shape discontinuity at the outer surface of the eye. (b) Histological landmark that probably marks the onset of the visible boundary. In that area, several corneal layers (c,d) terminate, scleral layers and blood vessels (g) start and corneal tissue (e) begins to merge with scleral tissue (f). (h) The iris extends below the limbus and the beginnings of the sclera into the iris root (i). (2) The cross-section shows the pupillary (A) and ciliary portions (B) for the different surfaces and layers of the iris. (d) Compared to the smooth and irregular boundary at the limbus, a forward extension of the inner layer, the pupillary ruff, and the sphincter muscle account for a clear circular boundary at the pupillary margin. (With permission from Remington (2004), Remington, LA: Clinical Anatomy of the Visual System, second edition, Butterworth-Heinemann, pp 27,41. Copyright © 2005 Elsevier. All rights reserved.)

iris boundary seen through the corneal limbus, the circular pupillary margin is a rather sharp edge. It is formed by the pupillary ruff, a dark-pigmented forward extension of the posterior tissue (Fig. 2.4(2)). The pupil appears black because most of the entering light is absorbed by the tissues of the inner eye. The pupil can appear red in an image when the eye is photographed in low-intensity ambient light under bright flash illumination. This so-called red-eye effect is caused by the large amount of light, reflected from the back of the eyeball in the direction of the camera when the flash is located near to the lens.

2.1.2 Geometric Eye Model

2.1.2.1 Schematic Eye Models

With the knowledge of shape and parameter distribution for the human eye, it becomes possible to construct eye models. Several so-called schematic eye models with different level of sophistication have been developed over the last

150 years; motivated by the aim to describe the imaging characteristics and performance of the eye as an optical system. There are several applications in research and development, e.g., the prediction of retinal images and adjustment from correcting lenses, and surgery.

Most of the early models, such as Gullstrand's No. 1 (exact) and No. 2 (simplified) eye (Gullstrand, 1909) and Le Grand's (1945) full theoretical eye (Le Grand and El Hage, 1980) can be referred to as paraxial models. This means that they only result in adequate accuracy near the optical axis. The eye is described by several spherical refracting surfaces. Due to their simplicity, paraxial models differ significantly from the physiological structure of the eye.

Since the last 40 years, more realistic anatomically inspired finite or wide-angle schematic models are proposed in order to overcome the paraxial limitations. These models provide a more reasonable prediction of on- and off-axis aberrations, e.g., by applying non-spherical refractive surfaces, and a lack of surface alignment along the optical axis. Examples include the finite model eyes of Lotmar (1971), Kooijman (1983), and Liou and Brennan (1997). Refer to Table 2.1 for an overview of parameter values.

Recently, Goncharov et al. (2008) and Sakamoto et al. (2008) proposed a method to create an individually parameterized eye model by optimizing a generic eye model through reverse ray-tracing using wavefront sensor data.

2.1.2.2 Applied Eye Model

Sophisticated schematic eye models are the outcome of a long research process to understand and model the complex individually varying optics of the human eye. For applications related to this dissertation, such as finding the position and orientation of an eye and recovering environmental structure from eye reflections, it is not necessary to deal with refractive surfaces of the inner eye. Often, it is not feasible to determine parameters of the individual eye geometry. Related works, therefore, usually apply a simple paraxial model with spherical or ellipsoid curvature for the outer surfaces of the eye.

Within this work, we follow this approach and assume that the eyeball can be approximated by two overlapping spheres with different radii and separated centers of curvature $\mathbf{C}$ and $\mathbf{E}$ as shown in Figure 2.1(c). The cornea is modeled as a spherical cap that is cut off from the corneal sphere by the limbus plane[1]. For the corneal sphere, we apply the common average radius of curvature r_{C} of 7.8 mm (Kaufman and Alm, 2003). The circular limbus marks the surface shape discontinuity at the intersection between corneal and eyeball sphere. For an adult, the radius of the limbus r_{L} averages approximately 5.5 mm (Nishino and Nayar, 2006). The iris has a slightly larger radius r_{I} of approximately 6 mm (Snell and Lemp, 1997). Since the visible part of the iris is bounded by

[1] For more complex aspherical representations of corneal shape as a spheroid cap, ellipsoid cap, or surface of revolution, refer to Atchison and Smith (2000); Baker (1943); Nishino and Nayar (2006); Nagamatsu et al. (2010).

the limbus circle, we assume its radius to be equal to r_{L}. The displacement $d_{\mathbf{LC}}$ between the centers of limbus circle and corneal sphere are obtained from the given known parameters as in

$$\begin{aligned} d_{\mathbf{LC}} &= \sqrt{r_{\mathrm{C}}^2 - r_{\mathrm{L}}^2} \\ &\approx 5.53 \text{ mm}. \end{aligned} \tag{2.2}$$

The height of the cornea is defined as the distance $d_{\mathbf{AL}}$ between the corneal apex **A** and the center of the circular limbus **L**. It is obtained as in

$$\begin{aligned} d_{\mathbf{AL}} &= r_{\mathrm{C}} - d_{\mathbf{LC}} \\ &\approx 2.27 \text{ mm}. \end{aligned} \tag{2.3}$$

All eye movements can be described as rotations around the geometric center of the eye **E**, located at a distance $d_{\mathbf{CE}}$ of approximately 5.70 mm posterior to the center of the corneal sphere (Remington, 2004). For this work, however, it is not necessary to model eye movements or the surface of the eyeball sphere. For reference, the applied parameter values are listed in Table 2.1. The gaze direction is assumed to be equal to the optical axis. If required, the offset to the visual axis may be obtained by additional personal calibration.

2.2 Eye Pose Estimation

The 3D pose of an eye describes the location and orientation of the eye model in the camera coordinate frame, where the origin $\mathbf{O} = (0, 0, 0)^{\mathrm{T}}$ is placed at the camera pupil.

All eye movements occur as rotations around three different axes of the eye that intersect at a fixed non-moving point, the approximate geometric center of the eye (Alpern, 1962), located at about 13.5 mm behind the apex of the cornea (Remington, 2004). The set of eye gaze directions is limited to a subset of anatomically possible positions, described by Donder's and Listing's law (Tweed and Vilis, 1990). Donder's law states that the gaze direction uniquely defines the orientation of the eye, which is independent of preceding positions. Listing's law explains that the set of valid eye positions is obtained from the primary position by a single rotation around an axis perpendicular to the gaze direction.

When changing gaze direction, usually first the head is moved to a comfortable pose from where detailed adjustment is done through eyeball rotations. Therefore, gaze direction is affected either by the described movements of the eyeball or by movements of the head. Head pose invariance in eye gaze tracking is achieved either by explicit head pose detection or by direct 3D eye modeling. This work employs the latter to estimate the 3D pose of the eyeball relative to the measurement device, a camera in this case.

2.2.1 Eye Detection and Tracking

The pose of the eye can be deduced from its shape, location, and other information in an image. In the following, we will give a survey on eye detection and tracking which involves finding the approximate eye region in an image, identifying meaningful eye features, and modeling their shape and location.

Classification of Methods. The corresponding methods for eye detection and tracking are mainly distinguished into shape-based, appearance-based, other, and hybrid methods. Shape-based methods consist of a geometric eye model and a similarity measure relating the model to the imaged eye. Typical shape features are contour features, such as pupil, iris, and eye-lid contours, or point features, such as pupil and iris centers or intersections of contour features. Appearance-based methods do not incorporate eye-specific geometric information, and detect and track eye or eye feature regions directly by their unique distribution of intensity values or filter responses in the image. This requires training of a model with large amounts of data from different subjects under various conditions. Other methods may detect unique characteristics of eyes such as symmetry, blinks, and motion. Hybrid methods combine different concepts to overcome their respective disadvantages. Eye detection and tracking methods can be also distinguished by their setup, regarding remote or rigid head-mounted camera-placement, passive or active controlled illumination, and the requirement for calibration or training data. For a recent survey refer to Hansen and Ji (2010).

Shape-based Methods. This work focuses on non-intrusive measurement with a single non-attached camera, not requiring active illumination and calibration or model training. The described characteristics are achieved by detecting and tracking a geometric eye model in the image using a shape-based method, followed by estimating the 3D pose of the eye. Shape-based methods are distinguished by which features they model and track: Simple methods only include the contours of the pupil or the iris. More complex methods also model the sclera, the eyelids or the eyebrows. This enables handling iris occlusion by eyelids, exploiting eyelid intersection at the corners of the eye, and modeling the relative alignment of features with an active shape model (Yuille et al., 1992; Xie et al., 1994; Lam and Yan, 1996; Cootes et al., 1995). Complex models also comprise further information about the eye that can be beneficial in eye pose estimation. The detection performance of shape-based methods can be improved by integration with appearance, reflection, or other characteristics in a hybrid method.

2.2.1.1 Iris and Pupil Contour Segmentation

Though there exists individual variation in eye shape and parameters, the visible iris and the pupil are often assumed to be circular. We follow this practice with the proposed geometric eye model, where cornea and eyeball have spherical curvature. Under perspective projection, a circle with arbitrary position and orientation in 3D projects to a general ellipse in an image (Semple and Kneebone, 1952; Hartley and Zisserman, 2003). Methods for pupil and iris segmentation are proposed in the context of eye tracking (Hansen and Ji, 2010), iris recognition (Bowyer et al., 2008; Shah and Ross, 2009; Matey et al., 2010), and medical image processing, and are either based on voting or model-fitting strategies.

Voting-based Methods. Voting-based methods accumulate votes for image locations where local features support a given model hypothesis. In case of an ellipse model, voting is performed in a five-dimensional parameter space from which the detected parameter values are obtained as the local maxima. A typical voting-based strategy may apply a Hough transform with an edge image (Ballard, 1981; Aguado et al., 1996; Guil and Zapata, 1997; Bennett et al., 1999). However, so far only circular features have been practically used with voting-based methods to either detect and track eyes at small gaze angles (Nixon, 1985; Young et al., 1995; Kothari and Mitchell, 1996) or segment the iris region from frontal eye images in biometrics (Wildes et al., 1996; Li et al., 2010).

Model-Fitting-based Methods. Model-fitting-based methods detect the elliptical contours directly in the image, either with a combined approach through a model-based contour detection or by first segmenting pupil or iris regions and subsequently fitting a model through the data points. While the majority of pupil and iris fitting methods supports only circular features, a few methods are proposed to deal with elliptical features, which is required for 3D eye pose estimation. Combined approaches directly fit a shape model to continuous features defined over the image, such as intensity gradients or the distance to edges: Nishino and Nayar (2006) generalize the circular integro-differential operator proposed by Daugman (1993) to elliptical features. The method detects the iris contour by non-linear maximization of an elliptical integral of intensity gradient magnitudes. Arvacheh and Tizhoosh (2006) also adapt the circular integro-differential operator and find the pupil and iris contour by fitting an active contour model. Iris occlusion is compensated by eyelid modeling. Hansen and Pece (2005) propose a likelihood model of the elliptical iris contour that incorporates also neighborhood information. The model is fit to the image through an EM and RANSAC optimization strategy that allows for multi hypothesis tracking using a particle filter and avoids explicit thresholds. Wu et al. (2007) propose a likelihood model of the iris

contour based on image intensity and gradients. It is used in conjunction with a complex 3D eye model to robustly track the non-occluded iris contour through a particle filter.

More common are two-step approaches that first detect the pupil or the iris without assuming any shape model. The result is a set of points either describing the feature area or its contour. Detection usually involves a cascade of image filters followed by an image-based search with a contour detection method. Area segmentation is commonly applied to pupil detection. The majority of methods uses active IR illumination to exploit the so-called red-eye effect (Miller et al., 1995; Nguyen et al., 2002; Agustin et al., 2006). On-axis illumination from a light source near the optical axis of the camera is reflected back through the pupil and produces a bright pupil area in the image. Off-axis illumination from a light source located farther from the optical axis does not reflect back and produces a dark pupil area. The difference between both images leads to robust and performant pupil detection and segmentation. Under active IR illumination, the contrast of the pupil contour is much higher compared to the contrast of the iris contour. Under visible light, the effect reverses (Grabowski et al., 2006). There exist only a few methods for pupil segmentation in visible light: Stiefelhagen et al. (1996, 1997) propose an iterative thresholding algorithm, first locating the face by a skin-color model, and subsequently locating the pupils as the darkest regions. Vezhnevets and Degtiareva (2003) locate the eye in an eye image using a specular highlight on the cornea. The pupil area is then segmented by detecting candidate pixels in the neighborhood. Yamazoe et al. (2008) propose a method for iris segmentation. They first localize the eye by facial-feature detection and tracking, and subsequently segment the eye into the three regions iris, sclera, and skin.

Another category of methods detects the pupil or iris contour instead of its area: Wang and Sung (2001, 2002) automatically extract the non-occluded iris contour from an eye image using a tailored vertical edge operator. (Colombo et al., 2007) explain an integrated approach for iris localization and tracking. Localization is performed using morphological operators and intensity histograms. Tracking between subsequent images detects vertical iris contours.

Several approaches detect the contours of pupil and iris along radial directions starting at a location near the pupil or iris center. To compensate for iris occlusion by eyelids and eyelashes, usually only the valid (horizontal) iris sectors are taken into account, or eyelids are explicitly modeled. Barry et al. (1997) first segment the pupil area by finding its corresponding peak in the histogram of an eye image. Both contours are then accurately detected along a set of radial strip ROIs starting from the pupil center. Morelande et al. (2002) detect the iris contour to describe the corneal limbus in videokeratoscopy. Obtaining the location of the corneal apex from the apparatus, the eye image is transformed into polar coordinates. The iris contour is then found along the radial direction with a gradient-based strategy. Iskander et al. (2004) improve and extend this method to find the pupil and iris contours in ordinary eye

images. Iskander (2006) argue that common non-parametric edge detection techniques have relatively low precision as the numerical gradient is sensitive to noise. They propose a parametric approach where a sigmoidal function is fit to radial intensity profiles of the iris contour.

Li et al. (2005) describe an iterative radial search strategy to detect the contour of the pupil in an eye image. Their *Starburst* algorithm starts at an initial guess for the pupil center that does not necessarily lie inside the pupil area, and detects edges along a number of radially extending rays. The result is iteratively refined until convergence. For subsequent frames, the location of the pupil center from the previous frame is used as initial guess. The method takes advantage of the high-contrast pupil contour under IR illumination but can also be applied to visible light images. Ryan et al. (2008) improve the *Starburst* algorithm to detect pupil and iris contours, evaluate multiple contour hypotheses, and compensate for eyelid and eyelash occlusion. Reale et al. (2010) track a close-up image of an eye using a PTZ camera and detect the iris contour with an approach similar to the *Starburst* algorithm. The initial guess for the iris center is obtained using a modification of the pupil segmentation algorithm proposed by Vezhnevets and Degtiareva (2003).

Since the pupil and iris contours are not exactly circles or ellipses and, thus, their image is not exactly an ellipse, active contour models (Kass et al., 1988) have been used to accurately detect the pupil and iris contours by fitting a deformable model. He et al. (2009) propose an accurate and fast iris contour detection method. After removing specular highlights from the eye image, the iris region is located with an Adaboost-cascade iris detector based on Haar-like features that is trained in advance. The pupil and iris contour are found by edge detection in radial direction and fit with a spline-based active contour model. The eyelids are explicitly modeled. Shah and Ross (2009) segment the pupil area in an eye image and fit its contour with a circular model. The iris contour is then searched and iteratively refined using geodesic active contours that combine the energy minimization approach of classical snakes with geometric active contours based on curve evolution (Caselles et al., 1997).

Ellipse Fitting. After the pupil or iris is identified in the image, an ellipse model needs to be fit through its contour. Most of the described methods either directly fit an ellipse model or describe a tailored fitting strategy. Generally, ellipse fitting through scattered data is achieved using a linear least squares method, such as the simple method described in Appendix A.4 or more sophisticated methods like the ones described by Halir and Flusser (1998) and Fitzgibbon et al. (1999). Robustness can be increased with outlier removal techniques such as RANSAC.

2.2.1.2 Eye Pose Estimation from Elliptical Feature Contour

Eye pose estimation requires known camera parameters, obtained for static cameras using standard methods (Hartley and Zisserman, 2003; Zhang, 2000; Bouguet, 2010), and for dynamic cameras by interpolation of calibration data or estimation from circular scene features, such as the shape of the visible iris (Wu et al., 2005b). If a dedicated calibration procedure cannot be performed, intrinsic parameters may be estimated directly from the shape of the visible iris (Johnson and Farid, 2007). Starting from the identified location of the ellipse in the image, and knowing camera intrinsics and radius of the circular feature, the 3D pose of the feature can be recovered. An inherent two-way ambiguity, however, leads to two possible sets of solutions, and needs to be resolved using additional geometry constraints. The estimated center point and surface normal of the base plane of either visible iris or pupil completely describe the optical axis of the eye and, therefore, the position and orientation of the geometric eye model in 3D.

2.2.1.3 Comparison between Iris- and Pupil-based Strategies

While the pose of the eye can be estimated from either pupil or iris contour, using the pupil involves several drawbacks compared to using the iris:

Radius The radius of the pupil varies in response to the light intensity at the retina due to the pupillary light reflex (Beatty and Lucero-Wagoner, 2000), and has to be determined separately. Regarding pose estimation of the pupil circle in 3D, an unknown radius defines two groups with an infinite number of parallel planes intersecting the back-projection cone at increasing distance from the camera. The unknown distance needs to be resolved using a corneal reflection from at least a single point light source at known position (Villanueva and Cabeza, 2007) or other knowledge.

Refraction The pupil as seen from the outside of the eye does not correspond to the real pupil or physical aperture of the eye. Light rays from the real pupil undergo refraction twice, between aqueous humor and cornea, and between cornea and air. Thus, the image of the pupil is a virtual image corresponding to the entrance pupil which is forward to and slightly larger than the real pupil (Atchison and Smith, 2000). Refraction causes bending of a light ray, and depends on the refractive indices of the involved media, the location of the surface point at which light enters the other medium, and the angle between incident light ray and surface normal. As this affects each ray individually, it leads to a non-linear displacement and distortion in the shape of the imaged pupil. Refraction is an important factor in eye pose estimation from pupil images and may account for errors $>1^\circ$ in gaze angle (Villanueva

and Cabeza, 2007). The 3D location of the real pupil contour and its center point can be calculated from the image of the virtual pupil under known shape, refractive indices, and location of the eye. The shape is defined by the geometric eye model using either population means or estimating individual parameter values (Shih et al., 2000; Ohno et al., 2002; Guestrin and Eizenman, 2006); the refractive indices of aqueous humor, cornea, and air are assumed as 1.336, 1.376, and 1 (Atchison and Smith, 2000) respectively; and the location of the corneal sphere is obtained from corneal reflections of at least two known point light sources (Shih et al., 2000; Guestrin and Eizenman, 2006; Villanueva and Cabeza, 2007). However, the applied values are only approximations, and additional hardware is required which involves further geometric calibration.

Shape and location The pupil is generally not centered at the optical axis of the eye, with its shape showing a relatively large non-circularity increasing with age (Wyatt, 1995; Rakshit and Monro, 2007; Atchison and Smith, 2000). On average, about half of the non-circularity of the pupil contour is described by its best fit ellipse. An accurate representation is obtained using a circular Fourier series where most of the contribution to shape is made by the first four or five harmonics. The shape of the pupil and the location relative to limbus and optical axis vary with incident illumination due to the pupillary light reflex, and largely vary among individuals.

Note, that there are also several drawbacks related to eye pose estimation using the iris contour, e.g.,

- occlusion by eyelids and eyelashes, especially at large gaze angles,
- refraction at large gaze angles,
- a low-contrast, varying, and non-smooth transition between corneal and scleral regions due to iris texture and blood vessels, and
- unknown individual radii of visible iris and limbus.

The majority of these effects, however, can be accounted for by using sophisticated algorithms, without the need of additional hardware and calibration. Therefore, this work focuses on iris-contour/limbus-based methods for eye pose estimation.

2.2.2 Limbus-based Eye Pose Estimation

We assume the visible iris contour to represent the corneal limbus. This is feasible since the iris is larger than the limbus and extends behind the beginnings of the sclera. The shape of the visible iris is, thus, defined by the

shape of the transparent cornea. Moreover, as the iris is located directly behind the limbus, their relative distance is negligible compared to the distance between eye and camera. The circular limbus in 3D is described by its center point $\mathbf{L} = (L_x, L_y, L_z)^{\mathrm{T}}$; the normal vector of its base plane $\mathbf{g} = (g_x, g_y, g_z)^{\mathrm{T}}$ corresponding to optical axis and gaze direction; and the radius r_{L}, a constant specified by the geometric eye model. Eye pose estimation aims in recovering these values from the visible iris contour represented by an ellipse in the image.

Before explaining the perspective projection of the limbus and introducing two methods for pose estimation under a full-perspective and a weak-perspective camera projection model, in the following, we want to survey directly related works for monocular eye pose estimation from the iris contour in visible light, without active illumination and additional hardware.

2.2.2.1 Related Work

Monocular Circle Pose Estimation. There exists a large body of works on closed-form solutions to the monocular reconstruction of circles under projective transformation, usually with application to external camera calibration, where it is necessary to identify the location and orientation of a camera relative to a calibration rig from various arrangements of planar circles. Dhome et al. (1990) recover the pose of an object of revolution from an image of a circular cross-section for which the supporting plane, the center, and the radius are known in an object coordinate system. Shiu and Ahmad (1989) present a solution for the pose estimation of 3D circular features on the basis of analytic geometry that also handles the position of spherical features. Safaee-Rad et al. (1991, 1992) present an integrated approach to the fitting of quadratic curves in an image that result from the projection of circular arcs. Based on analytic geometry, a solution is then given for the estimation of the 3D pose of the corresponding circular features, considering the two cases of known and unknown radius. Kanatani and Liu (1993) develop a new formulation of properties of conics with special emphasis on computational aspects in projective geometry. They describe the two cases where the conic results from either a circle or an ellipse with known shape. Chen and Huang (1999) propose a method for circle pose estimation based on two particular projected chords of a circle image. Chen et al. (2004) describe a two-circle algorithm that jointly estimates the pose of two coplanar circles and the unknown focal length of the camera. The approach is extended by Wu et al. (2005b) to allow for two parallel but non-coplanar circles, and applied to the estimation of the base planes of the two irides in a face image. Gurdjos et al. (2006) generalize the pose estimation of $N \geq 2$ parallel circles and describe the problem in terms of a system of linear equations, increasing applicability and numerical stability. Zheng et al. (2008) present a novel projective equation of a circle that naturally encodes the pose parameters. The theory is used for circle pose

estimation and to give an explanation on the reasonability of the two sets of solutions and their mutual relationship. The framework is applied to camera calibration by Zheng and Liu (2008) who estimate the pose of two coplanar circles and the focal length of the camera.

The works for eye pose estimation described in the following depend on the surveyed closed-form solutions for monocular reconstruction of circles under projective transformation and differ in the way they resolve the inherent ambiguity in the two sets of solutions.

Two Eyes. When both eyes are present in a face image, constraints on their relative pose have been applied to resolve the ambiguity. For example, if a person focuses on an object moving towards infinity, the gaze directions of both eyes become parallel. Based on this constraint, Wang and Sung (2001) assume the person is focusing on a sufficiently far-away object. They explain a two-circle algorithm where the correct solution for each eye is obtained as the one leading to a minimal angle between the surface normals of both eyes. According to their work, this is a valid strategy since the baseline between both eyes is generally much smaller than the distance to the focused object, and the angular deviation between the two correct normals is much smaller than the deviation among other combinations of normal directions.

Chen et al. (2004) also explain a two-circle algorithm applied to camera calibration. The algorithm uses two coplanar circles to simultaneously estimate the external parameters and the focal length of the camera. The coplanarity constraint determines a unique solution for focal length and plane normal. If the circle radii are unknown, their center points are reconstructed up to a scale ambiguity that is resolved if one of the radii is known. Simultaneously estimating the focal length enables the use of zoom lenses, for example, to track a close-up region of the eye (Oike et al., 2004). Wu et al. (2005a) apply this work to estimate the gaze direction of two eyes assuming the irides to be coplanar circles. Wu et al. (2005b) propose an extended two-circle algorithm allowing the two irides to be located on different but parallel planes.

Schnieders et al. (2010) extend the idea of display-camera calibration in Chapter 4 (Nitschke et al., 2009) for eye gaze tracking. The pose of a 3D display plane is estimated from the edges of screen content reflecting at the cornae of both eyes captured in a single face image. Their method for eye pose estimation is based on Chen et al. (2004), where the ambiguity is resolved by assuming the PoR to be located on the display. Under that assumption, the correct pair of eye poses is obtained as the one leading to minimal distance between the intersection points of the gaze rays with the display plane.

Single Eye. Wang and Sung (2002) and Wang et al. (2003, 2005b) state that the requirement of both eyes being visible in the same image in Wang and Sung (2001) leads to a relative low resolution for each particular eye region and an

increased error in eye pose estimation. To compensate for this, they introduce the so-called one-circle algorithm based on anthropometric knowledge. The eye model is extended by the upper and lower eyelid, intersecting in the two corners of the eye, and by modeling eye orientation as a rotation around the center of the eyeball. This enables formulation of a geometric constraint stating that the distances between each corner and the center of the eyeball are equal and independent of eye gaze. Under this constraint, the ambiguity for the back-projected limbus circle is resolved by selecting the solution that minimizes the deviation.

Johnson and Farid (2007) explain a method to identify digital forgeries, in the form of composite images, from inconsistencies in corneal reflections among persons originally photographed under different lighting conditions. Their work estimates the pose of a single eye under the assumption that each person looks into the camera. They formulate a non-linear optimization to simultaneously recover the limbus circle and its projective transform, based on the visible iris in the image. If the photograph is taken with an unknown camera, the focal length can be estimated using another non-linear optimization strategy.

Nishino and Nayar (2004a,b, 2006) and Nishino et al. (2005) do not propose an automatic way to resolve the ambiguity for the solution of the back-projected limbus circle. However, their work is important since it is the first comprehensive analysis of the visual information that is embedded within corneal reflections in an image of the human eye. There are two main differences to the other works in this section. First, the pose of the eye is estimated by a simple and straightforward method under the assumption of weak perspective projection, directly relating the parameters of the imaged ellipse to the pose of the 3D limbus circle. The method described in Section 2.2.2.3 is based on this approach. Second, the shape of the corneal surface is modeled as a spheroid (instead of a sphere) which better approximates the mean shape of the cornea discovered in anatomical studies.

Video-based Tracking. Wu et al. (2007) introduce an integrated 3D model of two eyes, consisting of two single-sphere eyeballs with circular irides and B-spline eyelid curves. After manually initializing the model to a face image of a frontal-looking person, the model is tracked in subsequent frames with a particle filter using a likelihood model for irides and eyelids. Yamazoe et al. (2008) follow a completely different strategy to estimate and track the gaze direction in a continuous sequence of video images based on structure-from-motion and bundle-adjustment. A set of keypoints is identified and tracked among subsequent face images. Assuming a rigid shape for the face, its 3D pose is reconstructed at every video frame. The imaged eye region is detected and segmented based on color and brightness assumptions. Knowing the approximate 3D position of the eye, the pose of a single-sphere eyeball model is optimized by minimizing the re-projection error to the imaged eye.

Reale et al. (2010) adopt a similar approach but add a five-point personal calibration to determine the individual size and visual axis offset for a single-sphere eyeball model. They obtain close-up images of a face with a PTZ camera system and track facial features to reconstruct the 3D pose of the head for estimating the positions of the two eyeball spheres. For each eye, the gaze direction is estimated by mapping the imaged iris contour onto the eyeball and accounting for the offset to the visual axis. Chen and Ji (2008) also rely on facial feature tracking to estimate the 3D pose of a rigid face model. Assuming known face direction and a rigid relation between eye corners and center, the position of an individually calibrated double-sphere eye model is estimated. The gaze direction is subsequently obtained from the imaged pupil center and the known distance between eyeball center and pupil.

2.2.2.2 Method for Perspective Projection

Let us now derive the perspective projection of a circle and apply it to recover the original 3D pose of the circular limbus from the elliptical contour of the visible iris in the image.

Perspective Projection of the Limbus. Without loss of generality let us assume the limbus to be centered at the origin $(0,0)^{\mathrm{T}}$ of its base plane. Any 2D point on its boundary, given in homogeneous coordinates as $\mathbf{p} = (p_x, p_y, 1)^{\mathrm{T}}$, satisfies the implicit equation

$$\mathbf{p}^{\mathrm{T}}\mathsf{Q}\mathbf{p} = 0, \tag{2.4}$$

where

$$\mathsf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -r_{\mathrm{L}}^2 \end{bmatrix} \tag{2.5}$$

is a symmetric matrix describing the circle.

The limbus is located with an arbitrary position and orientation in 3D that we want to recover from an image. When taking a photograph of an eye, the circular limbus is mapped by a projective transformation to an ellipse in the image (Semple and Kneebone, 1952; Hartley and Zisserman, 2003). This relation is expressed by

$$\mathbf{p}_{\mathrm{e}} = \mathsf{H}_{\mathrm{e}}\mathbf{p}, \tag{2.6}$$

where $\mathbf{p}_{\mathrm{e}}$ denotes a pixel on the boundary of the limbus ellipse in the image plane. The 3×3 matrix

$$\mathsf{H}_{\mathrm{e}} = \mathsf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}, \tag{2.7}$$

is a planar homography or collineation that describes the projective transformation. It is composed of a rotation, a translation $\mathbf{t} = \mathbf{L}$, and a projection with camera matrix K. The rotation is completely described by the first two columns $\mathbf{r}_1$ and $\mathbf{r}_2$ of a 3×3 rotation matrix R. This is easily verified since H_e does not map between arbitrary points in 3D but rather points located on 2D planes (planar subspaces). Without loss of generality assume the plane to be aligned with the xy-plane at $z = 0$ which eliminates the effect of $\mathbf{r}_3$ (Bradski and Kaehler, 2008).

Under projective transformation H_e, the circular limbus Q maps to the ellipse Q'_e in the image, given by

$$\begin{aligned} \mathsf{Q}'_\mathrm{e} &= \mathsf{H}_\mathrm{e}^{-\mathrm{T}} \mathsf{Q} \mathsf{H}_\mathrm{e}^{-1} \\ &= \left(\mathsf{K}^{-\mathrm{T}} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}^{-\mathrm{T}} \right) \mathsf{Q} \left(\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}^{-1} \mathsf{K}^{-1} \right). \end{aligned} \tag{2.8}$$

This is derived as follows

$$\begin{aligned} \mathbf{p}^\mathrm{T} \mathsf{Q} \mathbf{p} = 0 \quad \big| \mathbf{p} &= \mathsf{H}_\mathrm{e}^{-1} \mathbf{p}_\mathrm{e} \,, \\ \left(\mathsf{H}_\mathrm{e}^{-1} \mathbf{p}_\mathrm{e} \right)^\mathrm{T} \mathsf{Q} \left(\mathsf{H}_\mathrm{e}^{-1} \mathbf{p}_\mathrm{e} \right) &= 0, \\ \mathbf{p}_\mathrm{e}^\mathrm{T} \underbrace{\left(\mathsf{H}_\mathrm{e}^{-\mathrm{T}} \mathsf{Q} \mathsf{H}_\mathrm{e}^{-1} \right)}_{\mathsf{Q}'_\mathrm{e}} \mathbf{p}_\mathrm{e} &= 0. \end{aligned} \tag{2.9}$$

Such as limbus circle Q, also its image Q'_e is described by a symmetric matrix as in

$$\mathsf{Q}'_\mathrm{e} = \begin{bmatrix} A & B & D \\ B & C & E \\ D & E & F \end{bmatrix}, \tag{2.10}$$

where a pixel $\mathbf{p}_\mathrm{e}$ on its boundary satisfies

$$\mathbf{p}_\mathrm{e}^\mathrm{T} \mathsf{Q}'_\mathrm{e} \mathbf{p}_\mathrm{e} = 0. \tag{2.11}$$

Expanding equation (2.11) describes the ellipse in the normal form of a quadratic equation in two variables, given by

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0, \tag{2.12}$$

where A, B, C are not all zero and $B^2 - AC < 0$. If also $A = C \neq 0$ and $B = 0$ then Q'_e represents a circle, which is the case when the base plane of the limbus and the image plane are parallel. Refer to Appendix A for details on the representation of an ellipse.

Pose Estimation of the Limbus. Now, consider a supporting plane coordinate system with the same origin as the camera coordinate system, but the image plane aligned parallel to the limbus base plane with the z-axis being

perpendicular to both planes and representing the gaze direction (Fig. 2.5). Since both coordinate systems share the same origin, the transformation is described by a pure rotation R. A projective transformation H_c maps the circular limbus Q to the general circle Q'_c in the image[2], given by

$$\begin{aligned}\mathsf{Q}'_\mathrm{c} &= \mathsf{H}_\mathrm{c}^{-\mathrm{T}}\mathsf{Q}\mathsf{H}_\mathrm{c}^{-1} \\ &= \left(\mathsf{K}^{-\mathrm{T}}\begin{bmatrix}\mathbf{i}_1 & \mathbf{i}_2 & \mathbf{t}\end{bmatrix}^{-\mathrm{T}}\right)\mathsf{Q}\left(\begin{bmatrix}\mathbf{i}_1 & \mathbf{i}_2 & \mathbf{t}\end{bmatrix}^{-1}\mathsf{K}^{-1}\right),\end{aligned} \tag{2.13}$$

where $\mathbf{i}_1$ and $\mathbf{i}_2$ are the first two columns of the 3×3 identity matrix I_3. The derivation is carried out analogue to equation (2.9). Such as ellipse Q'_e, also circle Q'_c has the form of a symmetric matrix, given by

$$\mathsf{Q}'_\mathrm{c} = \begin{bmatrix} A & 0 & D \\ 0 & C & E \\ D & E & F \end{bmatrix}. \tag{2.14}$$

Since Q'_c represents a circle, $B = 0$ holds. A pixel $\mathbf{p}_\mathrm{c}$ on its boundary satisfies

$$\mathbf{p}_\mathrm{c}^{\mathrm{T}}\mathsf{Q}'_\mathrm{c}\mathbf{p}_\mathrm{c} = 0. \tag{2.15}$$

Removing the effect of camera matrix K from ellipse Q'_e and circle Q'_c creates their respective back-projection cones Q_e and Q_c as in

$$\begin{aligned}\mathsf{Q}_\mathrm{c} &= \mathsf{K}^{\mathrm{T}}\mathsf{Q}'_\mathrm{c}\mathsf{K}, \\ \mathsf{Q}_\mathrm{e} &= \mathsf{K}^{\mathrm{T}}\mathsf{Q}'_\mathrm{e}\mathsf{K}.\end{aligned} \tag{2.16}$$

Both describe the same cone surface and are related by the rotation R of their coordinate systems as in

$$\mathsf{Q}_\mathrm{e} = \mathsf{R}\mathsf{Q}_\mathrm{c}\mathsf{R}^{\mathrm{T}}. \tag{2.17}$$

Q_c, represented by

$$\mathsf{Q}_\mathrm{c} = \begin{bmatrix} \mathsf{I}_2 & -\mathbf{l}_\mathrm{c} \\ -\mathbf{l}_\mathrm{c} & \mathbf{l}_\mathrm{c}^{\mathrm{T}}\mathbf{l}_\mathrm{c} - r_\mathrm{c}^2 \end{bmatrix}, \tag{2.18}$$

where I_2 is the 2×2 identity matrix, describes the limbus circle with center and radius defined as in

$$\mathbf{l}_\mathrm{c} = \left(\frac{x_\mathrm{c}}{z_\mathrm{c}}, \frac{y_\mathrm{c}}{z_\mathrm{c}}\right)^{\mathrm{T}}, \quad r_\mathrm{c} = \frac{r_\mathsf{L}}{z_\mathrm{c}}. \tag{2.19}$$

[2]Note that this only holds when elements k_{12}, k_{21}, k_{31}, and k_{32} of camera matrix K are all zero. That means the camera model only defines focal length and principal point, with zero skew.

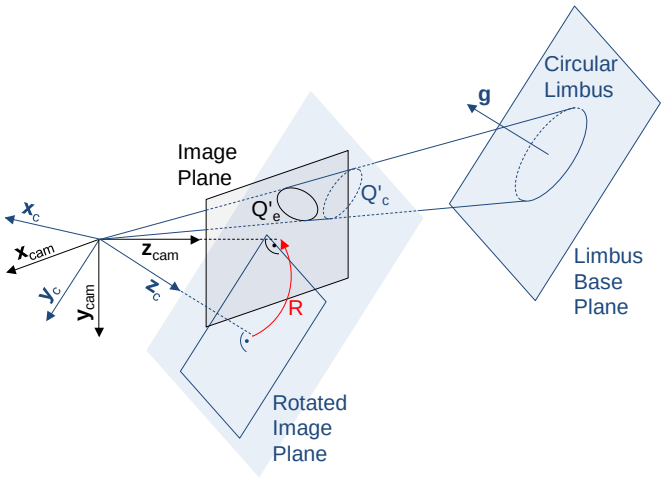


Figure 2.5: Eye pose estimation (perspective method). The 3D circular limbus with center $\mathbf{L}$ and normal direction $\mathbf{g}$ projects to a general ellipse Q'_{e} in the image. Now, consider a rotated image plane parallel to the limbus plane, where the limbus projects onto a general circle Q'_{c}. From the corresponding back-projection cone $\mathsf{Q}_{\mathrm{c}} = \mathsf{K}^{\mathrm{T}}\mathsf{Q}'_{\mathrm{c}}\mathsf{K}$ it is possible to extract the center position and normal direction of the limbus. It is obtained as in $\mathsf{Q}_{\mathrm{c}} = \mathsf{R}^{\mathrm{T}}\mathsf{Q}_{\mathrm{e}}\mathsf{R}$, where the rotation R can be recovered from Q_{e}.

At this point, the relation between the circular limbus on its base plane coordinate system Q', its projection onto the image plane as ellipse Q'_{e}, and its projection onto the image plane in a simplified supporting plane coordinate system as circle Q'_{c} are defined. We will now show how to recover rotation R that relates both projections and helps to obtain the position and orientation of the limbus in the camera coordinate system.

Let $\mathsf{Q}_{\mathrm{e}} = \mathsf{V}\mathsf{A}\mathsf{V}^{\mathrm{T}}$ describe the eigen decomposition that transforms Q_{e} into a diagonal matrix A representing its eigenvalues and an orthogonal matrix V with columns representing its eigenvectors, as in

$$\mathsf{A} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}, \quad \mathsf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}. \tag{2.20}$$

The geometric interpretation of this decomposition is that matrix V^{T} defines a rotation that transforms the general ellipse Q_{e} into an ellipse A, axis-aligned and centered at the origin of the image plane. Now, Q_{c} is given by

$$\begin{aligned} \mathsf{Q}_{\mathrm{c}} &= \mathsf{R}^{\mathrm{T}}\mathsf{Q}_{\mathrm{e}}\mathsf{R} \\ &= \left(\mathsf{R}^{\mathrm{T}}\mathsf{V}\right)\mathsf{A}\left(\mathsf{V}^{\mathrm{T}}\mathsf{R}\right). \end{aligned} \tag{2.21}$$

Since Q_{e} describes a real ellipse, all eigenvalues cannot have equal sign. Without loss of generality, it can therefore be assumed that

$$ab > 0, \quad ac < 0, \quad |a| \geq |b|. \tag{2.22}$$

Note that if necessary, the eigenvectors can be re-organized to meet these assumptions. Chen et al. (2004) derive a solution for $\mathsf{V}^{\mathrm{T}}\mathsf{R}$ from equation (2.21), given by

$$\mathsf{V}^{\mathrm{T}}\mathsf{R} = \begin{bmatrix} g\cos\alpha & \mathrm{S}_1 g\sin\alpha & \mathrm{S}_2 h \\ \sin\alpha & -\mathrm{S}_1\cos\alpha & 0 \\ \mathrm{S}_1\mathrm{S}_2 h\cos\alpha & \mathrm{S}_2 h\sin\alpha & -\mathrm{S}_1 g \end{bmatrix}, \tag{2.23}$$

where

$$g = \sqrt{\frac{b-c}{a-c}}, \quad h = \sqrt{\frac{a-b}{a-c}}. \tag{2.24}$$

Angle α represents the rotation of the limbus circle around the normal on its base plane. It can only be recovered when introducing further knowledge and, therefore, remains a free variable. S_1 and S_2 are two undetermined signs.

Now, circle Q_{c} can be computed by replacing $\mathsf{V}^{\mathrm{T}}\mathsf{R}$ in equation (2.21). Its center $\mathbf{l}_{\mathrm{c}}$ and radius r_{c} are obtained according to equation (2.19) as in

$$\mathbf{l}_{\mathrm{c}} = \begin{bmatrix} -\mathrm{S}_2\dfrac{\sqrt{(a-b)(b-c)}\cos\alpha}{b} \\ -\mathrm{S}_1\mathrm{S}_2\dfrac{\sqrt{(a-b)(b-c)}\sin\alpha}{b} \end{bmatrix}, \quad r_{\mathrm{c}} = \mathrm{S}_3\frac{\sqrt{-ac}}{b}, \tag{2.25}$$

where S_3 is another undetermined sign. The center of the circular limbus in camera coordinates is computed by applying rotation $\mathsf{R} = \mathsf{V}\,(\mathsf{V}^{\mathrm{T}}\mathsf{R})$, leading to

$$\mathbf{L} = z_{\mathrm{c}}\mathsf{R}\mathbf{l}_{\mathrm{c}} = z_{\mathrm{c}}\mathsf{V}\begin{bmatrix} \mathrm{S}_2 h\dfrac{c}{b} \\ 0 \\ -\mathrm{S}_1 g\dfrac{a}{b} \end{bmatrix}, \tag{2.26}$$

where the scale z_{c} is also given from equation (2.19) as in

$$z_{\mathrm{c}} = \frac{r_{\mathsf{L}}}{r_{\mathrm{c}}}. \tag{2.27}$$

Finally, the normal of the limbus base plane, describing the gaze direction, is obtained in the same way, as in

$$\mathbf{g} = \mathsf{R}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathsf{V}\begin{bmatrix} \mathrm{S}_2 h \\ 0 \\ -\mathrm{S}_1 g \end{bmatrix}. \tag{2.28}$$

Since there are three undetermined signs, the number of potential solutions for limbus position and orientation is $2^3 = 8$. The signs S_1 and S_3 can be resolved by implying the two constraints that $\mathbf{L}$ is located in front of the

camera with its normal $\mathbf{g}$ facing the camera. Further knowledge is necessary to break the remaining two-way ambiguity. Therefore, other works either apply anthropometric properties of the eyeball or constraints from the relation of the gaze directions of both eyes in the same face image. In Chapter 4 we introduce an approach to jointly resolve the ambiguities for multiple eyes using geometry constraints from a static scene that is reconstructed using corneal reflections.

2.2.2.3 Method for Weak Perspective Projection

The method described in the last section recovers the location and orientation of an eye from its image by computing the center $\mathbf{L}$ and the normal vector $\mathbf{g}$ of the circular limbus in 3D. The computation is accurate but complex since it involves a matrix diagonalization. A simpler method that acts under the assumption of weak perspective projection follows.

Perspective projection is a non-linear transformation, modeled as in

$$\begin{gathered}\mathsf{K}_{\mathrm{p}}\mathbf{P} = P_z\mathbf{p}_{\mathrm{p}}, \\ \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = P_z \begin{bmatrix} fP_x/P_z \\ fP_y/P_z \\ 1 \end{bmatrix} = P_z \begin{bmatrix} p_{\mathrm{p}u} \\ p_{\mathrm{p}v} \\ 1 \end{bmatrix},\end{gathered} \tag{2.29}$$

where $\mathbf{P} = (P_x, P_y, P_z, 1)^{\mathrm{T}}$ is a 3D point in homogeneous representation, $\mathbf{p}_{\mathrm{p}} = (p_{\mathrm{p}u}, p_{\mathrm{p}v}, 1)^{\mathrm{T}}$ its 2D projection in the image, and f the focal length of the camera. If an object has a relatively large depth, the effect of perspective distortion (foreshortening) becomes noticeable. That means the scaling varies within the projection of the same object. Weak perspective projection is an approximation that removes the perspective distortion, and is modeled as in

$$\begin{gathered}\mathsf{K}_{\mathrm{wp}}\mathbf{P} = \bar{z}\mathbf{p}_{\mathrm{wp}}, \\ \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & \bar{z} \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \bar{z} \begin{bmatrix} fP_x/\bar{z} \\ fP_y/\bar{z} \\ 1 \end{bmatrix} = \bar{z} \begin{bmatrix} p_{\mathrm{wp}u} \\ p_{\mathrm{wp}v} \\ 1 \end{bmatrix},\end{gathered} \tag{2.30}$$

representing a linear transformation where an orthographic projection onto a plane, located at the average depth of an object $\bar{z}$, is followed by a perspective projection from that plane (Fig. 2.6). The resulting projection in the image is denoted $\mathbf{p}_{\mathrm{wp}} = (p_{\mathrm{wp}u}, p_{\mathrm{wp}v}, 1)^{\mathrm{T}}$.

Let us assume weak perspective projection since the depth of the tilted limbus is much smaller than the distance between eye and camera. As in the full perspective case, the circular limbus projects to an ellipse that can be described by five parameters. These parameters are the center $\mathbf{l} = (l_u, l_v)^{\mathrm{T}}$,

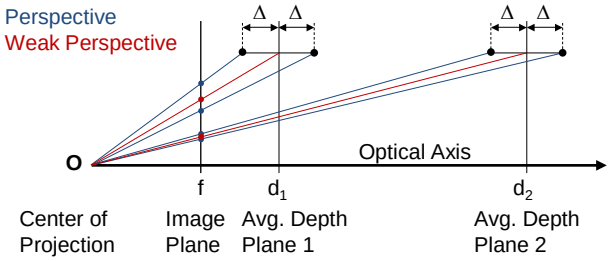


Figure 2.6: Comparison of perspective and weak perspective projection models. Weak perspective projection consists of an orthographic projection onto an average depth plane, followed by a perspective projection from the plane. A comparison of results from different depth planes demonstrates that the projection error $f\Delta/d$ decreases with increasing distance d from the center of projection.

the major and minor radii $r_{\max}$ and $r_{\min}$ and the rotation angle ϕ. This representation of an ellipse is obtained by applying the conversion in Appendix A to the representation as a quadratic equation used in the last section.

The 3D position of the limbus center $\mathbf{L}$ is estimated from the center of the ellipse $\mathbf{l}$ and the distance to the camera d as in

$$\mathbf{L} = d\left(\frac{l_u - c_{0u}}{f}, \frac{l_v - c_{0v}}{f}, 1\right)^{\mathrm{T}}, \quad d = f\frac{r_{\mathsf{L}}}{r_{\max}}, \tag{2.31}$$

where f is the focal length in pixels and $\mathbf{c}_0 = (c_{0u}, c_{0v})^{\mathrm{T}}$ the principal point. Figure 2.7 shows estimation results with increasing distance from the display-camera system.

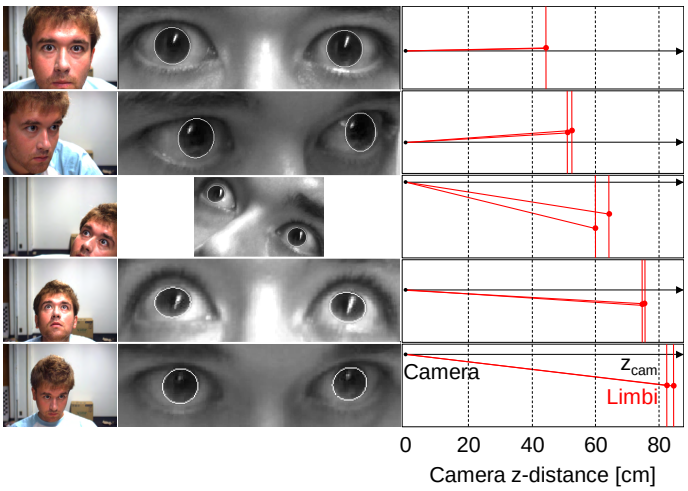
The orientation of the eye is described by its optical axis that is parallel to the surface normal of the limbus base plane. The normal direction is equal to gaze direction $\mathbf{g}$ and obtained as in

$$\mathbf{g} = \begin{bmatrix} \sin\tau\sin\phi \\ -\sin\tau\cos\phi \\ -\cos\tau \end{bmatrix}, \tag{2.32}$$

where $\phi \in [0, \pi]$ is already known as the rotation angle of the limbus ellipse in the image. Angle $\tau \in [0, \pi/2]$ corresponds to the tilt of the limbus base plane with respect to the image plane (Fig. 2.8). It is estimated from the shape of the ellipse up to a sign ambiguity as in

$$\tau = \pm\cos^{-1}\left(\frac{r_{\min}}{r_{\max}}\right). \tag{2.33}$$

The ambiguity is explained in the perspective method (eqs. (2.26),(2.28)) and can only be resolved by introducing further knowledge.



(a) Image (b) Detected irides (c) Estimated limbus centers

Figure 2.7: Experimental results for eye position estimation. For an increasing camera-eye distance, the figure shows (a) face image, (b) detected iris ellipses, and (c) estimated 3D limbus center positions.

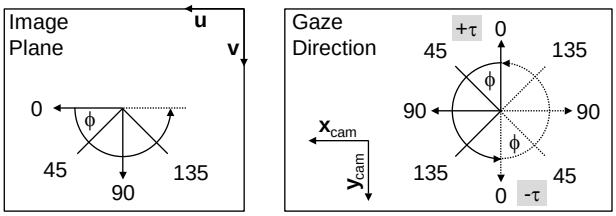


Figure 2.8: Detailed relation between rotation angle ϕ [deg] of the detected limbus ellipse and the resulting gaze direction in 3D. The gaze direction is shown in the xy-plane looking down along the negative z-axis. Since ellipse orientation ϕ is only defined in $[0, \pi)$ (instead of $[0, 2\pi)$) there exists a two-way ambiguity represented by an unknown sign for the tilt angle τ of the limbus plane.

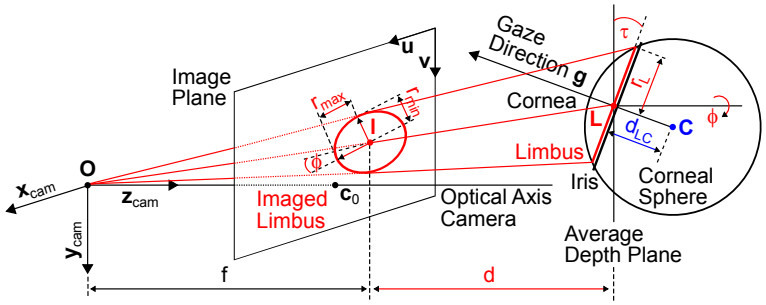


Figure 2.9: Eye pose estimation (weak-perspective method). The 3D position and orientation of the eye model is obtained from the imaged limbus that is described by an ellipse marking the contour of the visible iris.

2.2.3 Corneal Sphere Position

The cornea is modeled with spheric surface curvature. The corneal sphere is described by its radius r_{C} and center $\mathbf{C}$, located at distance $d_{\mathbf{LC}}$ from $\mathbf{L}$ (Fig. 2.9), and obtained as in

$$\mathbf{C} = \mathbf{L} - d_{\mathbf{LC}}\,\mathbf{g}. \tag{2.34}$$

A point $\mathbf{S} = (S_x, S_y, S_z)^{\mathrm{T}}$ on the corneal surface is modeled as in

$$\mathbf{S}\,(\phi, \theta) = \mathbf{C} + r_{\mathrm{C}} \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}, \tag{2.35}$$

with the two angles $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi]$. Knowing gaze direction $\mathbf{g}$, parallel to the optical axis, we are also able to construct the eyeball sphere with radius r_{E} around center $\mathbf{E}$ given by

$$\mathbf{E} = \mathbf{C} - d_{\mathbf{CE}}\,\mathbf{g}. \tag{2.36}$$

However, this is not necessary for this work.

Chapter 3

Light Transport at the Corneal Surface

This chapter builds on eye modeling and pose estimation to develop a theory of the light transport at the corneal surface.

Section 3.1 introduces a corneal reflection model for inverse light path construction to obtain the direction towards a light source from an imaged reflection.

Combining direction information obtained under multiple eye poses, Section 3.2 describes the triangulation of light paths to estimate the position of the corresponding light source.

Now consider the opposite case where the position of the light source is known and the location of the corresponding corneal reflection in the image is required. Section 3.3 studies this problem to estimate the position where light from the PoR reflects at the corneal surface into the camera. After formulating the problem, five methods are developed regarding available knowledge about the distance between eye and PoR along the gaze direction.

Computing the imaged reflection is necessary to look up information about the light source in an image. As features are usually sparse, information may not exist at the calculated location, and require approximation by inter- or extrapolation. The corresponding weights depend on the distances between the inverse reflection rays and the PoR. Section 3.4 derives a distance metric and develops three methods regarding available knowledge about the distance between eye and PoR along the gaze direction.

3.1 Corneal Reflection Model

Light that reaches the eye undergoes reflection and refraction at several transparent components—namely cornea, aqueous humor, lens, and vitreous humor—until finally reaching the fovea. The reflection at a particular transition is called the n-th Purkinje image (Duchowski, 2007). For the scope of this work, we only need to deal with the most prominent reflection at the outer surface of the cornea (the first Purkinje image) since later reflections along the light path cannot be detected without special hardware (Morimoto and Mimica, 2005). For what follows, we assume the pose of the eye model from Section 2.1 has been estimated using one of the techniques described in Section 2.2. Thus, the cornea is modeled as a sphere with known center $\mathbf{C}$ and radius r_{C}. We

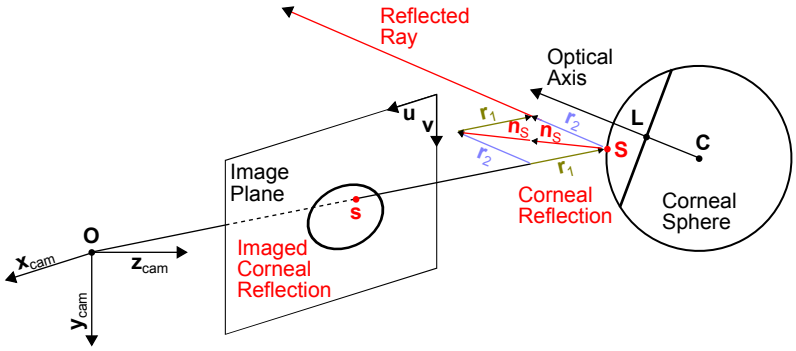


Figure 3.1: Inverse light path towards point light source. The back-projected light ray from the camera image intersects the corneal surface and reflects into the direction of the light source located at unknown distance.

will now develop a corneal reflection model used to calculate the inverse light path towards a point light source at an unknown distance (Fig. 3.1).

A point light source is located at an unknown position $\mathbf{P}$. Assuming the surface of the cornea to be a perfect mirror, light from $\mathbf{P}$ specularly reflects at surface point $\mathbf{S}$ into the direction of the camera. Taking an image of the eye captures the specular reflection as a bright patch (glint) located within the bounds of the visible iris. Let $\mathbf{s} = (s_u, s_v, 1)^{\mathrm{T}}$ denote the subpixel location of the patch centroid in the image. The corresponding location $\mathbf{s}'$ in the normalized image plane is obtained by removing the effect of camera matrix K as in

$$\mathbf{s}' = \mathsf{K}^{-1}\mathbf{s}. \tag{3.1}$$

Computing the normalized direction vector $\mathbf{r}_1 = \mathbf{s}'/\|\mathbf{s}'\|$, the point of reflection $\mathbf{S}$ on the corneal surface can be formulated as in

$$\mathbf{S} = t_1\mathbf{r}_1, \tag{3.2}$$

where t_1 is the unknown distance from the camera. To recover $\mathbf{S}$ we calculate the intersection with the corneal sphere by solving the quadratic equation

$$\|\mathbf{S} - \mathbf{C}\|^2 = r_{\mathrm{C}}^2. \tag{3.3}$$

Expanding and re-arranging leads to

$$t_1^2\mathbf{r}_1^2 - 2t_1\left(\mathbf{r}_1^2\mathbf{C}\right) + \mathbf{C}^2 - r_{\mathrm{C}}^2, \tag{3.4}$$

from which we construct the simplified quadratic formula

$$t_1 = \left(\mathbf{r}_1^2\mathbf{C}\right) \pm \sqrt{\left(\mathbf{r}_1^2\mathbf{C}\right)^2 - \mathbf{C}^2 + r_{\mathrm{C}}^2}. \tag{3.5}$$

The first intersection at the front side of the cornea is described by the smaller value of t_1. Knowing $\mathbf{S}$ and the corresponding surface normal $\mathbf{n_S} = \|\mathbf{S} - \mathbf{C}\|$,

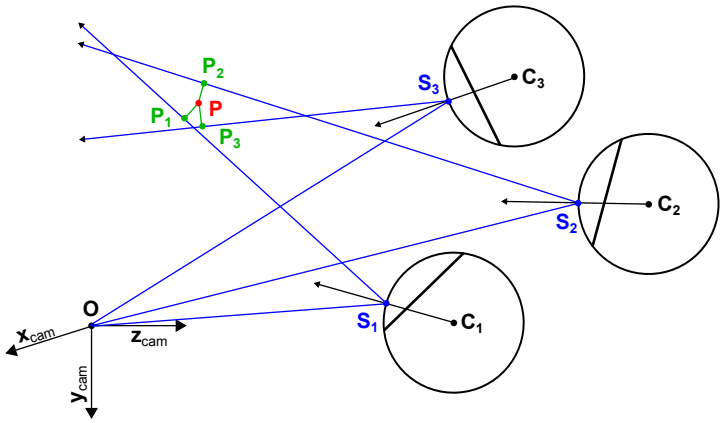


Figure 3.2: Estimation of point light source position $\mathbf{P}$ as the intersection of the set of inverse reflection rays. Since the rays generally do not intersect in a single point, we find the least-squares approximation as the point $\mathbf{P}$ with minimal distance to the set of rays.

the normalized direction vector $\mathbf{r}_2$ of the reflection ray is obtained by calculating the specular reflection as in

$$\mathbf{r}_2 = 2\left(-\mathbf{r}_1 \cdot \mathbf{n_S}\right)\mathbf{n_S} + \mathbf{r}_1. \tag{3.6}$$

The position of light source $\mathbf{P}$ then lies on the reflection ray extending from $\mathbf{S}$, defined as $\mathbf{P} = \mathbf{S} + t_2\mathbf{r}_2$, at unknown distance t_2.

3.2 Light Source Position Estimation

While the point light source remains static at position $\mathbf{P}$ we capture a set of images under varying eye poses. Applying the knowledge introduced so far we recover the corresponding eye poses and inverse reflection rays from $N \geq 2$ eye images. The unknown position of $\mathbf{P}$ is then obtained as the intersection of the inverse reflection rays. However, as a result of measurement errors and system model simplifications, the rays are generally skew and do not intersect. Thus, the task is to estimate the point $\mathbf{P}$ with minimal distance to the set of rays (Fig. 3.2).

Geometric approach for $N = 2$. There exists a simple geometric approach for the triangulation of two rays in 3D. The idea is to compute $\mathbf{P}$ as the midpoint of the shortest line connecting the two rays

$$\begin{aligned} \mathbf{P}_1 &= \mathbf{S}_1 + t_{21}\mathbf{r}_{21}, \\ \mathbf{P}_2 &= \mathbf{S}_2 + t_{22}\mathbf{r}_{22}. \end{aligned} \tag{3.7}$$

From the orthogonality constraint for the shortest connecting line we obtain the two equations

$$\begin{aligned}(\mathbf{P}_1 - \mathbf{P}_2) \cdot \mathbf{r}_{21} &= 0,\\ (\mathbf{P}_1 - \mathbf{P}_2) \cdot \mathbf{r}_{22} &= 0,\end{aligned} \tag{3.8}$$

that need to be solved for t_{21} and t_{22}. Inserting the ray equations (3.7) into the constraints (3.8) and expanding the dot product leads to

$$\begin{aligned}(\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{r}_{21} + t_{21}\,(\mathbf{r}_{21} \cdot \mathbf{r}_{21}) - t_{22}\,(\mathbf{r}_{22} \cdot \mathbf{r}_{21}) &= 0,\\ (\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{r}_{22} + t_{21}\,(\mathbf{r}_{21} \cdot \mathbf{r}_{22}) - t_{22}\,(\mathbf{r}_{22} \cdot \mathbf{r}_{22}) &= 0.\end{aligned} \tag{3.9}$$

Solving for t_{21}, back-substituting, and then solving for t_{22} gives

$$\begin{aligned}t_{21} &= \frac{((\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{r}_{22})\,(\mathbf{r}_{22} \cdot \mathbf{r}_{21}) - ((\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{r}_{21})\,(\mathbf{r}_{22} \cdot \mathbf{r}_{22})}{(\mathbf{r}_{21} \cdot \mathbf{r}_{21})\,(\mathbf{r}_{22} \cdot \mathbf{r}_{22}) - (\mathbf{r}_{22} \cdot \mathbf{r}_{21})\,(\mathbf{r}_{22} \cdot \mathbf{r}_{21})},\\ t_{22} &= \frac{((\mathbf{S}_1 - \mathbf{S}_2) \cdot \mathbf{r}_{22}) + t_{21}\,(\mathbf{r}_{22} \cdot \mathbf{r}_{21})}{(\mathbf{r}_{22} \cdot \mathbf{r}_{22})}.\end{aligned} \tag{3.10}$$

Finally, the searched point with minimal distance to both rays is obtained as

$$\mathbf{P} = \mathbf{P}_1 + \frac{\mathbf{P}_2 - \mathbf{P}_1}{2}. \tag{3.11}$$

Note that when the denominator t_{21} becomes zero both rays are parallel and do not intersect. Practically, this case does not occur since different eye poses result in different reflection directions. Nevertheless, it is beneficial to increase the baseline between the corneal spheres as this increases the denominator and, thus, numerical stability.

Algebraic approach for $N \geq 2$. In the general case, $\mathbf{P}$ can be obtained using matrix algebra as follows. At frame l, the distance between $\mathbf{P}$ and the nearest point on the ray $\mathbf{P}_l = \mathbf{S}_l + t_{2l}\mathbf{r}_{2l}$ is defined as

$$\|\mathbf{P}_l - \mathbf{P}\| = \frac{\|\mathbf{r}_{2l} \times (\mathbf{S}_l - \mathbf{P})\|}{\|\mathbf{r}_{2l}\|}. \tag{3.12}$$

Knowing $\|\mathbf{r}_{2l}\| = 1$ and re-arranging leads to

$$\|\mathbf{P}_l - \mathbf{P}\| = \left\|[\mathbf{r}_{2l}]_\times \mathbf{P} - \mathbf{r}_{2l} \times \mathbf{S}_l\right\|, \tag{3.13}$$

where $[\mathbf{r}_{2l}]_\times$ represents vector $\mathbf{r}_{2l}$ as a skew-symmetric matrix, given by

$$[\mathbf{r}_{2l}]_\times = \begin{bmatrix} 0 & -z_{\mathbf{r}_{2l}} & y_{\mathbf{r}_{2l}} \\ z_{\mathbf{r}_{2l}} & 0 & -x_{\mathbf{r}_{2l}} \\ -y_{\mathbf{r}_{2l}} & x_{\mathbf{r}_{2l}} & 0 \end{bmatrix}, \tag{3.14}$$

and expresses the cross product as a matrix multiplication. To solve for $\mathbf{P}$ we combine the N equations and formulate the problem as a least-squares minimization in the form $\|\mathsf{A}\mathbf{P} - \mathbf{b}\|$. Finally, point $\mathbf{P}$ is estimated through the pseudo-inverse as

$$\mathbf{P} = \left(\mathsf{A}^{\mathrm{T}}\mathsf{A}\right)^{-1}\mathsf{A}^{\mathrm{T}}\mathbf{b}, \tag{3.15}$$

$$\mathsf{A}_{3N\times 3} = \begin{bmatrix} [\mathbf{r}_{21}]_{\times} \\ \vdots \\ [\mathbf{r}_{2N}]_{\times} \end{bmatrix},$$

$$\mathbf{b}_{3N\times 1} = \begin{bmatrix} \mathbf{r}_{21} \times \mathbf{S}_1 \\ \vdots \\ \mathbf{r}_{2N} \times \mathbf{S}_N \end{bmatrix}.$$

3.3 Surface Reflection Position Estimation

The last section explains how to recover the position $\mathbf{P}$ of a point light source from inverse reflection rays obtained under multiple cornea positions. In this section we will examine the inverse case of this problem where the 3D position of a point $\mathbf{P}$ is known, and the corresponding point of reflection $\mathbf{S}$ on the corneal surface is unknown. In practice, this problem occurs in Chapter 5, where a person is gazing at a surface that contains a set of projected light sources. Capturing an image of the eye, the task is to find the pixel $\mathbf{s}$ that corresponds to the corneal reflection $\mathbf{S}$ of the gazed PoR, denoted as $\mathbf{P}$. Therefore, without loss of generality, let us assume $\mathbf{P}$ lies at distance $d_{\mathbf{CP}}$ from the center of the corneal sphere $\mathbf{C}$ along gaze direction $\mathbf{g}$, described by

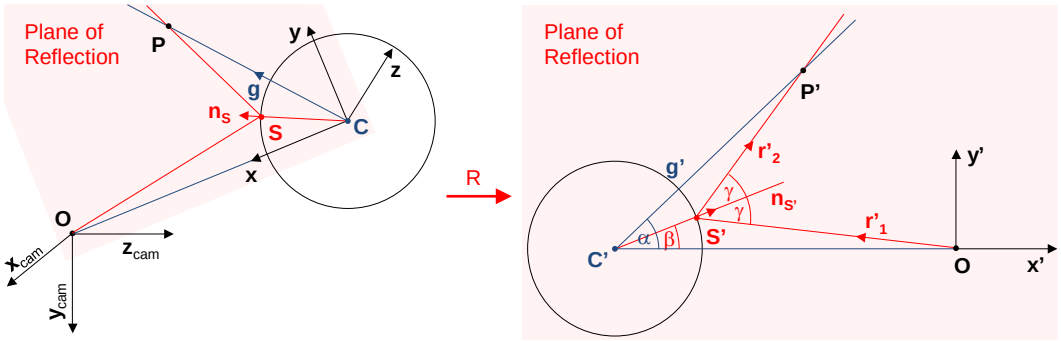
$$\mathbf{P} = \mathbf{C} + d_{\mathbf{CP}}\,\mathbf{g}, \quad d_{\mathbf{CP}} \geq r_{\mathrm{C}}. \tag{3.16}$$

The searched point of reflection on the corneal surface lies at distance r_{C} from center $\mathbf{C}$ along the normal direction $\mathbf{n}_{\mathbf{S}}$, described by

$$\mathbf{S} = \mathbf{C} + r_{\mathrm{C}}\,\mathbf{n}_{\mathbf{S}}. \tag{3.17}$$

3.3.1 Transformation of the Problem into the Plane

Directly using this parameterization leads to complex expressions that are difficult to handle. To simplify the problem we observe that the solutions for $\mathbf{S}$ are not arbitrarily distributed on the surface of the corneal sphere, but on a circular subspace, arising from the intersection with the plane containing the center of the camera $\mathbf{O}$, the center the corneal sphere $\mathbf{C}$, and the gazed point $\mathbf{P}$ (Fig. 3.3(a)). This is easily verified from the law of reflection stating that the incident ray, the reflected ray, and the surface normal at the point of reflection are coplanar. Since $\mathbf{O}$ lies on the incident ray, $\mathbf{P}$ on the reflected



(a) 3D Parameterization (b) 2D Parameterization

Figure 3.3: Transformation of the plane of reflection. Since incident ray, reflected ray, and surface normal are coplanar, a rotation R aligns the plane of reflection with the xy-plane of the camera at $z = 0$.

ray, and $\mathbf{C}$ on a ray along the inverse normal direction in $\mathbf{S}$, the points must be coplanar.

A rotation R, given by

$$\mathsf{R} = \begin{bmatrix} \mathbf{x}^{\mathrm{T}} \\ \mathbf{y}^{\mathrm{T}} \\ \mathbf{z}^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{x} = -\frac{\mathbf{C}}{\|\mathbf{C}\|}, \quad \mathbf{y} = \mathbf{z} \times \mathbf{x}, \quad \mathbf{z} = \frac{\mathbf{x} \times \mathbf{g}}{\|\mathbf{x} \times \mathbf{g}\|}, \tag{3.18}$$

aligns the plane of reflection with the xy-plane of the camera at $z = 0$ (Fig. 3.3(b)). In the following, we omit the z-coordinate for rotated quantities $\mathbf{X}' = \mathsf{R}\mathbf{X}$, obtaining

$$\begin{aligned} \mathbf{C}' &= \begin{bmatrix} -d_{\mathbf{OC}} \\ 0 \end{bmatrix}, \quad \mathbf{g}' = \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}, \quad \mathbf{n}_{\mathbf{S}'} = \begin{bmatrix} \cos\beta \\ \sin\beta \end{bmatrix}, \\ \mathbf{P}' &= \mathbf{C}' + d_{\mathbf{CP}}\,\mathbf{g}', \\ \mathbf{S}' &= \mathbf{C}' + r_{\mathrm{C}}\,\mathbf{n}_{\mathbf{S}'}. \end{aligned} \tag{3.19}$$

Angle $\alpha \in [0, \pi/2)$ is defined as in

$$\alpha = \cos^{-1} g'_x. \tag{3.20}$$

The searched angle $\beta \in [0, \alpha]$ depends on the distance to the PoR, becoming smaller with increasing distance.

3.3.2 Formulation of the Problem

After having obtained a simpler representation in the plane we will now explain how to compute the searched point of reflection $\mathbf{S}'$. At first, we choose a formulation for the problem. This should be done with much care as a wrong

choice easily leads to a complex expression which becomes difficult to handle. The task is to find the particular inverse reflection ray

$$L(\beta) = \mathbf{S}' + t_2 \mathbf{r}'_2, \tag{3.21}$$

that intersects the PoR $\mathbf{P}'$ at distance $d_{\mathbf{CP}}$ along the gaze ray. In case of an intersection the distance between $L(\beta)$ and $\mathbf{P}'$ is zero. Therefore, we search the particular β_0 that minimizes a distance function d, so that

$$\beta_0 = \arg\min_{\beta} d(\beta). \tag{3.22}$$

We define $d_{\mathbf{P}L}$ as the signed length of the line connecting $\mathbf{P}'$ and L, perpendicular to gaze direction $\mathbf{g}'$, as

$$d_{\mathbf{P}L}(d_{\mathbf{CP}}, \beta) = \frac{\mathbf{r}'_2 \times (\mathbf{P}' - \mathbf{S}')}{\mathbf{r}'_2 \cdot \mathbf{g}'}, \tag{3.23}$$

where the inverse light path is specified by the two direction vectors

$$\mathbf{r}'_1 = \frac{\mathbf{S}'}{\|\mathbf{S}'\|}, \quad \mathbf{r}'_2 = 2(-\mathbf{r}'_1 \cdot \mathbf{n}_{\mathbf{S}'})\,\mathbf{n}_{\mathbf{S}'} + \mathbf{r}'_1. \tag{3.24}$$

Note that this formulation of the distance between a point and a line, along a direction not parallel to the normal of the line (Fig. 3.4(a), blue), differs from the common distance definition (red). We choose this formulation since it expresses the deviation from the gaze ray. For practical application in eye gaze tracking we normalize the value of $d_{\mathbf{P}L}$ by the distance $d_{\mathbf{CP}}$ between circle center and PoR, and express the signed deviation in the visual angle α as in

$$d_\alpha(d_{\mathbf{CP}}, \beta) = \tan^{-1} \frac{d_{\mathbf{P}L}(d_{\mathbf{CP}}, \beta)}{d_{\mathbf{CP}}}. \tag{3.25}$$

3.3.3 Different Methods based on PoR

In practice, the detailed distance to the PoR along the gaze ray is often unknown. However, there probably exists some knowledge, such as a particular interval, or a relation to the distance between camera and eye. We will, therefore, explain the solutions for five different cases starting with the case of known a position for the PoR (Fig. 3.4(b)).

3.3.3.1 Method 1: Known $\mathbf{P}'$ with $d_{\mathbf{CP}} = d_{\mathbf{CP}_0}$

Without loss of generality, let us assume that camera center and PoR are located at the surface or outside the corneal sphere, with $d_{\mathbf{OC}} \geq r_{\mathrm{C}}$ and

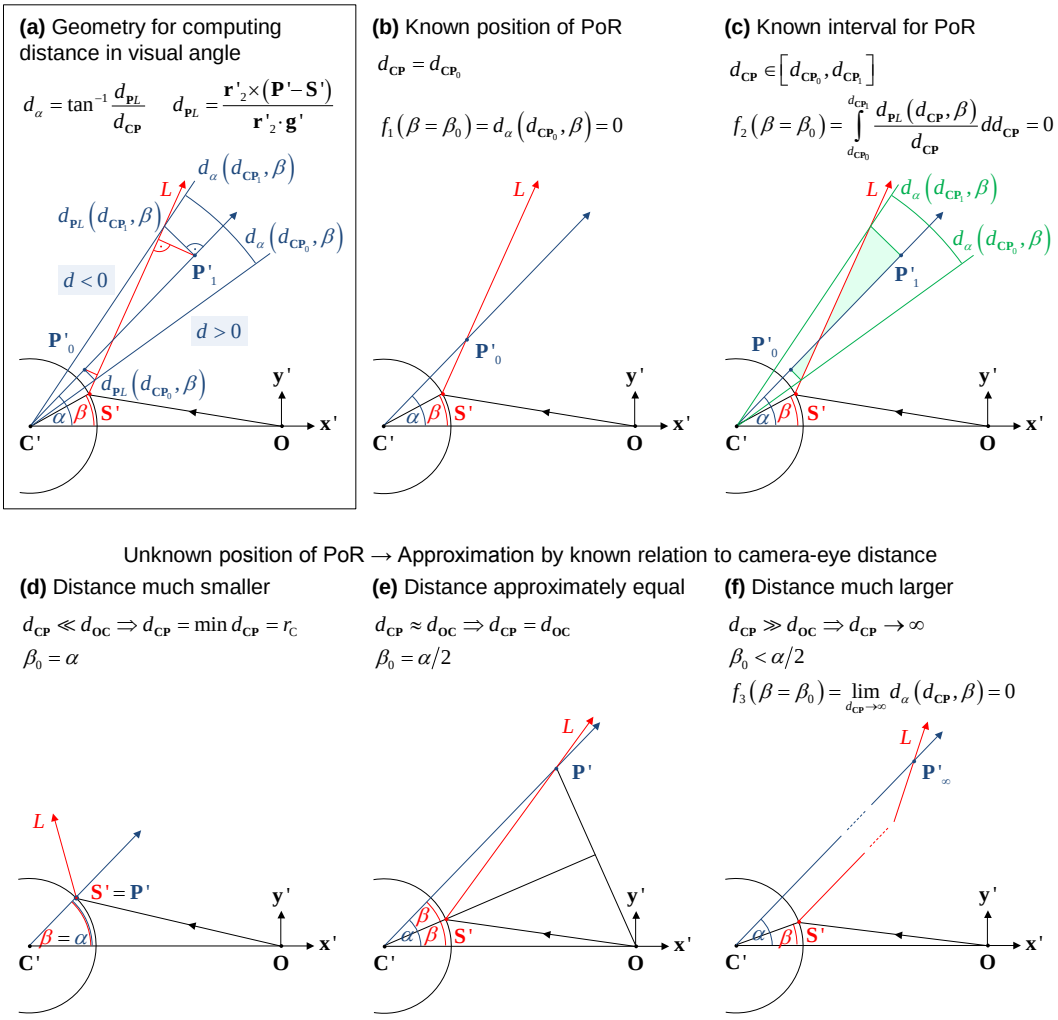


Figure 3.4: Comparison of different methods for estimating the point of reflection $\mathbf{S}'(\beta)$ on the surface of a circle in the plane of reflection, regarding the position of the gazed point of regard (PoR). (a) Geometry for computing the distance in visual angle which forms the basis for defining the objective functions f. (b) Method 1 covers the simplest case when the position of the PoR is known. (c) Method 2 covers the case when the position is unknown in a given interval, by minimizing the average distance to the gaze ray.
In case the approximate relation between PoR-eye distance and camera-eye distance is known instead of the actual PoR, we provide three different methods: (d) method 3 where the PoR-eye distance is much smaller, (e) method 4 where both distances are approximately equal, and (f) method 5 where the PoR-eye distance is much larger than the camera-eye distance.

$d_{\mathbf{CP}} \geq r_{\mathrm{C}}$. To find the reflection ray $L(\beta)$ that intersects $\mathbf{P}'_0$, with $d_{\mathbf{CP}} = d_{\mathbf{CP}_0}$, we search the β_0 where the distance function

$$\begin{aligned} f_1(\beta) &= d_\alpha(d_{\mathbf{CP}_0}, \beta) \\ &= \tan^{-1}\frac{d_{\mathbf{P}L}(d_{\mathbf{CP}_0}, \beta)}{d_{\mathbf{CP}_0}} \\ &= \tan^{-1}\left(\frac{\mathbf{r}'_2 \times (\mathbf{P}' - \mathbf{S}')}{d_{\mathbf{CP}_0}(\mathbf{r}'_2 \cdot \mathbf{g}')}\right) \end{aligned} \tag{3.26}$$

becomes zero. The solution

$$\beta_0 = \tan^{-1}\left(\frac{d_{\mathbf{CP}_0}\sin\alpha\,(d_{\mathbf{OC}} - 2R^2 d_{\mathbf{OC}} + R r_{\mathrm{C}})}{r_{\mathrm{C}}\,(d_{\mathbf{CP}_0}\cos\alpha + d_{\mathbf{OC}}) - 2R d_{\mathbf{CP}_0} d_{\mathbf{OC}}\cos\alpha}, R\right) \tag{3.27}$$

is expressed in terms of a root R of a fourth-order polynomial equation

$$P_4(x) = a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0 \tag{3.28}$$

with coefficients

$$\begin{aligned} a_4 &= 4d^2_{\mathbf{CP}_0} d^2_{\mathbf{OC}}, \\ a_3 &= -4 r_{\mathrm{C}} d_{\mathbf{CP}_0} d_{\mathbf{OC}}\,(d_{\mathbf{CP}_0} + d_{\mathbf{OC}}\cos\alpha), \\ a_2 &= -4d^2_{\mathbf{CP}_0} d^2_{\mathbf{OC}} + 2r^2_{\mathrm{C}} d_{\mathbf{CP}_0} d_{\mathbf{OC}}\cos\alpha + r^2_{\mathrm{C}}\left(d^2_{\mathbf{CP}_0} + d^2_{\mathbf{OC}}\right), \\ a_1 &= 4 r_{\mathrm{C}} d_{\mathbf{CP}_0} d_{\mathbf{OC}}\left(d_{\mathbf{OC}}\cos\alpha + d_{\mathbf{CP}_0} - 0.5 d_{\mathbf{CP}_0}\sin^2\alpha\right), \\ a_0 &= d^2_{\mathbf{CP}_0}\sin^2\alpha\left(d^2_{\mathbf{OC}} + r^2_{\mathrm{C}}\right) - r^2_{\mathrm{C}}\left(d^2_{\mathbf{CP}_0} + d^2_{\mathbf{OC}} + 2d_{\mathbf{CP}_0} d_{\mathbf{OC}}\cos\alpha\right). \end{aligned} \tag{3.29}$$

The four real roots x_{01}, x_{02}, x_{03}, and x_{04} of $P_4(x)$ are found using the algebraic method described in Appendix B, from where we obtain the corresponding solutions β_{01}, β_{02}, β_{03}, and β_{04} by back-substitution as root R into the equation of β_0. The final solution β_0 (Fig. 3.5) is selected as in

$$\beta_0 = \begin{cases} \beta_{01} & \text{if } \beta_{01} > \beta_{02}, \\ \beta_{02} & \text{otherwise.} \end{cases} \tag{3.30}$$

3.3.3.2 Method 2: Known Interval for $\mathbf{P}'$ with $d_{\mathbf{CP}} \in [d_{\mathbf{CP}_0}, d_{\mathbf{CP}_1}]$

Let us now generalize the problem to the case where we only know a lower and an upper bound for the distance of the PoR along the gaze ray, with $d_{\mathbf{CP}} \in [d_{\mathbf{CP}_0}, d_{\mathbf{CP}_1}]$ (Fig. 3.4(c)). Under this assumption, we minimize the average deviation in visual angle over the given interval by searching the β_0 where the distance function

$$f_2(\beta) = \int_{d_{\mathbf{CP}_0}}^{d_{\mathbf{CP}_1}} d_\alpha(d_{\mathbf{CP}}, \beta)\, dd_{\mathbf{CP}} \tag{3.31}$$

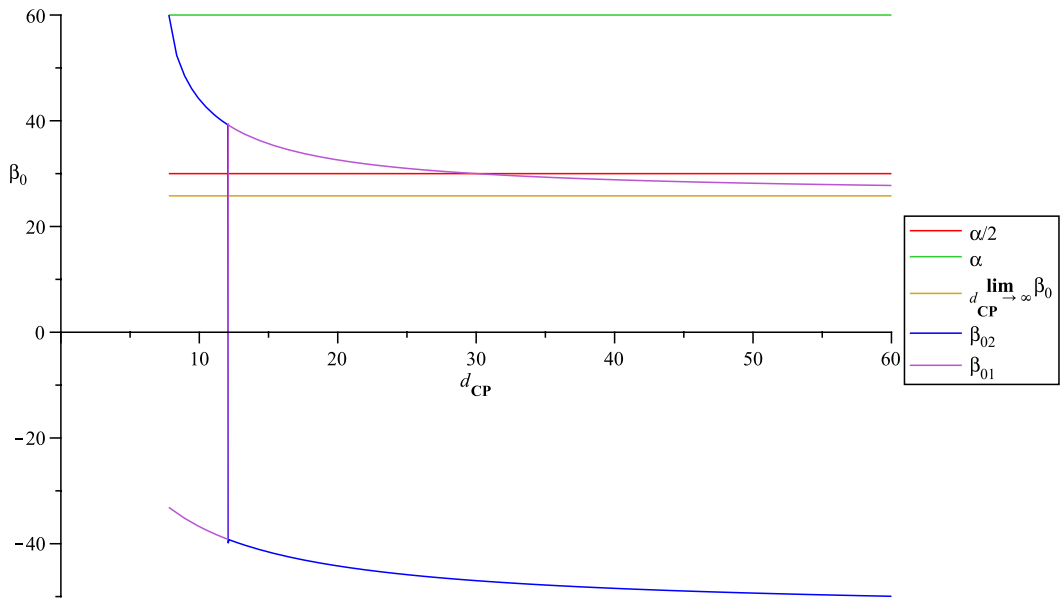


Figure 3.5: Solution function $\beta_0\,(d_{\mathbf{CP}})$ [deg] for increasing $d_{\mathbf{CP}}$ from 0 to 60 mm, where $\alpha = 60$ deg, $r_{\mathrm{C}} = 7.8$ mm, and $d_{\mathbf{OC}} = 30$ mm. (The values for $d_{\mathbf{CP}}$ and $d_{\mathbf{OC}}$ are uncommon but demonstrate the characteristic behavior of β_0.) Note that β_0 is only defined for $d_{\mathbf{CP}} \geq r_{\mathrm{C}}$. If $d_{\mathbf{OC}}$ becomes larger, the function becomes stretched in the positive horizontal direction. The two special cases occur at $d_{\mathbf{CP}} = r_{\mathrm{C}}$, where $\beta_0 = \alpha$, and at $d_{\mathbf{CP}} = d_{\mathbf{OC}}$, where $\beta_0 = \alpha/2$. The final choice between β_{01} and β_{02} depends on the relation to the inflection point of function β_0, where β_{02} is the appropriate value when $d_{\mathbf{CP}}$ is smaller, and β_{01} when it is larger than the value at the inflection point.

becomes zero. Note, that d_α represents a signed distance which is positive when the angle from $\mathbf{r}'_2$ to $(\mathbf{P}' - \mathbf{S}')$ is counterclockwise and negative otherwise. Another indicator for the sign of d_α is the distance $d_{\mathbf{CP}}$, where the sign is positive when the distance is smaller than the distance to the intersection with L, and negative otherwise. For the optimum value β_0, the intersection is located inside the interval. When $d_{\mathbf{CP}_1}$ approaches $d_{\mathbf{CP}_0}$, the value of f_2 converges to the value of f_1 with

$$\lim_{d_{\mathbf{CP}_1} \to d_{\mathbf{CP}_0}} f_2\,(\beta) = f_1\,(\beta)\,. \tag{3.32}$$

The solution involves handling the arc tangent in the definite integral over the distance function d_α. Removing the arc tangent from the expression leads to the same solution β_0, but reduces the complexity of the objective function

to

$$
\begin{aligned}
f_2(\beta) &= \int_{d_{\mathbf{CP}_0}}^{d_{\mathbf{CP}_1}} \frac{d_{\mathbf{P}L}(d_{\mathbf{CP}}, \beta)}{d_{\mathbf{CP}}}\, dd_{\mathbf{CP}} \\
&= \frac{\begin{array}{c} 2\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \sin\alpha \cos^2\beta - \left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \sin\alpha \\ -\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)\left(2 d_{\mathbf{OC}} \cos\alpha \sin\beta + r_{\mathrm{C}} \sin\alpha\right)\cos\beta \\ +r_{\mathrm{C}}\left(\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)\cos\alpha + \left(\ln d_{\mathbf{CP}_0} - \ln d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}}\right)\sin\beta \end{array}}{\begin{array}{c} r_{\mathrm{C}} \sin\alpha \sin\beta + d_{\mathbf{OC}} \cos\alpha - 2 d_{\mathbf{OC}} \cos\alpha \cos^2\beta \\ +\left(r_{\mathrm{C}} \cos\alpha - 2 d_{\mathbf{OC}} \sin\alpha \sin\beta\right)\cos\beta \end{array}}.
\end{aligned}
\tag{3.33}
$$

The solution

$$
\beta_0 = \tan^{-1}\left(\frac{\begin{array}{c}\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)\sin\alpha\left(d_{\mathbf{OC}} - 2R^2 d_{\mathbf{OC}} + R r_{\mathrm{C}}\right) \\ r_{\mathrm{C}}\left(\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)\cos\alpha + \left(\ln d_{\mathbf{CP}_0} - \ln d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}}\right) \\ -2R\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \cos\alpha \end{array}}, R\right)
\tag{3.34}
$$

for $f_2(\beta) = 0$ is obtained the same way as explained for f_1, where the coefficients of the fourth-order polynomial equation are defined as in

$$
\begin{aligned}
a_4 &= 4\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)^2 d_{\mathbf{OC}}^2, \\
a_3 &= -4 r_{\mathrm{C}}\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \\
&\quad \left(\left(\ln d_{\mathbf{CP}_0} - \ln d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \cos\alpha + d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right), \\
a_2 &= 2 r_{\mathrm{C}}^2\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)\left(\ln d_{\mathbf{CP}_0} - \ln d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \cos\alpha \\
&\quad + r_{\mathrm{C}}^2 d_{\mathbf{OC}}^2 \left(\ln d_{\mathbf{CP}_0}\right)^2 - 2 r_{\mathrm{C}}^2 d_{\mathbf{OC}}^2 \ln d_{\mathbf{CP}_0} \ln d_{\mathbf{CP}_1} + r_{\mathrm{C}}^2 d_{\mathbf{OC}}^2 \left(\ln d_{\mathbf{CP}_1}\right)^2 \\
&\quad + \left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)^2 \left(r_{\mathrm{C}}^2 - 4 d_{\mathbf{OC}}^2\right), \\
a_1 &= 4 r_{\mathrm{C}}\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \\
&\quad \left(\left(\ln d_{\mathbf{CP}_0} - \ln d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \cos\alpha - 0.5\left(\sin^2\alpha - 2\right)\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)\right), \\
a_0 &= \left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)^2 \left(d_{\mathbf{OC}}^2 + r_{\mathrm{C}}^2\right) \sin^2\alpha \\
&\quad - r_{\mathrm{C}}^2 \left(2\left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)\left(\ln d_{\mathbf{CP}_0} - \ln d_{\mathbf{CP}_1}\right) d_{\mathbf{OC}} \cos\alpha \right. \\
&\quad \left. + d_{\mathbf{OC}}^2 \left(\ln d_{\mathbf{CP}_0}\right)^2 - 2 d_{\mathbf{OC}}^2 \ln d_{\mathbf{CP}_0} \ln d_{\mathbf{CP}_1} + d_{\mathbf{OC}}^2 \left(\ln d_{\mathbf{CP}_1}\right)^2 \right. \\
&\quad \left. + \left(d_{\mathbf{CP}_0} - d_{\mathbf{CP}_1}\right)^2\right).
\end{aligned}
\tag{3.35}
$$

3.3.3.3 Method 3: Unknown P′ with $d_{\mathbf{CP}} \ll d_{\mathbf{OC}}$

We now discuss three special cases that can be used as an approximation if the position of the PoR is unknown, but the relation to the camera-eye distance is

known. If the PoR-eye distance is much smaller than the camera-eye distance, with $d_{\mathbf{CP}} \ll d_{\mathbf{OC}}$ (Fig. 3.4(d)), we use the approximation assumption

$$d_{\mathbf{CP}} = \min d_{\mathbf{CP}} = r_{\mathrm{C}}, \tag{3.36}$$

where the solution is given as in

$$\beta_0 = \alpha. \tag{3.37}$$

3.3.3.4 Method 4: Unknown $\mathbf{P}'$ with $d_{\mathbf{CP}} \approx d_{\mathbf{OC}}$

If the PoR-eye distance is approximately equal to the camera-eye distance, with $d_{\mathbf{CP}} \approx d_{\mathbf{OC}}$ (Fig. 3.4(e)), we use the approximation assumption

$$d_{\mathbf{CP}} = d_{\mathbf{OC}}, \tag{3.38}$$

where the solution is given as in

$$\beta_0 = \alpha/2. \tag{3.39}$$

3.3.3.5 Method 5: Unknown $\mathbf{P}'$ with $d_{\mathbf{CP}} \gg d_{\mathbf{OC}}$

If the PoR-eye distance is much larger then the camera-eye distance, with $d_{\mathbf{CP}} \gg d_{\mathbf{OC}}$ (Fig. 3.4(f)), we use the approximation assumption

$$d_{\mathbf{CP}} \to \infty, \tag{3.40}$$

where the reflection ray L will intersect the gaze ray in $\mathbf{P}'_{\infty}$ at infinity and both rays are parallel with

$$\mathbf{r}'_2(\beta_0) \parallel \mathbf{g}'. \tag{3.41}$$

Thus, we search for the β_0 where the distance function

$$\begin{aligned} f_3(\beta) &= \lim_{d_{\mathbf{CP}} \to \infty} d_\alpha(d_{\mathbf{CP}}, \beta) \\ &= \tan^{-1}\left(\frac{\begin{array}{c}\sin\alpha\left(-2d_{\mathbf{OC}}\cos^2\beta + r_{\mathrm{C}}\cos\beta + d_{\mathbf{OC}}\right) \\ -\cos\alpha\sin\beta\left(r_{\mathrm{C}} - 2d_{\mathbf{OC}}\cos\beta\right)\end{array}}{\begin{array}{c}\cos\alpha\left(-2d_{\mathbf{OC}}\cos^2\beta + r_{\mathrm{C}}\cos\beta + d_{\mathbf{OC}}\right) \\ +\sin\alpha\sin\beta\left(r_{\mathrm{C}} - 2d_{\mathbf{OC}}\cos\beta\right)\end{array}}\right) \end{aligned} \tag{3.42}$$

becomes zero. The solution $\beta_0 < \alpha/2$ for $f_3(\beta) = 0$ with

$$\beta_0 = \tan^{-1}\left(\frac{\tan\alpha\left(-2R^2 d_{\mathbf{OC}} + R r_{\mathrm{C}} + d_{\mathbf{OC}}\right)}{-2R d_{\mathbf{OC}} + r_{\mathrm{C}}}, R\right) \tag{3.43}$$

is obtained the same way as explained for f_1, where the coefficients of the fourth-order polynomial equation are defined as in

$$\begin{aligned} a_4 &= 4d_{\mathbf{OC}}^2 \left(\tan^2\alpha + 1\right), \\ a_3 &= -4r_{\mathrm{C}} d_{\mathbf{OC}} \left(\tan^2\alpha + 1\right), \\ a_2 &= \left(r_{\mathrm{C}}^2 - 4d_{\mathbf{OC}}^2\right) \left(\tan^2\alpha + 1\right), \\ a_1 &= 2r_{\mathrm{C}} d_{\mathbf{OC}} \left(\tan^2\alpha + 2\right), \\ a_0 &= d_{\mathbf{OC}}^2 \tan^2\alpha - r_{\mathrm{C}}^2. \end{aligned} \tag{3.44}$$

3.3.4 Back-Transformation from the Plane

Remember, that for simplification we aligned the plane of reflection with the xy-plane of the camera at $z = 0$ (Sec. 3.3.1). Now we need to transform the obtained point of reflection $\mathbf{S}'(\beta_0)$ back into 3D camera coordinates. This is done according to transformation

$$\mathbf{X} = \mathsf{R}^{\mathrm{T}}\mathbf{X}', \tag{3.45}$$

where $\mathbf{X}' = \left(X'_x, X'_y, 0\right)^{\mathrm{T}}$ represents a point in the plane, $\mathbf{X} = (X_x, X_y, X_z)^{\mathrm{T}}$ the corresponding transformed point, and R^{T} the inverse of the rotation defined in equation (3.18). The searched pixel $\mathbf{s}$ is obtained by projection into the image plane using camera matrix K according to

$$\mathbf{x} = \frac{1}{X_z}\mathsf{K}\mathbf{X}, \tag{3.46}$$

where $\mathbf{x} = (x_u, x_v, 1)^{\mathrm{T}}$ denotes the pixel in homogeneous coordinates.

3.4 Distance between Inverse Reflection Rays

The last section explained how to analytically solve for the point of reflection on the corneal surface for a given PoR in 3D. Now imagine the application of this result: Knowing the location where light from the PoR reflects at the cornea and projects into image we can obtain further information by image analysis. However, generally there exists only sparse information, defined for a particular subset of points on the corneal surface. To picture this, assume the eye image has been preprocessed by some feature-correspondence matching method, tracking environmental reflections on the cornea. The method will only identify a sparse set of features at certain discrete locations. When projecting the corneal reflection $\mathbf{S}$ of the PoR into the image, this may hit a location where no feature has been found. Moreover, this will generally be a subpixel location. Therefore, it is necessary to define an appropriate distance measure between neighboring feature points in the image, e.g., for interpolation.

The Euclidean distance between 2D points in the image plane does not provide an optimal measure since, in our case, each feature relates to a 3D point on a ray reflected at the corneal surface. The Euclidean distance between 3D points also does not provide an optimal measure since it does not account for eye position and gaze direction. Instead, we apply the distance in visual angle as in the last section. Having obtained the distance values for different neighboring locations, it is possible to compute the corresponding weights.

3.4.1 Formulation of the Problem

Assume the positions of the PoR $\mathbf{P}$, the point of reflection on the corneal surface $\mathbf{S}$, and the pixel in the image plane $\mathbf{s}$ are known. Further, assume a set of neighboring locations $\{\mathbf{s}_i \,|i = 1, \ldots, N\}$ are identified. We seek the distance in visual angle, regarding gaze ray and inverse reflection ray at $\mathbf{s}_i$, obtained using the method described in Section 3.2, and given as in

$$L_i = \mathbf{S}_i + t_{2i}\mathbf{r}_{2i}. \tag{3.47}$$

Since L_i is not necessarily coplanar with L in the plane of reflection, we cannot use the planar distance formulation introduced in the last section and, thus, apply its extension to three dimensions. Similar to the planar case, we define $d_{\mathbf{P}L}$, now as the absolute length of the line connecting $\mathbf{P}$ and L_i, perpendicular to gaze direction $\mathbf{g}$. To determine this line we need to find the intersection with reflection ray L_i, parameterized by t_{2i}. Since in 3D, there exists a plane of perpendicular directions, the line is known to lie in this plane. This gives rise to the following constraint from the plane equation

$$\begin{aligned} (L_i\,(t_{2i}) - \mathbf{P}) \cdot \mathbf{g} &= 0, \\ (\mathbf{S}_i + t_{2i}\mathbf{r}_{2i} - \mathbf{P}) \cdot \mathbf{g} &= 0, \end{aligned} \tag{3.48}$$

which holds when $\mathbf{r}_{2i}$ is not perpendicular to $\mathbf{g}$. Solving for t_{2i} yields

$$t_{2i} = \frac{(\mathbf{P} - \mathbf{S}_i) \cdot \mathbf{g}}{\mathbf{r}_{2i} \cdot \mathbf{g}}, \tag{3.49}$$

with the length $d_{\mathbf{P}L}$ of the connecting line given as in

$$\begin{aligned} d_{\mathbf{P}L}\,(d_{\mathbf{CP}}, L_i) &= \|L_i\,(t_{2i}) - \mathbf{P}\| \\ &= \left\| \mathbf{S}_i + \frac{(\mathbf{P} - \mathbf{S}_i) \cdot \mathbf{g}}{\mathbf{r}_{2i} \cdot \mathbf{g}} \mathbf{r}_{2i} - \mathbf{P} \right\|. \end{aligned} \tag{3.50}$$

The distance in visual angle is then obtained as

$$d_\alpha\,(d_{\mathbf{CP}}, L_i) = \tan^{-1} \frac{d_{\mathbf{P}L}\,(d_{\mathbf{CP}}, L_i)}{d_{\mathbf{CP}}}. \tag{3.51}$$

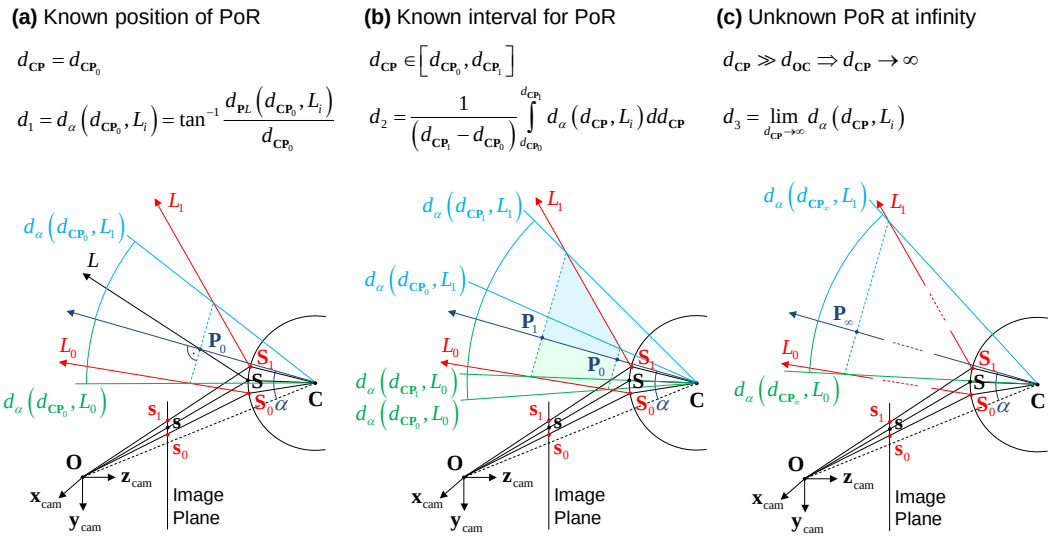


Figure 3.6: Comparison of different methods for estimating the distance between the PoR along the gaze ray and the inverse reflection ray towards a 3D feature. (a) Method 1 covers the simplest case when the position of the PoR is known. (b) Method 2 covers the case when the position is unknown in a given interval. (c) Method 3 covers the case when the position is unknown and the PoR-eye distance is much larger than the camera-eye distance.

3.4.2 Different Methods based on PoR

As done before, we assume that the position of the PoR along the gaze ray may be unknown and explain the solution for three different cases, starting with the case of known position (Fig. 3.6(a)).

3.4.2.1 Method 1: Known P with $d_{\mathbf{CP}} = d_{\mathbf{CP}_0}$

In case $d_{\mathbf{CP}} = d_{\mathbf{CP}_0}$, distance d_1 is simply defined as the distance in visual angle as in

$$d_1 = d_\alpha\left(d_{\mathbf{CP}_0}, L_i\right). \tag{3.52}$$

3.4.2.2 Method 2: Known Interval for P with $d_{\mathbf{CP}} \in [d_{\mathbf{CP}_0}, d_{\mathbf{CP}_1}]$

In case we only know the lower and the upper bound of the distance to the PoR along the gaze ray, with $d_{\mathbf{CP}} \in [d_{\mathbf{CP}_0}, d_{\mathbf{CP}_1}]$ (Fig. 3.6(b)), the average distance d_2 is obtained by computing the definite integral over the distance function in the given interval as in

$$d_2 = \frac{1}{d_{\mathbf{CP}_1} - d_{\mathbf{CP}_0}} \int_{d_{\mathbf{CP}_0}}^{d_{\mathbf{CP}_1}} d_\alpha\left(d_{\mathbf{CP}}, L_i\right)\, dd_{\mathbf{CP}}. \tag{3.53}$$

3.4.2.3 Method 3: Unknown P with $d_{\mathbf{CP}} \gg d_{\mathbf{OC}}$

If the distance to the PoR is unknown, but known to be equal or smaller than the camera-eye distance, method 1 or 2 can be applied. If the distance, however, is much larger, with $d_{\mathbf{CP}} \gg d_{\mathbf{OC}}$ (Fig. 3.6(c)), we use the approximation assumption

$$d_{\mathbf{CP}} \to \infty, \tag{3.54}$$

and obtain the distance d_3 as in

$$\begin{aligned} d_3 &= \lim_{d_{\mathbf{CP}} \to \infty} d_\alpha \left(d_{\mathbf{CP}}, L_i \right) \\ &= \tan^{-1} \left(\frac{\| \mathbf{r}_{2i} \times \mathbf{g} \|}{| \mathbf{r}_{2i} \cdot \mathbf{g} |} \right). \end{aligned} \tag{3.55}$$

Chapter 4

Display-Camera Calibration from Eye Reflections

This chapter applies the developed theory for eye pose estimation and light transport at the corneal surface to introduce a novel method for the calibration of the geometric relation in display-camera setups.

Section 4.1 provides an introduction to applications of display-camera setups and their required calibration, the principle and advantages of the proposed method, and the contributions of this work.

Section 4.2 surveys and discusses related work in geometric calibration and eye gaze tracking.

Section 4.3 then describes the proposed method to estimate the location of a display plane from marker reflections under multiple eye poses. The basic light transport theory in sections 3.1 and 3.2 is integrated with an optimization framework to improve the results of eye and display pose estimation by using known geometry constraints.

A prototype implementation is explained in Section 4.4 and subsequently applied in section. 4.5 to perform a comprehensive analysis of the applicability of eye reflection analysis for display-camera calibration. The gained understanding on the impact of factors regarding individual eye geometry, camera parameters, and geometric relation of the components also provides an answer to the general question about the accuracy that can be expected for scene reconstruction from multiple eye images.

Section 4.6 concludes this chapter, discussing results and findings, outlining potential implications on application scenarios and fields, stating limitations of the prototype implementation, and providing ideas for future work.

4.1 Introduction

Personal computers are turning more and more into multimedia processing machines that come with a large number of peripheral devices. One of these devices is a camera. Historically a tool for videoconferencing, advances in vision algorithms extend its area of application. As the camera is mounted on a PC, it can be related to its physical context. The main output device of a PC system is a (CRT or LCD) monitor. Together with a camera as input device, the monitor forms a controlled illumination system. Widely known are

projector-camera systems for scene reconstruction of lambertian surfaces (Battle et al., 1998) or novel display techniques (Bimber et al., 2008). While not every setup comes with a projector, the monitor facilitates a low-cost controlled illumination system, enabling a wide range of similar applications in non-professional everyday environments. In the past, there have been two major areas of application for display-camera systems: One is scene reconstruction to measure shape and reflectance properties from display reflections on the surface of an object. The other is human–computer interaction (HCI), where the content of the display is adapted according to information about a user, obtained from the camera.

Applications of Display-Camera Systems. In scene reconstruction, properties of static surfaces are recovered from series of images under varying screen illumination. Photometric stereo methods (Woodham, 1980) have been proposed to estimate the shape of lambertian (Clark, 2006, 2010; Funk and Yang, 2007; Schindler, 2008) or partially lambertian objects (Francken et al., 2008b). Compared to a digital video projector, monitor illumination is not directed and focused and is, therefore, ideal for coping with non-lambertian objects (Ihrke et al., 2008). Shape reconstruction methods for specular objects identify display-camera correspondences and estimate the respective surface normals (Tarini et al., 2005; Bonfort et al., 2006; Francken et al., 2008a; Nehab et al., 2008). Transparent and translucent objects are more difficult to handle, as only the minor part of the light is specularly reflected while the major part is refracted, enters the object, and might undergo subsurface scattering or reflection at the background (Kutulakos and Steger, 2008; Morris and Kutulakos, 2007). Display illumination has been further applied to analyze the complex light interaction with objects for scene compositing (Zongker et al., 1999) and relighting (Shah and Ross, 2009).

Vision-based user interfaces employ computer vision to "look at people" and perform tasks such as face recognition; head, face, eye, hand and body detection and tracking; facial expression and body movement analysis; and gesture, posture, and activity recognition (Turk, 2004; Porta, 2002; Jaimes and Sebe, 2007). Closely related with this work are eye gaze tracking techniques (Duchowski, 2007; Hansen and Ji, 2010). Often, the resulting information does not only passively affect display content but may also depend on information about the display itself. Such knowledge is usually obtained through calibration.

Calibration of Display-Camera Systems. Most of the described applications require calibration to find the relation between the display and the camera. There are mainly two forms of calibration, geometric and radiometric. Geometric calibration determines the internal camera parameters as well as the external display pose with respect to the camera (Bonfort et al., 2006;

Funk and Yang, 2007; Francken et al., 2009; Tarini et al., 2005). Radiometric calibration establishes the relation between the light emitted by the display and the light measured by the camera (Tarini et al., 2005; Funk and Yang, 2007; Francken et al., 2008b).

In this work, we focus on geometric calibration where we seek to find the pose of the display with respect to the camera for which internal parameters are known. If the screen is directly visible to the camera, such a calibration can easily be performed using standard techniques with correspondences obtained through patterns shown on the screen (e.g., checkerboard, structured light) (Hartley and Zisserman, 2003; Bouguet, 2010; Bradski and Kaehler, 2008). This is, however, not possible in the common case where the display and camera face a similar direction. In such a configuration, the calibration is achieved by analyzing reflections of a screen pattern from mirroring objects with known shape and pose (Sturm and Bonfort, 2006; Kumar et al., 2008). Methods have been proposed for planar (Bonfort et al., 2006; Funk and Yang, 2007) and spherical mirrors (Francken et al., 2009; Tarini et al., 2005). However, the process is cumbersome as it involves a special mirror as well as tedious physical user interaction.

Advantages of the Proposed Method. We describe a novel calibration technique that builds on the observation that the cornea of the human eye acts as a partial mirror and provides rich cues about environmental light. When exploring display-camera setups, we found that the reflected content is clearly visible in eye images of a person in front of the display. This provided the motivation for introducing the idea of using corneal reflections for display reconstruction. The foundation for the proposed approach lies in a combination of Nishino and Nayar's method for recovering the pose of an eye from its image (Nishino and Nayar, 2006) and Francken et al.'s method for screen-camera calibration using reflections on a freely moving spherical mirror (Francken et al., 2007). We discuss a thorough experimental evaluation of this strategy, regarding individual factors, display pose, eye position, and gaze direction, and show that it results in a large error and deviation due to the unknown geometry and size of the individual eye. To compensate for this, we introduce an optimization framework based on known geometry constraints in the setup, achieving considerable improvement that should be sufficient for many applications.

Our novel method makes display-camera calibration substantially more practical and leads to several benefits compared to previous approaches:

- Since no additional hardware is necessary the method is easily distributed and can be applied in existing off-the-shelf setups.
- Without interaction and awareness the calibration can be seamlessly performed by non-expert or disabled persons and children, or, in situations where it is not desired to disclose technical details.

- Accuracy increases with the number of images used. Nevertheless, the minimum requirement is a single face image. This enables online calibration of dynamic setups and allows applications such as camera tracking.
- The method does not only reconstruct the pose of the display, but also provides information about eye locations and gaze directions. This makes it ideal to realize human–computer interfaces based on eye gaze tracking.

Table 4.1 compares the features of the proposed method with previous approaches for display-camera calibration.

Contribution. The following contributions are achieved with this work:

- The idea is introduced, to analyze corneal reflections of computer monitor or projection screen illumination in eye images for reconstructing the position and orientation of the screen.
- To verify the proposed strategy, thorough experimental evaluation is conducted for the straightforward combination of eye pose estimation (Nishino and Nayar, 2006) and screen-camera calibration (Francken et al., 2007), which is found to result in a large error and deviation.
- To compensate for this, an optimization framework is introduced that jointly refines eye poses, reflection rays, and display pose subject to known geometry constraints in the setup.
- A large number of comprehensive experimental studies demonstrates that stable results can be obtained under varying conditions. A quantitative and qualitative comparison with spherical mirror ground truth is provided. The gained insights are not only applicable to the subject of this work, but could also be helpful when analyzing geometric reconstruction from eye reflections in general.
- A framework for physically correct rendering of synthetic eye images, with corneal reflections from environmental illumination, is designed. This provides a general tool to analyze the impact of specific system parameters on scene reconstruction from corneal reflections, especially when GT measurements are difficult to obtain as with parameters related to shape and reflection characteristics of the eye.
- The described developments and findings enable a novel method for the geometric calibration of display-camera setups that does not require special hardware, explicit user interaction or awareness, and allows online calibration.

Table 4.1: Feature matrix of geometric display-camera calibration methods.

	Checkerboard pattern	Planar mirror		Spherical mirror	This method
		Attached pattern	Plain		
	Hartley and Zisserman (2003), Bouguet (2010), Bradski and Kaehler (2008)	Funk and Yang (2007)	Bonfort et al. (2006)	Francken et al. (2009), Francken et al. (2007), Tarini et al. (2005)	
Catadioptric	−	+	+	+	+
Hardware	−	−	○	○	+
Interaction	−	−	−	○	+
Awareness	−	−	−	−	+
Online	−	−	−	−	+
Eye poses	−	−	−	−	+
Accuracy	+	+	+	+	○

4.2 Related Work

We will now review existing methods within the two fields related to this work. First, we discuss approaches for mirror-based geometric display calibration for the case when the screen is not directly visible to the camera. In the second part, we survey eye gaze tracking techniques exploiting corneal reflections either from display illumination or from other light sources in an arrangement that approximates the bounding quadrilateral of the screen.

These two fields have one property in common, the combination of a mirror and a lens that forms a catadioptric imaging system (Nishino and Nayar, 2006; Francken et al., 2009). While a camera has a single viewpoint, catadioptric systems can have a single or multiple viewpoints, depending on the shape of the mirror and its pose relative to the lens of the camera. Here, we cope with both kinds of systems, single viewpoint for planar and multiple viewpoints for convex mirror methods. Overviews of the optical properties of catadioptric systems are given by Baker and Nayar (1999); Geyer and Daniilidis (2001) and Swaminathan et al. (2006); Kuthirummal and Nayar (2006) respectively.

4.2.1 Geometric Display Calibration

Planar Mirror. Funk and Yang (2007) use a planar mirror and compute its pose from an additional pattern attached to it. The process is cumbersome as it involves a special prepared mirror, several known parameters, and physical user interaction. Bonfort et al. (2006) simplify the planar pose estimation by not requiring any marker attachments. Their first method uses a circular hard-drive platter with known interior and exterior radii. Its pose can be obtained from a single image of the circular boundaries that project to concentric ellipses. The second method (Bonfort et al., 2006; Sturm and Bonfort, 2006) uses at least three poses of an arbitrary planar mirror: At first, a virtual camera pose is calibrated with respect to the reflected screen in each mirror plane. Then, the planes themselves are recovered from the virtual camera poses.

Spherical Mirror. Tarini et al. (2005), and in more detail Francken et al. (2007), propose calibration techniques that use a spherical mirror of known size. The position of the mirror is uniquely determined from an image. Extracting the corners from the reflected screen allows to compute the corresponding light rays by inverse raytracing. The real corners are obtained by intersecting rays from different sphere positions. More recently, Francken et al. (2009) propose a refined approach using a time-series of Gray code patterns to recover a large number of correspondences from only a small number of images. They are able to increase accuracy while reducing the number of sphere positions. Applying a convex mirror also has the advantage that rays

are reflected from a wider field-of-view. This makes the corresponding methods ideal for the calibration of large-sized displays, for example, TV screens in home entertainment setups together with a game console and an attached camera. Spherical mirror techniques simplify the calibration process, however, still require special hardware and user interaction.

Our proposed method does not need any special hardware or interaction. It solely analyzes screen pattern reflections from eye images of a moving user to estimate the pose of the display. We employ a simple shape model to reconstruct the pose and reflection characteristics of the eye. Appropriate accuracy for this difficult geometrical configuration is achieved by non-linear optimization exploiting geometry constraints from screen size, planarity, and ray triangulation to adjust initial eye pose and reflection measurements.

4.2.2 Eye Gaze Tracking

Display illumination. There are only a few eye gaze tracking methods exploiting corneal reflections of screen illumination. Iqbal and Lee (2008) propose a method that identifies the reflection of a whole CRT monitor from its periodic flicker pattern using a high-speed camera at framerates larger than twice the screen refresh rate. The detected reflection patch is used to locate an eye in a face image. The patch centroid acts as a glint, where the PoR on the screen is obtained from the pupil-glint vector under a calibrated gaze-mapping. The drawbacks of this method are that it does not compensate for head movements and needs initial calibration of the mapping function. However, the method is interesting, since it is the first approach making use of screen illumination in eye gaze tracking. There are several differences to our proposed technique: The method requires special hardware in the form of a CRT monitor and a high-speed camera. The technique is purely image-based and does not handle geometric information about eye, camera, and display. Instead, it relies on a tedious calibration for a regression-based mapping that does not allowing head movement.

Recently, Schnieders et al. (2010) apply our proposed idea for reconstructing the pose of the display from corneal reflections (Nitschke et al., 2009) to eye gaze tracking. Instead of analyzing reflections of a special marker pattern, they directly use the curved edges from the reflection of a screen with bright content to recover the corresponding edges in 3D and, thus, the display plane. While this method applies a single face image with two eyes and assumes the user to look at the screen, the goal of our work is to provide a general method without constraining gaze direction, and to present a comprehensive performance analysis for scene reconstruction from eye reflections under multiple eye poses from different images regarding a large number of factors.

IR LED Markers. While this work focuses on passive eye pose estimation, the majority of eye gaze tracking techniques relies on active illumination. Commonly used are IR LEDs to create glints in both, image- and geometry-based techniques. A method that has received several focus over the last few years is the cross-ratios method originally introduced by Yoo et al. (2002) and Yoo and Chung (2005). Four LEDs are attached to the corners of the screen, leading to a similar arrangement as in the marker pattern that we use for the implementation of our method. The cross-ratios method describes an image-based remote eye gaze tracking method based on several simplification assumptions to allow slight head movement. The cornea is assumed to be flat; so when the user looks at the screen, the pupil center will be located within the bounding quadrilateral of the screen area created by the four glints. The approach then exploits the property of invariant cross-ratio of four points under projective transformation (Hartley and Zisserman, 2003) to estimate the PoR on the screen from the location of the pupil center within the reflected screen area. Since the technique is purely image-based and does not handle geometric information about eye, camera, and display, camera calibration is not required.

Several works propose extensions to the original method and achieve significant improvements in accuracy. Coutinho and Morimoto (2006) and Ko et al. (2008) introduce a personal calibration to estimate the angular offset between visual and optical axis. Ko et al. (2008) describe several modifications, accounting for robust detection of the specular highlights, replacing the cross-ratio with a more stable geometric transform for the mapping function, and compensating for errors introduced by head movement. Coutinho and Morimoto (2010) propose an extension to compensate for the dependency on the display-eye distance. Kang et al. (2008) carry out a theoretical analysis of the cross-ratios method that is found highly sensitive to the individual eye. They derive an analytic prediction of the subject-specific estimation bias and propose a compensation method.

The cross-ratios method relates to our proposed method as it exploits corneal reflections from screen corners. Nevertheless, it requires additional hardware in form of IR LEDs, and does not obtain any geometric information about the relation between camera, eyes, and screen.

4.3 Method

The controlled illumination system consists of three components: (1) a raster display device which acts as a light source, (2) the cornea of a human eye which reflects the light from the screen, and (3) a camera which captures the light reflected from the cornea. We assume a simple lightpath where the light is reflected only once at the outer surface of the cornea.

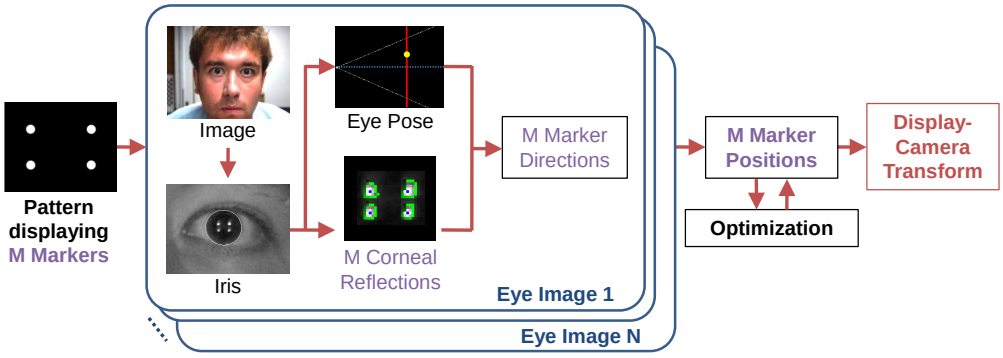


Figure 4.1: Display-camera calibration algorithm. A pattern with M markers is shown on the screen. The corresponding 3D positions in the camera coordinate frame are estimated by intersecting the inverse reflection rays obtained under N different eye poses. The results are improved with a non-linear optimization strategy that jointly refines eye poses, reflection rays, and display pose. Finally, the transformation between screen coordinates and locations on the estimated plane is computed from the M correspondence pairs.

4.3.1 Basic Algorithm

4.3.1.1 Overview

The geometric calibration algorithm (Fig. 4.1) computes the transformation between screen plane and its 3D estimate in the camera coordinate frame. For that purpose, the display shows a pattern encoding M pixel positions on the screen plane. The camera records images of a moving person facing the display, where the pattern reflection on the cornea is visible in an image. Detecting an eye and fitting an ellipse to the contour of the visible iris enables to recover the 3D position and orientation of the eye. After finding the M pattern correspondences from the imaged iris region, the inverse light rays are constructed and reflected at the recovered eye pose into the direction of the 3D screen plane. The actual positions are then computed from the intersections of N rays obtained under varying eye poses. Due to uncertainties in eye modeling and feature extraction, results from this basic approach may show a large error and deviation. The accuracy is, therefore, improved using a non-linear optimization strategy that jointly refines eye poses, reflection rays, and display pose subject to known geometry constraints in the setup. Finally, the transformation describing the relation between 2D screen coordinates and the estimated plane in 3D camera coordinates is obtained from the M correspondence pairs.

4.3.1.2 Display-Camera Transformation

A display is modeled as a screen plane containing the pixels $\mathbf{p} = (p_i, p_j)^{\mathrm{T}}$ that we want to describe as points $\mathbf{P} = (P_x, P_y, P_z)^{\mathrm{T}}$ in the camera coordinate system (Fig. 4.2). The transformation is expressed in homogeneous coordinates

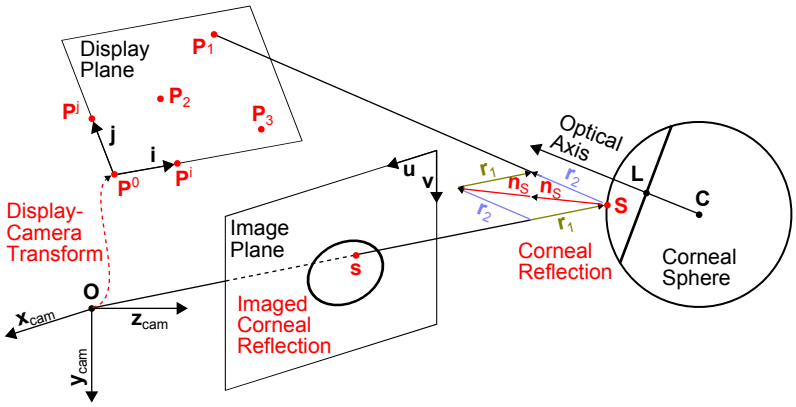


Figure 4.2: Setup for computing the direction towards a light source (on the display). The model is applied to estimate the 3D pose of the screen plane from a set of eye images showing corneal reflections of $M \geq 3$ displayed markers $\mathbf{P}_k$.

as in

$$\begin{aligned} \begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix} &= \mathsf{T} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}^{\mathrm{i}} - \mathbf{P}^0 & \mathbf{P}^{\mathrm{j}} - \mathbf{P}^0 & \mathbf{P}^0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}. \end{aligned} \tag{4.1}$$

Since T describes a planar homography, the inverse transformation from points $\mathbf{P}$ on the plane in camera coordinates to pixels $\mathbf{p}$ on the screen plane, is obtained with the inverse matrix T^{-1}. Matrix T itself depends on the three unknown correspondence pairs at $\mathbf{p}^0 = (0,0)^{\mathrm{T}}$, $\mathbf{p}^{\mathrm{i}} = (1,0)^{\mathrm{T}}$, and $\mathbf{p}^{\mathrm{j}} = (0,1)^{\mathrm{T}}$, and can be estimated from $M \geq 3$ arbitrary correspondence pairs where the points $\{\mathbf{P}_k \,|k = 1,\ldots,M\}$ are coplanar (and not collinear) on the screen plane. From the correspondences, we formulate the equation system $\mathsf{A}\mathbf{t} = \mathbf{b}$ as in

$$\begin{bmatrix} p_{1i} & p_{1j} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{1i} & p_{1j} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{1i} & p_{1j} & 1 \\ & & & & \vdots & & & & \\ p_{Mi} & p_{Mj} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{Mi} & p_{Mj} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{Mi} & p_{Mj} & 1 \end{bmatrix} \begin{bmatrix} t_{11} \\ t_{12} \\ t_{13} \\ t_{21} \\ t_{22} \\ t_{23} \\ t_{31} \\ t_{32} \\ t_{33} \end{bmatrix} = \begin{bmatrix} P_{1x} \\ P_{1y} \\ P_{1z} \\ \vdots \\ P_{Mx} \\ P_{My} \\ P_{Mz} \end{bmatrix}, \tag{4.2}$$

and solve for the vector $\mathbf{t} = (\mathsf{A}^{\mathrm{T}}\mathsf{A})^{-1}\,\mathsf{A}^{\mathrm{T}}\mathbf{b}$ containing the nine unknown matrix elements by computing the pseudo-inverse.

4.3.1.3 Correspondence Representation

To obtain the M required correspondence pairs, the display shows a pattern representing M screen locations, and the corresponding corneal reflections are identified from the eye images. We choose a simple direct representation, where a uniform filled circular marker is placed at each pixel location on a black background, and find the center of the corresponding reflection patch in the image. The relative spatial alignment of the markers is not affected by projection and reflection since the cornea has a convex shape. However, the method has an inherent ambiguity, where the index k of a reconstructed point $\mathbf{P}_k$ is unknown. This can be resolved if the reflected display is approximately aligned with the camera image plane (by initially adjusting the camera orientation around the viewing direction).

4.3.2 Optimization

Using the proposed algorithm to estimate light source positions (Sec. 3.2) leads to a larger error than using a spherical mirror of the same size, with the estimated positions usually being about halfway between the eyes and their true position accompanied by a large ray intersection error. This has several possible reasons. The two main sources of error are the following:

1. The individual shape and parameters of the eye are unknown. Distortion effects increase with gaze angle when reflections move away from the corneal apex towards the boundary. Moreover, the unknown radii of iris r_I and corneal limbus r_L influence the eye pose estimation.

2. It is difficult to exactly locate the contour of the iris which gradually dissolves into the sclera. Due to iris landmarks and blood vessels, this transition is not smooth (Iskander, 2006). Noisy measurements directly affect the orientation in eye pose estimation which itself is crucial for the overall accuracy.

4.3.2.1 Error Function from Known Geometry Constraints

The error for the reconstructed display can be largely decreased with a small modification of the estimated eye poses and imaged reflections. These serve well as initial guesses and are further adjusted by an optimization that minimizes a convex error function e defined as the weighted sum of three error terms as in

$$e = \frac{w_1 e_1 + w_2 e_2 + w_3 e_3}{w_1 + w_2 + w_3}. \tag{4.3}$$

The *intersection error* e_1 is defined as the average distance of the reflected light rays to their estimated intersection points as in

$$e_1 = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \|\mathbf{P}_{ij} - \mathbf{P}_i\|, \tag{4.4}$$

where $\mathbf{P}_{ij}$ is the point on ray j having minimal distance to the corresponding intersection point $\mathbf{P}_i$. If the absolute size of the screen plane is known, for example from a database of model specifications, the *size error* e_2 is defined as the average absolute error of the distances between all M estimated marker positions as in

$$e_2 = \frac{2}{M\,(M-1)} \sum_{i_1=1}^{M} \sum_{i_2=i_1+1}^{M} \left| \|\mathbf{P}_{i_1} - \mathbf{P}_{i_2}\| - \mathrm{GT}_{i_1,i_2} \right|, \tag{4.5}$$

where $\mathbf{P}_{i_1}$ and $\mathbf{P}_{i_2}$ are two estimated marker positions and GT_{i_1,i_2} the ground-truth distance obtained from the known display size. Finally, the *plane error* e_3 is defined as the average absolute deviation of the M estimated marker positions to their approximated best fit plane, containing the centroid, as in

$$e_3 = \frac{1}{M} \sum_{i=1}^{M} |\mathbf{P}_i \cdot \mathbf{n} + p|. \tag{4.6}$$

This plane is given by $\mathbf{P}_i \cdot \mathbf{n} + p = 0$ in Hessian normal form and obtained from orthogonal regression. In order to calculate it, the estimated light source positions are stacked into a matrix $\mathsf{A}_{M\times 4}$[1] as in

$$\mathsf{A} = \begin{bmatrix} \mathbf{P}_1^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \mathbf{P}_M^{\mathrm{T}} & 1 \end{bmatrix}. \tag{4.7}$$

The singular value decomposition (SVD) describes A as a product of three matrices, with $\mathsf{A} = \mathsf{U}\mathsf{D}\mathsf{V}^{\mathrm{T}}$, where $\mathsf{U}_{M\times 4}$ is a matrix with orthogonal columns, $\mathsf{V}_{4\times 4}$ is an orthogonal matrix whose columns are the singular vectors of A, and $\mathsf{D}_{4\times 4}$ is a diagonal matrix whose non-negative entries are the singular values. The plane unit vector $\mathbf{n}$ and distance from origin p are obtained as

$$\begin{bmatrix} \mathbf{n} \\ p \end{bmatrix} = \frac{1}{\sqrt{v_1^2 + v_2^2 + v_3^2}} \mathbf{v}, \tag{4.8}$$

where $\mathbf{v}$ is the singular vector corresponding to the smallest singular value.

[1] Note that this formulation of the SVD requires the number of rows to be equal or larger than the number of columns, with $M \geq 4$. If the plane is estimated from only $M = 3$ points, it is appropriate to extend A by adding a row of zeros to obtain a square matrix.

4.3.2.2 Optimization Strategy

The proposed optimization strategy comprises three subsequent steps:

1. First, we jointly optimize the estimated positions of all corneal sphere centers $\{\mathbf{C}_j \,|j = 1, \ldots, N\}$ with the weights for the error terms set as $w_1 = 0$, $w_2 = 1$, and $w_3 = 1$.

2. Next, we add the imaged reflection centroids $\{\mathbf{s}_{ij} \,|i = 1, \ldots, M; j = 1, \ldots, N\}$ and jointly optimize them together with the adjusted corneal sphere centers from the last step. The weights remain unchanged.

3. Finally, we set weight $w_1 = 2$ and repeat the last step.

Each step is performed using Powell's direction set method (Press et al., 2002) with the number of brackets set to nine, advancing to each subsequent step after convergence is achieved.

Evaluating the proposed technique, we found the adjustment for the corneal sphere projections to be small, where the projected contours correctly contain the imaged irides. Pupillary distances between left and right eye were matching their measured ground truth. This leads to the conclusion that a spheric corneal curvature model of constant size is a feasible assumption for the proposed approach. The technique performs robustly when adding small perturbations to the initial estimation, thus, enabling application with low quality hardware. This can be useful for developing robust solutions for eye gaze tracking. The technique, further, performs well at correcting calibration results obtained from a spherical mirror.

4.3.2.3 Resolving the Sign Ambiguity in Eye Orientation

Regarding 3D eye pose estimation in Section 2.2, there remains a sign ambiguity for the limbus tilt angle τ that could not be determined from only the ellipse parameters. However, after obtaining an initial guess for all eye poses and reflection centroids, it is possible to jointly estimate the missing signs $\{S(\tau_j) \in \{-1, 1\} \,|j = 1, \ldots, N\}$ by applying the geometry constraints introduced within this section. This is done by computing the value of size error e_2 under all 2^N sign combinations and selecting the one resulting in minimal error.

4.4 Implementation

The last section explained the general concepts behind the display-camera calibration algorithm. We will now explain the actual implementation in terms of chosen methods and algorithms to execute the different subtasks. The goal of this implementation is not a fully automatic prototype that can be applied

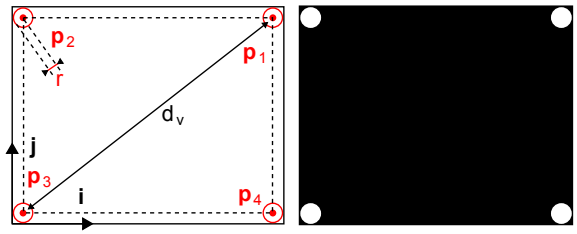


Figure 4.3: Marker pattern used for the implementation.

right away, but rather a research tool to support this first analysis on the applicability of eye reflections for display-camera calibration. The developed framework is applied for in-depth experiments explained in the next section. Figure 4.1 gives an overview of the calibration algorithm where each part will be covered in detail in the following.

4.4.1 Correspondence Representation

For the correspondences on the screen plane, we use a simple direct representation and draw 4 markers at the corners as white circles with radius r around each pixel location $\{\mathbf{p}_k \,|k = 1, \ldots, 4\}$ (Fig. 4.3). In the ideal case (with a point light source at infinity), r spans only a single pixel. However, it has to be set to some higher value to achieve a measurable camera response. On the other hand, r cannot be set too high so as not to overexpose imaged reflections. To resolve the ambiguity of unknown index k, when extracting corneal reflections of multiple markers from a single image, we assume camera and display are facing the user and are approximately aligned.

4.4.2 Image Acquisition

While the display shows the marker pattern, we capture face images of a moving person in front of the display, where each eye region exhibits corneal reflections from the pattern (Fig. 4.16). In order to eliminate spurious reflections from other light sources and increase the signal-to-noise ratio (SNR) for marker reflection patches, capturing is done in the absence of environmental light. To obtain enough camera response, we use maximum aperture and an exposure time of 133 ms. This leads to difficult capturing conditions: The opened aperture limits the depth of field, where objects appear sharp, to a depth range of approximately 5 cm. To prevent defocus blur, persons are required to restrict their movement to that depth plane. The high exposure time requires persons to keep their heads still while an image is recorded. To prevent motion blur, we support controlled head movement by using a low effective framerate of 1 FPS and by showing a small blue flashing mark that signals image capture.

To generally avoid quality degradation from blur, we analyze the sharpness of an eye region and remove images that fall below a certain threshold. In order to guide head movement, the result is then mapped to the frequency range between 500 and 5000 Hz to provide an audible feedback in form of a 100-ms beep sound through the PC speaker. The corresponding image processing is done as follows: At first, we find the eye regions by smoothing the face image with a Gaussian filter and subsequently applying the boosted cascade-classifier of Haar-like features contained with OpenCV (Bradski and Kaehler, 2008). We then calculate the sharpness for each rectangular bounding box using the method of Shen and Chen (2006), that is robust to low-contrast images, by calculating the energy ratio $E_{\mathrm{AC}}/E_{\mathrm{DC}}$ between high and low frequency band. Here, E_{DC} denotes the square of the constant-component coefficient and E_{AC} the sum of squares of the varying-component coefficients in the discrete cosine transform (DCT) of the image.

4.4.3 Eye Detection and Iris Contour Fitting

Detecting the contour of the iris comprises two tasks, eye detection and iris fitting. At first, the rough eye region is identified in the image. There exists a multitude of approaches for automatic eye detection and tracking; and a suitable one has to be selected based on the constraints of the particular system. An overview is given in Section 2.2.1. We apply a simple interactive strategy where an initial guess for the ellipse parameters is obtained from a user selecting four points on the contour of the imaged iris.

For automatic fitting, the image is transformed into a binary edge image by smoothing with a Gaussian filter and extracting edges with an adaptively-thresholded Canny edge detector. Starting from the initial guess, an accurate contour is estimated by iteratively minimizing the error function

$$eval = \sum_{(u,v)^{\mathrm{T}} \in E} D'(u,v), \tag{4.9}$$

$$D'(u,v) = \begin{cases} D(u,v) & \text{for } D(u,v) \leq D_{\max}, \\ D_{\max} & \text{otherwise,} \end{cases}$$

where E is the set of pixels representing the ellipse contour, and $D(x,y)$ is a particular pixel value in the distance-transformed binary edge image. For each value, we apply a constant upper bound $D_{\max}$ to reduce the effect of non-edge contour points on the estimation. This is necessary to robustly handle occlusions by the eyelids which especially occur at increasing gaze angles.

4.4.4 Eye Pose Estimation

Physiologically, the iris is located directly in front of the lens. Its outer structures extend behind the transparent cornea and the beginnings of the opaque

white sclera. The contour of the visible iris marks the corneal limbus which is the surface shape discontinuity where the corneal sphere transitions into the eyeball sphere. Having fit an ellipse to this contour, we use the method described in Section 2.2.2.3 to estimate the position and orientation of the circular limbus in 3D. The method is based on the weak-perspective projection model which is a feasible assumption since the depth of the tilted limbus is small compared to the camera-eye distance. The correct solution for the pose of the limbus is automatically selected from the two possible ones by jointly exploiting constraints of the display-camera geometry for multiple eyes as described in Section 4.3.2.3. Having obtained the pose of the limbus plane, we calculate the center of the corneal sphere located at some distance from the center of the limbus along the negative optical axis.

4.4.5 Correspondence Detection

To extract the marker reflections in image, we find the four intensity peaks within the boundary of the iris and segment the corresponding regions by connected-component analysis with an adaptive threshold for background subtraction. Having extracted the pixels corresponding to a particular region R_k, where $I(u,v)$ is the intensity value at pixel $(u,v)^{\mathrm{T}}$, we calculate the intensity centroid $\mathbf{s}_k$ of the region with subpixel accuracy as in

$$\mathbf{s}_k = \frac{\sum\limits_{(u,v)^{\mathrm{T}} \in R_k} I(u,v) \begin{bmatrix} u \\ v \end{bmatrix}}{\sum\limits_{(u,v)^{\mathrm{T}} \in R_k} I(u,v)}. \tag{4.10}$$

4.4.6 3D Display Reconstruction

Assume $N \geq 2$ eye images are captured for a moving person. Having obtained the eye poses and corresponding corneal reflections $\mathbf{s}_k$, we are able to construct the inverse reflection rays pointing towards the marker positions $\mathbf{P}_k$ on the display. Each position is estimated as explained in Section 3.2, as the intersection of reflection rays obtained for marker k under different eyes poses. The display-camera transformation is recovered from $M \geq 3$ reconstructed marker positions.

4.4.7 Optimization

Due to uncertainties in eye modeling and feature extraction, it turns out that the estimation error for the basic algorithm becomes large. We, therefore, apply the optimization strategy described in Section 4.3.2 to iteratively improve the reconstructed display plane, the positions of the estimated cornae, and the imaged reflection centroids. Beside obtaining a better estimate for the

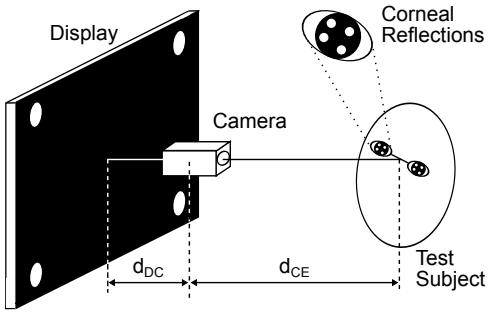


Figure 4.4: The experimental setup with display, camera, and test subject exhibiting corneal reflections from the light source pattern in both eyes.

display plane, the improved eye positions may be used for other applications such as eye gaze tracking.

4.5 Experiments

In the following section we explain several comprehensive experimental series that were conducted in order to thoroughly analyze the performance of the proposed method and to obtain further knowledge about scene reconstruction from eye reflections in general.

4.5.1 Single Eye

To verify the general feasibility of the approach, we first investigate the effect of display size, camera-eye distance, and individual eye anatomy experimentally using only off-the-shelf components.

4.5.1.1 Setup

The setup consists of a 57-in, 16:9 Epson Livingstation LS57P2 TFT LCD display and a Point Grey Flea2 camera with 1024×768 resolution. The intrinsic camera parameters are calibrated using OpenCV functions. The camera is placed at a distance d_{DC} of 15 cm in front of the display center. Test subjects are seated with the head fixed at increasing camera-eye distance d_{CE} from 25 to 95 cm in front of display and camera (Fig. 4.4). We display a time-series of patterns with diagonal size d_{v} increasing from 10 to 43 at steps of 1 in[2] and, for each pattern, capture a single image including both eyes. The radius r of the circular markers had to be linearly increased with diagonal size d_{v} from 0.5 to 3 in due to radiance attenuation at high viewing angles.

[2]Inches are used for better comparison with display sizes.

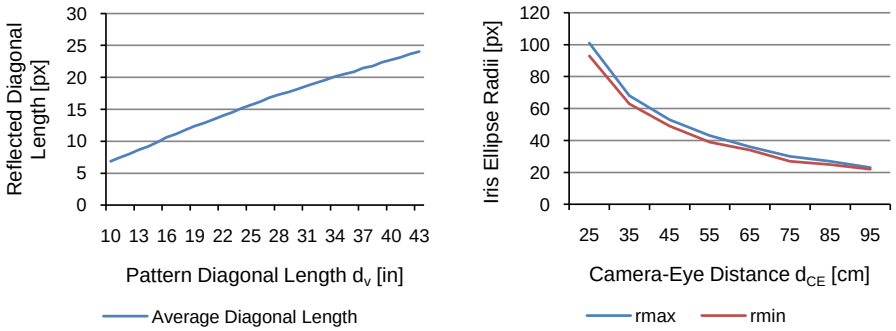


Figure 4.5: Results of image analysis. (left) The average size of the imaged pattern reflection increases with pattern size d_v (camera-eye distance d_{CE} fixed to 35 cm). The influence of corneal shape curvature can be noticed. (right) The resolution of the imaged iris decreases with increasing camera-eye distance d_{CE} (pattern size d_v fixed to 25 in).

4.5.1.2 Image Analysis

At a fixed distance d_{CE} of 35 cm, diagonal sizes for the imaged marker pattern reflections span a range of 7 to 24 pixels (Fig. 4.5, left). For large pattern sizes we notice a distortion of marker reflection patches from corneal curvature. This does, however, not lead to a measurable decrease in accuracy. The impact of decreasing resolution can be noticed with increasing camera-eye distance, where the size of the iris decreases from about 100 to 20 pixels (Fig. 4.5, right). Although the optical axes of eyes and camera are approximately aligned, the shape of the imaged irides is slightly elliptical, with the vertical axis constantly measuring 92% of the horizontal axis. This finding coincides with common anthropometric data (Sec. 2.1). Corneal reflection extraction works successful for typical distances up to 75 cm, begins to fail at larger distances for small pattern sizes, and completely fails above 85 cm. Refer to Figure 4.6 for results of image data evaluation.

4.5.1.3 Results

Each marker position on the display plane is estimated from only a single eye image by intersecting the corresponding inverse reflection ray with the display plane located at known GT display-eye distance $d_{DE} = d_{DC} + d_{CE}$. We define the estimation error as the signed deviation of the reconstructed pattern size from the GT d_v. Figure 4.7 shows experimental results. As corneal curvature can have a high local variation, we apply an especially large range for radius r_C. Its direct influence on the inverse reflection ray results in a high impact on performance. We obtain highest accuracy for all test subjects using a radius of 7.8 mm. Regarding camera-eye distance, we do not observe any significant deviation in results for common distances between 25 and 55 cm. At larger distances, the variance increases with noise in corneal reflection extraction resulting from low resolution. While the variance is low among multiple data

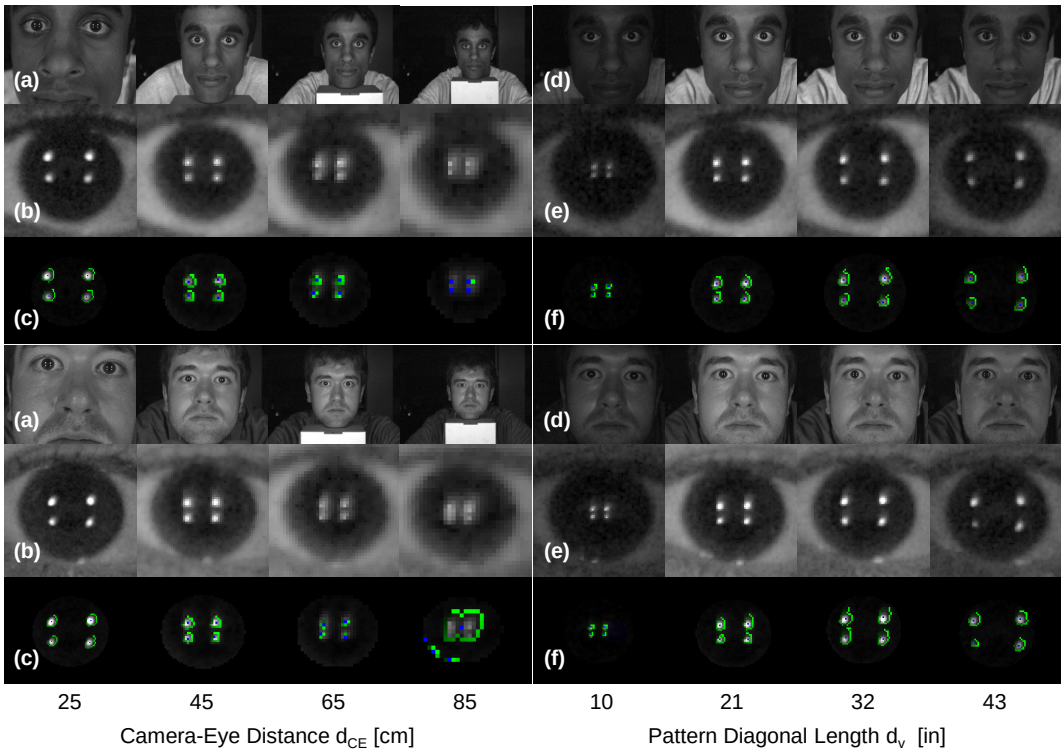


Figure 4.6: Face images of two test subjects. (a) Increasing camera-eye distance d_{CE} of 25, 45, 65, and 85 cm with a fixed pattern size d_{v} of 25 in. (b) Corresponding left eye images. The limbus size decreases from about 100 to 20 pixels. (c) Corneal reflection extraction results, showing the boundary of each reflection patch (green with intensity decreasing in clockwise direction) and the corresponding intensity centroid (blue). Extraction is successfully performed for common distances up to 75 cm, begins to fail for small pattern sizes at 85 cm, and completely fails at larger distances. (d) Increasing pattern size d_{v} of 10, 21, 32, and 43 in with a fixed camera-eye distance d_{CE} of 35 cm. (e) Corresponding left eye images. The size of the reflected pattern increases from 7 to 24 pixels. A small distortion of the reflection area, resulting from corneal curvature, can be noticed for large pattern sizes. Note that gamma correction is applied for better visibility to (a), (b), (d), and (e). (f) Corneal reflection extraction results.

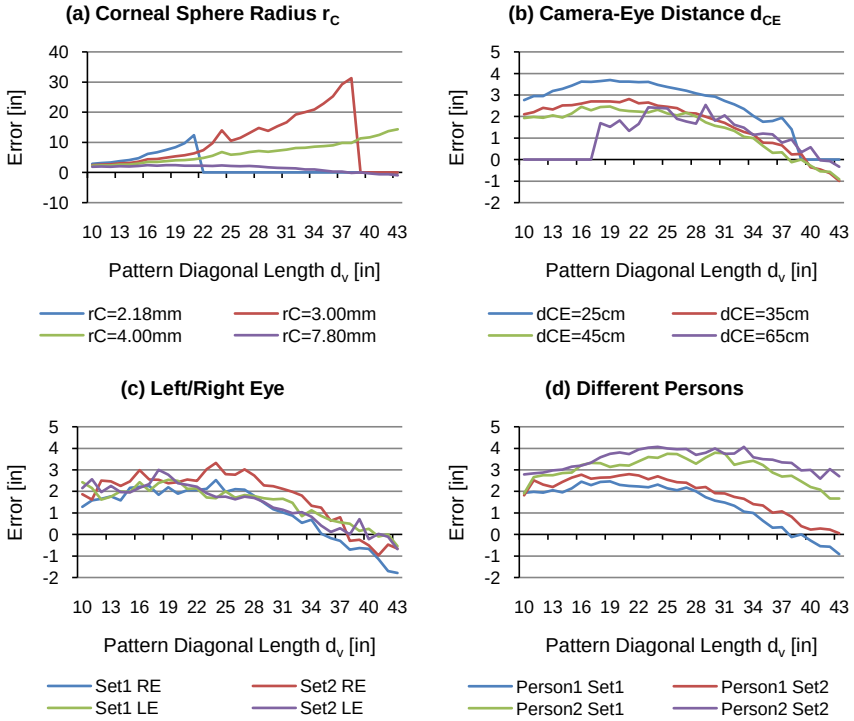


Figure 4.7: Results for the single eye image experiment. The error is obtained as the signed deviation between reconstructed and GT pattern sizes d_v for a range of 10 to 43 in, at a camera-eye distance d_CE of 45 cm (if not stated otherwise). (a) Increasing corneal sphere radius r_C. Small radii result in high errors and failing reflections at large pattern sizes. A value of 7.8 mm achieves the highest accuracy among all test subjects and is used for the following results. (b) Increasing camera-eye distance d_CE. Variance increases with distance because of deviations in iris fitting and reflection extraction due to low resolution. (c) A comparison between two data sets for left and right eye from the same subject does not show any significant deviation. (d) In contrast, a comparison between two data sets from two subjects shows significant deviation for results obtained from different persons, suggesting a correlation with individual eye anatomy.

sets of left and right eyes from the same subject, it is significantly higher among data sets of eyes from different subjects, increasing with pattern size. We relate this to differences in individual eye anatomy.

4.5.1.4 Discussion

The described experimental setup used only off-the-shelf components in controlled and unnatural conditions to conduct a first evaluation for recovering information from eye reflections. Results show a significant influence of individual eye geometry and a relatively large reconstruction error within 20% of pattern size d_{v}. Nevertheless, the obtained results are promising and successfully verify the basic feasibility of the proposed approach. The findings are summarized in the following.

Coinciding with common anthropometric data, the shape of imaged irides is found slightly elliptical and flattened in the vertical direction. In reconstruction, the corneal sphere radius that is applied with the eye model achieves the highest accuracy. No significant deviation is found for common display-eye distances and different datasets for the same subject. In contrast, a significant deviation is found for datasets from different subjects, relating to the individual eye geometry.

4.5.2 Two Eyes

The results obtained for the single eye experiment suggest that further evaluation is necessary. Different geometric eye models have to be tested in order to better approximate the eye shape. We are interested in common model parameter deviation and its respective effect on reflection estimation. As the aim is to apply the technique with off-the-shelf hardware, we need to consider practical problems such as image resolution or noise. To evaluate and understand these effects, we introduce a test framework using synthetic data.

4.5.2.1 Rendering Framework

It is not a trivial task to create valid synthetic data. There are several requirements that have to be fulfilled:

- The complete scene consisting of display, eyes, and camera has to be modeled.
- The synthetic system parameters have to resemble the real system parameters.
- The parameters have to be adjustable to generate experimental data.
- The modeling has to be physically correct.

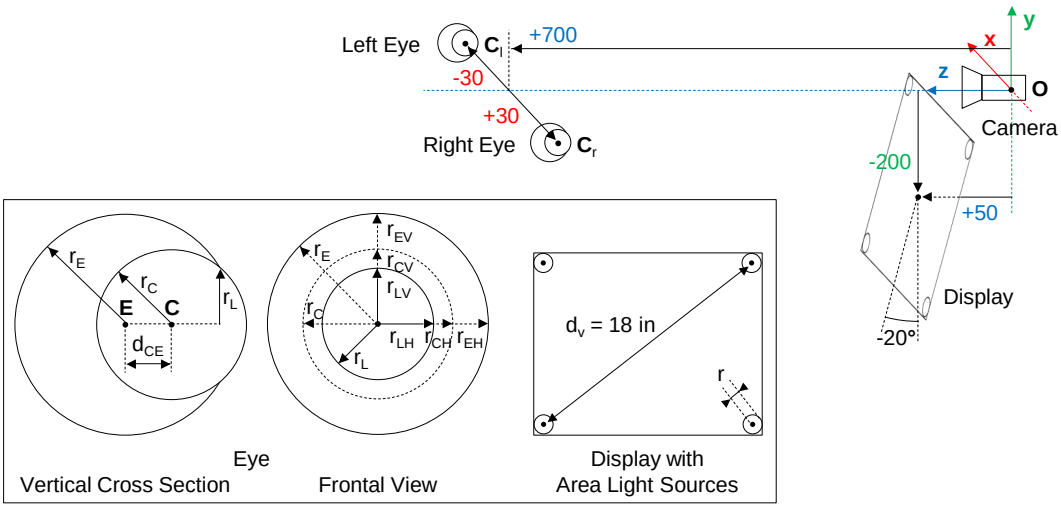


Figure 4.8: Scene model for rendering synthetic image data using the `pbrt` framework. Modeling parameter values and relationship between display, eyes, and camera are chosen as to reproduce the physical setup used for the real experiments. The eye model is constructed of two overlapping ellipsoids to allow analysis of asphericity. The display is represented as a planar arrangement of spherical area light sources, resembling the marker pattern.

Designing a complete physically correct eye reflection modeling system from scratch is an ambitious task and out of the scope of this work. Instead, we create an experimental framework based on the free physically based renderer `pbrt` (Pharr and Humphreys, 2004) that implements a state-of-the art ray-tracer for photorealistic rendering. It is distributed with sources and, thus, can be extended and re-compiled. This work applies `pbrt` version 1.03[3].

4.5.2.2 Scene Model

We create a scene model that closely resembles the setup for the real experiments in order to obtain comparable results. The camera is placed at the origin of a left-handed coordinate system looking along the positive z-axis. The eyes are located at $z = 700$ mm with a baseline separation of 60 mm. The display is modeled as its planar light source pattern, and located below and in front of the camera, slightly tilted towards the positive z-axis. See Figure 4.8 for an overview of the setup.

Display. The display is represented by its planar pattern. The corresponding markers are modeled as spherical area light sources with radius $r = 0.25$ in, in a rectangular arrangement with diagonal separation $d_v = 18$ in.

[3]By the time of writing, `pbrt` version 2 has been released together with the second edition of the book.

Eye. The algorithm for eye reflection analysis explained in this work uses the eye model derived in Section 2.1, where the eyeball and the cornea are represented as two intersecting spheres. While the shape of the real eye varies individually, its average shape is very close to two intersecting ellipsoids, slightly flattened in the vertical direction. In order to analyze the effect of this asphericity we represent the eyeball and the cornea as ellipsoids with horizontal, vertical, and curvature radii, $r_{\mathrm{EH}}/r_{\mathrm{CH}}$, $r_{\mathrm{EV}}/r_{\mathrm{CV}}$, and $r_{\mathrm{E}}/r_{\mathrm{C}}$ respectively. Since an ellipsoid shape model is not available in `pbrt` version 1.03 we implement it as a plug-in extension into the renderer. Note that the radii for eyeball and cornea, in a particular dimension, are correlated with a constant ratio.

In the synthetic eye model, the center of the cornea $\mathbf{C}$ marks the origin of the eye coordinate system that is align with the camera coordinate system, where the gaze direction points towards $-z$. The center of the eyeball $\mathbf{E}$ is located at distance $d_{\mathbf{CE}}$ from the center of the cornea along the optical axis. The eyeball surface is modeled using material "uber" with a small roughness value of 0.001, and equal diffuse and glossy reflection coefficients $k_{\mathrm{d}} = k_{\mathrm{s}} = 0.8$. This creates a white shiny "kitchen-sink"-like appearance. The corneal surface is modeled using the "translucent" material with roughness 0, fraction of light reflected 0.2, and specular reflection coefficient $k_{\mathrm{s}} = 1.0$. The iris is not modeled explicitly, but emerges as the base plane of the part of the corneal shape that is not occluded by the eyeball. Thus, the iris boundary is equal to the limbus. The color of the iris is controlled by the diffuse reflection coefficient of the cornea, where the default value $k_{\mathrm{d}} = 0$ represents a black iris without any impact on corneal reflection extraction. As to not affect the appearance of the iris, the part of the eyeball in front of the limbus is removed, creating an ellipsoid cap.

Camera. The rendered scene is captured by a virtual camera with similar specifications as the Point Grey Flea2G camera used for the multiple eyes experiments. We set the resolution $r_x = 2448$, $r_y = 2048$ pixels, and the field of view $fov_y = 10.50°$. The intrinsic parameters for the virtual camera are required to run the display-camera calibration algorithm on synthetic data. For a real camera, we need to perform a calibration, e.g., by recording images of a known calibration rig, detecting known point-correspondences, and minimizing the re-projection error in the image. For the virtual camera, this is not necessary since its behavior is described by the imaging model of `pbrt` in Pharr and Humphreys (2004, pp 255).

The image plane is aligned parallel to the xy-plane at $z = 1$, with the image bounded by the camera viewing frustum, where x_{l}, x_{r}, y_{b}, and y_{t} denote the left, right, bottom, and top coordinates respectively. The field of view fov corresponds to the smallest of both image dimensions with resolution r and range $[-1, 1]$ on the image plane. In case of a square image, we have $x_{\mathrm{l}} = -1$, $x_{\mathrm{r}} = 1$, $y_{\mathrm{b}} = -1$, $y_{\mathrm{t}} = 1$, and $r = r_x = r_y$. In case of a rectangular image, we

generalize this, and distinguish two cases based on image orientation as in

$$\begin{cases} x_\mathrm{l} \leftarrow a x_\mathrm{l},\ x_\mathrm{r} \leftarrow a x_\mathrm{r},\ r = r_y & \text{if } r_x \geq r_y, \\ y_\mathrm{b} \leftarrow \frac{1}{a} y_\mathrm{b},\ y_\mathrm{t} \leftarrow \frac{1}{a} y_\mathrm{t},\ r = r_x & \text{otherwise,} \end{cases} \tag{4.11}$$

where $a = r_x / r_y$ is the aspect ratio.

Camera matrix K as derived from the given imaging geometry as in

$$\begin{aligned} \mathsf{K} &= \begin{bmatrix} f & 0 & c_{0u} \\ 0 & f & c_{0v} \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{r_x}{x_\mathrm{r} - x_\mathrm{l}}\right) \frac{1}{\tan(fov/2)} & 0 & -x_\mathrm{l} \left(\frac{r_x}{x_\mathrm{r} - x_\mathrm{l}}\right) \\ 0 & \left(\frac{r_y}{y_\mathrm{t} - y_\mathrm{b}}\right) \frac{1}{\tan(fov/2)} & -y_\mathrm{b} \left(\frac{r_y}{y_\mathrm{t} - y_\mathrm{b}}\right) \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{r}{2}\right) \frac{1}{\tan(fov/2)} & 0 & \left(\frac{r_x}{2}\right) \\ 0 & \left(\frac{r}{2}\right) \frac{1}{\tan(fov/2)} & \left(\frac{r_y}{2}\right) \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned} \tag{4.12}$$

where f is the focal length and $\mathbf{c}_0 = (c_{0u}, c_{0v})^\mathrm{T}$ the principal point. Since the camera is modeled as an ideal pinhole camera, non-linear lens distortions do not occur.

4.5.2.3 Setup

We perform experiments for varying eye shape, iris color, marker size, image resolution, and image noise. The data for each experiment is an image series generated from a `pbrt` scene description according to the format specification in Pharr and Humphreys (2004, pp 911). Since we use high quality settings for resolution, pixel sampling, and surface integration, the rendering time for each image amounts to approximately 15 min on a Intel Core2 Duo CPU.

Display-camera calibration is performed using the standard algorithm explained in Section 4.4. Where applicable, we compare the results obtained with and without eye pose estimation. Using the known GT eye position allows to separately analyze the effect of parameter variation on inverse light path and reconstructed display. On the other hand, using eye pose estimation increases the overall error, but better describes reality. Comparing both results allows to analyze the effect of eye pose estimation.

Accuracy is estimated using the three measures

$$
\begin{aligned}
e_{\mathrm{C}} &= \frac{1}{2}\sum_{i=1}^{2} \|\mathbf{C}_i - \mathrm{GT}\left(\mathbf{C}_i\right)\| \\
e_{\mathrm{P}} &= \frac{1}{4}\sum_{k=1}^{4} \|\mathbf{P}_k - \mathrm{GT}\left(\mathbf{P}_k\right)\| \\
e_{\mathrm{S}} &= \frac{1}{2}\sum_{l=1}^{2} \|d_l - \mathrm{GT}\left(d_l\right)\|
\end{aligned}
\tag{4.13}
$$

where the *cornea position error* e_{C} describes the average deviation between reconstructed and GT corneal center positions $\mathbf{C}$, the *display pose error* e_{P} describes the average deviation between reconstructed and GT marker positions $\mathbf{P}$, and the *display size error* e_{S} describes the average absolute deviation between reconstructed and GT diagonal lengths d_{v}.

4.5.2.4 Corneal Shape

Corneal shape is an important parameter that we want to analyze in multiple experiments. Regarding the calibration algorithm, we set the values for limbus and corneal sphere radii $r_{\mathrm{L}} = 5.75$ and $r_{\mathrm{C}} = 7.80$ mm.

Corneal Sphere Radius r_{C}. At first, we represent the eye as two overlapping spheres and vary the scale according to radius $r_{\mathrm{C}} \in [7.65, 7.95]$ with steps of 0.05 mm. As the whole eye model scales, other parameter values are calculated accordingly. Important parameters in rendering are the distance between eyeball and cornea $d_{\mathbf{CE}}$; and in eye pose estimation, the distance between limbus and cornea $d_{\mathbf{LC}}$, and the size of the limbus r_{L}. Refer to Table 4.2(a) for an overview of all eye shape parameter values in rendering. Figure 4.9 shows the results. As expected, the error becomes minimal when parameter values are equal to the values chosen for the estimation algorithm. Interestingly, the impact of r_{C} turns out rather small.

Corneal Ellipsoid Radii r_{CH} and r_{CV}. We now test the effect of asphericity. Starting with spherical shape, we either decrease the vertical ellipsoid radius r_{CV} or increase the horizontal ellipsoid radius r_{CH}. The actual values for all eye model parameters are based on varying limbus radii according to common anthropometric ranges (Snell and Lemp, 1997; Kaufman and Alm, 2003): We independently vary the vertical and horizontal limbus radii, $r_{\mathrm{LV}} \in [5.25, 5.75]$ and $r_{\mathrm{LH}} \in [5.75, 6.25]$, with steps of 0.10 mm. The boundary of the iris is equal to the limbus. Since other parameters remain constant at their average value, the curvature at the corneal apex r_{C}, the distance $d_{\mathbf{CE}}$, and the distance $d_{\mathbf{LC}}$ are not affected. Refer to Table 4.2(b) for an overview of eye shape parameter values.

Table 4.2: Eye parameter variation [mm] in synthetic experiments on corneal shape.

r_{LH}	r_{CH}	r_{EH}	r_{LV}	r_{CV}	r_{EV}	r_{L}	r_{C}	r_{E}	$d_{\mathbf{CE}}$
(a) Corneal sphere radius r_{C}									
r_{L}	r_{C}	r_{E}	r_{L}	r_{C}	r_{E}	5.64	**7.65**	11.28	4.60
						5.68	**7.70**	11.35	4.63
						5.71	**7.75**	11.43	4.66
						5.75	**7.80**	11.50	4.69
						5.79	**7.85**	11.57	4.72
						5.82	**7.90**	11.65	4.75
						5.86	**7.95**	11.72	4.78
(b) Corneal ellipsoid radii r_{CH}, r_{CV}									
5.75	7.80	11.50	**5.25**	7.44	10.97	—	**7.80**	11.50	4.69
			5.35	7.51	11.07	—			
			5.45	7.58	11.18	—			
			5.55	7.65	11.28	—			
			5.65	7.73	11.39	—			
			5.75	7.80	11.50	—			
5.85	7.87	11.61				—			
5.95	7.95	11.72				—			
6.05	8.02	11.83				—			
6.15	8.10	11.94				—			
6.25	8.18	12.05				—			

Note: The three groups of values separated by vertical lines corresponding to the three dimensions of an ellipsoid. The values for a particular configuration (row) and group are calculated from the bold-marked parameter value. The corresponding ranges are based on common anthropometric variation.

(a) A varying corneal sphere radius r_{C} changes the size of the eye equally in all three dimensions.

(b) Keeping the curvature at the corneal apex r_{C} fixed, we either decrease the size in the vertical dimension or increase the size in the horizontal dimension.

The estimation error does not only depend on varying eye parameters but also on the relative arrangement of eye positions in the xy-plane of the camera coordinate frame: Assume an equal shape variation is applied to both eyes. Then, a variation in a direction parallel to the baseline of both eyes has a higher impact than a variation in any other direction. To explain this effect, we perform evaluation under three different eye arrangements, with a baseline in x-direction ($x = \pm 30$, $y = 0$ mm), y-direction ($x = 0$, $y = \pm 30$ mm) and both, xy-directions ($x = \pm 30$, $y = \pm 30$ mm). The latter two cases correspond to rather uncommon face positions and might in fact be taken from multiple images.

Figure 4.10 shows the results. As in the last experiment, the error becomes minimal for a spherical shape at parameter values equal to the constants of

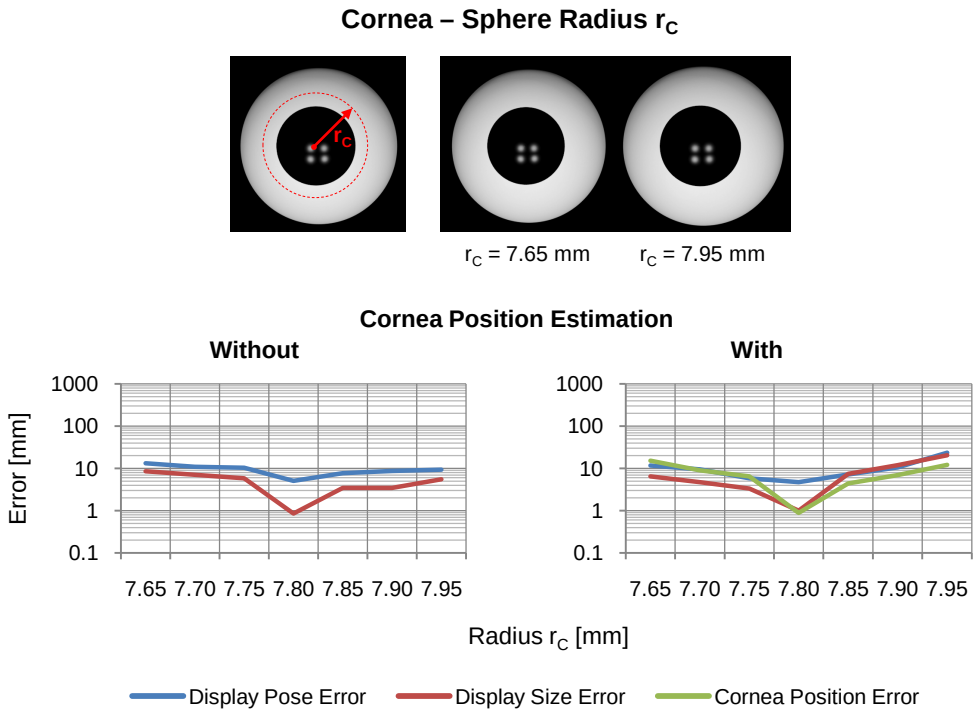


Figure 4.9: Result for experiment on synthetic data for corneal size variation. The eye is represented as two overlapping spheres with varying corneal sphere radius r_C. Other shape parameters are calculated accordingly. (top) Rendered right eye with minimum and maximum size. (bottom) Reconstruction error, with and without eye pose estimation.

the geometric eye model assumed for reconstruction. Comparing the different arrangements, without eye pose estimation, we observe the explained effect where the error is larger for variation in a direction parallel to the baseline. With a baseline in xy-directions, the effects overlap and significantly decrease the average error. Applying eye pose estimation, the result becomes affected by the varying iris contour. The eye pose behaves corresponding to the properties of the algorithm under weak-perspective projection (Sec. 2.2.2.3), where a varying minor radius for the limbus ellipse affects only eye orientation, but a varying major radius affects orientation and position. With eye pose estimation, the overall display reconstruction error increases. The display pose error becomes independent of eye arrangement. The display size error behaves more differentiated, however, with the same effect mentioned before where an arrangement in xy-directions significantly decreases the average error.

4.5.2.5 Iris Color

Caused by an increasing density of light-absorbing pigment in the iris tissue the color of normal irides among individuals ranges from gray to green, blue,

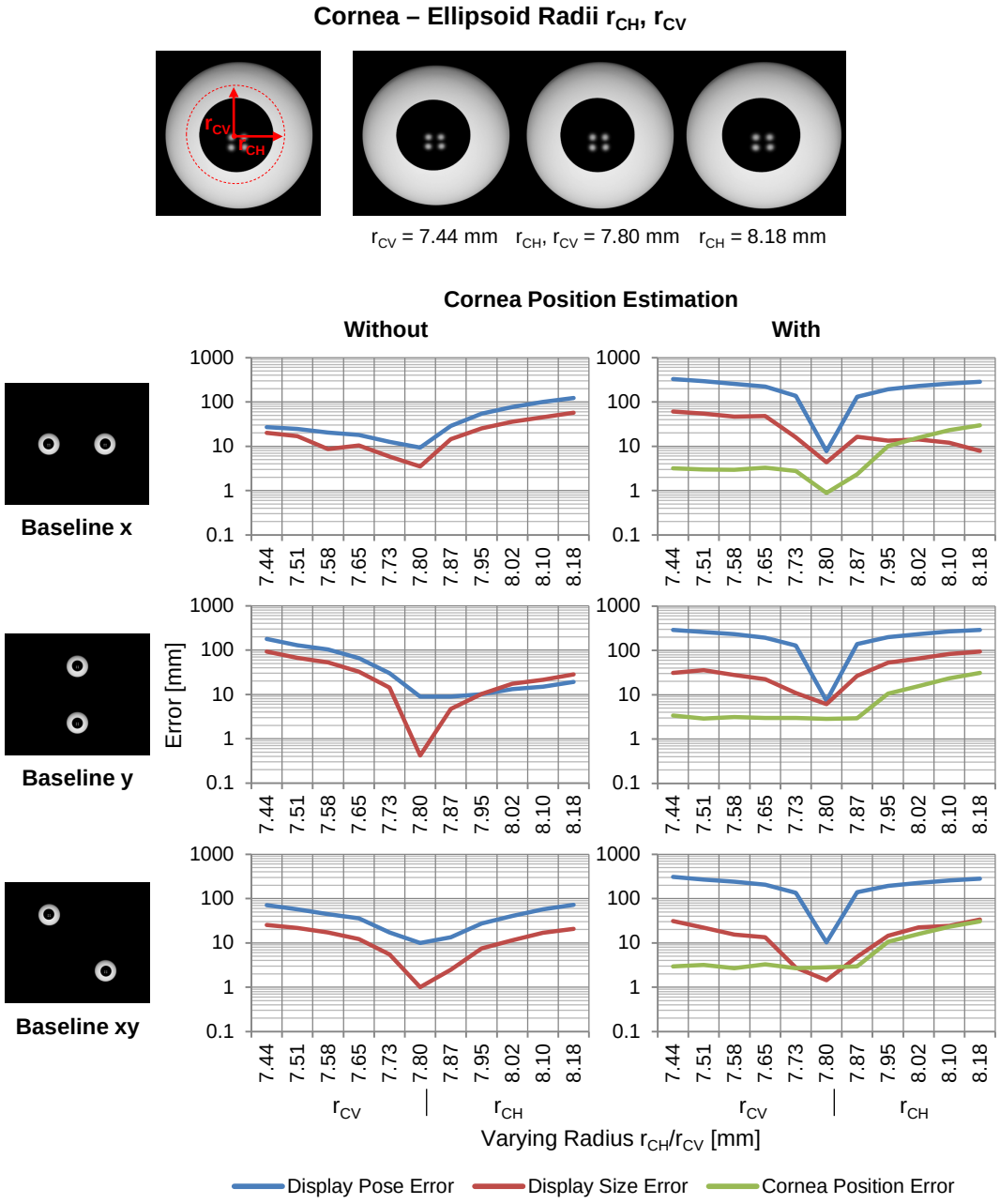


Figure 4.10: Result for experiment on synthetic data for corneal shape variation. The eye is represented as two overlapping ellipsoids with varying either the vertical radius r_{CV} or the horizontal radius r_{CV}. Other shape parameters are calculated accordingly. (top) Rendered right eye with minimum, average, and maximum radii. (bottom) Reconstruction error, with and without eye pose estimation, for three different eye arrangements, with baseline in x-, y-, and xy-directions.

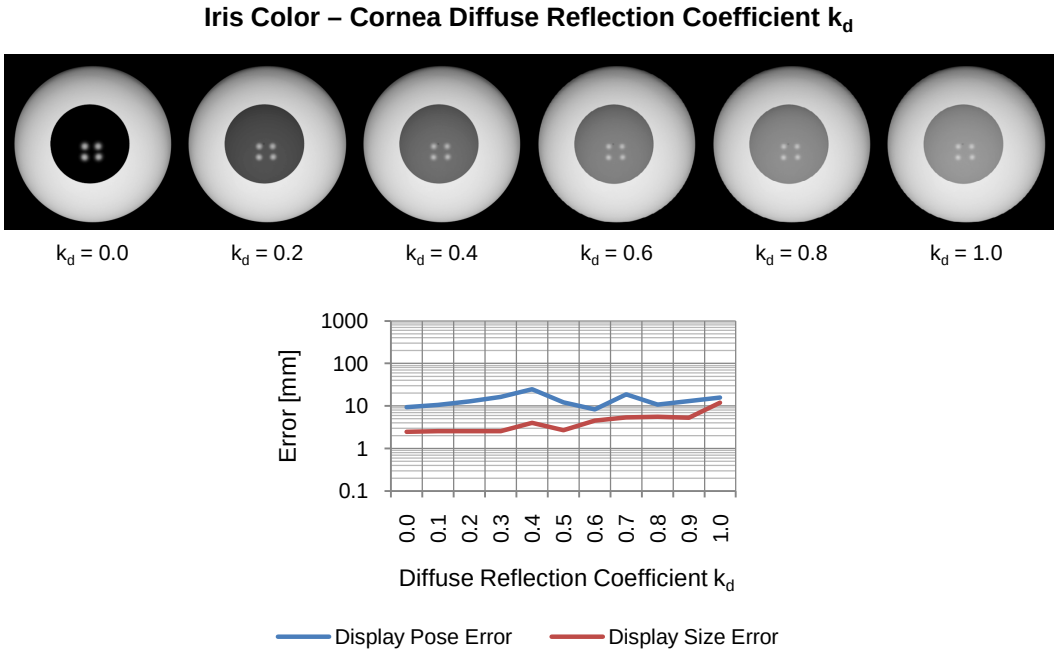


Figure 4.11: Result for experiment on synthetic data for iris color variation. The iris is not rendered explicitly, but emerges as the base plane of the limbus where cornea and eyeball intersect. Therefore, varying iris color is simulated by increasing the diffuse reflection coefficient of the cornea k_d. (top) Rendered right eye with increasing k_d. (bottom) Reconstruction error without eye pose estimation.

brown, and black. To simulate this effect, we vary the diffuse reflection coefficient of the rendered cornea $k_\mathrm{d} \in [0.0, 1.0]$ with steps of 0.1. Since iris color variation does not significantly influence robust iris contour fitting, we only discuss an analysis without eye pose estimation.

Figure 4.11 shows the results. An increasing iris reflectivity leads to increasing background subtraction threshold for the connected component analysis in corneal reflection finding, which decreases the size of each marker reflection patch. While the display size error increases with higher coefficient values, we do not measure any significant influence on the average display pose error. However, variance slightly increases for both error measures.

4.5.2.6 Light Source Size

The display pattern shows four white circular markers with radius r. In the ideal case, the marker spans only a single pixel. In reality, it needs to be set to some higher value to achieve a measurable camera response, which leads to increasing size, distortion, and probably overexposure for each imaged reflection patch. We simulate this effect by increasing the radius of the spherical area light sources representing the markers $r \in [0.1, 1.0]$ with steps of 0.1 in.

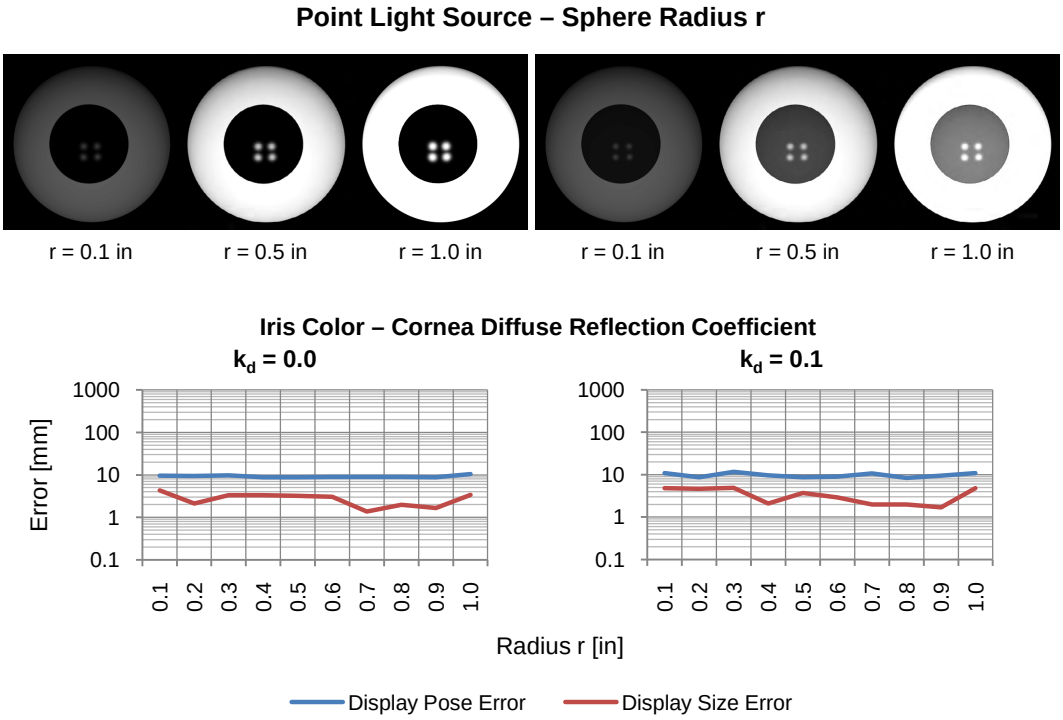


Figure 4.12: Result for experiment on synthetic data for light source size variation. The markers on the display plane are rendered as spherical area light sources. The experiment simulates varying marker size by increasing light source radius r. (left),(right) Results for different simulated iris colors. (top) Rendered right eye with increasing radius r. (bottom) Reconstruction error without eye pose estimation.

Because an increasing light source size increases the amount of light reflected from the iris, we simulate two different iris colors with diffuse corneal reflection coefficients $k_\mathrm{d} = 0.0$ and $k_\mathrm{d} = 0.1$. Eye pose estimation is omitted since variations in iris color and reflectance do not have any significant impact on iris contour fitting.

Figure 4.12 shows the results. We do not measure any significant impact of light source size on the estimation. Comparing different iris colors shows the same effect discussed before, where a brighter color leads to a slight increase in error variance.

4.5.2.7 Camera Quality

Understanding the impact of imaging parameters is important since the calibration algorithm is potentially applied with low-quality camera hardware. We perform experiments for varying image resolution and noise.

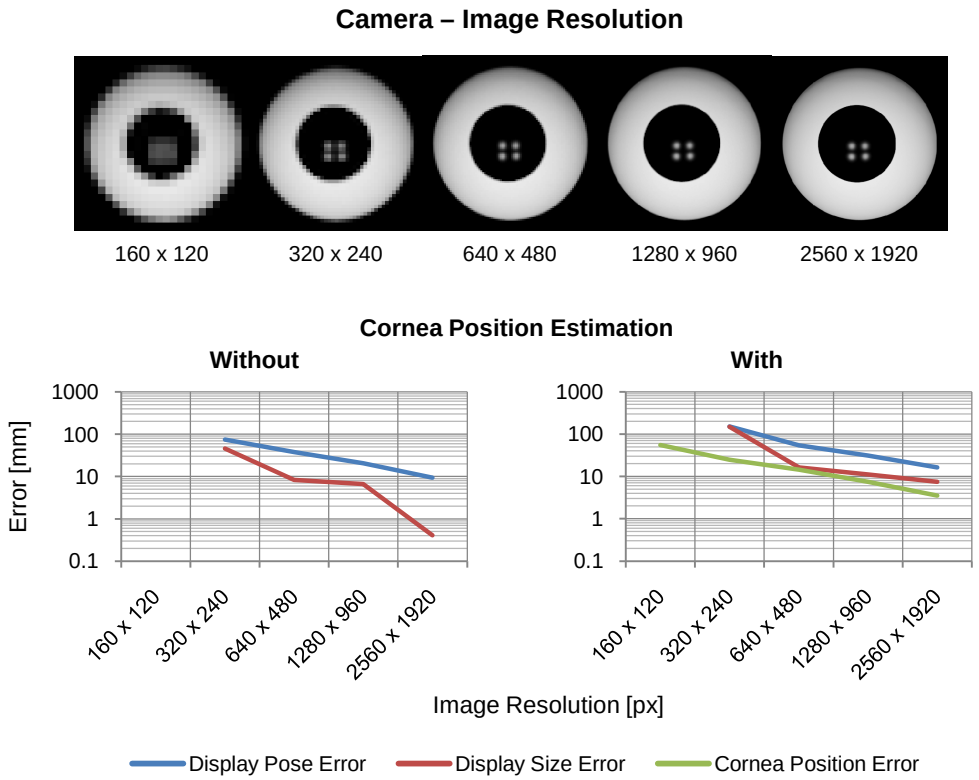


Figure 4.13: Result for experiment on synthetic data for image resolution variation. To examine the effect of varying spatial sampling we render images with increasing resolution. (top) Rendered right eye with increasing resolution. (bottom) Reconstruction error with and without eye pose estimation.

Image Resolution. Image resolution describes the number and size of the pixels and, thus, the spatial sampling distance in the image plane. A decreasing resolution increases the threshold for the size of captured details and acts as a low-pass filter on light from smaller structures that integrates at a particular pixel. To examine this effect we render images with increasing resolution, starting at 160×120 and doubling dimensions until reaching a maximum of 2560×1920 which compares to the resolution of the Point Grey Flea2G camera used for multiple eyes experiments. As resolution affects both, iris contour fitting and corneal reflection extraction, we compare results with and without eye pose estimation. Note that camera parameters change with resolution, and have to be adjusted for data processing.

Figure 4.13 shows the results. At the lowest resolution, corneal reflection extraction fails because the whole pattern blurs into a single patch. With increasing resolution, the overall error exponentially decreases and converges. The inflection point of the error function represents the best trade-off between resolution and accuracy. It depends on the particular arrangement of display,

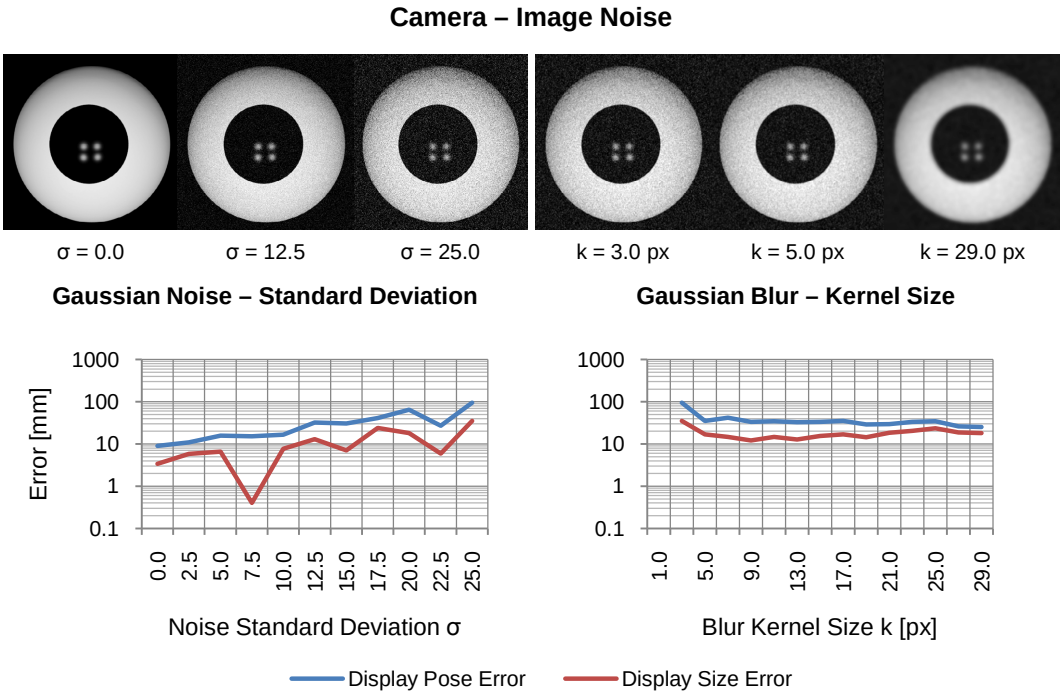


Figure 4.14: Result for experiment on synthetic data for image noise variation. To study the effect of noise we generate a series of images from the same source image by adding independent Gaussian noise with increasing standard deviation σ. Due to failing corneal reflection extraction at $\sigma > 5.0$, we apply a Gaussian blur to the raw images. (left, top) Rendered right eye with increasing standard deviation. (left, bottom) Reconstruction error without eye pose estimation, for minimal successful blur kernel size 3×3. (right, top) Rendered right eye with increasing blur kernel size at $\sigma = 25.0$. (right, bottom) Corresponding reconstruction error without eye pose estimation.

eyes, and camera, and lies at 640×480 for the current setup.

Image Noise. Image noise refers to a random variation of pixel intensity produced by the sensor and circuitry of the camera. It is a mixture of different types of noise, regarding source, characteristics, and probability distribution. The most common type is independent additive Gaussian noise (Gonzalez and Woods, 2007; Boncelet, 2005), that will be used as a placeholder to study the effects of an unknown mixture. Therefore, we render a single image with default parameters and generate a series of images by adding independent Gaussian noise with increasing standard deviation $\sigma \in [0.0, 25.0]$ at steps of 2.5.

Figure 4.14 shows the results. Since we did not observe any significant impact on iris contour fitting, we only show results without eye pose estimation. Because of failing corneal reflection extraction for $\sigma > 5.0$, we apply a Gaussian blur to the raw images before further processing. The complete data

series could be successfully processed with the minimum filter kernel size of 3×3. The error mean and variance increase exponentially with noise. The figure further shows the effect of increasing filter kernel size on the result obtained from the image having maximum noise with $\sigma = 25.0$, a relatively high value that can be regarded as an upper bound for common use cases. The error rapidly decreases at small kernel sizes, then reaches its minimum, stabilizes, and slightly increases at larger kernel sizes. Because kernel size has an impact on time performance, we note that small values between 5×5 and 9×9 are sufficient.

4.5.2.8 Discussion

With the described experimental series we analyzed the effect of several important parameters, in a controlled environment using synthetic data generated by physically based rendering. Anatomic parameters are difficult to examine using real data, because these require measurement with complex machinery and occur with uncontrolled variation, making it necessary to perform large-scale experiments for statistical evaluation. Imaging parameters are important as these especially affect results in low-quality hardware; and the proposed algorithm is intended for non-professional setups. Let us now summarize the findings.

If the cornea has spherical curvature, as assumed by the majority of eye models, the impact of its radius is relatively small. The real cornea, however, does not have a spherical curvature. Thus, we analyzed the effect of asphericity by varying only the radius in vertical or horizontal dimension which creates an ellipsoid shape. The impact is about one magnitude larger than the combined variation in all three dimensions. We further noticed a dependency on the arrangement of the applied eye positions. The average error can be significantly decreased when sampling non-degenerate eye positions with variation in both, x- and y-dimensions.

We simulated the effect of brighter iris colors by adding an increasing diffuse reflectivity to the specular cornea. As a result, the contrast for specular pattern reflections decreases. The mean reconstruction error remains relatively low, with a slight increase in variance. Note that we did not model iris texture which can have an additional effect on pattern extraction and becomes more evident in bright irides.

Increasing the size of the display markers may become necessary to achieve a measurable response in setups with diffuse environmental illumination, low quality image sensors, or short exposure at high framerates. On the other hand, this invalidates the point-light-source assumption and creates a larger marker reflection patch with distortion. More light is diffusely reflected from the iris, which has a larger effect for brighter irides and slightly increases the error variance as explained before. As a result, we did not find any significant impact for increasing marker size.

The negative effect of imaging parameters on the result is disproportionally larger under low specifications and rapidly becomes smaller with increasing quality. We observed convergence for the error with increasing image resolution. The inflection point of the error function marks the best trade-off between resolution and accuracy, and lies at VGA resolution for the current setup representing a common usage scenario. A linear increase in the standard deviation of the image noise causes an exponential increase in the error. To avoid failing image processing, a noise reduction strategy should be applied to the raw contaminated data. We found a simple Gaussian blur with minimal kernel size to be effective, even for the maximum amount of tested noise that lies above common scenarios. The error rapidly decreases with increasing kernel size, and achieves highest accuracy and convergence for kernel sizes between 5×5 and 9×9.

4.5.3 Multiple Eyes

Previous single-eye experiments employed a simple setup together with the basic algorithm, to verify the general feasibility of performing geometric display calibration from corneal reflections. While the results are promising, the error remains relatively large. Subsequent two-eyes experiments examined the impact of different parameters under controlled synthetic conditions. The results are not only important to display calibration, but may also help to understand scene reconstruction from eye reflections in general. Interpreting the findings, however, we predict a high error for common parameter variations and system configurations, suggesting the basic algorithm to be insufficient for accurate scene reconstruction.

Available geometric knowledge for the reconstructed scene enables an optimization strategy that achieves highly improved results. The following experiments are based on that algorithm assuming known display size. To simulate conditions of practice we apply more than two images, capturing eyes under general varying poses.

4.5.3.1 Setup

The setup comprises a 19-in display with 1280×1024 (5:4) resolution, 250-cd/m^2 brightness, 500:1 contrast ratio, and $170°/170°$ (H/V) viewing angles. We use a Point Grey Flea2 camera at 2448×2048 resolution mounted on a Fujinon HF35SA-1 lens with viewing angles $14°35'/10°58'$. Intrinsic camera parameters are calibrated using OpenCV functions. The camera is placed at about 30 cm above and behind the display. Test subjects are seated with their faces positioned about 50–60 cm in front of the display (Fig. 4.15(a)). The pattern uses a constant marker radius r of 0.25 in (6.35 mm) creating a pattern with diagonal size d_v of 18 in (457.2 mm) (Fig. 4.3). We capture face images of a test subject moving in front of the display. Regarding eye

orientation, we apply random measurements with a tilt angle $\tau < 20°$. As shown in experiment 4, calibration accuracy does not vary significantly within this range, however, decreases rapidly at larger angles.

4.5.3.2 Experiment 1: Test Subjects

Experimental verification was performed with 11 test subjects. Table 4.3 shows that there is no significant correlation between individual parameters such as age, body height, and pupillary distance. We acquired data sets of 10 face images per subject that are used for reconstruction (Fig. 4.16). The imaged irides have an average diameter of 160 pixels with the reflected screen occupying about 30×25 pixels (Fig. 4.17(a)–(d)).

Experimental results for the accuracy of reconstructed display poses are found in Table 4.4 and Figure 4.18. After optimization, the standard deviation between the center points of the estimated screen planes decreased considerably from 81.60 to 10.28 mm. It is important to note that no statistical significance could be observed between test subjects with normal eyes, near-sightedness uncorrected, and corrected with contact lenses (Tab. 4.5). This means that the method can be applied to any of these conditions despite them having an impact on corneal shape. Figures 4.15(b)–(d) show results where the reconstructed screen matches the real one given in Figure 4.15(a). The average and standard deviation of corneal sphere position adjustment in optimization are <0.01/0.67, 0.02/0.39, and 8.94/11.92 mm for x, y, and z-coordinates respectively. This shows that the error in corneal position estimation is largest along the depth direction.

4.5.3.3 Experiment 2: Display Pose

We mounted the display on a turntable, operated by a Chuo Seiki QT-CM2 stage controller, and took data sets of 10 face images of a single person at discrete display orientations of 0°, 10°, 20°, 30°, and 40° that act as ground truth (Fig. 4.19(a)). We further took data sets for a large and a small spherical mirror with 20 images per mirror. The large mirror with a radius of 25.4 mm is similar in size to the one used by Francken et al. (2007) and acts as ground truth. The small mirror with a radius of 7.9 mm is similar in size to the corneal sphere. This makes it possible to independently analyze the errors from small reflector size and unknown shape difference (asphericity). The eye pose estimation algorithm was adapted for the mirror. The imaged large and small mirrors have an average diameter of about 650 and 210 pixels with the reflected screen occupying about 95×75 and 30×25 pixels respectively (Fig. 4.17(e)–(h)).

Accuracy is estimated using two measures: The position error e_{P} describes the deviation in the center position of the display. The orientation error e_{O} describes the deviation in the normal direction of the display. Let $\mathbf{X}_{\mathrm{GT}}$ and

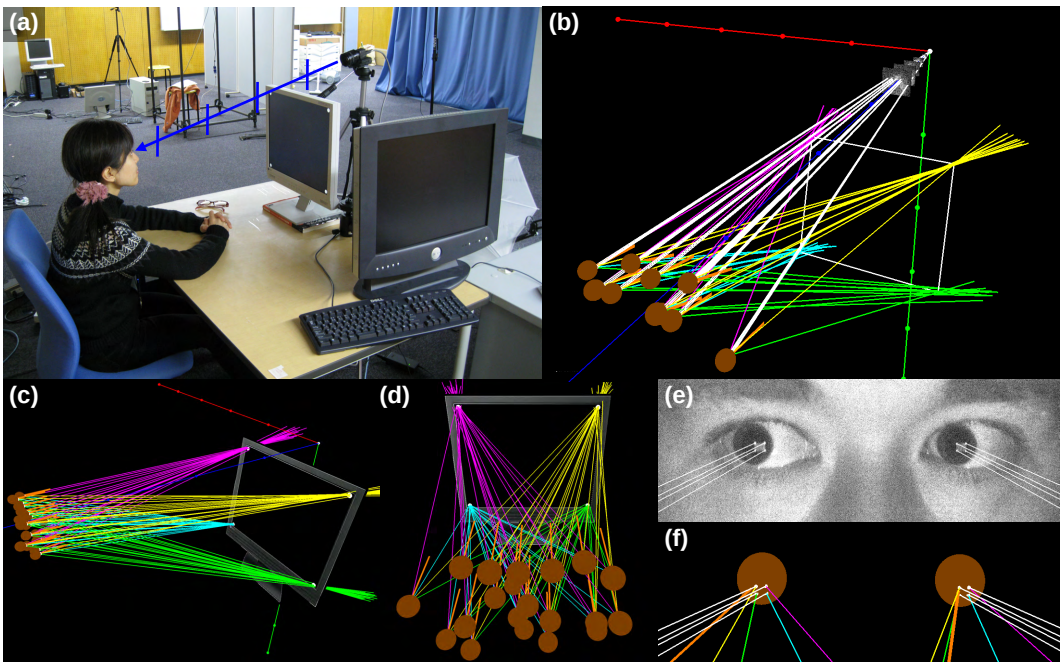


Figure 4.15: (a) An experimental setup. The camera is placed about 30 cm distance above and behind a 19-in display. Test subjects are seated about 50–60 cm in front of the display that shows a static pattern with $M = 4$ markers. The blue line indicates the camera z-axis, with marks at each multiple of 20 cm. (b) A display reconstructed from 10 eye poses of a single person, rendered from a similar view as shown in (a). The camera coordinate system is indicated as a white origin with three colored axes. Each mark along the axes indicates a multiple of 20 cm. The back-projected light rays (white) penetrate the rendered image planes and reflect at the corneal spheres (brown) towards the four markers on the screen (yellow, green, turquoise, pink). Corneal orientations (eye gaze) are indicated by orange lines. (c),(d) A display reconstructed from 20 eye poses with a display-eye distance d_{DE} of 65 cm (experiment 3). (e) An image of two eyes with pattern reflections and back-projected light rays. Intensity scaling is applied for better visibility. The effect of display contrast ratio (black level) can be noticed from the reflections. (f) The estimated corneal spheres with light rays and reflections. Note that the radius of the corneal sphere r_{C} is larger than the radius of the limbus r_{L} which marks the boundary of the visible part of the iris.

Table 4.3: Personal statistics of test subjects.

Gender	m	m	m	m	m	m	m	m	m	m	f
Age	**21**	**23**	**23**	**24**	**24**	**24**	**24**	**29**	**35**	**38**	**41**
Height	166	168	183	161	167	171	174	176	166	179	154
Pupillary Distance	6.3	6.5	6.4	6.5	5.8	6.6	6.6	5.9	6.7	7.1	6.3
Myopia		y	y	y		y	y		y		
Contact Lens				y							

Note: The datasets are ordered by age. There is no significant correlation between parameters. Myopia (near-sightedness) occurs in six subjects. Five of them wear glasses that were taken off for the experiments. A single subject wears contact lenses that were kept on.

Figure 4.16: Example images of moving subjects, acquired in the absence of environmental light. Intensity scaling is applied for better visibility.

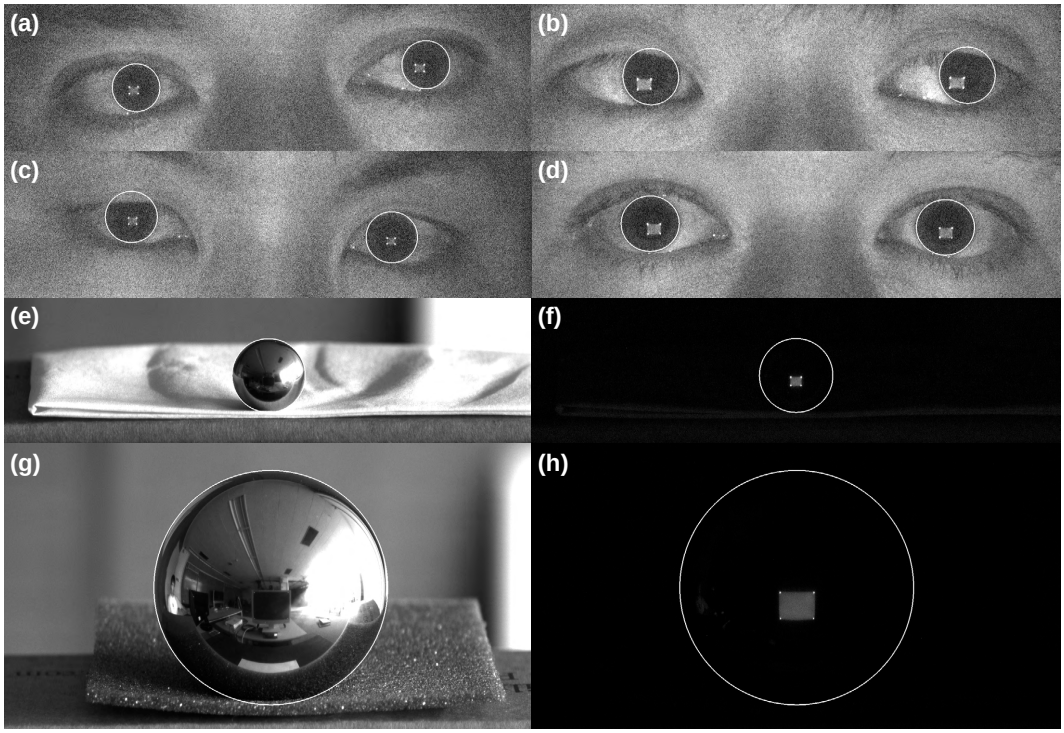


Figure 4.17: Iris and mirror contour fitting. (a)–(d) Cropped facial images with iris ellipse fitting results. Intensity scaling is applied for better visibility. (e)–(h) A pair of corresponding shape and reflection images with fitting results for each spherical mirror.

Table 4.4: Experimental results for multiple subjects experiment (1), comparing results before and after optimization.

	(a) Statistical Error of Reconstructed Display						(b) Residual Errors after Optimization					
	Top-Right [mm]	Top-Left [mm]	Bottom-Left [mm]	Bottom-Right [mm]	Center [mm]	Orientation [deg]	Intersection [mm]		Size [mm]		Plane [mm]	
	Stddev	Stddev	Stddev	Stddev	Stddev	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev
Pre-Opt	93.04	82.25	81.53	93.85	81.60	16.37						
Opt	**10.22**	**10.37**	**16.76**	**14.63**	**10.28**	**3.90**	**4.53**	**2.03**	**2.27**	**1.71**	**0.07**	**0.10**

(a) The statistical error for the reconstructed display is considerably lower after optimization. The position error is indicated as the standard deviation of the estimated marker positions and the calculated center points. The orientation error is calculated as the standard deviation of the angles between the plane normals.

(b) The residual errors after optimization. The intersection, screen-size, and plane error terms e_1, e_2, and e_3 are minimized well.

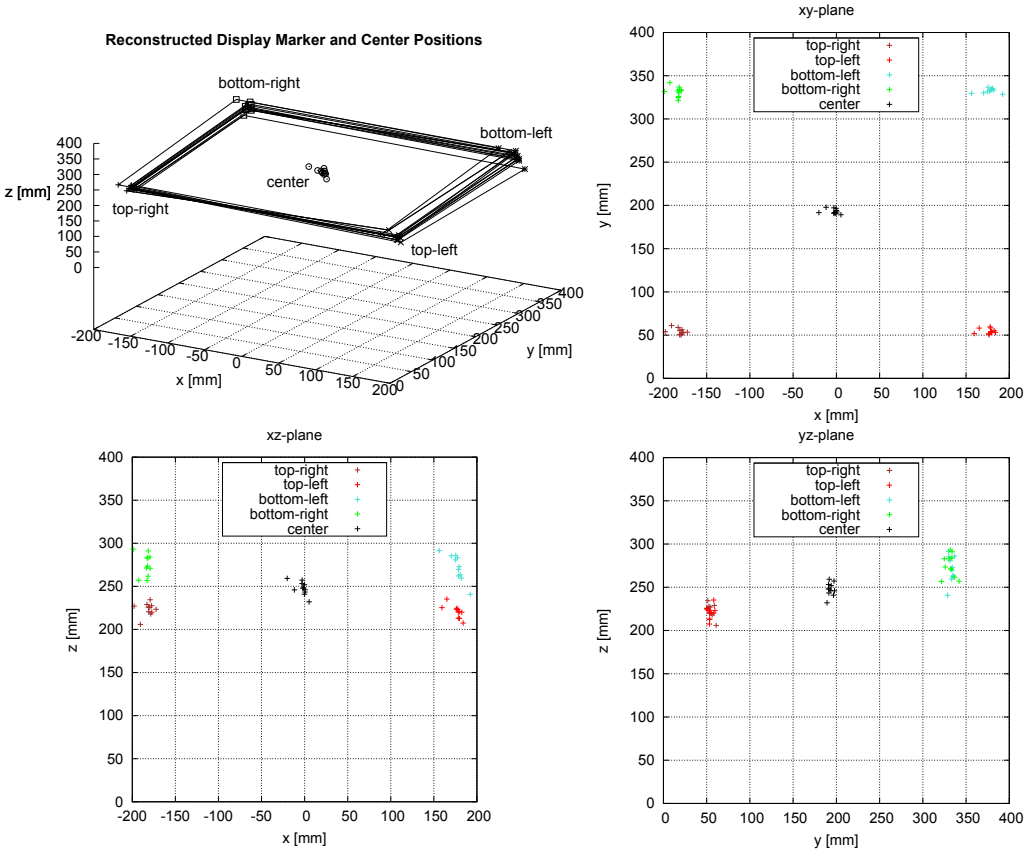


Figure 4.18: Experimental results for multiple subjects experiment (1), showing the distribution of estimated marker positions and display center points. For the 3D scatterplot (top, left), each black rectangle indicates the marker bounding box estimated from the set of face images obtained for a particular individual. There is no significant deviation between results from test subjects with normal eyes, near-sightedness uncorrected, and corrected with contact lenses.

Table 4.5: Statistical significance of eye condition in multiple subjects experiment (1).

	Top-Right			Top-Left			Bottom-Left			Bottom-Right			Center		
	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z
P-value	0.144	0.905	0.356	0.273	0.850	0.056	0.053	0.010	**0.006**	0.590	0.119	0.358	0.196	0.387	0.052
Bonferroni	1.727	10.863	4.268	3.275	10.197	0.678	0.631	0.116	**0.075**	7.083	1.423	4.298	0.588	1.160	0.155
Holm	1.007	0.905	1.778	1.638	1.699	0.508	0.526	0.106	**0.075**	1.771	0.948	1.433	0.392	0.387	0.155

Note: This table shows the results of multiple analysis-of-variance tests (ANOVA) carried out separately for each coordinate of the four estimated marker positions and the calculated center point. Each test is a single-factor ANOVA regarding eye condition, divided into the three levels: normal, uncorrected near-sighted, and corrected with contact lenses. The first row shows corresponding p-values with $p \in [0, 1]$. Lower values imply that the result has a higher statistical significance (less likely to be achieved randomly). However, in the present case there are multiple $\{p_i \,|\, i = 1, \ldots, 15\}$ from different comparisons that have to be adjusted in order to draw valid conclusions. The second and third rows show Bonferroni- and Holm-adjustment respectively, that take into account the result of all the coordinates simultaneously. Measurements from the calculated center point are treated separately. Truncation to $[0, 1]$ is omitted. The smallest p_i-value is marked bold. None of the adjusted p_i are statistically significant at $\alpha = 0.05$.

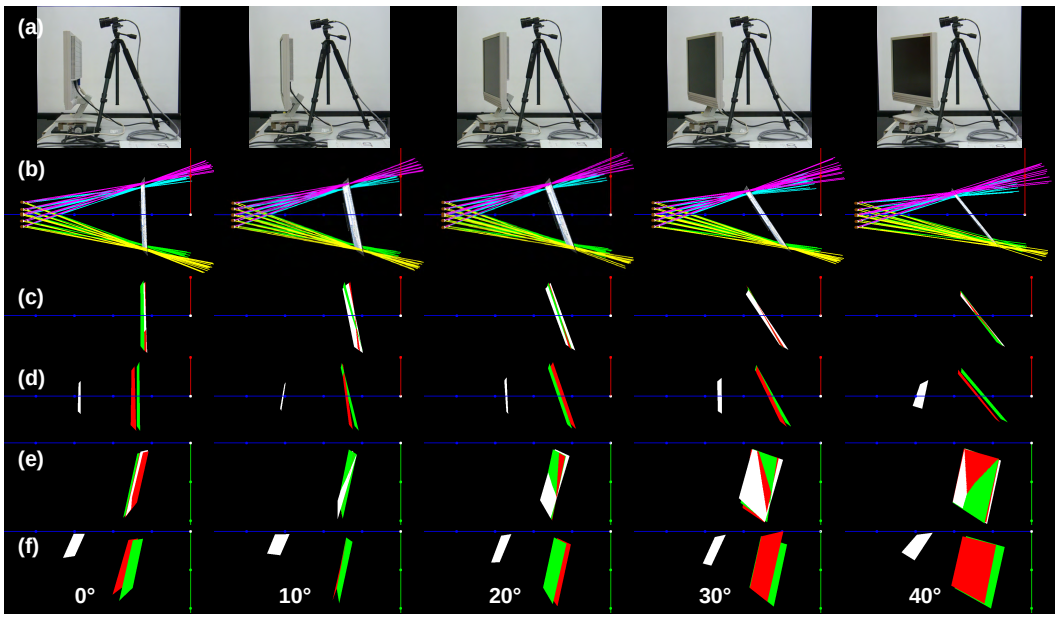


Figure 4.19: Experimental results for display orientation experiment (2). (a) Experimental setup with varying display pose. Each column corresponds to a particular display orientation (from left to right: 0°, 10°, 20°, 30°, and 40°). (b) The displays are reconstructed from eye reflections after optimization, rendered by looking down onto the xz-plane (red, blue) of the camera frame. Each mark along the axes indicates a multiple of 20 cm. (c) Comparison between the three displays reconstructed from eyes (white), large (green), and small spherical mirrors (red) after optimization. All three planes are estimated with a low position and orientation error. (d) Same as (c), but before optimization. The screen planes from mirror reflections are estimated slightly smaller and rotated, in front of the actual position, having a screen size error e_2 of about 25 mm. In contrast, the screen plane from eye reflections is estimated incorrectly with an average of 319.4 mm in front of the one obtained as ground truth using the large mirror, having an error e_2 of about 200 mm. (e)–(f) Same as (c)–(d), rendered similar to (a) by looking from the side onto the yz-plane (green, blue).

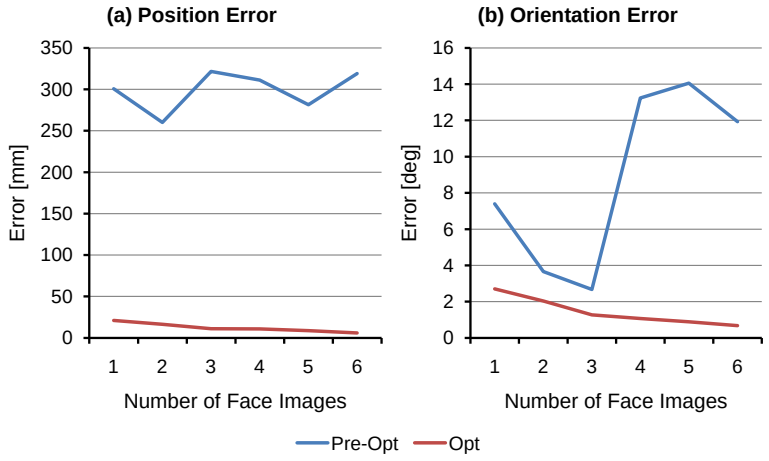


Figure 4.20: Experimental results for display orientation experiment (2) at 0°. The (a) position and (b) orientation error describe the deviation from the ground-truth estimate obtained using the large mirror. An increasing number of face images in calibration leads to a decreasing error and convergence towards the final result.

$\mathbf{n}_{\text{GT}}$ denote the ground truth for the display center point and the normal direction, and let $\mathbf{X}_{\text{eye}}$ and $\mathbf{n}_{\text{eye}}$ denote the estimates obtained using eyes, then the error is defined as in

$$e_{\text{P}} = \|\mathbf{X}_{\text{eye}} - \mathbf{X}_{\text{GT}}\|, \tag{4.14}$$

$$e_{\text{O}} = \cos^{-1}(\mathbf{n}_{\text{eye}} \cdot \mathbf{n}_{\text{GT}}). \tag{4.15}$$

Experimental results for the accuracy of reconstructed display poses are found in Table 4.6 and Figure 4.21. After optimization, the result from the eyes correctly matches the actual display pose and is only slightly worse than the result from the small mirror. It outperforms the results obtained from both spherical mirrors before optimization, which coincides with the results for the method in Francken et al. (2007), shown in Francken et al. (2009). Figure 4.19 offers a detailed visual comparison. The strategy further achieves a decreasing error and convergence with increasing number of face images used for calibration (Fig. 4.20).

4.5.3.4 Experiment 3: Display-Eye Distance

In this experiment we evaluate an increasing distance between display-camera setup and eye positions. Calibration is performed at six discrete display-eye distances d_{DE} of 35, 65, 95, 125, 155, and 185 cm. The interval and step size are chosen with some considerations: A minimum distance of 35 cm is necessary to capture both eyes within a single face image and to allow some head movement. A distance of 65 cm is approximately the one used in experiments 1 and 2 and, thus, enables for comparison. A maximum distance of 185 cm is rather uncommon for a 19-in display, but is tested in order

Table 4.6: Experimental results for display orientation experiment (2), comparing the accuracy obtained from eyes, small, and large spherical mirror ($r = 7.93$ and $r = 25.4$ mm) before and after optimization.

		(a) Error to GT		(b) Error to Mirror L				(c) Residual Errors after Optimization					
		Orientation [deg]		Position [mm]		Orientation [deg]		Intersection [mm]		Size [mm]		Plane [mm]	
		Avg	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev
Pre-Opt	Eye	17.33	10.02	319.44	27.11	11.87	8.32						
	Mirror S	1.90	1.21	21.70	7.41	1.83	1.53						
	Mirror L	1.34	1.28	0.00	0.00	0.00	0.00						
Opt	**Eye**	**0.82**	**0.37**	**11.18**	**4.42**	**0.69**	**0.45**	**7.76**	**0.28**	**0.37**	**0.37**	**0.02**	**0.02**
	Mirror S	0.49	0.30	9.95	4.59	0.63	0.39	3.99	0.70	0.09	0.15	0.01	0.01
	Mirror L	0.31	0.10	0.00	0.00	0.00	0.00	1.43	0.26	0.03	0.03	0.01	0.00

(a) Orientation error to the turntable ground truth. The orientation is computed as the difference in the normal direction of the display obtained from the eyes to the one from the large mirror at 0°. After optimization, the error for the eye calibration becomes very similar to the errors obtained for the mirrors.

(b) Position and orientation error to the estimate obtained from the large mirror at the same turntable orientation.

(c) Residual errors after optimization. The intersection error e_1 increases for small mirror and eyes as deviation in model parameters has a higher impact. It further increases for the eyes as the actual shape deviates from that of a sphere. The screen size and plane error e_2 and e_3 can be minimized well.

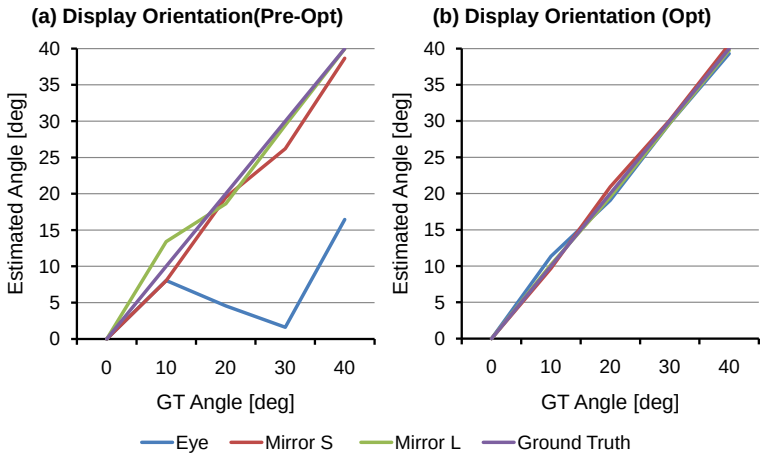


Figure 4.21: Experimental results for display orientation experiment (2). Comparison of orientation angles estimated from eyes and spherical mirrors, (a) before and (b) after optimization, with the turntable ground truth.

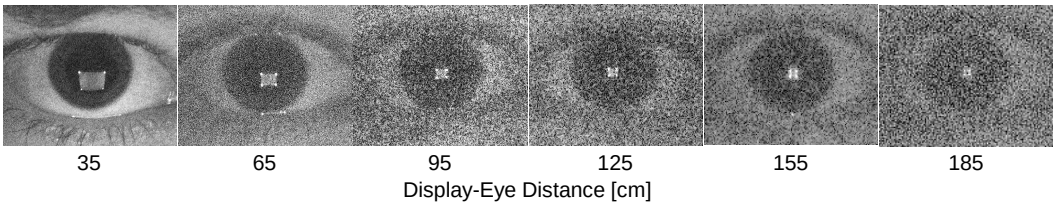


Figure 4.22: Image data for the display-eye distance experiment (3). The sequence shows the test subject's right eye at increasing distance from display-camera setup. Intensity scaling is applied for better visibility. Due to low illumination and small display size, larger distances result in difficult conditions and image data.

to evaluate the theoretical limitation. It also marks the upper bound for obtaining usable data for this setup, since it results in very difficult conditions and image data, as can be seen in Figure 4.22. Due to the inverse-square law for light intensity, the marker radius needed to be increased to $r = 0.5$ in at 155 and 185 cm. At larger distances, it was not possible to extract independent centroids for each marker since the distance between their reflection patches becomes too small. Also, it was not possible to perform ellipse fitting to detect the iris boundary due to the large amount of noise.

A particular calibration result was obtained from a data set of 10 face images capturing both eyes of a single person. For statistical evaluation, we took 10 independent trials at each distance (600 images in total). Calibration was also performed for the large spherical mirror using data sets of 20 images. The optimized mirror result at 35 cm has been found to be most accurate and is used as ground truth. The camera is fixed above and approximately 10 cm behind the display, where camera parameters are calibrated separately at each eye distance due to necessary focus adjustment.

Experimental results obtained before and after optimization are shown in Table 4.7 and Figure 4.23. Along the whole range, optimization can effectively reduce the overall error that increases with distance. Results show relatively good accuracy for the near range. At 95 cm, the average position error amounts to 17.72 mm, and the average orientation error to 7.04°. At larger distances, we observe a sharp increase, especially for the position error. Depending on the intended use, the accuracy should be sufficient for distances up to 1 m. However, regarding the size of the display, real usage scenarios will probably not involve distances larger than 65 cm.

4.5.3.5 Experiment 4: Gaze Angle

The following experiment studies the influence of eye orientation. For four test subjects, calibration was performed at six discrete limbus tilt angles τ of 0°, 6°, 12°, 18°, 24°, and 30°. Each corresponding data set comprises eight face images capturing both eyes, distributed evenly along the whole range of eye rotation angles ϕ with steps at 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°.

Table 4.7: Experimental results for display-eye distance experiment (3), comparing the accuracy obtained from eyes before and after optimization.

		(a) Error to GT				(b) Residual Errors after Optimization					
	Display-Eye	Position [mm]		Orientation [deg]		Intersection [mm]		Size [mm]		Plane [mm]	
	Distance [mm]	Avg	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev
Pre-Opt	350	204.74	7.39	29.04	3.54						
	650	401.80	20.71	18.48	6.15						
	950	576.97	60.17	15.24	4.36						
	1250	759.58	76.25	11.92	4.95						
	1550	1156.03	77.58	27.82	12.93						
	1850	1361.72	94.43	34.26	16.02						
Opt	**350**	**1.39**	**0.54**	**0.38**	**0.25**	**4.33**	**0.33**	**0.09**	**0.15**	**0.00**	**0.00**
	650	**6.79**	**2.38**	**2.16**	**1.19**	**9.27**	**1.30**	**0.25**	**0.30**	**0.01**	**0.01**
	950	**17.72**	**4.23**	**7.04**	**3.11**	**12.35**	**1.71**	**0.98**	**0.89**	**0.02**	**0.01**
	1250	**91.47**	**29.70**	**15.85**	**5.50**	**19.18**	**3.50**	**3.81**	**3.50**	**0.02**	**0.02**
	1550	**227.59**	**79.19**	**30.33**	**7.30**	**24.45**	**4.82**	**4.75**	**5.33**	**0.02**	**0.02**
	1850	**491.34**	**66.40**	**42.28**	**5.09**	**24.99**	**6.95**	**11.19**	**5.80**	**0.03**	**0.02**

Note: Errors are computed using the estimate from the large mirror at display-eye distance d_{DE} of 35 cm as ground truth.

(a) Position error in the center position of the display.

(b) Orientation error in the normal direction of the display.

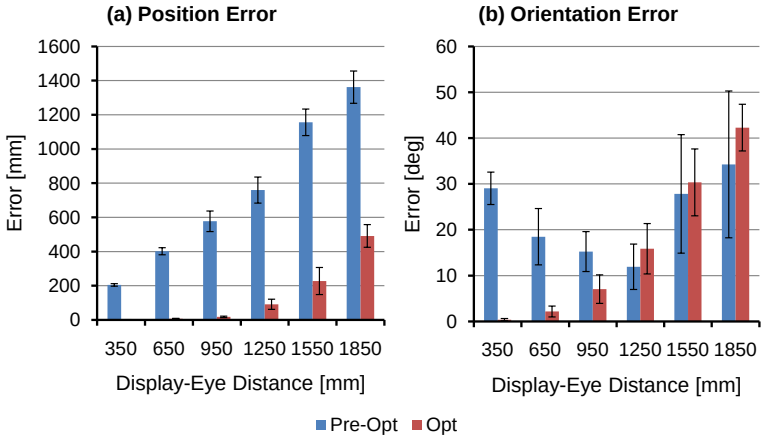


Figure 4.23: Experimental results for display-eye distance experiment (3). Plot of the (a) position and (b) orientation error in Table 4.7. Columns represent the average error, black bars the standard deviation. The error increases with distance. Usable results are achieved for display-eye distances up to 100 cm (camera-eye distance of 110 cm), from where the error increases rapidly.

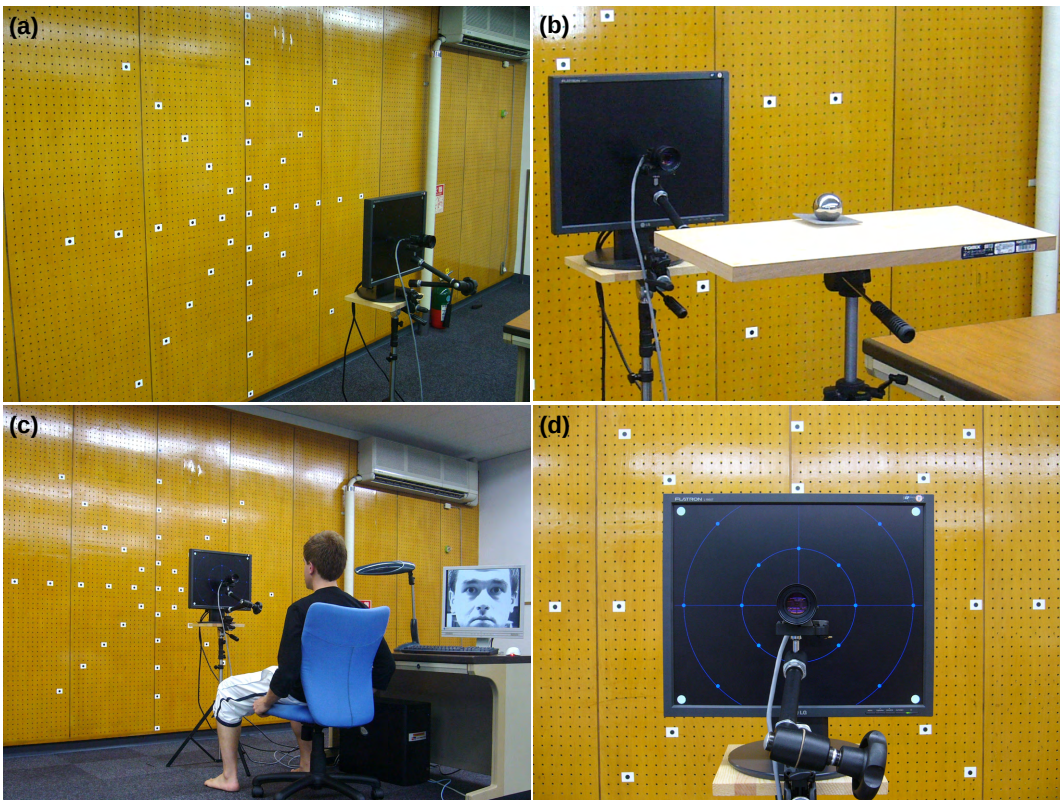


Figure 4.24: Setup for the gaze-angle experiment (4). (a) Gaze markers are attached to the wall. The radial direction represents increasing limbus tilt angles τ of 0°, 6°, 12°, 18°, 24°, and 30°. The polar steps are evenly distributed according to eye rotation angles ϕ of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°. The camera is fixed in front of the display center with its optical axis perpendicular to display plane and wall. (b) The ground truth is obtained from the large spherical mirror at display-eye distance. (c) A test subject in front of the setup. (d) Gaze markers, occluded from the perspective of the test subjects, are rendered on the display.

Figure 4.24 explains the experimental setup. Refer to Figure 2.8 for details about the representation of eye orientation. Gaze markers are attached to the wall according to the distribution of gaze directions for test subjects who are located at a distance of 170 cm. The display is mounted on a tripod. The camera is fixed with an adjustable arm at 10 cm in front of the display center and 100 cm in front of the center marker for $(\phi, \tau) = (0°, 0°)$ with its optical axis perpendicular to display plane and wall. The ground-truth display pose is obtained from 15 images of the large spherical mirror, distributed evenly along the camera field of view at a distance of approximately 80 cm in front of the display. Under the same conditions, data sets are acquired for moving test subjects gazing at the respective markers. Gaze markers for angles τ of 0°, 6°, and 12°, occluded from the perspective of the test subjects, are rendered with dark blue color in order to not affect the measurements.

The setup is designed with some considerations: The camera is placed at a rather uncommon location in front of the display center. With that configuration, there is no bias regarding angle ϕ since each gaze marker, corresponding to the same angle τ, and each of the four display calibration markers have equal distance to the camera. The display-camera setup is designed to minimize occlusion of the gaze markers. The maximum tilt angle τ of 30° marks the upper bound for the marker reflections to be located on the corneal surface. The subject-wall distance is maximized in order to reduce the impact of slight head movement on the gaze direction; where 170 cm is the largest distance that allows attaching the center marker at a comfortable sitting height.

Figure 4.25 shows captured eye images for the whole set of gaze directions. The iris detection produces applicable results. For an increasing tilt angle τ the angular error in gaze direction amounts to 3.62°, 3.96°, 3.79°, 3.36°, 3.01°, and 4.94° (RMSE). The reflected marker pattern moves towards the corneal boundary. For one subject, several markers at 30° were already out of the corneal boundary and have not been used for calibration. Distortion increases for the whole pattern as well as for each reflected marker patch. Both effects vary according to individual differences in corneal shape.

Experimental results obtained before and after optimization are shown in Table 4.8 and Figure 4.26. Optimization can effectively reduce the overall error along the whole range of tilt angles τ. Estimation error is lowest for 6°–18°, slightly larger for 24°, and large for 30°, which is a result of pattern distortions and deviations in corneal curvature and shape. Interestingly, the error at 0° goes up slightly.

Consider the common case where the user is located at a distance of 60 cm in front of the center of a 19-in display. The camera is placed either above or below the display. For iris detection, the maximum gaze angle when looking at the display amounts to approximately 17°/27° (H/V). For display calibration, the maximum angular deviation for the marker reflections from the corneal apex amounts to approximately 17°/14°. The values are within the applicable limits of eye and display pose estimation.

4.5.3.6 Discussion

Previous single-eye experiments employed a simple, unnatural setup together with the basic algorithm, to verify the general feasibility of performing geometric display calibration from corneal reflections. Results, however, show a relatively large error. To gain detailed understanding, subsequent two-eyes experiments applied synthetic data to simulate the impact of parameters that are difficult to control and measure in practice. Results show a high impact of individual eye geometry where common parameter variation leads to large errors. This suggests that the basic algorithm in combination with a static spherical eye model is not sufficient for accurate calibration.

To improve accuracy while relying on the simple geometric eye model, an

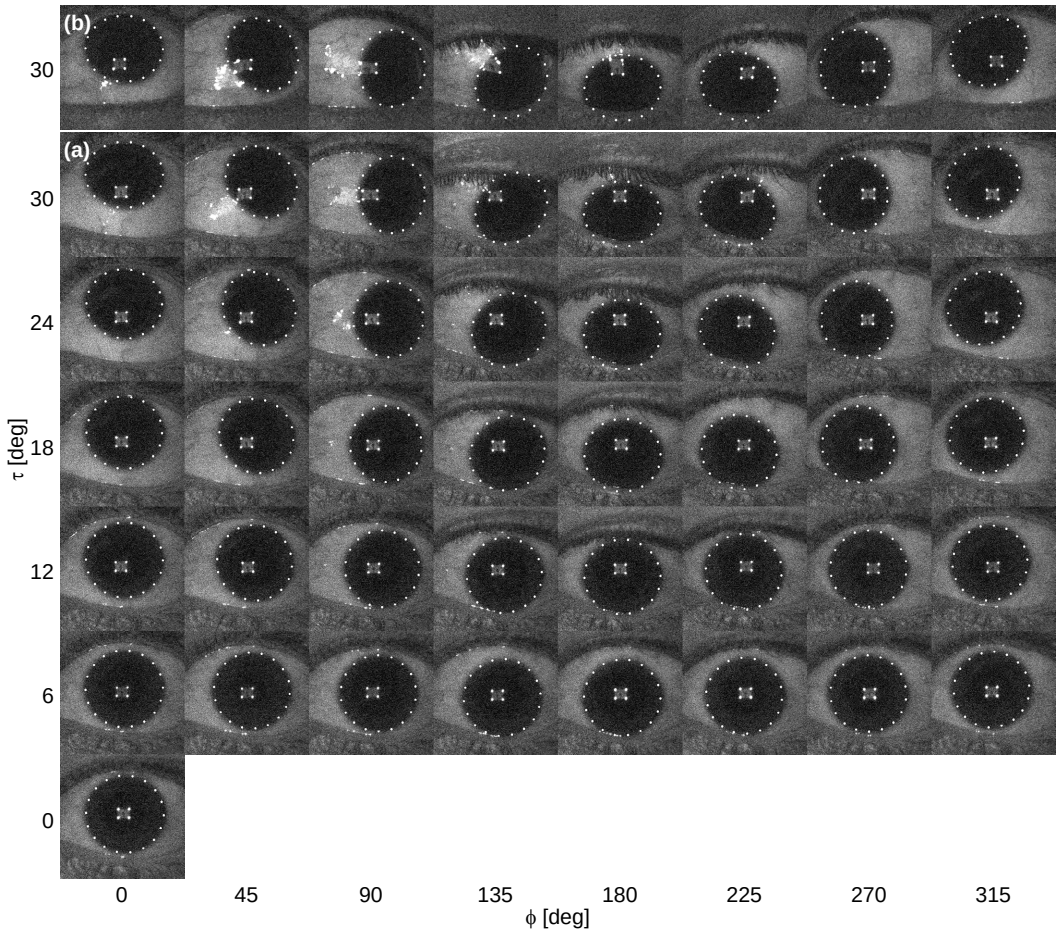


Figure 4.25: Image data for the gaze-angle experiment (4). (a) Imaged irides of the left eye of a single test subject for the set of eye orientation angles ϕ and limbus tilt angles τ. Intensity scaling is applied for better visibility. The head was fixed, with the midpoint of the eyes baseline centered in the camera image. To not only show iris deformation but also translation of its imaged position, each image patch is centered at the invisible back-projected center of the estimated corneal sphere $\mathbf{C}$. With increasing angle τ, the reflected marker pattern moves towards the corneal boundary. Distortion increases for the whole pattern as well as for each reflected marker patch. This leads to a rapidly increasing display pose estimation error for reflections near the corneal boundary. Above 30°, some of the markers move out of the corneal boundary and cannot be detected. (b) Imaged irides of the left eye of another test subject at 30°. Comparison shows larger distortion and missing reflections, which relates to individual differences in corneal shape.

Table 4.8: Experimental results for gaze-angle experiment (4), comparing the accuracy obtained from eyes before and after optimization.

		(a) Error to GT				(b) Residual Errors after Optimization					
	Gaze Angle τ	Position [mm]		Orientation [deg]		Intersection [mm]		Size [mm]		Plane [mm]	
	[deg]	Avg	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev	Avg	Stddev
Pre-Opt	0	517.49	129.31	5.11	4.90						
	6	628.25	51.17	8.15	2.64						
	12	641.89	36.79	5.96	4.78						
	18	646.57	4.18	5.48	2.74						
	24	644.88	23.15	12.24	3.06						
	30	705.37	50.02	24.02	11.08						
Opt	**0**	**24.91**	**16.56**	**3.78**	**2.32**	**11.13**	**1.73**	**0.22**	**0.22**	**0.01**	**0.01**
	6	**4.95**	**4.84**	**2.86**	**1.28**	**10.57**	**1.41**	**0.18**	**0.26**	**0.01**	**0.00**
	12	**7.57**	**4.55**	**1.88**	**0.63**	**10.44**	**0.54**	**0.06**	**0.08**	**0.00**	**0.00**
	18	**15.58**	**9.45**	**2.11**	**1.36**	**11.99**	**1.04**	**0.08**	**0.07**	**0.01**	**0.01**
	24	**30.61**	**10.93**	**3.71**	**2.69**	**15.76**	**4.11**	**0.03**	**0.03**	**0.02**	**0.02**
	30	**108.28**	**38.65**	**12.11**	**8.78**	**21.96**	**6.72**	**1.90**	**0.91**	**0.03**	**0.02**

Note: Errors are computed using the estimate from the large mirror as ground truth.

(a) Position error in the center position of the display.

(b) Orientation error in the normal direction of the display.

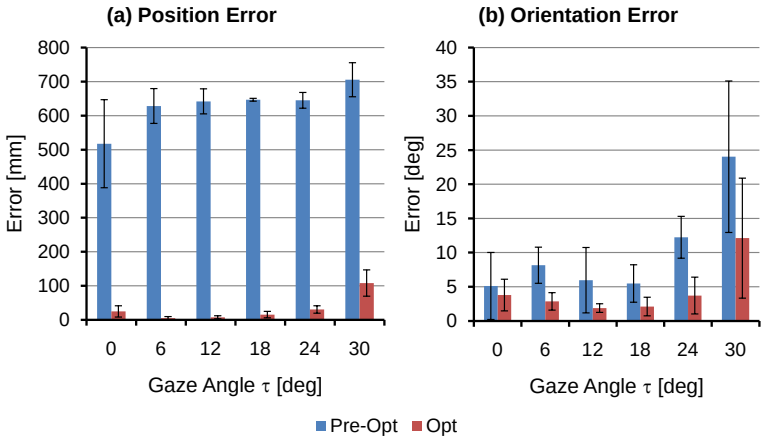


Figure 4.26: Experimental results for gaze-angle experiment (4). Plot of the (a) position and (b) orientation error shown in Table 4.8. The error is lowest for 6°–18° and slightly larger for 0° and 24°. It rapidly increases with larger angles. For the common case, where a user is looking at a 19-in display from a distance of 60 cm, the maximal gaze angle and deviation of the reflection from the corneal apex are within the applicable limits of eye and display pose estimation.

optimization strategy is proposed that exploits knowledge of scene geometry to adjust eye positions and scene structure. The described multiple-eyes experiments applied more than two eye images per calibration to examine the performance of this strategy. Different studies were designed to analyze the impact of individual factors, display pose, eye position, and gaze direction, representing common influences under conditions of practice. Let us now summarize the findings.

In experiment 1, optimization largely decreases the standard deviation of results among different subjects and eliminates the statistical significance of eye condition (normal, uncorrected near-sighted, corrected with contact lenses). This shows that the influence of individual factors can be successfully reduced. For experiments 2–4, ground truth was acquired from two spherical mirrors of different size to distinguish the impact of size from the one of asphericity. In accordance with the results obtained from synthetic data, asphericity is found to have a high impact, which can, however, be successfully compensated by optimization. The resulting error remains stable under display orientation and low gaze angles, and increases with gaze angle and display-eye distance. Nevertheless, the applicable range is found to exceed common use cases.

An important observation is that using optimization, accuracy increases with the number of images until finally achieving convergence. Furthermore, the strategy also gains improvement in spherical mirror based methods. The results and findings of this study are not only important for display-camera calibration, but also for the understanding of eye-scene relation in general, and the optimization potential of introducing geometry constraints.

4.6 Conclusion

4.6.1 Discussion

This chapter proposed and verified the idea for geometric calibration of display-camera setups from corneal reflections in eye images. A four-step approach was described that involves (1) detecting the eye pose from an image of the eye, (2) computing the light source position from corneal reflections under different eye poses, (3) estimating the pose of the display from multiple light sources represented by markers on a screen pattern, and (4) optimizing the result by refining eye poses and reflections rays. Compared to previous approaches, the concepts and findings achieved with this work remove the requirement for special hardware, explicit user interaction, and awareness, and allow for online execution.

In experimental evaluation, a framework was described for the physically correct rendering of synthetic eye images. The framework was applied to analyze the impact of system parameters for which ground-truth measure-

ments are difficult to obtain. Using real and synthetic data, a large number of comprehensive experimental studies was performed, showing that a straightforward combination of eye pose estimation (Nishino and Nayar, 2006) and screen-camera calibration (Francken et al., 2007) leads to a large error, which relates to the unknown geometry and size of the individual eye. Findings are as follows:

- The results show a significant influence of individual eye geometry.
- The impact of increasing corneal asphericity is about one magnitude larger than an equal variation in the radius of a spherical cornea. The error can be reduced by sampling a large spatial distribution of cornea positions parallel to the image plane.
- No significant impact is observed from parameters related to the appearance of specular highlights on the cornea, such as the size and intensity of light sources and the diffuse reflectivity of the iris.
- Regarding camera specifications, low resolution has a disproportional large effect. Accuracy is, however, rapidly increasing and converging at standard resolutions of current video hardware, that is, therefore, sufficient for common display-camera arrangements. A linear increase in image noise results in a larger non-linear decrease in accuracy. The effect is successfully compensated by applying standard noise reduction techniques.

It has further been shown that results can be considerably improved by introducing an optimization framework that jointly refines eye poses, reflection rays, and display pose subject to known geometry constraints (ray intersection distance, screen size, screen planarity). Particular achievements from this strategy are:

- Compared to a simple geometric reconstruction, the error can be considerably decreased.
- An increased tolerance to noise allows using commodity hardware in conjunction with a simple geometric eye model that does not require individual calibration.
- A former ambiguity in recovering the full range of eye orientations is automatically resolved.
- The strategy also achieves improvements in previous spherical-mirror-based methods.

Results from a large number of comprehensive experimental studies demonstrate the effectiveness of the approach where stable results are obtained under varying conditions. The findings could also be helpful to geometric reconstruction from eye reflections in general. Important conclusions are the following:

- Despite individual differences in eye shape, the deviation in results from different persons can be considerably decreased. Varying eye conditions (normal, uncorrected near-sighted, corrected with contact lenses) do not show any statistical significance.
- The absolute error to the ground truth can be considerably decreased. Convergence is achieved with an increasing number of images.
- The error increases with distance from the display-camera setup. Nevertheless, the applicable range for this method has been found to exceed common use cases.
- The error remains stable with increasing display orientation and gaze angle for the applicable range of this method which also covers common use cases. Larger angles lead to distortions for reflections in the corneal periphery and should therefore be avoided.

4.6.2 Implications

With the proposed method, we established and verified the integration of eye reflection analysis with display-camera systems. Despite the difficult working conditions, the results are good and should be sufficient for many applications. We believe that this work has the potential to facilitate novel developments in the community and helps to generally increase usability and acceptance of applications "outside the laboratory". The unique characteristics of the method enable applications in novel scenarios and system configurations. An overview of potential implications is given in the following.

Calibration-free Applications Since calibration is achieved implicitly without requiring interaction and awareness, the method can be applied where a dedicated calibration procedure is not possible. This could be for any of several possible reasons: A lack of time, when attention is required for other tasks such as re-arranging the display in driver-assistance systems or at the workplace. Second, a lack of ability, when working with non-experts, physically/mentally disabled people, or children and infants (Guestrin and Eizenman, 2008; Franchak et al., 2010; Gredebäck et al., 2010; Noris et al., 2010). Other reasons could include hiding technical details of the system or seamlessly integrating with art decors.

Dynamic Setups The calibration does not require interaction and may be performed online. This allows applications where the relation between display and camera is changing. Examples for changing camera pose include hand-held video cameras, and PTZ cameras in surveillance and vision-based interfaces. Examples for changing display pose include

hand-held/mobile devices, and projection displays, such as head-up displays in cars (HUD) or special displays in augmented reality (Bimber et al., 2005).

The proposed method recovers the geometric relation between display, camera, and eyes. This can be beneficial for applications in different fields. An overview of potential implications is given in the following.

Human–Computer Interaction The method enables improved calibration-free remote eye gaze tracking where the PoR is obtained by relating gaze direction and display plane. Screen-based eye gaze tracking has many applications in different fields (Duchowski, 2007; Hammoud, 2008; Hansen and Ji, 2010). However, there is no restriction to planar screens: A 3D PoR on an arbitrary surface is obtained when the estimated eye pose is related with a model of the environment or an image-based environment map, explained more in detail in Chapter 5. Eye gaze tracking may be further combined with eye reflection analysis for scene reconstruction or eye pose refinement.

Surveillance and Security It has been shown that display content can be recovered from reflections in the eyes of a person in front of a PC, imaged from far-away locations (Backes et al., 2008, 2009). The quality of the result may be improved by undistortion, which requires knowledge about display pose, eye pose and shape. Furthermore, it is possible to extend this to real-time monitoring of the interaction with mobile devices. Knowledge of eye gaze and display reflections may also be beneficial for technical improvement and to introduce context information in iris-based biometric systems (Daugman, 2004; Bowyer et al., 2008).

Photometric Stereo There exist several works using display-camera systems for scene reconstruction by photometric stereo (Woodham, 1980). Knowledge about the distribution of environmental light sources is important and can be recovered from an image of the eye (Nishino and Nayar, 2006; Tsumura et al., 2003; Johnson and Farid, 2007). The proposed method lays the foundation to exploit this information in the context of display-camera setups, leading to performance improvement in calibration and application.

Medicine Analyzing the relation between display content and eye poses can help to diagnose patterns related to physical and mental degrading of the visual and motor system. After having detected a particular condition, corrective actions may be provided through adaption of the displayed content. Moreover, such information could also be used in order to detect and support correct 3D perception with auto-stereoscopic displays (Hoffman et al., 2008; Lambooij et al., 2009).

4.6.3 Limitations

The scope of the described implementation is to provide an in-depth analysis of the applicability of eye reflections for display-camera calibration. There are limitations when using the implementation in its current form within real conditions. Details for a fully automatic calibration procedure largely depend on the requirements of each particular setup. Necessary extensions include

- a strategy for calibrating camera parameters, e.g., directly from eye images (Johnson and Farid, 2007),
- a technique for tracking a first guess for the eye region in a video (Hansen and Ji, 2010),
- a scheme for discarding unusable frames that do not include an eye, have too low quality, or relate to configurations known to result in decreased accuracy,
- a pattern architecture that increases information throughput and allows robust reflection extraction in the presence of environmental light and varying iris colors (Wang et al., 2008), e.g., using coded markers,
- a more accurate geometric model for the surface of the cornea, using an aspheric model based on anthropometric statistics or on parameterizing an individual shape by exploiting display reflections of more complex patterns,
- an extension to suppress complex light interaction at different layers of the eye and to handle more complicated light paths (Kutulakos and Steger, 2008), e.g., occurring when users wear glasses, and
- a strategy for calibrating display and camera photometric properties, e.g., by analyzing eye reflections.

4.6.4 Future Work

Beside the described limitations that need to be tackled to turn the current prototype into a practical system, there is requirement for future research. In the following, a survey is given about possible display-camera correspondence coding strategies.

4.6.4.1 Correspondence Coding Strategy

In order to obtain the M required pairs of pixels in 2D screen coordinates and points in 3D camera coordinates, the display shows a pattern representing the screen locations. The corresponding corneal reflections are identified from eye images of a person facing the screen. There exist different strategies to

represent the locations by display patterns; each one having advantages and disadvantages. The choice is a trade-off depending on the requirements of the particular application scenario.

Direct representation. This is the simplest strategy and the one chosen for the current implementation. The display pattern contains white circular markers centered at particular pixel locations on black background. Corresponding specular reflections are detected from an eye image. The spatial arrangement of the markers is not affected by projection and reflection since the cornea has a convex shape. While this method is simple, it lacks accuracy when marker regions become larger, and does not handle reflections from other light sources. To account for this, screen locations can be represented with more detailed structures, e.g., intersections of lines or corners in a checkerboard pattern. However, such an approach requires sufficient resolution. Direct methods suffer from an inherent ambiguity where the index k of a reconstructed marker is unknown. This can be resolved when the reflected display is approximately aligned with the camera image plane (by initially adjusting camera orientation around its optical axis).

Coded representation. To automatically resolve the ambiguity, further information needs to be added to the pattern. A straightforward way is to use coded structured light techniques, which assign a unique codeword to each pixel or region of pixels in a larger non-periodic area of the pattern. An overview of different methods, comparing characteristics and technical details, is found in Battle et al. (1998); Salvi et al. (2004); for more recent extensions see Salvi et al. (2010). Coded structured light can also help to generally increase robustness under challenging conditions such as high spatial resolution, image noise, low image resolution, environmental light, and superimposed iris features. According to Salvi et al. (2004), three main strategies can be distinguished: direct coding, space-multiplex coding, and time-multiplex coding.

Direct coding simply represents each location by its unique intensity or color value. Care has to be taken that the whole intensity range can be detected from an image and that the inter-value distance is sufficient for discrimination, especially in the case of superimposed iris features. Thus, color coding is best applied when environmental light and spacial resolution in the display are low, and image quality is high.

Space-multiplex coding represents the codeword for each location by the unique intensity or color variation in its neighborhood. An advantage to direct coding is the lower number of required intensity values, allowing for an increased inter-value distance. This, however, is obtained at the cost of robustness since each codeword depends on information in the

neighborhood and, thus, a single error affects multiple locations. Space-multiplex coding can cope with several amount of diffuse environmental light, but also limits spatial resolution and requires high quality images.

Time-multiplex coding distributes the codeword for each location into a sequence of patterns along the temporal domain. The approach does not suffer from most of the drawbacks of direct and space-multiplex coding. It achieves the highest inter-value distance, and accounts for environmental light, superimposed iris features, and low quality images while allowing high spatial resolution. The major drawback is that the method uses multiple frames, limiting its application to static scenes. Since the eyes are moving, time-multiplex coding requires to capture at high framerates and to compensate for eye movements.

Extensions. The presence of corneal reflections from environmental light can make it difficult to distinguish these from screen reflections. In such a case, reflection features from all sources can first be matched among multiple images and reconstructed. Additional geometry constraints are then used to remove the outliers. For example, a simple distance thresholding would be effective for the common use case where the display is the nearest light source to the eyes.

In order to increase robustness to ambient light, a bright uniformly colored fullscreen pattern may be used (Francken et al., 2007). After detecting the reflected screen patch (e.g., by using a tailored color thresholding approach (Wang et al., 2008)), the display plane can be reconstructed either from its corners or edges (Schnieders et al., 2010).

Using screen patterns restricts the system to static scenarios where calibration and application are separate steps. Directly identifying and matching salient features (Sugano et al., 2010) between screen content and eye reflections could, however, allow online calibration of dynamic setups, for example, with a PTZ camera tracking a close-up region of the eye at high resolution.

Chapter 5

Calibration-free Non-Intrusive Eye Gaze Tracking in Arbitrary Environments

This chapter applies the developed theory for eye pose estimation and light transport at the corneal surface to introduce a novel system architecture and method for geometric-calibration-free eye gaze tracking in arbitrary, geometric and photometric complex, environments.

Section 5.1 provides an introduction to problems in state-of-the-art remote eye gaze tracking and corneal reflection analysis, the principle and advantages of the proposed method, and the contributions of this work.

Section 5.2 surveys and discusses related work in eye gaze tracking for arbitrary environments, approaches without requiring geometric calibration, and the use of controlled illumination to assign information to environment locations that can be recovered from images.

Section 5.3 continues with describing the proposed system architecture and method to track the PoR of an observer in an environment image. The basic light transport theory from sections 3.3 and 3.4 is integrated with the analysis of corneal reflections of projected invisible structured light to define an image-based mapping between eye and environment images. This also provides a solution to the general problem of accurate and robust feature matching among multiple eye images.

A prototype implementation is explained in Section 5.4 and subsequently applied in Section 5.5 to analyze the performance of coded structured light recovery from corneal reflections, and light path estimation between PoR, cornea, and camera.

Section 5.6 concludes this chapter, discussing results and findings, outlining potential implications on application scenarios and fields, stating limitations of the prototype implementation, and providing ideas for future work.

5.1 Introduction

While recent developments in the field of remote eye gaze tracking are promising, state-of-the-art techniques are still far from being unobtrusive and usable for practical applications. There are different characteristics that restrict their

use to labor-intensive controlled laboratory conditions with experienced instructors and trained users. Moreover, techniques still have a high degree of intrusiveness from setup requirements and operation restrictions due to their technical approach in combination with hardware limitations.

Proposed Method. This work proposes a novel system architecture that overcomes several limitations of existing eye gaze tracking techniques; specifically it removes the need for geometric calibration and enables application with arbitrary dynamic scenes. The system uses two cameras, an environment camera capturing the gazed scene and a non-attached eye camera featuring high-resolution or a PTZ tracking architecture to capture a close-up view of the eye. The task is to identify the PoR in the image of the environment camera. For current systems, this involves a geometric calibration where a 3D model of the scene has to be obtained and aligned in the coordinate frame of the eye. Since this is a complex task, eye tracking so far is restricted to planar surfaces.

Geometric-Calibration-free Image-based Mapping. With this work, a method is developed to estimate the corneal reflection of the PoR and map it to the environment image, based on a large number of keypoint matches. While keypoint matching among camera views from different perspectives is a well studied problem (Tuytelaars and Mikolajczyk, 2008), the methods cannot be directly applied to corneal reflections. Specific problems are described in the following.

- Specular corneal reflections are superimposed with reflections from other structures of the eye. The iris texture becomes especially disturbing under high intensity illumination or bright textures. Although Wang et al. (2005a, 2008) describe a method to separate corneal reflections and iris texture, the general problem is ill-posed where results may not be appropriate for feature matching.

- The cornea is not a perfect mirror. The low reflectivity of less than 1% (Kaufman and Alm, 2003) results in a strong compression for the dynamic range of reflected illumination.

- The eye has a curved surface. While common feature matching methods for perspective camera images can be applied to planar mirror reflections (Sturm and Bonfort, 2006), this is not possible with the distortions caused by a curved mirror. Special catadioptric techniques have to be used (Hansen et al., 2007; Scaramuzza et al., 2008). A strategy may involve projecting corneal reflections image into perspective images, perform feature matching and transform the result back. Nevertheless, this can result in high errors due to errors in eye pose estimation and an unknown shape for the individual cornea.

In order to robustly obtain a large number of keypoint matches between environment and eye images under severe conditions in arbitrary environments, we apply invisible/imperceptible coded structured light (CSL). A projector projects structured light patterns onto the scene, assigning a unique visual codewords to each surface location. Synchronized with pattern projection, the environment camera captures the illuminated scene, and the eye camera captures the corresponding corneal reflection. The codewords are then decoded in each camera view uniquely identifying corresponding surface points. Using either imperceptible projection with a standard digital video projector at high framerates or a special invisible (infrared, IR) light projector, the dynamic code is not perceived by the user, but recovered from the camera images. With this procedure, a large number of informative reflections is robustly obtained over a wide area of the corneal surface. The pose of the eye is estimated from an image and applied to locate the corneal reflection of the PoR in the image. The detected correspondences in eye and environment images then define a mapping for obtaining the location of the PoR in the environment image. In case of a known geometric relation between environment camera and projector, a 3D model of scene and PoR can be recovered.

Advantages of the Proposed Method. The described strategy involves the following advantages to state-of-the-art remote eye tracking systems:

Calibration-free Conventional systems need a geometric calibration to determine the relation between camera and scene model. Because reconstructing and aligning a 3D surface model is a complex task, systems are generally restricted to planar surfaces. The calibration is done manually, or automatically by identifying corneal reflections from active light sources (Ko et al., 2008) or scene features (Nitschke et al., 2009; Schnieders et al., 2010). Nevertheless, manual calibration is time consuming, installing light sources is often not feasible, and scene features are commonly sparse and cannot be detected robustly. The proposed method also analyzes scene features, however, in the form of robust projected markers with a high spatial resolution and surface coverage. Since a distinct calibration is not required the system can also be applied with dynamic camera-projector setups, e.g., in tracking scenarios.

Attachment-free Conventional systems require geometric calibration and can, therefore, not be applied to dynamic setups. To achieve eye gaze tracking for a moving user, a head-attached setup using a combination of eye and environment camera is commonly applied (Babcock and Pelz, 2004; Li et al., 2006). Compared to the proposed method, these systems require a gaze-mapping calibration on a planar surface: As the user moves and looks at other surfaces, the calibrated mapping becomes invalid and the error increases. Novel system designs aim in eliminating

the environment camera using a special curved half-transparent mirror to combine eye and mirror-reflected environment in a single camera view (Mori et al., 2010). While this removes the need for calibration, the architecture requires setup and becomes more intrusive since the mirror is placed in front of the eye. The proposed method does not require calibration or head attachments and can, therefore, be applied with moving users in non-obtrusive scenarios.

Arbitrary environments Conventional systems generally assume a planar surface restricting their application to locations such as computer monitors, projection canvases, white- and blackboards, or planar objects on tables and walls. The proposed approach naturally supports surfaces with arbitrary geometry forming the majority of our environment. This enables more realistic scenarios for diagnostic applications, for example in human factor analysis and marketing research; and interactive applications with selective and gaze-contingent user interfaces for ubiquitous and ambient environments.

Free head movement While conventional systems claim to support free head movements, this is not completely true, due to several restrictions for gaze-angle and viewing volume. The systems often rely on corneal reflections and fail in cases when not all reflections are detected. This easily happens at large gaze angles when reflections move outside the corneal boundary and disappear. The same effect restricts the volume of valid head poses. The proposed method generates a large number of reflections covering a wide area of the environment and the corneal surface. Missing parts of the reflection patch at large angles and viewing volumes do not cause fail or reduced accuracy since the effect only happens for the periphery, but not the area where the user is gazing.

Challenging conditions Conventional systems place several assumptions on the setup, limiting the applicability to controlled laboratory conditions with experienced and trained users. It can be hard to set up the tracking environment and obtain reliable data for a particular user. Schnipke and Todd (2000) describe difficulties relating to illumination conditions, camera placement, calibration process and eye conditions. The proposed system increases these tolerances, and especially allows environmental light, reduced image quality, and capturing at high framerates.

Improved accuracy Conventional systems estimate the pose of the eye and intersect the gaze direction with the surface to obtain the PoR. It is easily seen that the estimation error scales with the distance between eye and surface. The proposed method is not affected by this effect since measurement is done image-based. After estimating the pose and

gaze of the eye, the method calculates the reflection of the PoR in the image. Since the location is obtained by projection, the value converges with increasing distance compensating for the described effect. Note, that distance has an effect when the method is applied with non-planar surfaces due to decreasing spatial resolution in the image. However, the effect is compensated by the use of coded structured light which allows for high spatial resolution.

Contribution. The following contributions are achieved with this work (Fig. 5.1):

- A novel eye gaze tracking architecture is proposed that integrates a range of advantages not achieved with existing work.
  - The method works naturally and automatically with arbitrary surface geometries.
  - The absence of geometric calibration allows easy setup and dynamic pose adjustment of system components during runtime, e.g., for tracking scenarios involving a PTZ camera, projector or other mechanisms.
  - The method achieves non-intrusive application without user awareness by not requiring body-attachments and allowing larger tolerances for gaze angles, operation volume and motion.
- Up to our knowledge, the proposed approach is the first to apply structured light projection to eye gaze tracking.
  - The projected coded feature points define a calibration-free relation between eye and environment camera views. The information for missing locations can be interpolated if located within the convex hull of detected features or extrapolated otherwise.
  - The use of projection causes reflections to span a wider area on the corneal surface than achieved with point light sources in front of the user.
  - The use of coded structured light allows robust detection under high spatial resolution. Experimental results verify robustness under challenging conditions, such as short exposure times, image noise and environmental light.
  - The use of imperceptible or invisible structured light lets the dynamic code projection not be perceived by human observers. Experimental results show that double-frame codes, required for imperceptible projection, also allow for increased accuracy.

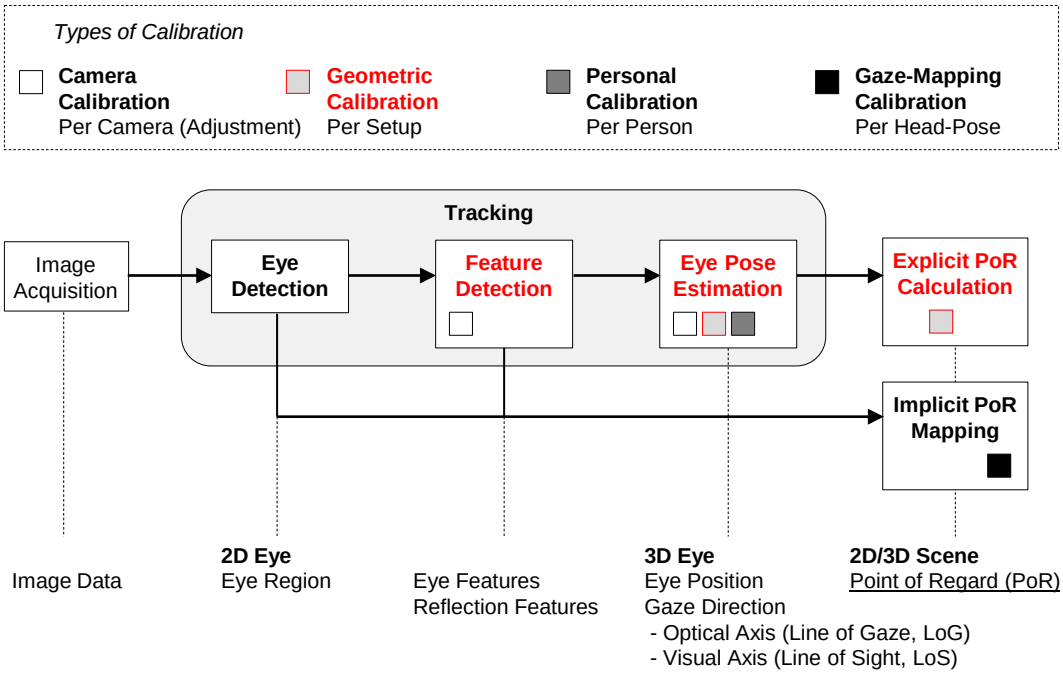


Figure 5.1: Video-based eye gaze tracking pipeline. Contributions of this work are shown in red. There exist different types of parameters sets determined by calibration. The task is to detect and track the PoR from video images of an eye. Two main strategies can be distinguished, explicit and implicit. Explicit tracking, also known as remote eye gaze tracking, reconstructs the explicit eye geometry and obtains the PoR by intersecting the gaze direction with the 3D scene. Implicit tracking obtains the PoR using a head-pose dependent mapping from imaged eye appearance or features. This work introduces several novelties for explicit tracking regarding detection of reflection features, optimization of eye poses, and simplification of PoR calculation for arbitrary surfaces by eliminating geometric calibration.

The proposed system architecture allows for increased applicability not possible with existing techniques. Due to an easy setup, and tolerance to environmental conditions with same time increased accuracy, it has the potential to make eye tracking available to non-professional users in everyday environments. Furthermore, due to absence of calibration, body-attachments and tolerance to operation conditions it enables practical applications generally requiring unobtrusiveness. This enables natural and unbiased conditions in diagnostic scenarios or interactive interfaces for ubiquitous and ambient environments.

5.2 Related Work

Let us now discuss relations of the proposed approach to works in eye gaze tracking and other fields aiming in partially solving similar problems.

5.2.1 Eye Gaze Tracking in Arbitrary Environments

Stationary Eye Gaze Tracking. Traditional stationary systems perform a gaze-mapping calibration that establishes an implicit mapping between eye image data and PoR location on a destination surface (Merchant et al., 1974; Stampe, 1993; Morimoto and Mimica, 2005). Since the mapping is head-pose dependent, the head is either kept fixed (e.g., using a chin rest or bite bar) or movement is compensated (Kolakowski and Pelz, 2006; Karmali and Shelhamer, 2006; Zhu and Ji, 2007; Li et al., 2008). When applying systems based on gaze-mapping calibration to arbitrary environments, a large amount of sampling points is required to describe the local depth variation in the destination surface. Since this is commonly not feasible, stationary systems are, by design, limited to planar surfaces. Furthermore, the placement of calibration markers requires either physical modification or controlled surfaces, such as computer monitors and projection screens on canvases, white- and blackboards, tables, walls and other planar objects.

Remote Eye Gaze Tracking. Recently, remote systems are introduced to achieve head-pose invariance by explicitly modeling the 3D geometry of camera, eyes and scene. Systems, therefore, rely on different kinds of calibration, namely personal calibration of individual eye model parameters, camera calibration, and geometric calibration of the relation between cameras, light sources and scene. The eye is tracked by a remote stationary high-resolution camera or a dynamic camera system (Kim et al., 2004; Oike et al., 2004; Yoo and Chung, 2005; Reale et al., 2010). The gaze direction is obtained either using a passive or an active-light method. Passive methods are commonly based on tracking the visible iris contour as the methods described in this work (Sec. 2.2) (Wang and Sung, 2001, 2002; Wu et al., 2005b; Nishino and Nayar, 2006; Wu et al., 2007; Chen and Ji, 2008; Yamazoe et al., 2008; Schnieders et al., 2010; Reale et al., 2010). To increase accuracy and robustness, the majority of remote methods applies active illumination in form of IR LEDs, commonly based on the pupil center and corneal reflections (PCCR) technique (Shih et al., 2000; Ohno et al., 2002; Guestrin and Eizenman, 2006; Villanueva and Cabeza, 2007; Zhu and Ji, 2007; Villanueva et al., 2009). After computing the gaze direction, the PoR is obtained through geometric modeling involving a 3D model of known pose for the destination surface. Such a model is, however, unavailable for arbitrary environments what commonly restricts systems to planar surfaces. To support arbitrary environments, cameras are added to capture the scene from multiple viewpoints. This allows to estimate the 2D location of the PoR in the view of each single camera, and to obtain its corresponding 3D location by triangulation using multiple cameras (Smart Eye AB, 2011b; SR Research Ltd., 2011a; Seeing Machines Inc., 2011). Such an approach, however, requires additional hardware, probably in a rigid alignment with additional interactive geometric calibration.

Wearable Eye Gaze Tracking. Also recently, wearable head-mounted systems are introduced to combine mobile tracking in arbitrary environments with head-pose invariance between head and scene (Babcock and Pelz, 2004; Li et al., 2006; Wagner et al., 2006). Similar to remote trackers the basic approach combines an eye camera with an environment camera where the 2D location of the PoR is be estimated. The corresponding 3D location is obtained by triangulation using either multiple cameras or tracking scene features in a single moving camera (Munn and Pelz, 2008; Takemura et al., 2010). The difference is that the cameras are rigidly attached to a head-mount that itself is rigidly mounted on the head of the user. While wearable trackers are especially developed for application in arbitrary environments the systems commonly rely on gaze-mapping calibration which, by design, limits them to planar surfaces as in stationary systems: The mapping is calibrated from gaze markers projected onto a planar surface in front of the user. As soon as the user moves away from the surface, the mapping is invalidated by the changing depth. This requires for compensation strategies such as parallax correction (Tobii Technology AB, 2011b). As described with remote trackers, arbitrary environments are supported when using explicit geometric modeling which, however, requires camera, geometry and personal calibration. There exist head-mounted systems naturally allowing for arbitrary environments by exploiting special optical constructions to align eye and scene information (Mackworth and Thomas, 1962; Mori et al., 2010). These systems, however, introduce further issues such as high intrusiveness or reduced data quality.

The proposed approach is an attachment-free remote tracker where a 2D PoR is located in the view of an environment camera. While existing approaches require an offline interactive geometric calibration and are, thus, limited to static camera placement or known parameters for a dynamic camera system, the proposed approach applies structured light to automatically determine the geometric relation between eye and environment cameras in an online process, naturally enabling for arbitrary environments with dynamic camera placement.

5.2.2 Geometric-Calibration-free Eye Gaze Tracking

Geometric Calibration. Compared to systems relying on implicit gaze-mapping calibration, explicit geometric modeling involves several advantages such as head-pose invariance, absence of error accumulation and support of arbitrary environments. This, however, requires camera, personal and geometric calibration. While camera calibration is required only when parameters are modified and personal calibration is required once per subject and achieved already with a single marker (Villanueva and Cabeza, 2008), geometric calibration remains a challenge.

Exploiting Corneal Reflections. There are different strategies to avoid explicit geometric calibration. One particular exploits corneal reflections of environmental light to directly relate eye and environment geometry. Existing approaches, however, involve different issues: Extraction of environmental light is affected by superimposed iris texture, low corneal reflectivity, errors in pose estimation and distortions from unknown corneal shape for a single eye, and unreliable and sparse feature matching between multiple eyes (Nishino and Nayar, 2006). Controlled illumination enables more robust and accurate extraction, but commonly uses only a sparse set of features. Existing systems are, therefore, restricted to planar surfaces, either determining explicit pose (Nitschke et al., 2009; Schnieders et al., 2010) or an implicit mapping (Yoo and Chung, 2005; Coutinho and Morimoto, 2006, 2010; Ko et al., 2008; Kang et al., 2008). Eye gaze tracking in arbitrary environments has been realized by attaching IR tags to scene objects (Smith et al., 2005). While such an approach works without geometric calibration, it involves intrusive modification of the environment, tedious manual configuration, and is, by design, restricted to a small number of pre-defined objects. As explained, there also exist wearable head-mounted systems that use special optical constructions to avoid geometric calibration, but these come with further issues (Mackworth and Thomas, 1962; Mori et al., 2010).

Exploiting Assumptions. Explicit geometric modeling intuitively describes the physical problem, supports for head-pose variance and can be used with arbitrary surfaces. On the other hand, systems may be difficult to develop, involve error accumulation from model approximation or loss of important information from early decisions. In contrast, appearance-based methods, also known as image template or holistic methods, directly model and track eye gaze, based on the photometric appearance of eye image patches characterized by the distribution of intensity values or filter responses. Appearance-based methods are often used to simplify system development. Nevertheless, this is achieved at the expense of a gaze-mapping calibration to train a model that maps eye image patches to gaze information. To avoid an interactive calibration in eye gaze tracking with planar computer screens, automatic methods are proposed that evaluate other available information exploiting assumptions on their relationship with eye gaze. Sugano et al. (2008) perform gaze-mapping calibration by evaluating mouse-click locations on the screen under the assumption that the user gazes the corresponding locations. To apply the same strategy with passive scenarios where a user watches screen content without interaction, Sugano et al. (2010) describe an approach based on visual saliency which is the distinct subjective perceptual quality of visual information that immediately attracts human attention (Koch and Ullman, 1985). They perform gaze-mapping calibration by treating saliency maps of display content as probability distributions for the observers gaze.

The proposed approach relates eye and environment geometry by exploiting corneal reflections of controlled illumination. The novelty is that it applies invisible coded structured light projected with high spatial resolution onto a wide surface area. This enables for accurate, robust and dense extraction of scene features from corneal reflections and environment cameras which is a requirement for automatic geometric calibration of arbitrary surfaces.

5.2.3 Optical Encoding of Environment Locations

To avoid a dedicated geometric calibration and model reconstruction for arbitrary surfaces, corneal reflections of environmental light are analyzed to automatically relate eye gaze and environment locations with support for dynamic setups. Solely relying on the passive light distribution, however, causes less robust and accurate extraction and low quality environment information (Nishino and Nayar, 2006).

Intrusive LED- and Screen-based Techniques. Active illumination can be used to compensate for this and has been applied to eye gaze tracking with planar screens using screen illumination (Nitschke et al., 2009; Schnieders et al., 2010) or attaching IR LEDs (Yoo and Chung, 2005; Coutinho and Morimoto, 2006, 2010; Ko et al., 2008; Kang et al., 2008). Controlled illumination enables optical encoding and transmission of information, and allows for dynamic parameter modification. Such coded structured light techniques have been widely used for assigning information to environment locations. Attaching controlled IR LEDs as tags to scene objects permits identification by detecting unique temporary codes based on blinking patterns (Sakata et al., 2002). There are attempts to program and control IR tags automatically using attached sensors in conjunction with projected light where the information is encoded by slight variation of the content from a digital video projector or by using a specially designed high-frequency projector (Nii et al., 2005; Lee et al., 2005). IR tags have been exploited in wearable head-mounted eye gaze tracking to detect when the user gazes a tagged object without the need for calibration (Smith et al., 2005). Beside representing scene locations, special array configurations of IR LEDs have been applied as structured light in gaze direction estimation to reject false corneal reflections and increase robustness in case of lost reflections at large gaze angles (Hua et al., 2006; Li et al., 2007). Note, that "structured light" in this context refers to the geometric alignment of the LED array. There are several issues related with the described approaches for active illumination: Special illumination patterns on computer screens are intrusive and interfere with content. Attaching LEDs to scene locations is intrusive and limited to a small number of locations as system complexity rapidly increases.

Non-Intrusive Projection-based Techniques. To avoid the described complications and enable further advantages, coded structured light is commonly projected using off-the-shelf or specially designed digital video projectors (Battle et al., 1998; Salvi et al., 2004, 2010). Non-intrusiveness is accomplished by either exploiting invisible or imperceptible structured light (Fofi et al., 2004). Invisible structured light operates in wavelengths outside the visible spectrum. Infrared light is mainly applied as it can be detected with common imaging sensors when IR-block filters are removed. Zhang et al. (2008) use this approach to project IR tags in the form of temporary-coded blinking dots projected onto selected scene locations. While being invisible, the tags are detected in camera images enabling lookup of object-specific information stored in a database. A problem remains with projection of IR light: Since the lamps of digital video projectors only emit a small amount of IR light which is further reduced with improvements in lamp engineering, the straightforward strategy of using an IR-pass filter results in low intensity and a waste of energy. Lee et al. (2007) design and implement a high-resolution, scalable and general-purpose solution to combine IR and visible light with common projector designs. Therefore, they replace the light source of a DLP projector with an array of IR and visible-light LEDs synchronized with the DMD of the projector.

Another technique to realize non-intrusive illumination is imperceptible structured light where a sequence of alternating light pattern and complement (inverse pattern) is projected using visible light. If the sequence exceeds the critical flicker frequency (CFF) (Watson, 1986), the dynamic content is visually integrated over time and appears as a static white homogeneous illumination area with average projected intensity. While imperceptible for human observers, the pattern is detected in images of a synchronized camera enabling for information transmission (Raskar et al., 1998; Livingston, 1998; Waschbüsch et al., 2005). Several research aims in combining imperceptible structured light and content projection within a single projector (Cotting et al., 2004; Grundhöfer et al., 2007; Bimber et al., 2008). Others use imperceptible structured light projection to track light sensors (Lee et al., 2005). A recent overview of high-speed synchronized vision systems, required for visual information transmission is found in Kagami (2010).

The proposed approach uses either invisible or imperceptible structured light projection to assign unique codes to environment locations. Detecting and matching codes between environment and eye images, geometric information is automatically obtained without the need for dedicated calibration. The use of coded structured light projection, therefore, enables for several benefits such as flexibility supporting arbitrary environments, attachment-free remote tracking, dynamic scenarios, wide area coverage and high-spatial resolution; robustness to practical conditions and eye appearance; and accuracy through image-based matching.

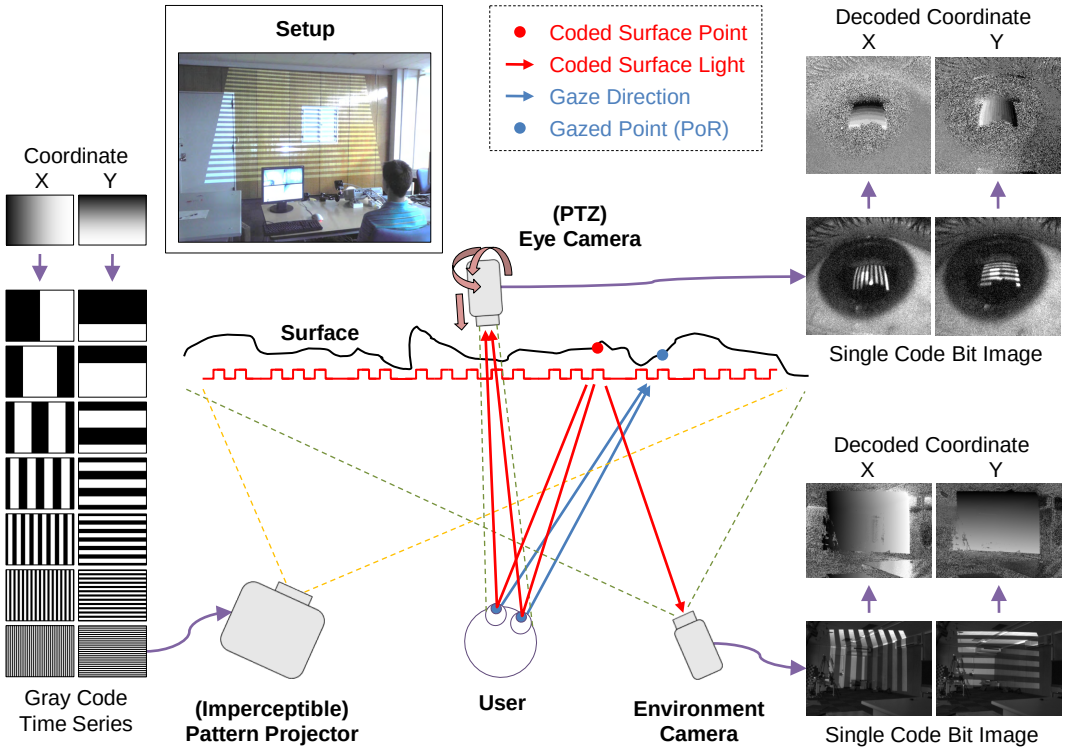


Figure 5.2: Setup for eye gaze tracking using coded structured light. A user is gazing an unknown PoR (blue) on an arbitrary surface. An eye camera tracks a close-up image of the user's eye. An environment camera captures a view of the gazed surface. Coded structured light is projected onto the surface (red) where it reflects towards environment camera and eye, from where it again reflects into the eye camera. Correspondences are then obtained by decoding the reflections.

5.3 Method

In the following we explain the building blocks of the proposed eye gaze tracking architecture. Figure 5.2 explains setup and method. A user gazes an unknown location on an arbitrary surface. The task is to locate the PoR in an image of the surface captured by an environment camera[1]: An eye camera tracks a close-up image of the user's eye from where the pose of the eye is estimated and the corneal reflection of the PoR is calculated. Coded structured light is then used to obtain robust correspondence information between surface points, their images, and the images of their corneal reflections. A projector projects a pattern or time series of patterns onto the surface where the light reflects towards environment camera and eye, from where it again reflects into the eye camera. Capturing the illuminated scene with synchronized cameras, the correspondences can be decoded from the images. Based on that information, the known location of the reflected PoR in the eye image

[1]The approach directly allows for multiple environment cameras.

is mapped into the environment image. When the external relation between projector and environment camera is known, a surface model can be recovered by active stereo where the PoR is located in 3D.

5.3.1 Projector-Camera Synchronization

The gaze tracking system is based on projector-camera synchronization modeled in Figure 5.3. A concurrent implementation is favored in order to maximize framerate which is limited by the slowest component of the system. In the following we discuss the different factors.

The projector refresh rate is determined by model and resolution, and usually lies at around 85 Hz for common digital video projectors and 120 Hz for stereo projectors. Higher framerates are possible with professional projectors and special projection devices.

The camera framerate depends on model, resolution and data bus. Synchronization requires asynchronous frame grabbing triggered either using a hardware or software signal. Although, asynchronous mode has an impact on framerate, current IEEE 1394b or USB 3.0 machine vision cameras generally achieve framerates exceeding the refresh rate of digital video projectors.

The pattern generation time is the time required to render a pattern into the graphics buffer. It is usually negligible since rendering on current systems is fast. If coding is complex, patterns can be pre-calculated and stored in memory since the sequence is usually static or follows a dynamic combination of a finite number of patterns.

The frame processing time comprises triggering, image exposure and processing for all cameras. While most of the image processing steps are not complex, automatic eye detection and iris contour fitting may become the bottleneck of the overall system. If required, image processing performance can be largely improved using GPU acceleration.

5.3.2 Correspondence Representation

The task is to uniquely identify the image locations of particular scene points in different camera views. This is usually achieved with a keypoint-matching technique, such as the Scale-invariant Feature Transform (SIFT) (Lowe, 2004). Using a purely image-based passive strategy has the advantage of not requiring additional hardware and being non-intrusive to the environment. However, this is achieved at the expense of robustness where accuracy largely depends on the application scenario. To become more independent, the described system takes an active strategy where each surface point is represented with a unique

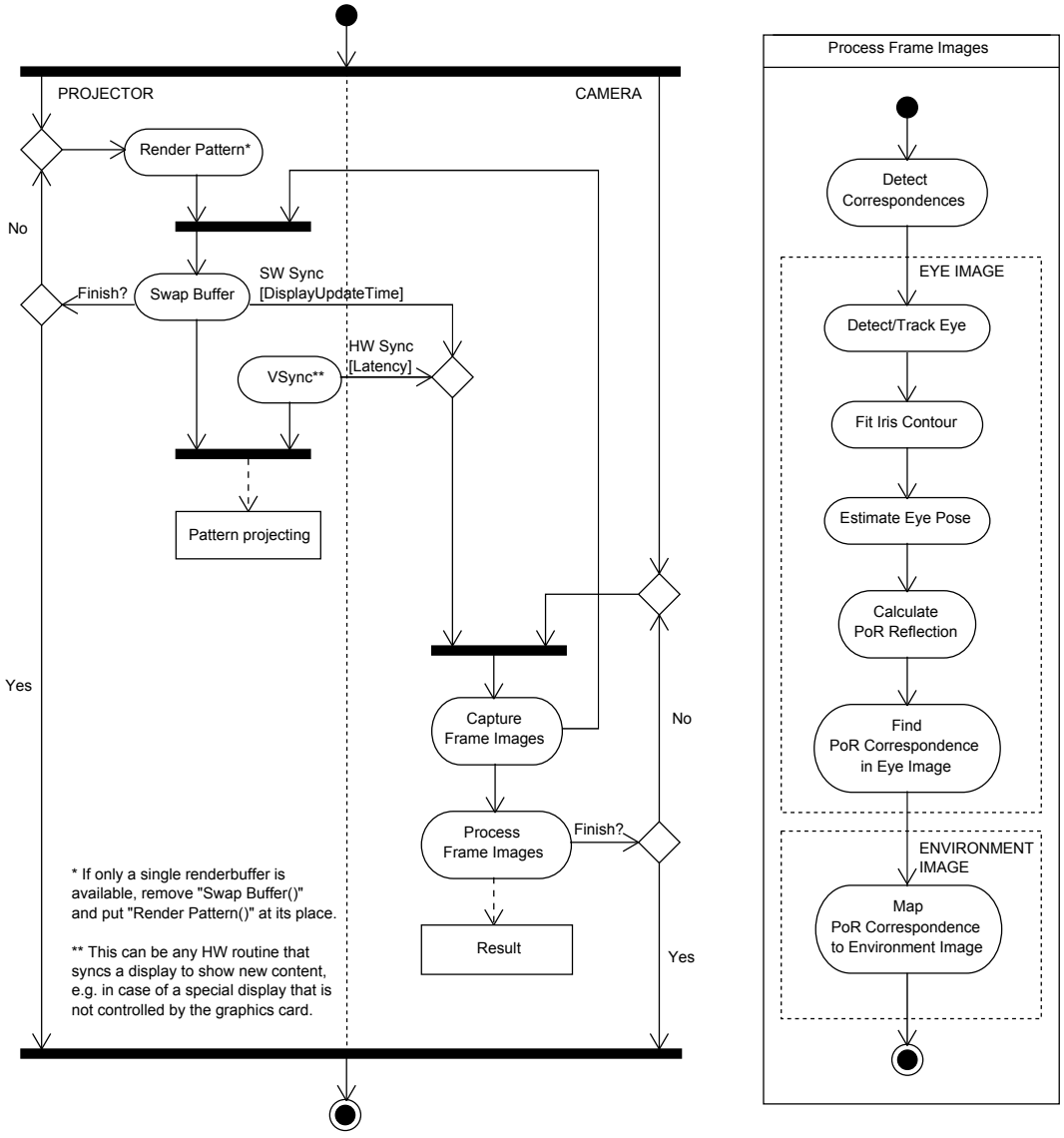


Figure 5.3: Eye gaze tracking algorithm. (left) Multi-threaded generic projector-camera synchronization with real-time image processing showing projector-related components top-left and camera-related components bottom-right. The projector is modeled as a common digital video projector operated through the graphics card (as opposed to a specialized illumination device). The function `RenderPattern()` corresponds to a double-buffer scenario, where idle-time is used to render the pattern into the back-buffer. In case of a single-buffer scenario, `RenderPattern()` replaces `SwapBuffer()`. After the new pattern is displayed by the projector, camera frame grabbing is triggered by either a software or a hardware synchronization signal. In case of software synchronization, there is no knowledge about the events, when the buffer switch occurs and the updated content is shown. Therefore, the trigger is sent after a pre-defined delay `DisplayUpdateTime`. In case of hardware synchronization, the trigger signal is taken from the `VSync` pin of the graphics port signaling the updated content is sent to the projector. Frame grabbing is done after a projector-dependent delay `Latency`. (right) The function `ProcessFrameImages()` contains the entire eye gaze tracking algorithm. First, correspondences are decoded for all images. Second, the gazed PoR is recovered from the eye image, and third, mapped into the environment image.

codeword using coded structured light. In the most general case the codewords are visually encoded and projected by some light-emitting focusable device. Capturing the illumination information in an image or a sequence of images, each scene point is uniquely identified from its recovered codeword.

Coded Structured Light. There are different pattern codification strategies, such as direct, spatial-multiplex and time-multiplex coding. An overview comparing characteristics and technical details is found in Battle et al. (1998); Salvi et al. (2004); for more recent extensions see Salvi et al. (2010). Section 4.6.4.1 discusses the application to display-camera calibration. The general choice is a time-multiplex approach, since it offers increased robustness under challenging conditions, including environmental light, bright iris color, high framerates and low image resolution. Motion blur from eye movements along the time series can be handled by motion compensation based on eye pose estimation. The effect of such motion blur is an integration of information from different codewords at a particular image location, decreasing spatial resolution. An approach balancing between time- and spatial-multiplex coding can be used when a higher effective framerate is required.

Invisible/Imperceptible Structured Light. A further requirement for the system is that projection is not noticed by the user. This is achieved by using invisible or imperceptible structured light. Invisible structured light requires a special projection device operating outside the range of the visible spectrum. In order to use it together with common camera hardware, IR light is usually applied. Imperceptible structured light uses visible light, but projects a sequence of alternating pattern and complement at high framerates (Fofi et al., 2004). The user still perceives the homogeneous illumination, but does not notice the patterns. The same effect is used to restore the surface texture in the environment camera by integrating two complementary frames. Non-intrusive illumination allows the projector to be combined with other light sources for common room illumination. High refresh rates are available with stereo projectors gradually turning into off-the-shelf devices with increasing popularity of 3D visualization. Imperceptible pattern projection doubles the number of required frames, but on the other hand enables more robust decoding methods.

5.3.3 Image Acquisition

While the projector illuminates the surface with a particular pattern an image is recorded with both, environment and eye camera. The environment camera captures the primary reflection from the surface, the eye camera the secondary reflection at the cornea. In case of time-multiplex coding, the system needs to operate at high framerates because a single effective frame requires processing a complete time series.

5.3.4 Correspondence Detection

The projected correspondence information is recovered for eye and environment views. The result may be enhanced by reducing noise and filling holes using an optimization strategy, probably applying knowledge about geometric and photometric properties of the surface. Since decoding is applied to each image location, it is further necessary to segment meaningful locations that relating to illuminated surface areas.

5.3.5 Eye Pose Estimation and PoR Computation

Applying an eye pose estimation method from the ones described in Section 2.2, two possible solutions are obtained. For each solution, the corneal reflection of the PoR is computed using an appropriate method from Section 3.3. The correct solution for the pose of the eye is selected as the one resulting in minimum distance between the computed reflection of the PoR and the center of the coded area in the eye image.

5.3.6 PoR Mapping

After computing the corneal reflection of the PoR in the eye image the codeword corresponding to the PoR can be looked-up. The corresponding surface location in the environment image is obtained by codeword matching. Missing information in the mapping can be calculated by interpolation, in case the pixel is located in the convex hull of recovered correspondences, or by extrapolation otherwise.

5.3.7 3D Scene Reconstruction

In case of a known geometric relation between projector and environment camera, a 3D surface model of the scene including the PoR may be reconstructed by means of active stereo (Salvi et al., 2010). The geometric relation can be automatically calibrated up to a scale ambiguity using the following steps: Assume the camera center is the origin, with camera and projector projection matrices P_C and P_P given as

$$\mathsf{P}_\mathrm{C} = \mathsf{K}_\mathrm{C}\left[\mathsf{I}\,|\mathbf{0}\right], \quad \mathsf{P}_\mathrm{P} = \mathsf{K}_\mathrm{P}\left[\mathsf{R}\,|\mathbf{t}\right]. \tag{5.1}$$

1. Compute the fundamental matrix F from the decoded correspondence pairs between camera and projector (Hartley and Zisserman, 2003, pp 279).

2. Compute the intrinsic matrices of camera and projector, K_C and K_P, using some auto-calibration method, such as Drareni et al. (2009).

3. Compute the essential matrix E from the fundamental matrix as in $\mathsf{E} = \mathsf{K}_\mathrm{P}^\mathrm{T}\mathsf{F}\mathsf{K}_\mathrm{C}$ (Hartley and Zisserman, 2003, p 257).

4. Decompose the essential matrix $\mathsf{E} = [\mathbf{t}]_\times \mathsf{R}$ into rotation and translation components (Hartley and Zisserman, 2003, pp 258).

5.4 Implementation

The last section introduced the general concepts behind the eye gaze tracking algorithm. Let us now describe the actual implementation in terms of applied methods and algorithms to execute different subtasks. The purpose of this implementation is a prototype that can be used within two modes: The interactive mode provides an accurate assisted method for eye detection. It is used as a research tool to support a first experimental analysis for recovering and mapping the corneal reflection of a PoR among eye and environment images. The automatic method provides a less accurate but non-assisted method for eye detection. It is used as a demonstration system.

5.4.1 Projector-Camera Synchronization

The implementation of a real-time synchronized projector-camera system follows the model in Figure 5.3. Because pattern projection and image processing are realized as concurrent processes, the camera is required to support asynchronous access by either a hardware or a software trigger signal. Projection is accomplished with a common digital video projector operated via graphics pipeline using OpenGL (OpenGL, 2011), and synchronizing image acquisition with content update. Enabling `VSync` (vertical sync) ensures that content update is synchronized with the vertical refresh-rate of the projector. Double-buffering allows rendering of the following pattern while keeping the currently displayed pattern consistent.

Since OpenGL commands are non-blocking where the actual execution can be delayed, we need a means of synchronization ensuring that rendering is accomplished and that the result is available in the back-buffer. A call to `glFinish()` does that by waiting and returning when previous OpenGL functions are completed. More fine-grained synchronization for command completion is offered with the extensions `ARB_sync` (OpenGL, 2009) introduced in OpenGL version 3.2, or `NV_fence` (OpenGL, 2008) introduced for Nvidia chipsets earlier in OpenGL version 1.2.1.

After completing image acquisition for the previous frame we issue the `SwapBuffer()`-command of the GUI subsystem requesting the rendered pattern in the back-buffer to be displayed. We need to wait until the display is updated before performing image acquisition for the next frame. If the camera supports only software triggering we are restricted to software synchronization. Since OpenGL currently does not provide information about the display update we assume a maximum value for the time duration and initialize a system timer that fires a trigger event after a timeout `DisplayUpdateTime`.

The required maximum time duration is measured in an automatic calibration step in advance. If the camera supports hardware triggering we are able to apply a more accurate method using hardware synchronization where we grab the `VSync` pin of the graphics port signaling when the updated framebuffer content is sent to the projector, and send a trigger signal to the camera. There might be an additional projector-dependent latency that is also measured in an automatic calibration step in advance. The delay `Latency` is either implemented in the controller that transforms the `VSync` signal into a triggering signal, or into the camera itself.

The described implementation uses a single PC. Increasing the number of cameras or the complexity of the processing algorithm may require to distribute the system onto multiple PCs. In that case, additional communication and synchronization strategies are required.

5.4.2 Correspondence Representation

To uniquely identify the location of a surface point in different camera views, a common off-the-shelf digital video projector illuminates the surface with patterns assigning a unique codeword to a pixel or a region of pixels. A time-multiplex coding approach is chosen as it offers the best accuracy and scalability. This is achieved at the expense of effective framerate since the codewords are spread across a time series of pattern frames.

Binary Code. To realize the maximum SNR we use a binary encoding where each frame contributes a single bit to the codewords in parallel (Posdamer and Altschuler, 1982). Each projector pixel is represented by its x- and y-coordinate encoded into two binary codewords

$$\begin{aligned} x \in \{1,\dots,R_x\} &\rightarrow (x_1,\dots,x_{N_x})^{\mathrm{T}},\ x_i \in \{0,1\},\ N_x = \lceil \log_2 R_x \rceil, \\ y \in \{1,\dots,R_y\} &\rightarrow \left(y_1,\dots,y_{N_y}\right)^{\mathrm{T}},\ y_j \in \{0,1\},\ N_y = \lceil \log_2 R_y \rceil, \end{aligned} \tag{5.2}$$

where R_x and R_y denote the resolution of the projector. N_x and N_y are the required number of bits to represent the whole range of coordinate values. Using less bits results in the effect that neighboring pixels obtain an equal codeword decreasing spatial resolution. The number of frames in a time series required to transmit both coordinate values is the sum $N_x + N_y$.

Reflected Binary Gray Code. Instead of directly projecting the binary coded patterns we rearrange the bits of each codeword in a way that adjacent numbers only differ in a single bit. This enables error detection and correction under the assumption of continuos surfaces and comes at no additional cost. The obtained representation is referred to as reflected binary Gray code, or short Gray code (Gray, 1953; Weisstein, 2011c). Figure 5.2 shows the encoding of pixel values into a series of Gray code patterns (Inokuchi et al., 1984).

The binary values 0 and 1 are visually represented by the lowest and highest projectable intensities. On the camera side, the codewords are recovered after a complete sequence of patterns is projected and corresponding images are captured.

Double-Frame Pattern. We use an alternating sequence of positive and negative patterns, doubling the number of required frames to $2\,(N_x + N_y)$. This strategy allows for robust double-frame thresholding and makes the dynamic pattern imperceptible to a human observer when projected at high framerates (Fofi et al., 2004).

5.4.3 Image Acquisition

While the projector illuminates the surface with a particular pattern we capture an image with both, environment and eye camera. We do not apply tracking using a PTZ camera. The aperture of the lens is manually adjusted for a particular setup to maximize the amount of captured light while still keeping the depth-of-field large enough to eliminate focus blur. The use of binary encoding increases robustness under environmental light what relaxes conditions compared to the implementation of the display-camera calibration in Section 4.4.

5.4.4 Correspondence Detection

Horizontal and vertical coordinates of the projector are encoded into two N-bit codewords, requiring N frames under invisible projection and $2N$ frames under imperceptible projection. The pattern for a single frame t contains binary information where each surface point is either illuminated or not. This information is decoded from the camera image using binarization where the binary value b_t at each pixel is decided by thresholding (Sezgin and Sankur, 2004). We will now explain different methods.

Zero Thresholding. For imperceptible pattern projection using an alternating sequence of positive and negative patterns, double-frame thresholding is simple and robust. The binary value is obtained as in

$$b_t = \begin{cases} 1 & \text{if } \left(I_t - I_t^-\right) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5.3}$$

where I_t and I_t^- denote the intensity values at the pixel in the pattern and complement image respectively.

With invisible pattern projection, single-frame thresholding can be applied. A constant threshold for all pixels is usually not recommended since the imaged intensity does not only depend on projected intensity but also on surface properties, eye appearance and the geometric relationship between projector, surface, eye and cameras. To obtain a pixel-dependent threshold let us first introduce several required per-pixel statistics, calculated over the intensity values in a time series. These are, the minimal and maximal intensity $I_{\min}$ and $I_{\max}$ as in

$$I_{\min} = \min_{t=1}^{N} I_t, \quad I_{\max} = \max_{t=1}^{N} I_t, \tag{5.4}$$

and the minimal and maximal absolute intensity difference $\Delta I_{\min}$ and $\Delta I_{\max}$ between subsequent frames as in

$$\Delta I_{\min} = \min_{t=1}^{N-1} \Delta I_t, \quad \Delta I_{\max} = \max_{t=1}^{N-1} \Delta I_t, \quad \Delta I_t = \left| I_t - I_{t+1} \right|. \tag{5.5}$$

Mid-Range Thresholding. This is the simplest thresholding method using the mid-range of intensity values as threshold IT with

$$IT = \frac{1}{2} \left(I_{\max} + I_{\min} \right), \tag{5.6}$$

where the binary value is obtained by thresholding the corresponding intensity value as in

$$b_t = \begin{cases} 1 & \text{if } I_t > IT, \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

Otsu Thresholding. For this thresholding method we compute the threshold by applying Otsu's image thresholding method (Otsu, 1979) to the series of intensity values. It calculates the optimum threshold IT_{Otsu} by finding the binary classification that minimizes the combined spread (intra-class variance) for two classes of intensity values. The binary value is then obtained by thresholding the corresponding intensity value as

$$b_t = \begin{cases} 1 & \text{if } I_t > IT_{\text{Otsu}}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.8}$$

Mid-Range Difference Thresholding. This method is similar to the mid-range thresholding method, but works on the absolute differences between intensity values in subsequent frames instead of the raw intensity values. Thus, we calculate the mid-range of difference values as threshold ΔIT with

$$\Delta IT = \frac{1}{2} \left(\Delta I_{\max} + \Delta I_{\min} \right). \tag{5.9}$$

The binary value is then obtained by thresholding the corresponding absolute difference between subsequent intensity values as

$$b_{t,t=N} = \begin{cases} 1 & \text{if } I_t > IT, \\ 0 & \text{otherwise,} \end{cases} \quad b_{t,t<N} = \begin{cases} \overline{b_{t+1}} & \text{if } \Delta I_t > \Delta IT, \\ b_{t+1} & \text{otherwise.} \end{cases} \tag{5.10}$$

After thresholding a time series corresponding to a particular surface location we obtain N binary values b_t which we integrate into a N-bit binary number. Since a coded representation is used for transmission, the number needs to be decoded from binary reflected Gray code by a simple rearrangement of the bits. Finally, the decoded binary number is converted into a decimal number representing the recovered projector coordinate. Figures 5.4 and 5.5 show an example series of projected patterns, captured images and recovered results.

Identification of Coded Pixels corresponding to Direct Light Paths. The current implementation recovers a codeword for every pixel in the image without testing if the pixel is actually affected by projected light and corresponds to a direct light path. This leads to noisy pixels around the valid areas in the recovered correspondence maps. If filtering is required, different strategies have to be used for environment and eye image.

The environment image captures diffuse first-order reflections from the surface. Illuminated pixels are recovered by capturing two additional images with full projector illumination and without illumination. The mask of illuminated pixels is obtained by thresholding the difference image. Note that this does not filter pixels corresponding to inter-reflections.

The eye image captures specular second-order reflections from the cornea. To filter correct matches, we first apply the same method as used for the environment image. However, there may still remain pixels from different parts of the face illuminated by surface reflections. We notice that the cornea has the highest specularity and remove other reflections that fall below a particular level of specularity. To measure specularity, we use the fact that reflections can be seen as a convolution of incoming light and BRDF kernel of the material (Ramamoorthi and Hanrahan, 2001). Diffuse materials cause blur for finer details where specular materials preserve them. Thus, the intensity difference between pattern and complement across a time series will converge faster to zero for diffuse reflectors than for specular reflectors.

5.4.5 Eye Detection and Iris Contour Fitting

Detecting the contour of the visible iris consists of two tasks, a rough iris detection or tracking between subsequent frames and a detailed fitting of the

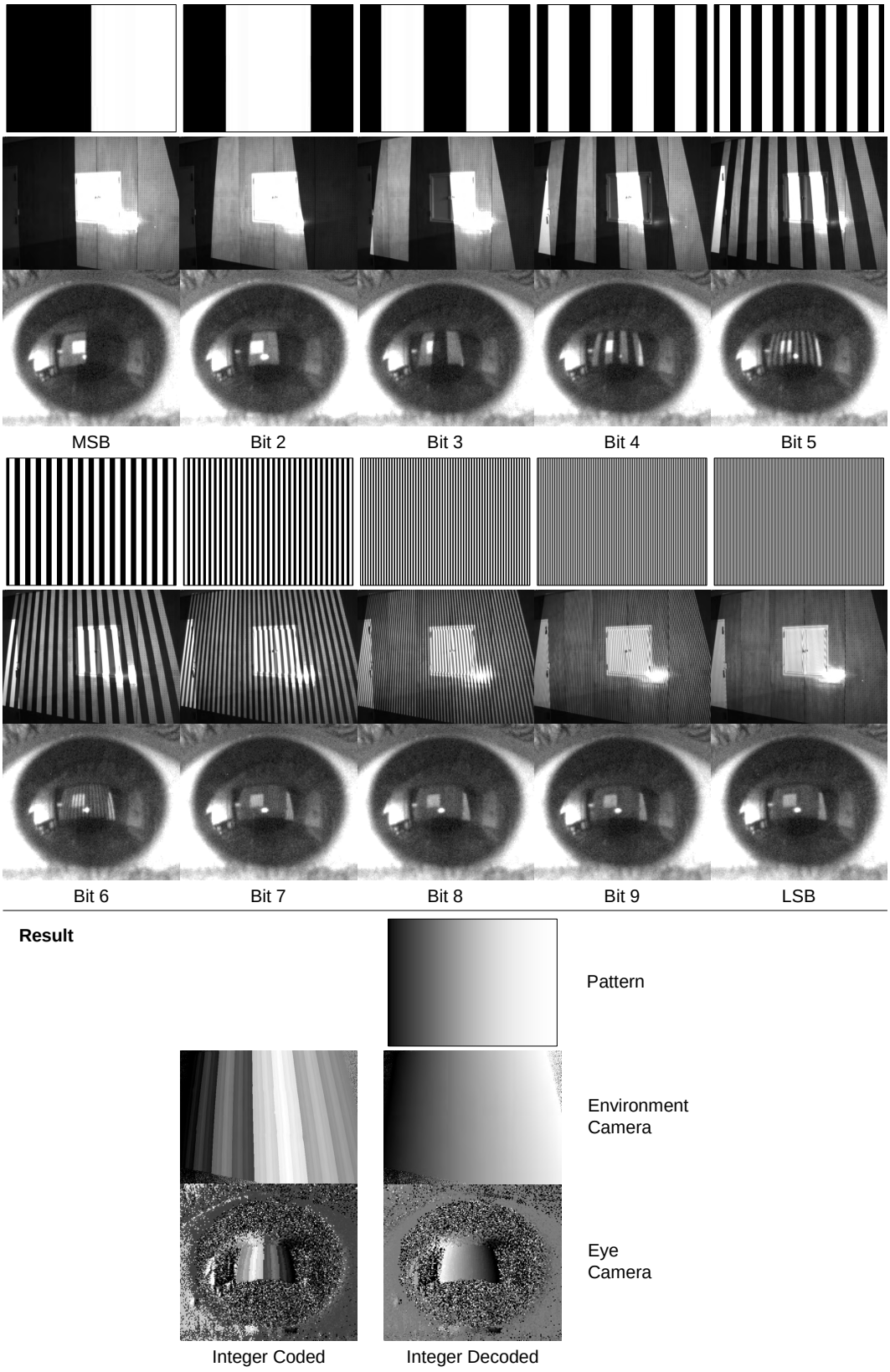


Figure 5.4: Projector x-coordinate correspondences from coded structured light. (top) Projected pattern, environment camera image and cropped eye camera image for each positive frame of a 10-bit Gray code. (bottom) Gray-coded and decoded result after integrating all binarized frames.

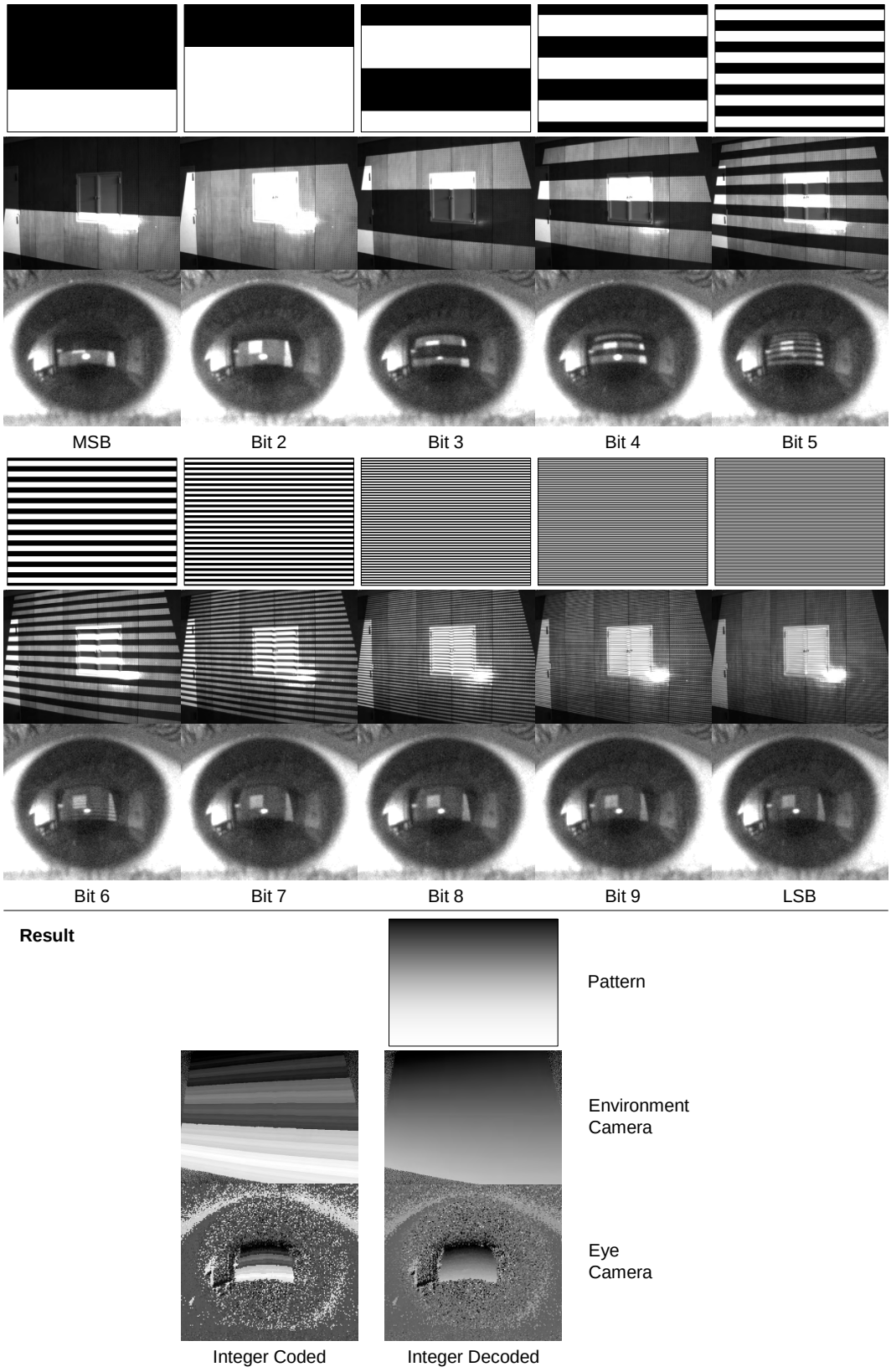


Figure 5.5: Projector y-coordinate correspondences from coded structured light. (top) Projected pattern, environment camera image and cropped eye camera image for each positive frame of a 10-bit Gray code. (bottom) Gray-coded and decoded result after integrating all binarized frames.

iris contour. Obtaining accurate results is a difficult problem. We therefore provide two implementations for solving the first task, catering different purposes.

Interactive Method. To create an accurate estimate for the iris contour used for experimental evaluation, we apply the same strategy as described with the implementation of the display-camera calibration in Section 4.4. We first project a sequence of patterns and capture the corresponding camera images. Processing is then done offline. For each eye image, we manually specify an initial guess for the iris ellipse by selecting four points on its boundary. The estimate is then automatically refined by iteratively minimizing the integral of edge-distances along the arc of the ellipse.

Automatic Method. Opposed to performing accurate offline experiments, the purpose of this method is to create an automatic demonstration system for the eye tracking algorithm. Processing is done online. At the first frame, we again manually specify an initial guess for the iris ellipse. At subsequent frames we apply an automatic model-based tracking strategy using the Condensation algorithm (Conditional Density Propagation) (Isard and Blake, 1998) to track the initial guess of the iris ellipse. The result is automatically refined using the same technique as with the interactive method.

5.4.6 Eye Pose Estimation and PoR Computation

The pose of the eye is estimated from the contour of the imaged iris. For the implementation of the display-camera calibration in Section 4.4 we use the weak-perspective method described in Section 2.2.2.3. For the current implementation of the eye gaze tracking algorithm we do not use an automatic zoom camera and, therefore, place the camera relatively near to the eye. Under that configuration we cannot assume weak-perspective projection and instead use the perspective projection method described in Section 2.2.2.2. We obtain two possible solutions, and for each, compute the corresponding position on the corneal surface where the gazed PoR reflects. The appropriate method is selected from the ones described in Section 3.3 based on the knowledge about the distance of the PoR. The correct solution for the pose of the eye is automatically determined as the solution resulting in minimum distance between computed PoR reflection and center of the reflected pattern in the image. Knowing the position and orientation of the limbus plane we model the corneal sphere with the center at a defined distance along the negative optical axis. Figure 5.6(2) shows results using the less accurate automatic method for eye tracking.

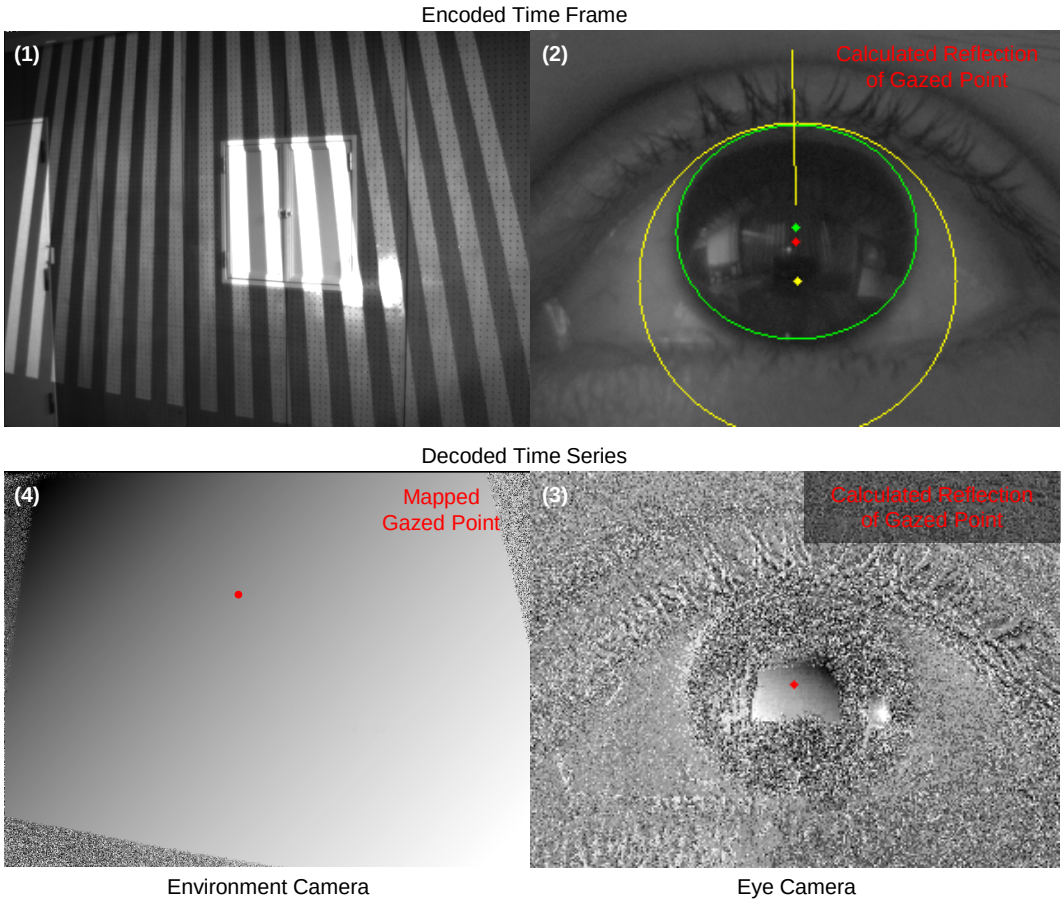


Figure 5.6: Estimation result of eye gaze tracking. (1) An environment image showing a particular frame of the time series code. (2) The cropped eye region in the corresponding eye image with superimposed result from particle-filter based eye-model tracking showing iris/limbus contour and center (green), corneal sphere contour and center (yellow), gaze ray from corneal apex (yellow), and corneal reflection of PoR (red). On closer inspection, the reflection of the illuminated surface shown in (1) can be discovered. (3) The recovered correspondences for the eye region decoded from a complete time series including the current frame. The corneal reflection of the PoR lies within the coded area, suggesting that the user is gazing a point within the coded area on the wall. (4) The resulting PoR in the environment image is obtained by code-based mapping from the eye image. Further processing of these results may involve elimination of invalid decoded locations observed as noise in the correspondence images (3) and (4), elimination of the binary code illumination from the surface texture in (1), reconstruction of a depth map for the surface, and object recognition using 2D or 3D data.

5.4.7 PoR Mapping

Let us assume the user is gazing an illuminated surface point $\mathbf{P}$ for which the corresponding projector pixel is successfully decoded. $\mathbf{p}^{\text{eye}} = (p_u^{\text{eye}}, p_v^{\text{eye}})^{\text{T}}$ denotes the location of gazed surface point in the eye image. The corresponding location in the projector image $\mathbf{p} = (p_x, p_y)^{\text{T}}$ is obtained by looking up the

coordinates in both decoded eye correspondence images I^{eye} as

$$\begin{aligned} p_x &= I_x^{\text{eye}}\left(\mathbf{p}^{\text{eye}}\right), \\ p_y &= I_y^{\text{eye}}\left(\mathbf{p}^{\text{eye}}\right). \end{aligned} \tag{5.11}$$

Now we want to find the location $\mathbf{p}^{\text{env}} = (p_u^{\text{env}}, p_v^{\text{env}})^{\text{T}}$ in the environment image corresponding to that projector pixel. The location is obtained by searching through all decoded projector correspondences in the environment image I^{env} and selecting the one with minimal absolute distance

$$d\left(\mathbf{p}^{\text{env}}\right) = \sqrt{\Delta I_x\left(\mathbf{p}^{\text{env}}\right)^2 + \Delta I_y\left(\mathbf{p}^{\text{env}}\right)^2}, \tag{5.12}$$

where

$$\begin{aligned} \Delta I_x^{\text{env}}\left(\mathbf{p}^{\text{env}}\right) &= I_x^{\text{env}}\left(\mathbf{p}^{\text{env}}\right) - p_x, \\ \Delta I_y^{\text{env}}\left(\mathbf{p}^{\text{env}}\right) &= I_y^{\text{env}}\left(\mathbf{p}^{\text{env}}\right) - p_y, \end{aligned} \tag{5.13}$$

are the differences in decoded projector coordinates between the reflection of the PoR in the eye camera image and a location in the environment image. The resulting location of the PoR in the environment image is the result of the eye gaze tracking algorithm. Figure 5.6 explains the mapping for an example setup.

5.5 Experiments

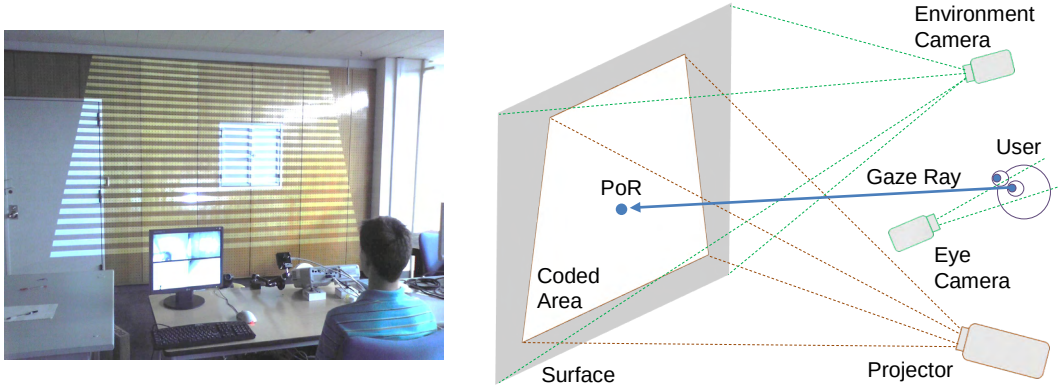
5.5.1 Setup

The setup (Fig. 5.7) employs an Epson PowerLite 410W short-throw (wide FOV) projector with 1280×800 (16:10) resolution, 2000 lumens brightness, and contrast ratio 500:1, located at a distance of approximately 3 m in front of a planar surface where it creates an illumination area of approximately 3.2×2 m. The projector continuously shows a time series of 10-bit binary Gray code patterns at a maximum refresh rate of 85 FPS.

Image grabbing is accomplished using two Point Grey Dragonfly Express cameras with 640×480 resolution. Each one is connected to a separate IEEE 1394b interface to allow triggering up to 168 FPS[2]. The camera parameters are calibrated using OpenCV functions. The environment camera is mounted on a Cosmicar/Pentax H416 (C60402) lens (4.2 mm, F1.6, $60.27° \times 47.87°$ FOV) at a camera-surface distance of approximately 3 m in front of the center of the illuminated area that is completely captured in the image.

A test subject is seated at a distance of 3 m in front of the surface. The eye camera is mounted on a Spacecom JF7.5M-2 lens (7.5 mm, F1.4, $30.80° \times 23.75°$ FOV) at a camera-eye distance d_{CE} of approximately 8 cm

[2]This is the maximum framerate for this camera model in asynchronous (trigger) mode.



(a) Photo of the setup (b) Corresponding schematic diagram

Figure 5.7: Experimental setup for eye gaze tracking using coded structured light.

on a table, and located slightly below an eye of the subject to not occlude the view of the surface. The camera captures close-up eye images under fixed head-position and varying gaze directions.

5.5.2 Results

In the following we analyze the performance of correspondence detection from corneal reflections using coded structured light under different thresholding methods, high framerates and environmental light. A fourth experiment analyzes the performance of eye pose estimation and PoR reflection calculation.

5.5.2.1 Thresholding Method

Raw image thresholding is the algorithmic step deciding the value of each bit in the time series and, with that, the error of the final result. Since there is a large deviation in the performance of different methods, the choice is important. We compare the four methods explained in Section 5.4.

We capture a series of eye images corresponding to a 10-bit Gray code with alternating pattern and complement frames. A large exposure time of 33.33 ms is applied to not contaminate the images with additional noise and introduce a bias. Figure 5.8 compares the results for different thresholding methods. Coded results are presented to allow a detailed visual inspection under an increased intensity distance between decoded neighboring coordinate values. The number of successfully decoded effective bits N_{eff} is calculated as in

$$N_{\text{eff}} = \left\lceil \log_2 \frac{R}{\Delta_{\min}} \right\rceil , \tag{5.14}$$

where R denotes the resolution of the encoded coordinate and $\Delta_{\min}$ the minimum distance between decoded coordinate values. Note, that noise reduction and outlier removal should be performed before calculating this measure.

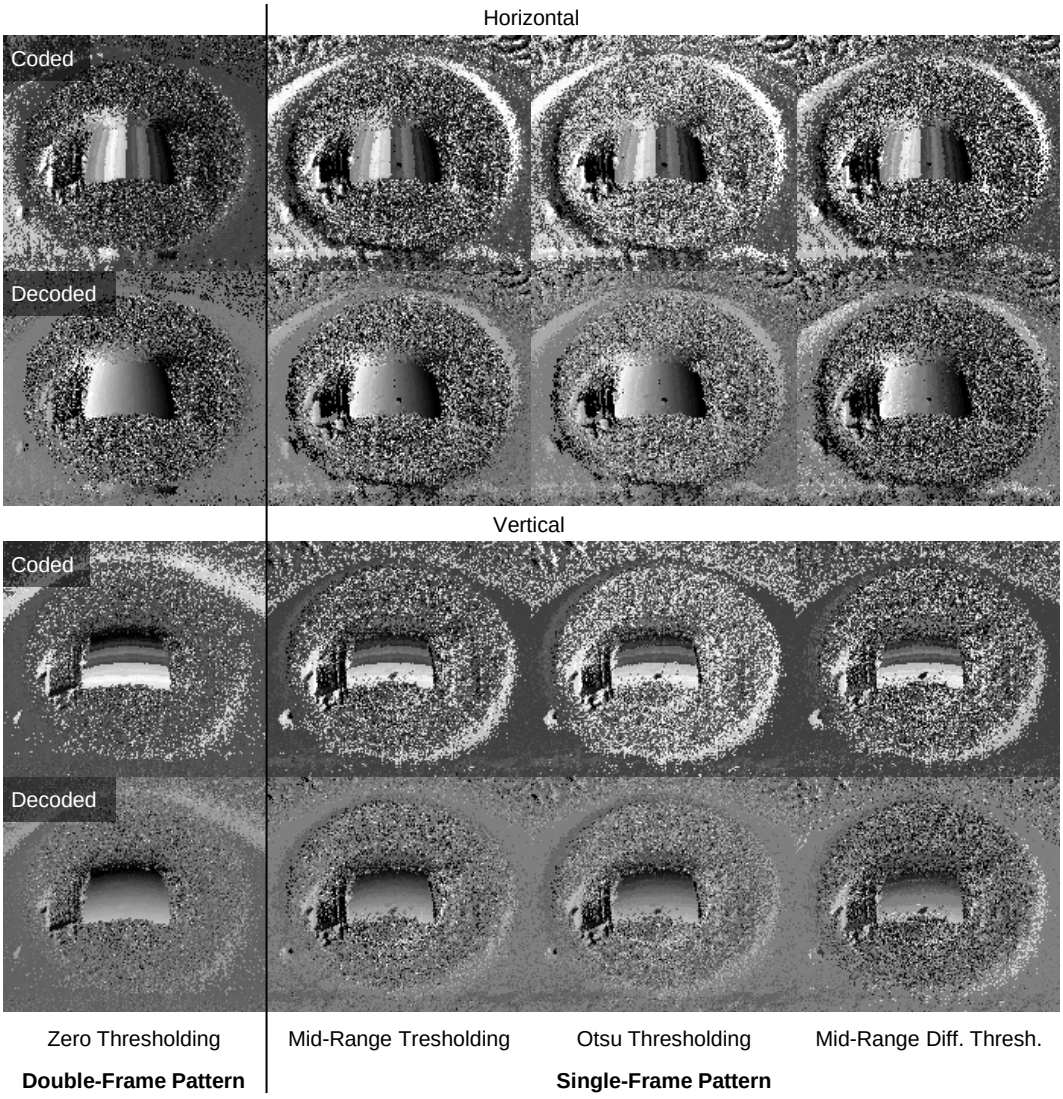


Figure 5.8: Experimental results for coded structured light correspondence detection comparing different thresholding methods. Results for cropped eye image regions corresponding to 10-bit Gray code captured with exposure time of 33.33 ms. (rows) Coded and decoded result after integrating the thresholded frames. (columns) Different thresholding methods, from left to right, (1) double-frame zero thresholding method (default), (2)–(4) single-frame methods based on pixel intensity statistics calculated over time series, (2) mid-range thresholding, (3) Otsu thresholding, and (4) mid-range difference thresholding.

For the described scenario up to $N_{\text{eff}} = 6$ bits are recovered for all methods, however, with large deviation in noise and accuracy. The double-frame zero thresholding method is most robust to noise and reduced contrast and, thus, achieves the best performance. It is the natural choice in case of imperceptible projection where a sequence of alternating pattern and complement is required. In case of invisible projection, the reconstruction framerate can be doubled by projecting only the pattern frames. This reduces the duration of a time series and, with that, motion blur for the result. Nevertheless, single-frame methods are based on calculating statistics over the time series of intensity values which introduces ambiguity and causes a larger error. Comparing the single-frame methods, mid-range thresholding performs best, followed by Otsu thresholding with slightly worse performance, and mid-range difference thresholding achieving the worst result. In terms of computational complexity, mid-range thresholding is also the most efficient of the three methods.

5.5.2.2 High Framerate

The user should not be aware of the pattern projection. With a visible-light projector this is achieved by alternating pattern and complement at high framerates. On the camera side this requirement results in short exposure times with increased noise level. To maximize the amount of captured light and increase the SNR, practical setups best apply a camera with a large image sensor. However, fast machine vision cameras usually come with a small sensor—1/3 in for the Point Grey Dragonfly Express. To handle this condition, we completely open the aperture and carefully adjust the depth of field to remove focus blur.

We want to analyze the impact of increasing framerates on the decoded result. Since software synchronization limits the system to 15 FPS, we simulate the effect of increasing framerate by decreasing exposure time from 33.33 ms (30 FPS) to 6.25 ms (160 FPS). Figure 5.9 compares both cases. At short exposure, raw images and decoded result contain a lot of noise. Nevertheless, the noise can be removed by applying a simple smoothing to the raw images. We tested Gaussian and median filtering with blur kernel size of 3×3, and obtain the best results using Gaussian filtering. On the other hand, images from long exposure contain less noise but motion blur. The quality of the result is comparable to the one from short exposure with noise reduction. Applying noise reduction for long exposure images is not necessary and does not lead to an improvement. Unlike expected we do not observe a performance impact of short exposure times $\geq$6.25 ms.

Due to the long duration of a complete time series with software synchronization the eye is not guaranteed to remain static. Eye movement leads to inconsistent pixels that do not correspond to the same object point over a complete time series, introducing motion blur into the decoded result which causes decrease in spatial resolution. A possible solution lies in model-based

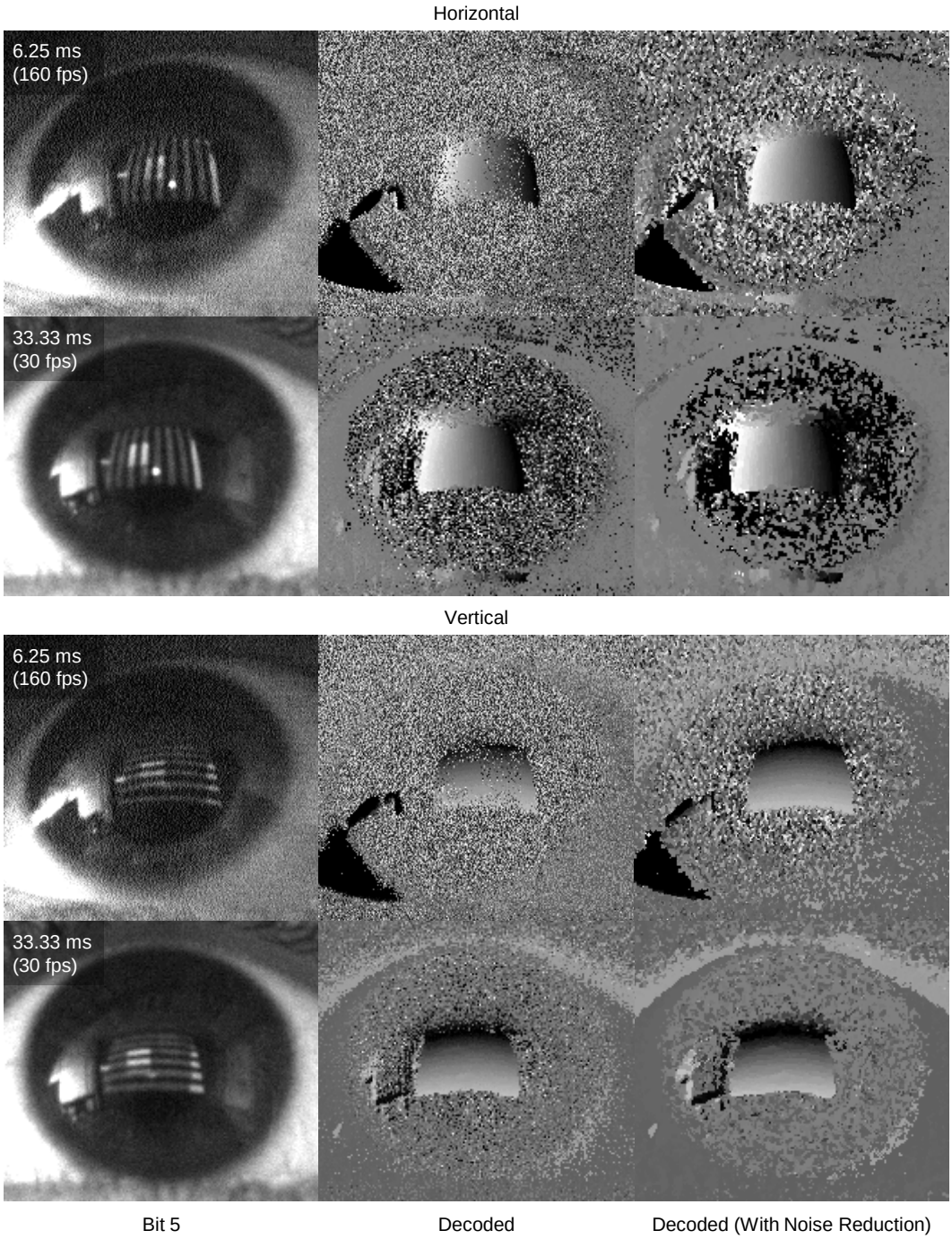


Figure 5.9: Experimental results for coded structured light correspondence detection comparing different exposure times. Results for cropped eye image regions corresponding to 10-bit Gray code captured with 6.25 ms (short) and 33.33 ms (long) exposure times. (short) The noise in the result is removed by applying a 3×3 Gaussian filter to the raw images. (long) Images contain less noise but motion blur. The quality is comparable to the one from short exposure with noise reduction.

motion compensation using estimated eye poses to register light paths between frames.

5.5.2.3 Environmental Light

Environmental light causes corneal reflections that interfere with reflections from controlled light sources and, therefore, has a direct influence on algorithms for eye feature detection and reflection analysis. Different effects can be distinguished: Diffuse illumination creates corneal and iris reflections that decrease contrast and SNR over the whole eye region, which leads to increased estimation noise. The iris texture produces a spatially varying effect. Direct illumination from distinct light sources create specular corneal reflection glints that disturb techniques based on glint detection, such as active light eye gaze tracking or the implementation of the display-camera calibration in Section 4.4. Therefore, eye analysis usually assumes low or absent environmental illumination.

The proposed strategy employs coded structured light to increase robustness to environmental light. We analyze this behavior by varying the number of area light sources at the ceiling of the experimental room from complete darkness to bright conditions. Figures 5.10 and 5.11 compare the results for encoded x- and y-coordinates respectively. To sample the whole dynamic range without changing camera parameters, we use a short exposure of 6.25 ms. With increasing light intensity, not only image noise decreases but also contrast. This leads to an overall increase in estimation noise, largely removed by applying a Gaussian filter to the raw images as suggested previously. While the performance still decreases with increasing light we observe that even at maximum light the correspondences are largely recovered. This result is due to robust double-frame thresholding that works with low contrast, as long as projector illumination produces a measurable camera response. Remaining noise and holes may be removed with an additional optimization strategy, for example, using graph-cuts on a Markov random field (MRF) representation.

We conclude that correspondence detection using coded structured light is affected by an increasing amount of environmental light. However, it still produces feasible results with potential for further optimization, in cases where simple methods completely fail.

5.5.2.4 Eye Pose Estimation and PoR Calculation

After several experiments on the recovery of coded correspondences from corneal reflections, we now want to analyze the performance of calculating the corneal reflection of the PoR in the eye image. Both results are necessary to determine the location of the PoR in the environment image.

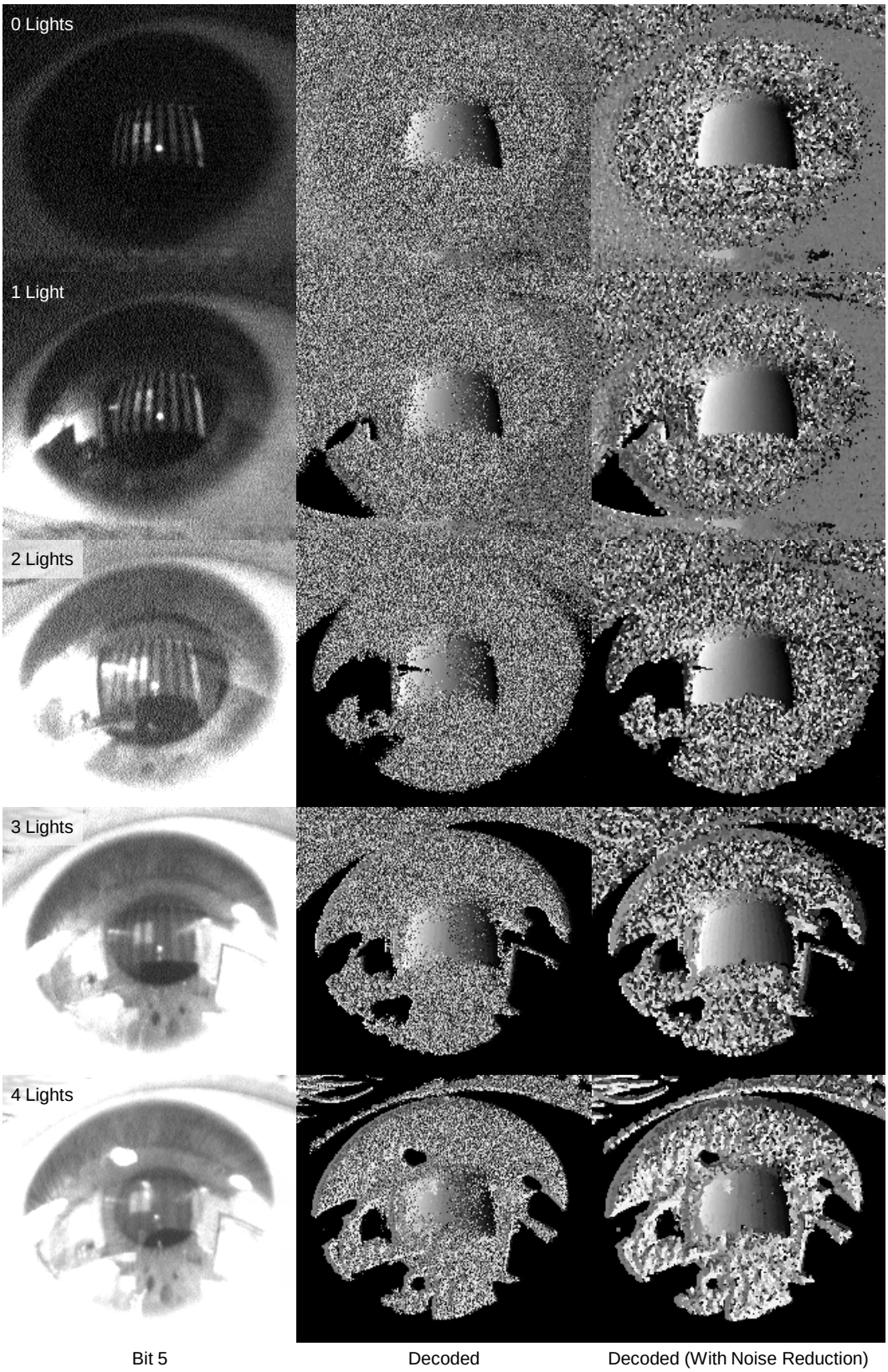


Figure 5.10: Experimental results for coded structured light correspondence detection under increasing environmental light (x-coordinate). (left) Single frame of 10-bit Gray code captured with short exposure time 6.25 ms. (middle) Estimation noise increases with light intensity. (right) Noise is largely reduced when applying a 3×3 Gaussian filter to the raw images.

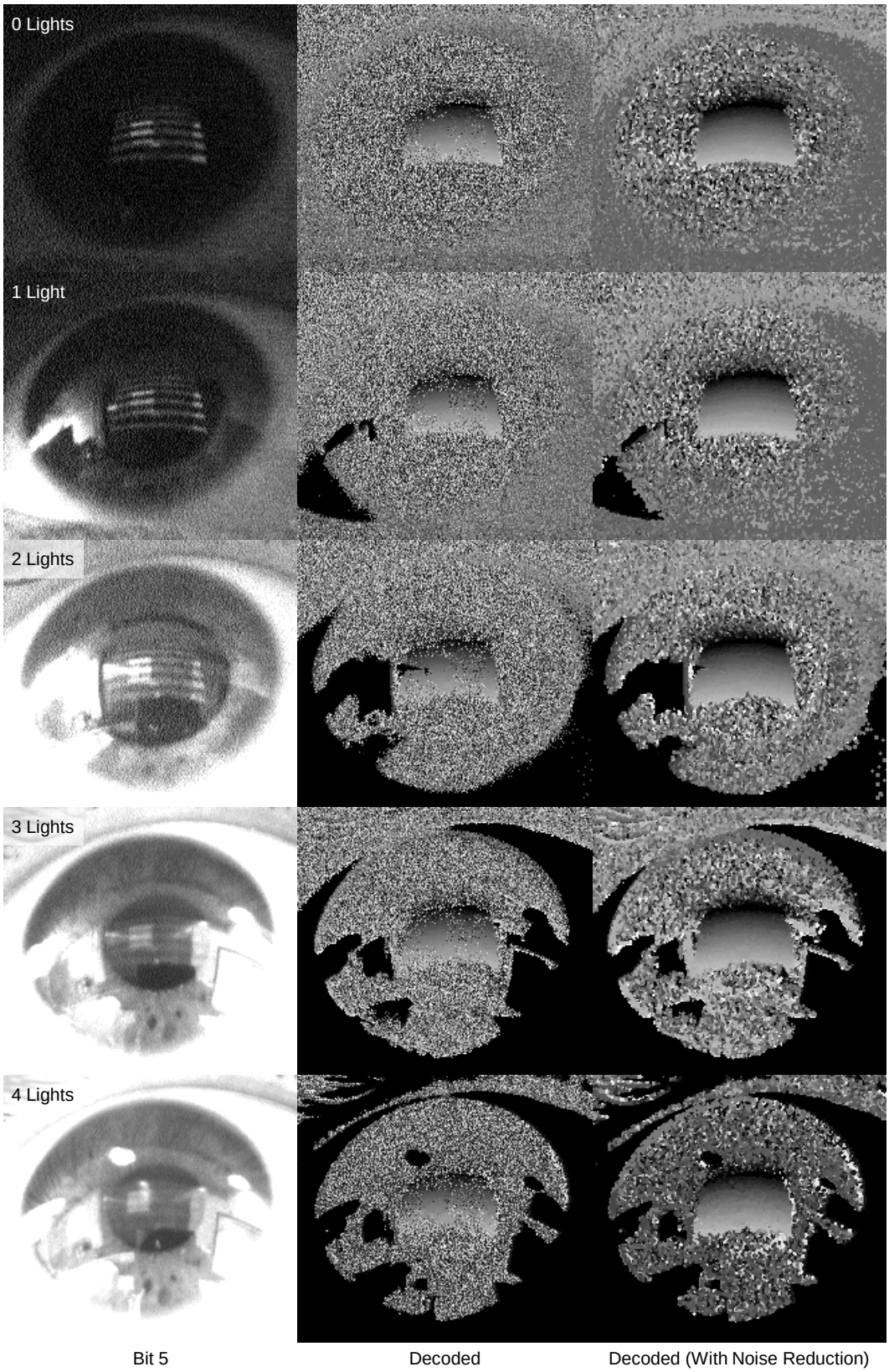


Figure 5.11: Experimental results for coded structured light correspondence detection under increasing environmental light (y-coordinate). (left) Single frame of 10-bit Gray code captured with short exposure time 6.25 ms. (middle) Estimation noise increases with light intensity. (right) Noise is largely reduced when applying a 3×3 Gaussian filter to the raw images.

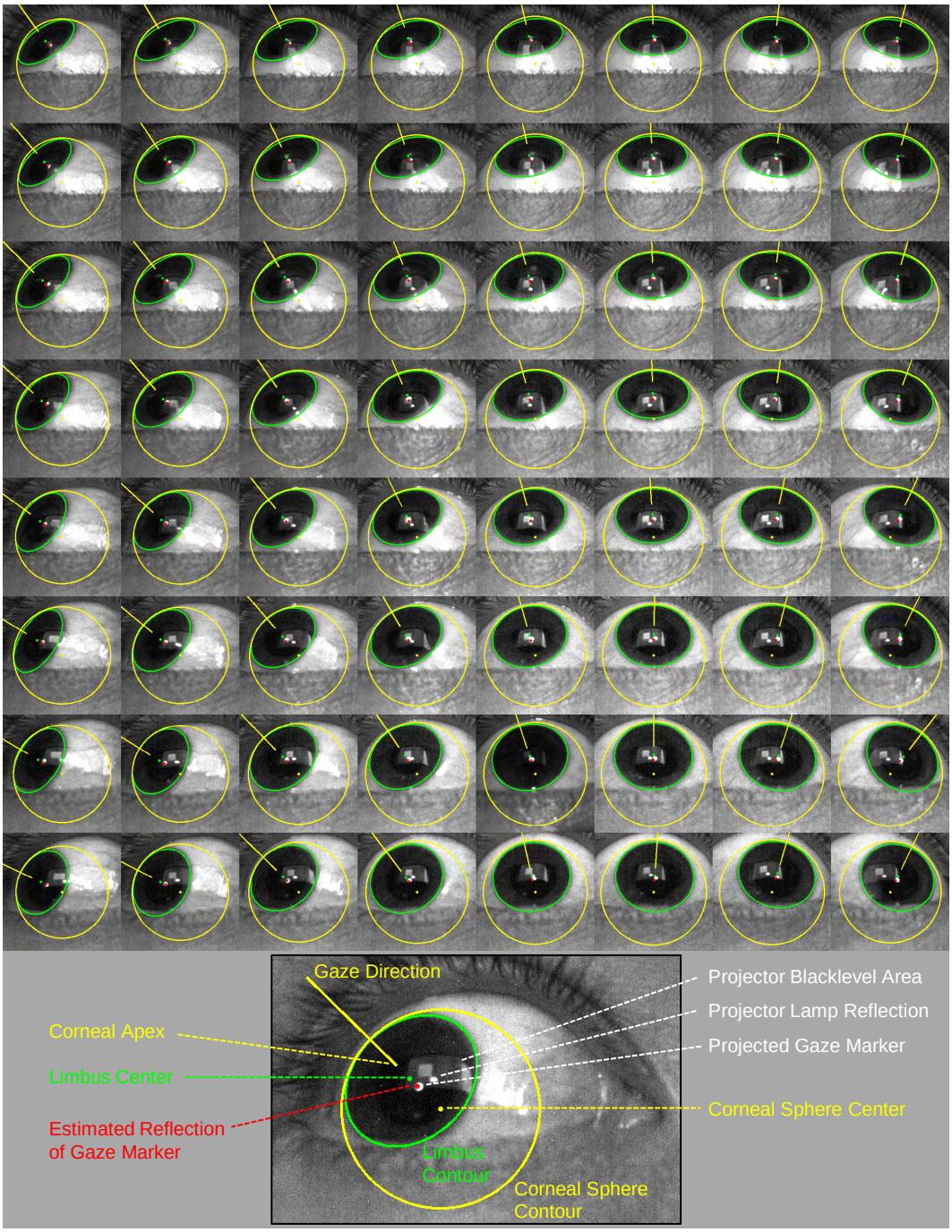


Figure 5.12: Experimental results for calculating corneal reflection of PoR. The figure shows cropped eye regions corresponding to 8×8 projected circular gaze markers on a planar wall in front of a test subject. For each image, we estimate the pose of the eye and the reflection of the PoR (red dot) that should match the ground-truth reflection of the gazed marker (white moving blob).

We project a sequence of a white circular gaze marker sweeping over 8×8 locations on the illuminated surface area. For each location, we capture the right eye of a test subject about 3 m in front of the surface and estimate the pose of the eye. Figure 5.12 shows the results for the 64 cropped eye regions centered at the projected contours of their estimated corneal spheres.

For each estimated eye pose we calculate the corresponding corneal reflection of the PoR. Section 3.3 describes five distinct methods depending on the knowledge about the distance between eye and PoR. Because the camera-eye distance d_{CE} of approximately 5 cm is much smaller than the eye-surface distance, we apply method five. The result is rendered as a red dot that should match the white moving reflection blob from the gaze marker. Due to noise in eye pose estimation we observe a varying deviation of the calculated PoR from the GT. However, the deviation never exceeds a threshold of 25 cm relating to a visual angle of 5°, with a mean of <3°.

5.5.2.5 Discussion

Correspondence detection using coded structured light performs robustly under practical requirements of high framerate and environmental light where simple corneal reflection analysis usually fails or results in high noise levels. Under severe conditions with noticeable reconstruction noise, a simple Gaussian filter applied to the raw images can largely improve the result. In the experimental scenario, where the major axis of the iris ellipse averages 150 px and the pattern reflection spans an area of approximately 55×35 px, up to $N_{\mathrm{eff}} = 6$ bits are effectively recovered for all tested conditions and methods.

Double-frame thresholding achieves the best results. It is the method of choice for imperceptible projection alternating pattern and complement. However, it doubles the duration of a time series, introducing motion blur as the eye does not remain static. This decreases spatial resolution, corresponding to a smaller number of effectively recovered bits. If that property is measured the number of encoding bits may be adaptively adjusted to maximize the effective framerate. Motion blur may also be removed with eye-model-based motion compensation. Under invisible projection, single-frame methods can be used to increase framerate and reduce motion blur. These methods, however, result in a higher overall error. Simple mid-range thresholding is found to perform best in terms of accuracy and computational complexity.

Computing the image location where the gazed PoR reflects at the cornea depends on the accuracy of eye pose estimation and the assumed distance between eye and PoR. Due to noise in eye pose estimation the accuracy of the obtained location usually varies. In the experimental scenario the deviation from the GT measures a maximum of <5° and a mean of <3° in visual angle.

5.6 Conclusion

5.6.1 Discussion

We proposed a novel architecture and method enabling eye gaze tracking in arbitrary dynamic environments without a dedicated geometric calibration. It achieves non-intrusiveness without user awareness by not requiring body-attachments and allowing larger tolerances for gaze angles, operation volume and movement speeds. The system comprises a remote camera for tracking a close-up view of an eye, and an environment camera capturing the scene where gaze should be detected. The task is to identify the location of the PoR in the environment camera image. The basic principle is described by the following findings:

- The corneal reflection of the PoR in the eye image can be calculated from the estimated pose of the eye.
- A mapping between eye and environment camera views can be obtained from identifying keypoint matches in corneal reflection and environment images corresponding to equal scene locations.
- The PoR in the environment camera image can be calculated based on a mapping defined by the obtained keypoint matches.

We explained that keypoint matching between eye and environment images is difficult due to the overlay of iris texture, the low reflectivity and the curved surface of the cornea. We further showed that a dense wide-area set of matches can be robustly detected using coded structured light projection. The pattern illumination is not noticed by the user and, therefore, does not distract operation, either using imperceptible projection where a fast sequence of alternating pattern and complement is projected with a standard digital video projector, or invisible projection where patterns are projected with a special device operating in IR light. The correspondences may be further applied to reconstruct a 3D surface model of the environment and obtain the 3D location of the PoR.

A number of comprehensive experimental studies demonstrated the effectiveness of the proposed approach analyzing its characteristics under varying conditions. Important conclusions are:

- The use of coded structured light enables robust correspondence detection under the practical requirements of environmental light, high framerates, and image degradation, where simple active-light methods would fail.
- Image thresholding to recover projected information from the image is the most crucial step in correspondence decoding.

  - Double-frame thresholding provides inherent robustness. It is required for imperceptible visible-light projection. It is generally recommended under challenging conditions when half effective framerate and reduced spacial resolution from eye movement are acceptable.
  - Single-frame thresholding is a non-robust and ambiguous strategy, and only recommended under low framerate and large eye movement without motion compensation.

- The wide area, the high spatial resolution, and the robustness of the detected correspondences allow for an accurate mapping between eye and environment images.

- Eye pose estimation has the highest impact on the overall result. Passive limbus-based eye pose estimation suffers from an increased noise and a larger error compared to hardware-intense methods using active light and multiple cameras. Nevertheless, the error for the current study does not exceed 5° in visual angle.

5.6.2 Implications

We believe that this work has the potential to facilitate novel developments in the community and helps to generally increase usability and acceptance of applications "outside the laboratory". The method provides a unique combination of characteristics, enhancing usability in conventional applications and enabling eye gaze tracking for novel scenarios and tasks.

Arbitrary Environments The system is the first to automatically support arbitrary scene geometry and robustly work under challenging illumination conditions. These characteristics enable a paradigm-shift in eye gaze tracking, commonly limited to controlled planar surfaces such as monitors or projection screens. Achieving robustness along with an easy setup, eye gaze tracking becomes available to everyday environments without the need of experienced supervision.

Calibration-free Applications Since an interactive geometric calibration is not necessary, the system can be applied in situations that do not allow for a dedicated calibration procedure. This could be for any of several possible reasons: A lack of time, when attention is required for the task where eye tracking should be applied to; a lack of ability, in unsupervised conditions involving non-expert users, physically/mentally disabled persons, and children; or a prevention of awareness, where either seamless integration is required, or technical details and indicators of operation cannot be exposed.

Dynamic Setups The absence of a dedicated calibration procedure allows applications with changing geometric relation between system components. Examples include user tracking, mobile systems, and dynamic objects in real environments.

The proposed method provides robust unobtrusive eye gaze tracking in real environments, automatically determines scene structure, and links both information. This may be beneficial for applications in different fields:

Human–Computer Interaction Requirements and characteristics of conventional systems limit eye gaze tracking to laboratory environments with experiences users. The proposed system has the potential to facilitate practical interfaces due to its easy setup and unobtrusive operation, enabling many applications (Duchowski, 2002; Hammoud, 2008; Hansen and Ji, 2010). In some situations, eye tracking can support or even be superior to conventional interaction techniques, resulting in faster and less fatiguing task completion. Advancements in computational systems, devices and architectures, demand for novel forms of interaction. The proposed method supports ad-hoc usage in uncontrolled dynamic scenarios with arbitrary surface geometry and illumination conditions, enabling seamless gaze-based interaction in ubiquitous and ambient living spaces. Practicable and robust eye gaze tracking can further help elderly and disabled people to maintain their independence in various areas, making it an essential tool for communication and mastering the daily life (Donegan et al., 2005; Daunys et al., 2006).

Children and Infants Eye gaze tracking with children and infants is important for diagnostic applications, such as studies on visual exploration related to provided stimuli or in natural interaction (Gredebäck et al., 2010). Gaze information is important for understanding the process of human development, and for detecting possible deficits, disorders or disease patterns, such as Autism (Boraston and Blakemore, 2007). Conventional systems, however, cannot be applied to children and infants since tedious calibration or head fixation are not feasible, and head-mounted trackers are too large and heavy. Specially developed solutions still require setup efforts and experienced operators (Guestrin and Eizenman, 2008; Franchak et al., 2010; Gredebäck et al., 2010; Noris et al., 2010). The proposed architecture combines novel features to support application with children and infants, for example remote operation without body-attachments and head fixation, absence of calibration, applicability in arbitrary environments, and reduced obtrusiveness through relaxed operation conditions.

Diagnostic Studies In the past, eye tracking has been applied for diagnostic studies in many disciplines including cognitive science, psychology,

medicine, industrial engineering and marketing research (Duchowski, 2002). The requirement that measurement equipment and conditions do not interfere with the task, or do not affect the subject, often leads to problems with conventional techniques where experimental setups are compromised or results become biased. The proposed technique enables unobtrusive application for diagnostic studies in natural environments where easy setup and operation allow researchers to focus on their target. Linking gaze information with scene structure in arbitrary environments allows visual and interactive data exploration, possibly leading to novel insights and understanding.

Data Mining Diagnostic studies in eye tracking commonly observe reactions to predefined stimuli. With the availability of eye gaze tracking in arbitrary environments, the resulting data can be exploited with machine learning and data mining techniques. Such an approach could be beneficial for applications in several fields, such as the understanding of human problem solving in algorithm design; the understanding of human behavior, cognitive and affective states in machine and robot interaction; and the analysis of the correlation between gaze trajectory and scene information in cognitive science, psychology, and medicine.

Practical Conditions The proposed technique can be applied to practical problems in natural dynamic environments, such as industrial engineering and human factors analysis with real non-planar objects; marketing research and advertising analysis in physical shopping situations; surveillance and security applications under remote and imperceptible conditions; and driver assistance systems with seamless integration of near and far gaze targets.

5.6.3 Limitations

The scope of the explained implementation is to provide a first proof-of-concept and allow experimental verification of the key characteristics of the proposed technique. There are several limitations that need to be considered when applying the described prototype within real conditions. Implementation details for a comprehensive system also largely depend on the requirements of the particular application. Necessary extensions include

- a strategy for calibrating the internal parameters of the cameras (Hartley and Zisserman, 2003; Zhang, 2000; Bouguet, 2010),
- a PTZ camera system to track a close-up region of an eye with corresponding calibration strategy (Oike et al., 2004; Yoo and Chung, 2005; Reale et al., 2010),

- a parameterization for the Condensation algorithm (Isard and Blake, 1998) or a different algorithm for robust and accurate eye feature tracking in a video sequence,
- a more accurate eye pose estimation, probably using an active-light approach by integrating reconstructed scene geometry and a combination of eye features, such as pupil and iris center/contour, and iris texture under high resolution,
- a more accurate geometric model for the surface of the cornea, either using an aspheric model based on anthropometric statistics, or parameterizing an individual shape by exploiting the projected correspondences,
- a choice of an appropriate pattern coding strategy, probably supporting for spatio-temporal tradeoff (Battle et al., 1998; Salvi et al., 2004, 2010),
- a strategy for robust pattern decoding to reduce noise and holes, probably by integrating context information such as iris texture (Wang et al., 2008) and geometric scene constraints,
- a strategy for eye motion compensation in temporal codes based on eye pose and reflection modeling, and
- an extension to handle more complicated light paths, e.g., when users wear glasses.

5.6.4 Future Work

Beside the described limitations that need to be tackled to turn the current prototype into a practical system, there is requirement for future research. Four concrete ideas are outlined in the following.

5.6.4.1 Correspondence Coding Strategy

The described implementation uses a binary time-multiplex coding strategy. The major advantage is that this achieves the highest possible SNR and spatial resolution. The disadvantage, however, is that the information for a single effective frame is distributed along the temporal domain into a series of pattern frames, requiring the geometric relation between cameras, eyes and scene to remain static for the duration of a series. Regarding the proposed architecture for uncalibrated eye gaze tracking in arbitrary environments, the geometry can change at any time due to camera movement, eye movement, and scene deformation. This produces motion blur among the series of images where different light paths—carrying information from different codewords—mix at a particular image pixel. The effect acts as a low-pass filter, leading to a

decrease of spatial resolution for the recovered coded area and, thus, a decrease of spatial resolution in PoR estimation. However, the method will still work.

Possible solutions include motion compensation, increase of framerate, or shift from time- towards space-multiplex coding. In case of a fixed relation between camera and scene with the eye being the only moving object, a model-based motion compensation strategy using eye pose information can be applied to reconstruct the light path geometry and re-arrange the code information. To generally compensate for motion, the simplest strategy is to increase the framerate of the projector-camera system. This mainly depends on the capability of the hardware as decoding is found to work effective with short exposure times. Another strategy, that can be combined with increased framerate, is to dynamically balance between time- and space-multiplex coding based on the degree of motion, for example by evaluating the error rate in decoding. A pure space-multiplex strategy encodes all information into a single pattern and does not suffer from the described effects of motion blur. Nevertheless, this is commonly achieved at the cost of SNR and spatial resolution what may be crucial when dealing with difficult geometric and photometric characteristics in arbitrary environments and eye modeling. A possible solution are novel one-shot methods dynamically adapting coding parameters based on system requirements (Koninckx and Van Gool, 2006; Sagawa et al., 2009; Salvi et al., 2010).

5.6.4.2 Invisible-Light Pattern Projector

For a practical system it is necessary that the user is not distracted by pattern projection. Using imperceptible techniques with standard digital video projectors we observe that the dynamic pattern is still evident at 120 Hz, the maximum refresh rate of common stereoscopic projectors. To examine this effect at higher speeds we assembled a row of white-light LEDs, each one equipped with a manual-focus lens. The LEDs can be independently triggered by a programmable controller with framerates up to 500 Hz. The projection device is placed in front of a white planar surface. We asked test subjects to identify noticeable artifacts while randomly switching between a time series of alternating pattern and complement, and a permanent illumination with average pattern intensity. Gradually increasing framerate we found that flicker becomes hardly noticeable above 200 Hz, however, was still perceived above 300 Hz by a few subjects. Beside the perceptibility of flicker, we generally found projector illumination more distractive then common room illumination due to characteristics such as color temperature, sharp boundary, and shadows.

While the observed effects require elaboration and careful adjustment to set up imperceptible pattern projection, another strategy is to use invisible light. Infrared light is appealing as it is contained in the spectrum of projector illumination and can be recorded with cameras, while at the same time be-

ing invisible to human observers. To evaluate the feasibility of this approach we removed the IR cut-off filter of the camera and mounted an IR pass filter in front of a standard digital video projector with a brightness of 4000 lumens. We found the intensity of the remaining IR component highly insufficient for the intended usage. It is, therefore, necessary to develop a special high-intensity IR projecting device. A promising architecture is the described approach using a matrix of IR LEDs allowing for independent adjustment of orientation and focus to provide a flexible tradeoff between projection area and spatial resolution.

5.6.4.3 Eye Pose Estimation based on Coded Structured Light

The current implementation of eye pose estimation employs a passive limbus-based strategy suffering from an increased error compared to conventional active systems using at least two point light sources (Shih et al., 2000; Guestrin and Eizenman, 2006). While providing better results, these methods need additional hardware and involve a dedicated geometric calibration, or a fixed arrangement of light sources in an apparatus where a small baseline can again cause large errors. Nevertheless, there is a way to use the advantages of additional light sources in combination with the proposed architecture, to increase the accuracy in eye pose estimation:

Reconstructing 3D scene structure from correspondences as explained in Section 5.3.7 creates a depth map in the view of the environment camera. Aligning the surface model with respect to the corneal reflections in the eye image (e.g., by minimizing the re-projection error), the point-based model acts as a large set of constraints that can be exploited for eye pose estimation or parameterization of an individual shape model of the corneal surface. The advantages of coded structured light projection can achieve performance improvement to conventional methods based on active light.

The gaze direction is recovered from the estimated center of the cornea and an additional point on the optical axis, such as the centers of pupil and iris. The pupil comes with the advantage of being tracked and segmented with high accuracy and robustness due to its sharp edge, where the rough location is identified from the detected correspondences within the boundary of the iris region.

5.6.4.4 Tracking Additional Information

The task in eye gaze tracking is to track the PoR revealing the particular location a person is looking at. Light from that location enters the eye and projects onto the fovea, the part of the retina that achieves the maximum acuity of vision due to the highest concentration of cone photoreceptors. Foveal vision enables a person to gather detailed information directly related to accomplishing a particular mission or goal. Nevertheless, the fovea only subtends

the central 2° of the visual field, where the entire field spans approximately 180° in the horizontal and 130° in the vertical direction (Duchowski, 2007, pp 30).

The fovea is enclosed by the parafovea as the zone with high acuity, extending to about 4–5°, followed by a sudden drop-off in acuity for areas beyond. These peripheral areas are important for certain functions in visual perception, such as the recognition of known structures or the identification of similar structures and movements. Peripheral vision further delivers context information for visual perception and helps in planning eye movement to control detailed vision. It is, therefore, necessary to not only track the PoR but also provide the location of surrounding zones of vision to develop a comprehensive understanding of human visual perception. Using the proposed eye gaze tracking architecture the entire human visual field may be extracted within arbitrary environments.

Chapter 6
Conclusion

This work combines a range of contributions for corneal reflection and environment relation analysis from multiple eye images. Two methods were proposed to solve a combination of problems for practical application in display-camera calibration and eye gaze tracking. Integrated with these methods are the solutions to two general problems in corneal reflection analysis and scene reconstruction that can be relevant to other work. It follows a short summary of the contribution and findings made by this work.

Display-camera Calibration from Eye Reflections. A novel technique was proposed for calibrating the geometric relation of display-camera setups using corneal reflections. Since extensive experimental evaluation showed that the straightforward geometric reconstruction results in a large error, an optimization framework that exploits geometry constraints in the scene was developed. The technique makes display-camera calibration substantially more practical and leads to several benefits compared to previous works:

- Since no additional hardware is necessary the method can be distributed online and applied with existing off-the-shelf setups.
- The calibration is performed automatically without interaction and awareness, enabling non-expert or disabled persons and children, or, situations where it is not desired to disclose technical details.
- Accuracy increases with the number of images used. Nevertheless, the minimum requirement is a single face image. This enables online calibration of dynamic setups and allows applications such as camera tracking.
- Beside reconstructing the pose of the display, the method also estimates eye locations. This enables to realize display-based eye gaze tracking systems without dedicated calibration.

The technique increases the number of potential application scenarios. Moreover, the developed optimization framework can be of general relevance to other work when knowledge about the scene in form of geometry constraints is available.

Calibration-free Non-intrusive Eye Gaze Tracking in Arbitrary Environments. A novel system architecture for eye gaze tracking was proposed to overcome several limitations of existing techniques. It involves the following advantages to state-of-the-art remote eye gaze tracking:

Calibration-free A geometric calibration to obtain and align a 3D surface model with the camera-eye reference frame is not required. The information is automatically obtained through coded structured light. Moreover, this approach supports scenarios with dynamic scene structure and hardware poses.

Attachment-free Head attachments as applied in portable eye gaze tracking systems are not necessary. Additionally, since gaze-mapping calibration is not required, there is no error accumulation with changes in head-camera relation and no performance impact when the user moves away from an initial surface. The system is non-intrusive and does not require awareness.

Arbitrary environment A planar surface, such as computer monitor, projection canvas, and wall is not required. Instead, surfaces with arbitrary geometry forming the majority of our environment are naturally supported. This enables more realistic scenarios, and future ubiquitous and ambient environments.

Free head movement Compared to stationary systems, relaxed operation conditions in terms of gaze-angle and viewing volume allow for free head-movement.

Challenging conditions A flexible use of coded structured light allows increased tolerances for environmental light and image quality. The benefits are more reliable data, an easy setup, and practical application conditions. This enables eye gaze tracking under natural conditions, for non-professional and untrained users.

Improved accuracy The PoR is calculated with an image-based method. It is not affected by the distance to the scene resulting in an improved accuracy. Moreover, the inherent negative effect of distance on image resolution can be compensated with the high spatial resolution obtained using coded structured light.

The proposed architecture and method have the potential to increase the usability and acceptance of eye gaze tracking "outside the laboratory" under conditions of practice. A range of application scenarios and implications was discussed.

Accuracy of Scene Reconstruction from Eye Images. Applying the display-camera calibration framework, a large number of comprehensive experimental studies were conducted to understand the impact of individual system parameters on the overall accuracy using real and synthetic data.

- Individual eye geometry:
  - There is a significant impact of individual eye geometry.
  - The impact of aspherical corneal size variation in one particular dimension is about one magnitude higher than the impact of spherical size variation in all three dimensions. The error can be reduced by increasing the spatial distribution of cornea positions parallel to the image plane.
- Camera specifications and image quality:
  - With increasing image resolution, accuracy increases non-linearly and converges. Resolutions provided by current hardware achieve sufficiently low impact for common setups.
  - With increasing noise, accuracy decreases non-linearly. The effect can be compensated by applying noise reduction techniques.
- Geometric relation between camera, eyes, and scene:
  - With increasing distance between eyes and camera/scene, the error proportionally increases.
  - With increasing gaze angle, the error first increases gradually, and then rapidly from approximately 25°.
  - Due to the large overall error, an increasing number of eye images does not lead to a considerable increase in accuracy.

The findings provide a tool to assess the quality that can be expected for a particular setup and provide an aid for the decision to where compensation strategies are best applied.

The large overall error obtained with straightforward geometric modeling lead to the development of an optimization framework performing joint refinement of eye poses and scene structure using known geometry constraints in the scene. Further evaluation was conducted to analyze the resulting performance improvement:

- Optimization of results:
  - The accuracy in scene reconstruction and eye pose estimation is largely improved.
  - The tolerance to system parameters is considerably increased.

  - The error decreases and converges with an increasing number of eye images.

- An inherent two-way ambiguity in eye pose estimation is resolved.

The accuracy in corneal reflection analysis and geometric modeling highly depends on the quality of image-based eye pose estimation. While allowing application with off-the-shelf hardware in everyday environments, passive methods are sensitive because of unknown eye parameters and errors in pupil/iris detection. They generally do not achieve the accuracy and automation possible with active-light methods that should be favored when compatible with the application scenario.

More complex geometric eye models have to be tested in order to better approximate the shape of the eye as the eyeball is slightly flattened in the vertical plane (Snell and Lemp, 1997) and corneal topology is complex (Bogan et al., 1990). According to anatomic studies and experimental evaluation within this work, it can be beneficial to model the eye as two intersecting ellipsoids and include its radii as shape parameters in the proposed optimization framework. The aspect ratio of the eyeball can be calibrated from a single iris image where the user looks directly into the camera. Strategies for calibration of each individual's eye geometry may lead to further improvements.

It is also interesting to consider a reorganization of the system: This work estimates the pose of a pre-defined constant eye model and applies corneal reflection analysis to the reconstruction of an unknown scene structure. Related to the idea in Section 5.6.4.3, robust corneal reflection analysis in conjunction with a known scene structure may be beneficial to improve eye pose and parameter estimation itself.

Accurate and Robust Correspondence Matching among Multiple Eye and Scene Images. An accurate and robust strategy for matching correspondences among multiple eye images is important for accurate and dense geometric modeling. This work proposed a solution based on coded structured light projection. Particular advantages of this strategy are the following:

- A higher inherent accuracy is achieved compared to methods relying on passive feature extraction and stereo epipolar geometry. The method is purely image-based and, thus, independent of intrinsic and extrinsic calibration, scene geometry, and complexity of the underlying light path.

- A considerable increase in robustness is achieved compared to passive feature matching in eye images, suffering from superimposed iris texture, dynamic range compression, and geometric distortion.

- A considerable increase in robustness is achieved under challenging practical conditions, such as environmental light, short exposure, and image noise.
- A dense area of matches is achieved through increased spatial resolution and interpolation/extrapolation techniques.
- A wide area on the corneal surface is covered. Compared to common point light sources with small baseline, light is flexibly projected into a wide environment with the ability for stacking multiple projectors.
- Using imperceptible techniques or invisible light, the dynamic projection is not perceived by human observers. Imperceptible codes can be removed from the image data to restore the raw texture of the scene without additional equipment.

Matching is accurately performed for more than two eye images, which is the limitation of the previous approach Nishino and Nayar (2004b, 2006). Moreover, it naturally supports eye and scene images allowing to integrate information from eye image processing with high-quality images data.

This thesis proposed the idea to analyze corneal reflections of environmental light to relate the individual (eye) with its environment. We discussed that this creates a basis for lightweight non-intrusive techniques that require tracking of camera, eye, and scene pose, and reconstruction of eye and scene properties. Integrating all information into a comprehensive framework can lead to mutual benefit for different tasks.

We showed implications to a wide range of areas including human–computer interaction, scene/object reconstruction, surveillance/forensics, medicine, cognitive science, psychology, industrial engineering, and marketing research. The established link between eye and environment information can lead to novel insights when analyzed with data mining techniques, and provide a capable basis for future smart sensors in an ambient environment.

There still remains a large number of problems to be solved in order to achieve the overall goal of accurately and robustly estimating geometry and light field from multiple eye and scene images, with dynamic camera and scene pose, and non-rigid eye and scene structure. However, with the novelties and findings of this work we make a step into that direction.

Appendix A

Ellipse

A.1 General Equation

A circle with radius r, centered at the origin of the coordinate system $(0, 0)$, can be described by the canonical implicit equation

$$\frac{x^2 + y^2}{r^2} = 1, \tag{A.1}$$

where (x, y) denote the coordinates of a point lying on the circle. A circle is a special form of an ellipse where the semi-major axis a and the semi-minor axis b are equal to radius r. Thus, replacing r leads to the canonical implicit equation of an ellipse where the major axis is aligned with the x-axis

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{A.2}$$

General Implicit Equation. Any arbitrary ellipse can be obtained from the canonical ellipse by a rotation and a translation, leading to the general implicit equation

$$\frac{\left((x - x_0)\cos\phi + (y - y_0)\sin\phi\right)^2}{a^2} + \frac{\left(-(x - x_0)\sin\phi + (y - y_0)\cos\phi\right)^2}{b^2} = 1, \tag{A.3}$$

where (x_0, y_0) denote the coordinates of the center and ϕ is the counterclockwise angle of rotation from the x-axis to the major axis of the ellipse.

General Parametric Form. The corresponding parametric form of an ellipse is expressed as in

$$\begin{aligned} x(t) &= x_0 + a\cos t\cos\phi - b\sin t\sin\phi, \\ y(t) &= y_0 + a\cos t\sin\phi + b\sin t\cos\phi, \end{aligned} \tag{A.4}$$

where $t \in [0, 2\pi)$ parameterizes the counterclockwise path along its arc.

A.2 Ellipse as a Conic Section

Ellipses are closed curves that represent the bounded case of the conic sections. These are the curves that arise from the intersection of a circular cone and a plane not passing through its apex. In Cartesian coordinates, a conic section can be described by a quadratic equation in two variables

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0, \tag{A.5}$$

where A, B, C are not all zero. The value of the discriminant $B^2 - AC$ classifies the type of the conic section. If the conic is non-degenerate, and if $B^2 - AC < 0$, the equation represents an ellipse. If also $A = C$ and $B = 0$, the equation represents a circle. Expansion and re-organization of the general implicit equation of the ellipse (eq. (A.3)) leads to the equation of the conic section (eq. (A.5)), where the coefficients are given as in

$$\begin{aligned}
A &= \frac{(\cos\phi)^2}{a^2} + \frac{(\sin\phi)^2}{b^2}, \\
B &= \frac{\sin\phi\cos\phi}{a^2} - \frac{\sin\phi\cos\phi}{b^2}, \\
C &= \frac{(\sin\phi)^2}{a^2} + \frac{(\cos\phi)^2}{b^2}, \\
D &= \frac{(-x_0\cos\phi - y_0\sin\phi)\cos\phi}{a^2} - \frac{(x_0\sin\phi - y_0\cos\phi)\sin\phi}{b^2}, \\
E &= \frac{(-x_0\cos\phi - y_0\sin\phi)\sin\phi}{a^2} + \frac{(x_0\sin\phi - y_0\cos\phi)\cos\phi}{b^2}, \\
F &= \frac{(-x_0\cos\phi - y_0\sin\phi)^2}{a^2} + \frac{(x_0\sin\phi - y_0\cos\phi)^2}{b^2} - 1.
\end{aligned} \tag{A.6}$$

Matrix Notation. Using homogeneous coordinates, the equation can be described in matrix notation as

$$\mathbf{x}^{\mathrm{T}}\mathsf{Q}\mathbf{x} = 0, \tag{A.7}$$

where

$$\mathsf{Q} = \begin{bmatrix} A & B & D \\ B & C & E \\ D & E & F \end{bmatrix} \tag{A.8}$$

is a symmetric matrix representing a conic section, and $\mathbf{x} = (x, y, 1)^{\mathrm{T}}$ is a point on its boundary. The discriminant to determine the type of conic section is represented as in

$$B^2 - AC = -\begin{vmatrix} A & B \\ B & C \end{vmatrix}. \tag{A.9}$$

Ellipse Parameters. If a conic section is an ellipse, the five parameters a, b, x_0, y_0, and ϕ can be recovered from equation (A.5) (Weisstein, 2011b). Let us therefore define

$$\begin{aligned} \Delta &= \begin{vmatrix} A & B & D \\ B & C & F \\ D & F & G \end{vmatrix}, \\ J &= \begin{vmatrix} A & B \\ B & C \end{vmatrix}, \\ I &= A + C. \end{aligned} \tag{A.10}$$

Assuming the conic section is an ellipse, it holds $\Delta \neq 0$, $J > 0$, and $\Delta / I < 0$. Further assuming that the ellipse is non-degenerate (i.e., it is not a circle, so $A \neq C$, and it is not a point, so $J = AC - B^2 \neq 0$), then the lengths of the semi-axes a and b are

$$\begin{aligned} a &= \sqrt{\frac{2\left(AF^2 + CD^2 + GB^2 - 2BDF - ACG\right)}{\left(B^2 - AC\right)\left[\sqrt{(A-C)^2 + 4B^2} - (A+C)\right]}}, \\ b &= \sqrt{\frac{2\left(AF^2 + CD^2 + GB^2 - 2BDF - ACG\right)}{\left(B^2 - AC\right)\left[-\sqrt{(A-C)^2 + 4B^2} - (A+C)\right]}}, \end{aligned} \tag{A.11}$$

the center of the ellipse (x_0, y_0) is given by

$$\begin{aligned} x_0 &= \frac{CD - BF}{B^2 - AC}, \\ y_0 &= \frac{AF - BD}{B^2 - AC}, \end{aligned} \tag{A.12}$$

and the counterclockwise angle of rotation from the x-axis to the major axis of the ellipse is

$$\phi = \begin{cases} 0 & \text{if } B = 0 \text{ and } A < C, \\ \frac{1}{2}\pi & \text{if } B = 0 \text{ and } A > C, \\ \frac{1}{2}\cot^{-1}\left(\frac{A-C}{2B}\right) & \text{if } B \neq 0 \text{ and } A < C, \\ \frac{1}{2}\pi + \frac{1}{2}\cot^{-1}\left(\frac{A-C}{2B}\right) & \text{if } B \neq 0 \text{ and } A > C. \end{cases} \tag{A.13}$$

A.3 Degrees of Freedom

An arbitrary ellipse in the plane has five degrees of freedom, defining its position, orientation, shape, and scale. These degrees are represented, for

example, by the coefficients a, b, x_0, y_0, and ϕ. The symmetric matrix Q representing a general conic section has six different elements. However, only their five ratios are unique since adding a scale parameter s does not change the equation $s\mathbf{x}^{\mathrm{T}}\mathsf{Q}\mathbf{x} = 0$. Therefore, the five degrees of a general conic section can be represented by the coefficients A', B', C', D', and E', which are the coefficients in equation (A.5) normalized by $1/F$.

A.4 Least Squares Estimation from a Set of Points

Since an arbitrary ellipse has five degrees of freedom and is represented by a quadratic equation in two variables, five coplanar but non-collinear points on its boundary are sufficient to uniquely determine the conic coefficients. It follows a simple algorithm to estimate a linear least-squares best-fit ellipse from $N \geq 5$ points, representing (probably noisy) measurements of its boundary.

Each point (x_i, y_i) places one constraint on the conic coefficients which can be written by re-organizing equation (A.5) as

$$\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \end{bmatrix} \mathbf{q} = 0, \tag{A.14}$$

where $\mathbf{q} = (A, B, C, D, E, F)^{\mathrm{T}}$ is a vector containing the unknown coefficients. Stacking all N constraints leads to a homogeneous system of linear equations (in the unknown coefficients)

$$\mathsf{A}\mathbf{q} = 0, \tag{A.15}$$

where matrix $\mathsf{A}_{N\times 6}$ is given as in

$$\mathsf{A} = \begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ & & \vdots & & & \\ x_N^2 & x_N y_N & y_N^2 & x_N & y_N & 1 \end{bmatrix}. \tag{A.16}$$

The conic vector $\mathbf{q}$ is the null space of matrix A. In the more general case that allows the points to contain noise, $\mathbf{q}$ is the vector that minimizes $\|\mathsf{A}\mathbf{q}\|$. The least squares solution $\hat{\mathbf{q}}$ of the equation system is the eigenvector corresponding to the smallest eigenvalue of the square matrix $\mathsf{A}^{\mathrm{T}}\mathsf{A}$. It is obtained as the last column of matrix V where $\mathsf{A} = \mathsf{U}\mathsf{D}\mathsf{V}^{\mathrm{T}}$ is the Singular Value Decomposition (SVD) of A (Hartley and Zisserman, 2003).

For more sophisticated approaches on closed-form solutions for least-squares fitting of an ellipse to noisy data points refer to Halir and Flusser (1998); Fitzgibbon et al. (1999).

Appendix B

Real-Valued Solution of 4th-order Polynomial Equation

Given a univariate polynomial equation of degree n in normal form as

$$P_n(x) = \mathbf{A}^{\mathrm{T}}\mathbf{X} = 0, \tag{B.1}$$

where

$$\mathbf{A}_{(n+1)\times 1} = (a_0, a_1, \ldots, a_{n-1}, a_n)^{\mathrm{T}} \tag{B.2}$$

is the vector of coefficients and

$$\mathbf{X}_{(n+1)\times 1} = \left(1, x, \ldots, x^{n-1}, x^n\right)^{\mathrm{T}} \tag{B.3}$$

the vector of powers of its variable x. The equality holds only for certain values $x_1, x_2, \ldots, x_n$ of x, that are called the solutions or the roots of the equation. For degree at most four, there exists an algebraic method with a formula or a finite sequence of formulas to obtain the solutions. For higher degree, numerical methods are used to obtain approximate solutions. While the methods for degree up to two are rather simple, the methods for degree three and four are more elaborate and have different properties. Since necessary for this work, we will explain a method to obtain the real roots for polynomials of degree three and four, where the coefficients are also real-valued.

B.1 Quartic Equation

It follows an algebraic method (Weisstein, 2011d; Bronshtein et al., 2007) for calculating the real roots of a quartic polynomial by solving the quartic equation, given in normal form as in

$$a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0. \tag{B.4}$$

Without loss of generality assume coefficient $a_4 = 1$ by multiplying the entire equation with $1/a_4$. If all coefficients are real, then there exist 0, 2, or 4 real solutions. There does not exist any analytic formula to directly solve the quartic equation. It is, however, possible to solve for the roots in terms of the roots of the *resolvent cubic* equation (Terr, 2010) defined as in

$$y^3 + b_2y^2 + b_1y + b_0 = 0, \tag{B.5}$$

with

$$\begin{aligned} b_2 &= -a_2, \\ b_1 &= a_1 a_3 - 4a_0, \\ b_0 &= 4a_0 a_2 - a_1^2 - a_0 a_3^2, \end{aligned} \tag{B.6}$$

that can be solved using the cubic formula explained in Appendix B.2. As result, we obtain its first real root, which is either y_{11}, y_{12}, or y_{13}, referred to as y_1 for what follows. The four roots of the original quartic equation then coincide with the roots of the quadratic equation

$$z^2 + c_1 z + c_0 = 0, \tag{B.7}$$

with the coefficients

$$\begin{aligned} c_1 &= \frac{1}{2}\left(a_3 \pm \sqrt{a_3^2 - 4a_2 + 4y_1}\right), \\ c_0 &= \frac{1}{2}\left(y_1 \pm \sqrt{y_1^2 - 4a_0}\right), \end{aligned} \tag{B.8}$$

given as

$$\begin{aligned} x_{1,2} &= -\frac{1}{4}a_3 + \frac{1}{2}R \pm \frac{1}{2}D, \\ x_{3,4} &= -\frac{1}{4}a_3 - \frac{1}{2}R \pm \frac{1}{2}E, \end{aligned} \tag{B.9}$$

where

$$R = \sqrt{\frac{1}{4}a_3^2 - a_2 + y_1},$$

$$D = \begin{cases} \sqrt{\frac{3}{4}a_3^2 - R^2 - 2a_2 + \frac{1}{4}\left(4a_3 a_2 - 8a_1 - a_3^3\right) R^{-1}} & \text{if } R \neq 0, \\ \sqrt{\frac{3}{4}a_3^2 - 2a_2 + 2\sqrt{y_1^2 - 4a_0}} & \text{if } R = 0, \end{cases} \tag{B.10}$$

$$E = \begin{cases} \sqrt{\frac{3}{4}a_3^2 - R^2 - 2a_2 - \frac{1}{4}\left(4a_3 a_2 - 8a_1 - a_3^3\right) R^{-1}} & \text{if } R \neq 0, \\ \sqrt{\frac{3}{4}a_3^2 - 2a_2 - 2\sqrt{y_1^2 - 4a_0}} & \text{if } R = 0. \end{cases}$$

B.2 Cubic Equation

It follows an algebraic method (Bronshtein et al., 2007; Weisstein, 2011a) for calculating the real roots of a cubic polynomial by solving the cubic equation, given in normal form as in

$$a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0. \tag{B.11}$$

Without loss of generality assume coefficient $a_3 = 1$ by multiplying the entire equation with $1/a_3$. The number of real solutions depend on the value of discriminant

$$D = q^2 + p^3, \tag{B.12}$$

where

$$q = \frac{2a_2^3 - 9a_2a_1 + 27a_0}{54}, \quad p = \frac{3a_1 - a_2^2}{9}. \tag{B.13}$$

Defining

$$r = \operatorname{sgn} q \sqrt{|p|}, \tag{B.14}$$

the following cases can be distinguished:

- For $D \leq 0$, there exist the three real solutions

$$\begin{aligned} x_{11} &= -2r \cos\left(\frac{1}{3}\cos^{-1}\frac{q}{r^3}\right) - \frac{1}{3}a_2, \\ x_{12} &= +2r \cos\left(\frac{1}{3}\left(\pi - \cos^{-1}\frac{q}{r^3}\right)\right) - \frac{1}{3}a_2, \\ x_{13} &= +2r \cos\left(\frac{1}{3}\left(\pi + \cos^{-1}\frac{q}{r^3}\right)\right) - \frac{1}{3}a_2, \end{aligned} \tag{B.15}$$

  where

  - for $D < 0$, all three solutions are different,
  - for $D = 0$ and $p^3 = -q^2 \neq 0$, two solutions are equal, and
  - for $D = 0$ and $p = q = 0$, all three solutions are equal.

- For $D > 0$ and $p < 0$, there exists the single real solution

$$x_2 = -2r \cosh\left(\frac{1}{3}\cosh^{-1}\frac{q}{r^3}\right) - \frac{1}{3}a_2. \tag{B.16}$$

- For $D > 0$ and $p > 0$, there exists the single real solution

$$x_3 = -2r \sinh\left(\frac{1}{3}\sinh^{-1}\frac{q}{r^3}\right) - \frac{1}{3}a_2. \tag{B.17}$$

Bibliography

Aguado, A. S., Montiel, M. E., and Nixon, M. S. (1996). On using directional information for parameter space decomposition in ellipse detection. *Pattern Recogn.*, 29(3):369–381.

Agustin, J. S., Villanueva, A., and Cabeza, R. (2006). Pupil brightness variation as a function of gaze direction. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, page 49.

Alpern, M. (1962). *The Eye Vol. 3*, chapter Movements of the eyes, page 42. Academic Press, New York.

Arvacheh, E. M. and Tizhoosh, H. R. (2006). IRIS segmentation: Detecting pupil, limbus and eyelids. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 2453–2456.

Atchison, D. A. and Smith, G. (2000). *Optics of the Human Eye*. Butterworth-Heinemann, Edinburgh.

Babcock, J. S. and Pelz, J. B. (2004). Building a lightweight eyetracking headgear. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 109–114.

Backes, M., Chen, T., Dürmuth, M., Lensch, H. P. A., and Welk, M. (2009). Tempest in a teapot: Compromising reflections revisited. In *Proc. IEEE Symposium on Security and Privacy (SP)*, pages 315–327.

Backes, M., Dürmuth, M., and Unruh, D. (2008). Compromising reflections-or-How to read LCD monitors around the corner. In *Proc. IEEE Symposium on Security and Privacy (SP)*, pages 158–169.

Baker, S. and Nayar, S. K. (1999). A theory of single-viewpoint catadioptric image formation. *Int. J. Comput. Vision*, 35(2):175–196.

Baker, T. Y. (1943). Ray tracing through non-spherical surfaces. *Proc. Phys. Soc.*, 55(5):361–364.

Ballard, D. H. (1981). Generalizing the hough transform to detect arbitrary shapes. *Pattern Recogn.*, 13(2):111–122.

Barry, J. C., Pongs, U. M., and Hillen, W. (1997). Algorithm for purkinje images I and IV and limbus centre localization. *Comput. Biol. Med.*, 27(6):515–531.

Battle, J., Mouaddib, E., and Salvi, J. (1998). Recent progress in coded structured light as a technique to solve the correspondence problem: A survey. *Pattern Recogn.*, 31(7):963–982.

Beatty, J. and Lucero-Wagoner, B. (2000). *Handbook of Psychophysiology*, chapter The Pupillary System, pages 142–162. Cambridge University Press, 2nd edition.

Bennett, N., Burridge, R., and Saito, N. (1999). A method to detect and characterize ellipses using the hough transform. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(7):652–657.

Beymer, D. and Flickner, M. (2003). Eye gaze tracking using an active stereo head. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 451–458.

Bimber, O., Iwai, D., Wetzstein, G., and Grundhöfer, A. (2008). The visual computing of projector-camera systems. *Comput. Graph. Forum*, 27(8):2219–2245.

Bimber, O., Wetzstein, G., Emmerling, A., and Nitschke, C. (2005). Enabling view-dependent stereoscopic projection in real environments. In *Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 14–23.

Bogan, S. J., Waring III, G. O., Ibrahim, O., Drews, C., and Curtis, L. (1990). Classification of normal corneal topography based on computer-assisted videokeratography. *Arch. Ophthalmol.*, 108(7):945–949.

Boncelet, C. (2005). *Handbook of Image and Video Processing*, chapter Image Noise Models, pages 397–410. Academic Press, 2nd edition.

Bonfort, T., Sturm, P., and Gargallo, P. (2006). General specular surface triangulation. In *Proc. Asian Conference on Computer Vision (ACCV)*, pages 872–881.

Boraston, Z. and Blakemore, S.-J. (2007). The application of eye-tracking technology in the study of autism. *J. Physiol.*, 581(3):893–898.

Bouguet, J.-Y. (2010). Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/. Last accessed on January 9, 2011.

Bowyer, K. W., Hollingsworth, K., and Flynn, P. J. (2008). Image understanding for iris biometrics: A survey. *Comput. Vis. Image Underst.*, 110(2):281–307.

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV*. O'Reilly Media, Inc., Sebastopol, CA.

Bronshtein, I. N., Semendyayev, K. A., Musiol, G., and Muehlig, H. (2007). *Handbook of Mathematics*. Springer, 5th edition.

Caselles, V., Kimmel, R., and Sapiro, G. (1997). Geodesic active contours. *Int. J. Comput. Vision*, 22(1):61–79.

Chen, J. and Ji, Q. (2008). 3D gaze estimation with a single camera without ir illumination. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pages 1–4.

Chen, Q., Wu, H., and Wada, T. (2004). Camera calibration with two arbitrary coplanar circles. In *Proc. European Conference on Computer Vision (ECCV)*, pages 521–532.

Chen, Z. and Huang, J.-B. (1999). A vision-based method for the circle pose determination with a direct geometric interpretation. *IEEE Trans. Robot. Autom.*, 15(6):1135–1140.

Clark, J. J. (2006). Photometric stereo with nearby planar distributed illuminants. In *Proc. IEEE Canadian Conference on Computer and Robot Vision (CRV)*, pages 16–23.

Clark, J. J. (2010). Photometric stereo using LCD displays. *Image Vision Comput.*, 28(4):704–714.

Colombo, C., Comanducci, D., and Bimbo, A. D. (2007). Robust tracking and remapping of eye appearance with passive computer vision. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(4):2:1–2:20.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models–their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59.

Cotting, D., Naef, M., Gross, M., and Fuchs, H. (2004). Embedding imperceptible patterns into projected images for simultaneous acquisition and display. In *Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 100–109.

Coutinho, F. L. and Morimoto, C. H. (2006). Free head motion eye gaze tracking using a single camera and multiple light sources. In *Proc. Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pages 171–178.

Coutinho, F. L. and Morimoto, C. H. (2010). A depth compensation method for cross-ratio based eye tracking. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 137–140.

Crick, R. P. and Khaw, P. T. (2003). *A Textbook of Clinical Ophthalmology*. World Scientific, Singapore, 3rd edition.

Daugman, J. (2004). How iris recognition works. *IEEE Trans. Circuits Syst. Video Technol.*, 14(1):21–30.

Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(11):1148–1161.

Daunys, G., Istance, H., Bates, R., Signorile, I., Corno, F., Garbo, A., Farinetti, L., Holmqvist, E., Buchholz, M., Joos, M., Hansen, J., MacKay, D., Eskillson, R., and Majaranta, P. (2006). Cogain: D3.2 report on features of the different systems and development needs. Technical report, European Union Network of Excellence COGAIN (contract no. IST-2003-511598) of the 6th Framework Programme.

Dhome, M., Lapresté, J.-T., Rives, G., and Richetin, M. (1990). Spatial localization of modelled objects of revolution in monocular perspective vision. In *Proc. European Conference on Computer Vision (ECCV)*, pages 475–485.

Ditchburn, R. W. and Ginsborg, B. L. (1953). Involuntary eye movements during fixation. *J. Physiol.*, 119(1):1–17.

Donders, F. C. (1864). *On the Anomalies of Accommodation and Refraction of the Eye*. The New Sydenham Society.

Donegan, M., Oosthuizen, L., Bates, R., Daunys, G., Hansen, J., Joos, M., Majaranta, P., and Signorile, I. (2005). Cogain: D3.1 user requirements report, with observations of difficulties users are experiencing. Technical report, European Union Network of Excellence COGAIN (contract no. IST-2003-511598) of the 6th Framework Programme.

Drareni, J., Roy, S., and Sturm, P. (2009). Geometric video projector auto-calibration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 39–46.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behav. Res. Meth. Instrum. Comput.*, 34(4):455–470.

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag London, Ltd., 2nd edition.

Duchowski, A. T., Cournia, N., and Murphy, H. (2004). Gaze-contingent displays: A review. *CyberPsychol. Behav.*, 7(6):621–634.

Dumas, B., Lalanne, D., and Oviatt, S. (2009). Multimodal interfaces: A survey of principles, models and frameworks. In *Proc. Human Machine Interaction*, pages 3–26.

Dunne, M. C., Royston, J. M., and Barnes, D. A. (1992). Normal variations of the posterior corneal surface. *Acta Ophthalmol.*, 70(2):255–261.

Edmund, C. and Sjøntoft, E. (1985). The central-peripheral radius of the normal corneal curvature. A photokeratoscopic study. *Acta Ophthalmol.*, 63(6):670–677.

El Hage, S. G. and Berny, F. (1973). Contribution of the crystalline lens to the spherical aberration of the eye. *J. Opt. Soc. Am.*, 63(2):205–211.

Escudero-Sanz, I. and Navarro, R. (1999). Off-axis aberrations of a wide-angle schematic eye model. *J. Opt. Soc. Am. A*, 16(8):1881–1891.

Fasel, B. and Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recogn.*, 36(1):259–275.

Fitzgibbon, A., Pilu, M., and Fisher, R. B. (1999). Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(5):476–480.

Fofi, D., Sliwa, T., and Voisin, Y. (2004). A comparative survey on invisible structured light. In *Proc. SPIE 5303 Machine Vision Applications in Industrial Inspection XII*, pages 90–98.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., and Adolph, K. E. (2010). Head-mounted eye-tracking of infants' natural interactions: A new method. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 21–27.

Francken, Y., Cuypers, T., Mertens, T., Gielis, J., and Bekaert, P. (2008a). High quality mesostructure acquisition using specularities. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7.

Francken, Y., Hermans, C., and Bekaert, P. (2007). Screen-camera calibration using a spherical mirror. In *Proc. IEEE Canadian Conference on Computer and Robot Vision (CRV)*, pages 11–20.

Francken, Y., Hermans, C., and Bekaert, P. (2009). Screen-camera calibration using Gray codes. In *Proc. IEEE Canadian Conference on Computer and Robot Vision (CRV)*, pages 155–161.

Francken, Y., Hermans, C., Cuypers, T., and Bekaert, P. (2008b). Fast normal map acquisition using an LCD screen emitting gradient patterns. In *Proc. IEEE Canadian Conference on Computer and Robot Vision (CRV)*, pages 189–195.

Funk, N. and Yang, Y.-H. (2007). Using a raster display for photometric stereo. In *Proc. IEEE Canadian Conference on Computer and Robot Vision (CRV)*, pages 201–207.

Geyer, C. and Daniilidis, K. (2001). Catadioptric projective geometry. *Int. J. Comput. Vision*, 45(3):223–243.

Goncharov, A. V., Nowakowski, M., Sheehan, M. T., and Dainty, C. (2008). Reconstruction of the optical system of the human eye with reverse ray-tracing. *Opt. Express*, 16(3):1692–1703.

Gonzalez, R. C. and Woods, R. E. (2007). *Digital Image Processing*. Prentice Hall, 3rd edition.

Grabowski, K., Sankowski, W., Zubert, M., and Napieralska, M. (2006). Reliable iris localization method with application to iris recognition in near infrared light. In *Proc. IEEE International Conference on Mixed Design of Integrated Circuits and System (MIXDES)*, pages 684–687.

Gray, F. (1953). Pulse code communication.

Gredebäck, G., Johnson, S., and von Hofsten, C. (2010). Eye tracking in infancy research. *Dev. Neuropsychol.*, 35(1):1–19.

Grundhöfer, A., Seeger, M., Hantsch, F., and Bimber, O. (2007). Dynamic adaptation of projected imperceptible codes. In *Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–10.

Guestrin, E. D. and Eizenman, M. (2006). General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Trans. Biomed. Eng.*, 53(6):1124–1133.

Guestrin, E. D. and Eizenman, M. (2008). Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 267–274.

Guil, N. and Zapata, E. L. (1997). Lower order circle and ellipse hough transform. *Pattern Recognition*, 30(10):1729–1744.

Guillon, M., Lydon, D. P., and Wilson, C. (1986). Corneal topography: A clinical model. *Ophthalmic Physiol. Opt.*, 6(1):47–56.

Gullstrand, A. (1909). *Helmholtz's Handbuch der Physiologischen Optik*, volume 1, chapter Appendices II and IV, pages 301–358, 382–415. Voss Hamburg.

Gurdjos, P., Sturm, P., and Wu, Y. (2006). Euclidean structure from $N \geq 2$ parallel circles: Theory and algorithms. In *Proc. European Conference on Computer Vision (ECCV)*, pages 238–252.

Halir, R. and Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In *Proc. International Conference in Central Europe on Computer Graphics and Visualization (WSCG)*, pages 125–132.

Halstead, M. A., Barsky, B. A., Klein, S. A., and Mandell, R. B. (1996). Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals. In *Proc. ACM SIGGRAPH*, pages 335–342.

Hammoud, R. I. (2008). *Passive Eye Monitoring: Algorithms, Applications and Experiments*. Springer-Verlag Berlin-Heidelberg.

Hansen, D. W. and Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):478–500.

Hansen, D. W. and Pece, A. E. C. (2005). Eye tracking in the wild. *Comput. Vis. Image Underst.*, 98(1):155–181.

Hansen, P., Corke, P., Boles, W., and Daniilidis, K. (2007). Scale invariant feature matching with wide angle images. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1689–1694.

Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, 2nd edition.

He, Z., Tan, T., Sun, Z., and Qiu, X. (2009). Toward accurate and fast iris segmentation for iris biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(9):1670–1684.

Hill, R. (2002). Retina identification. In Jain, A. K., Bolle, R., and Pankanti, S., editors, *Biometrics*, pages 123–141. Springer.

Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.*, 8(3):1–30.

Hsu, R. L., Abdel-Mottaleb, M., and Jain, A. K. (2002). Face detection in color images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):696–706.

Hua, H., Krishnaswamy, P., and Rolland, J. P. (2006). Video-based eye-tracking methods and algorithms in head-mounted displays. *Opt. Express*, 14(10):4328–4350.

Ihrke, I., Kutulakos, K. N., Lensch, H. P. A., Magnor, M., and Heidrich, W. (2008). State of the art in transparent and specular object reconstruction. In *STAR Proc. Eurographics*.

Inokuchi, S., Sato, K., and Matsuda, F. (1984). Range imaging system for 3-D object recognition. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pages 806–808.

Iqbal, N. and Lee, S.-Y. (2008). A study on human gaze estimation using screen reflection. In *Proc. International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*, pages 104–111.

Isard, M. and Blake, A. (1998). Condensation–conditional density propagation for visual tracking. *Int. J. Comput. Vision*, 29:5–28.

Iskander, D. (2006). A parametric approach to measuring limbus corneae from digital images. *IEEE Trans. Biomed. Eng.*, 53(6):1134–1140.

Iskander, D. R., Collins, M. J., Mioschek, S., and Trunk, M. (2004). Automatic pupillometry from digital images. *IEEE Trans. Biomed. Eng.*, 51(9):1619–1627.

Jaimes, A. and Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Comput. Vis. Image Underst.*, 108(1-2):116–134.

Johnson, M. K. and Farid, H. (2007). Exposing digital forgeries through specular highlights on the eye. In *Proc. International Workshop on Information Hiding (IH)*, pages 311–325.

Kagami, S. (2010). High-speed vision systems and projectors for real-time perception of the world. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 100–107.

Kanatani, K. and Liu, W. (1993). 3D interpretation of conics and orthogonality. *CVGIP Image Underst.*, 58(3):286–301.

Kang, J. J., Eizenman, M., Guestrin, E. D., and Eizenman, E. (2008). Investigation of the cross-ratios method for point-of-gaze estimation. *IEEE Trans. Biomed. Eng.*, 55(9):2293–2302.

Karmali, F. and Shelhamer, M. (2006). Compensating for camera translation in video eye movement recordings by tracking a landmark selected automatically by a genetic algorithm. In *Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS)*, pages 5298–5301.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *Int. J. Comput. Vision*, 1(4):321–331.

Kaufman, P. L. and Alm, A. (2003). *Adler's Physiology of the Eye: Clinical Application*. Mosby, Inc., St. Louis, MO, 10th edition.

Khurana, A. K. (2007). *Comprehensive Ophthalmology*, volume 4. New Age International (P) Ltd., New Delhi.

Kiely, P. M., Smith, G., and Carney, L. G. (1982). The mean shape of the human cornea. *Opt. Acta*, 29:1027–1040.

Kim, S. M., Sked, M., and Ji, Q. (2004). Non-intrusive eye gaze tracking under natural head movements. In *Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS)*, pages 2271–2274.

Ko, Y. J., Lee, E. C., and Park, K. R. (2008). A robust gaze detection method by compensating for facial movements based on corneal specularities. *Pattern Recogn. Lett.*, 29(10):1474–1485.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum. Neurobiol.*, 4(4):219–227.

Kolakowski, S. M. and Pelz, J. B. (2006). Compensating for eye tracker camera movement. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 79–85.

Koninckx, T. P. and Van Gool, L. (2006). Real-time range acquisition by adaptive structured light. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):432–445.

Kooijman, A. C. (1983). Light distribution on the retina of a wide-angle theoretical eye. *J. Opt. Soc. Am.*, 73(11):1544–1550.

Koretz, J. F., Kaufman, P. L., Neider, M. W., and Goeckner, P. A. (1989). Accommodation and presbyopia in the human eye–aging of the anterior segment. *Vision Res.*, 29(12):1685–1692.

Kothari, R. and Mitchell, J. L. (1996). Detection of eye locations in unconstrained visual images. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 519–522.

Kris, C. (1960). *Medical Physics Vol. III*, volume 3, chapter Vision: Electrooculography, pages 692–700. Year Book Publishers, Chicago.

Kumar, R. K., Ilie, A., Frahm, J.-M., and Pollefeys, M. (2008). Simple calibration of non-overlapping cameras with a mirror. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7.

Kuthirummal, S. and Nayar, S. K. (2006). Multiview radial catadioptric imaging for scene capture. In *Proc. ACM SIGGRAPH*, pages 916–923.

Kutulakos, K. N. and Steger, E. (2008). A theory of refractive and specular 3D shape by light-path triangulation. *Int. J. Comput. Vision*, 76(1):13–29.

Lam, A. K. C. and Douthwaite, W. A. (1997). Measurement of posterior corneal asphericity on Hong Kong Chinese: A pilot study. *Ophthalmic Physiol. Opt.*, 17(4):348–356.

Lam, K.-M. and Yan, H. (1996). Locating and extracting the eye in human face images. *Pattern Recogn.*, 29(5):771–779.

Lam, M. W. Y. and Baranoski, G. V. G. (2006). A predictive light transport model for the human iris. *Comput. Graph. Forum*, 25(3):359–368.

Lambooij, M., IJsselsteijn, W., Fortuin, M., and Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3):1–14.

Le Grand, Y. and El Hage, S. G. (1980). *Physiological Optics*. Springer Berlin.

Lee, J. C., Hudson, S., and Dietz, P. (2007). Hybrid infrared and visible light projection for location tracking. In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*, pages 57–60.

Lee, J. C., Hudson, S. E., Summet, J. W., and Dietz, P. H. (2005). Moveable interactive projected displays using projector based tracking. In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*, pages 63–72.

Lefohn, A., Budge, B., Shirley, P., Caruso, R., and Reinhard, E. (2003). An ocularist's approach to human iris synthesis. *IEEE Comput. Graphics Appl.*, 23(6):70–75.

Li, D., Babcock, J., and Parkhurst, D. J. (2006). openEyes: a low-cost head-mounted eye-tracking solution. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 95–100.

Li, D., Winfield, D., and Parkhurst, D. J. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition - Workshops (CVPRW)*, pages 79–86.

Li, F., Kolakowski, S., and Pelz, J. (2007). Using structured illumination to enhance video-based eye tracking. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 373–376.

Li, F., Munn, S., and Pelz, J. (2008). A model-based approach to video-based eye tracking. *J. Mod. Optic.*, 55(4):503–531.

Li, P., Liu, X., Xiao, L., and Song, Q. (2010). Robust and accurate iris segmentation in very noisy iris images. *Image and Vision Computing*, 28(2):246–253. Segmentation of Visible Wavelength Iris Images Captured At-a-distance and On-the-move.

Liou, H. L. and Brennan, N. A. (1997). Anatomically accurate, finite model eye for optical modeling. *J. Opt. Soc. Am. A*, 14(8):1684–1695.

Liu, Z., Huang, A. J., and Pflugfelder, S. C. (1999). Evaluation of corneal thickness and topography in normal eyes using the orbscan corneal topography system. *Br. J. Ophthalmol.*, 83(7):774–778.

Livingston, M. A. (1998). *Vision-based tracking with dynamic structured light for video see-through augmented reality*. PhD thesis, The University of North Carolina at Chapel Hill. AAI9914867.

Lotmar, W. (1971). Theoretical eye model with aspherics. *J. Opt. Soc. Am.*, 61(11):1522–1529.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110.

Lowe, R. F. and Clark, B. A. (1973). Posterior corneal curvature. Correlations in normal eyes and in eyes involved with primary angle-closure glaucoma. *Br. J. Ophthalmol.*, 57(7):464–470.

Mackworth, N. H. and Thomas, E. L. (1962). Head-mounted eye-marker camera. *J. Opt. Soc. Am.*, 52(6):713–716.

Mandell, R. B. (1996). A guide to videokeratography. *Int. Contact Lens Clin.*, 23(6):205–228.

Mandell, R. B. and St Helen, R. (1971). Mathematical model of the corneal contour. *Br. J. Physiol. Optic.*, 26:183–197.

Mangold International GmbH (2011a). A-S-L Mangold Mobile Eye. http://www.mangold-international.com/eye-tracking/head-mounted/overview.html. Last accessed on January 9, 2011.

Mangold International GmbH (2011b). MangoldVision Eye Tracker. http://www.mangold-international.com/eye-tracking/stationary/overview.html. Last accessed on January 9, 2011.

Marg, E. (1951). Development of electro-oculography: Standing potential of the eye in registration of eye movement. *Arch. Ophthalmol.*, 45(2):169–185.

Matey, J. R., Broussard, R., and Kennell, L. (2010). Iris image segmentation and sub-optimal images. *Image Vision Comput.*, 28(2):215–222.

Merchant, J., Morrissette, R., and Porterfield, J. L. (1974). Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Trans. Biomed. Eng.*, 21(4):309–317.

Miller, J. M., Hall, H. L., Greivenkamp, J. E., and Guyton, D. L. (1995). Quantification of the brückner test for strabismus. *Invest. Ophthalmol. Vis. Sci.*, 36(5):897–905.

Morelande, M. R., Iskander, D. R., Collins, M. J., and Franklin, R. (2002). Automatic estimation of the corneal limbus in videokeratoscopy. *IEEE Transactions on Biomedical Engineering*, 49(12):1617–1625.

Mori, H., Sumiya, E., Mashita, T., Kiyokawa, K., and Takemura, H. (2010). A wide-view parallax-free eye-mark recorder with a hyperboloidal half-silvered mirror and appearance-based gaze estimation. *IEEE Trans. Vis. Comput. Graph.*, PP(99):1–1.

Morimoto, C. H. and Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.*, 98(1):4–24.

Morris, N. and Kutulakos, K. (2007). Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1–8.

Muhammad, N., Fofi, D., and Ainouz, S. (2009). Current state of the art of vision based SLAM. In *Proc. SPIE Image Processing: Machine Vision Applications II*, volume 7251, page 72510F.

Munn, S. M. and Pelz, J. B. (2008). 3D point-of-regard, position and head orientation from a portable monocular video-based eye tracker. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 181–188.

Nagamatsu, T., Iwamoto, Y., Kamahara, J., Tanaka, N., and Yamamoto, M. (2010). Gaze estimation method based on an aspherical model of the cornea: surface of revolution about the optical axis of the eye. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 255–258.

Nayar, S. K. (1988). Sphereo: Determining depth using two specular spheres and a single camera. In *Proc. SPIE Conference on Optics, Illumination, and Image Sensing for Machine Vision III*, pages 245–254.

Nehab, D., Weyrich, T., and Rusinkiewicz, S. (2008). Dense 3D reconstruction from specularity consistency. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7.

Nene, S. A. and Nayar, S. K. (1998). Stereo with mirrors. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1087–1094.

Nguyen, K., Wagner, C., Koons, D., and Flickner, M. (2002). Differences in the infrared bright pupil response of human eyes. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 133–138.

Nii, H., Sugimoto, M., and Inami, M. (2005). Smart light-ultra high speed projector for spatial multiplexing optical transmission. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 95–102.

Nishino, K., Belhumeur, P. N., and Nayar, S. K. (2005). Using eye reflections for face recognition under varying illumination. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 519–526.

Nishino, K. and Nayar, S. K. (2004a). Eyes for relighting. In *Proc. ACM SIGGRAPH*, pages 704–711.

Nishino, K. and Nayar, S. K. (2004b). The world in an eye. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Nishino, K. and Nayar, S. K. (2006). Corneal imaging system: Environment from eyes. *Int. J. Comput. Vision*, 70(1):23–40.

Nitschke, C., Nakazawa, A., and Takemura, H. (2009). Display-camera calibration from eye reflections. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1226–1233.

Nixon, M. (1985). Eye spacing measurement for facial recognition. In *Proc. SPIE Vol. 575 Applications of Digital Image Processing VIII*, pages 279–285.

Noris, B., Keller, J.-B., and Billard, A. (2010). A wearable gaze tracking system for children in unconstrained environments. *Comput. Vis. Image Underst.*, In Press, Accepted Manuscript:–.

Ohno, T., Mukawa, N., and Yoshikawa, A. (2002). Freegaze: A gaze tracking system for everyday gaze interaction. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 125–132.

Oike, H., Kato, T., Wada, T., and Wu, H. (2004). A high-performance active camera system for taking clear images (in japanese). In *Proc. CVIM-144*, volume 2004, pages 71–78.

OpenGL (2008). OpenGL Extension `NV_fence`. http://www.opengl.org/registry/specs/NV/fence.txt. Last accessed on January 9, 2011.

OpenGL (2009). OpenGL Extension `ARB_sync`. http://www.opengl.org/registry/specs/ARB/sync.txt. Last accessed on January 9, 2011.

OpenGL (2011). OpenGL. http://www.opengl.org. Last accessed on January 9, 2011.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.*, 9(1):62–66.

Pajdla, T., Svoboda, T., and Hlavác, V. (2001). *Panoramic Vision: Sensors, Theory, and Applications*, chapter Epipolar geometry of central panoramic cameras, pages 85–114. Springer, New York.

Pamplona, V. F., Oliveira, M. M., and Baranoski, G. V. G. (2009). Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Trans. Graph.*, 28(4):1–12.

Patel, S., Marshall, J., and Fitzke, F. W. (1993). Shape and radius of posterior corneal surface. *Refract. Corneal Surg.*, 9(3):173–181.

Pharr, M. and Humphreys, G. (2004). *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. http://www.pbrt.org. Last accessed on January 9, 2011.

Porta, M. (2002). Vision-based user interfaces: Methods and applications. *Int. J. Hum. Comput. Stud.*, 57(1):27–73.

Posdamer, J. and Altschuler, M. (1982). Surface measurement by space-encoded projected beam systems. *Comput. Graph. Image Process.*, 18(1):1–17.

Press, W. H., Vetterling, W. T., Teukolsky, S. A., and Flannery, B. P. (2002). *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press, New York, NY, 2nd edition.

Rakshit, S. and Monro, D. M. (2007). Pupil shape description using fourier series. In *Proc. IEEE Workshop on Signal Processing Applications for Public Security and Forensics (SAFE)*, pages 1–4.

Ramamoorthi, R. and Hanrahan, P. (2001). A signal-processing framework for inverse rendering. In *Proc. ACM SIGGRAPH*, pages 117–128.

Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. (1998). The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proc. ACM SIGGRAPH*, pages 179–188.

Reale, M., Hung, T., and Yin, L. (2010). Viewing direction estimation based on 3D eyeball construction for HRI. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 24–31.

Remington, L. A. (2004). *Clinical Anatomy of the Visual System*. Butterworth-Heinemann, 2nd edition.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., and Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *J. Opt. Soc. Am.*, 43(6):495–501.

Ryan, W. J., Woodard, D. L., Duchowski, A. T., and Birchfield, S. T. (2008). Adapting starburst for elliptical iris segmentation. In *Proc. IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–7.

Safaee-Rad, R., Tchoukanov, I., Benhabib, B., and Smith, K. C. (1991). Accurate parameter estimation of quadratic curves from grey-level images. *CVGIP: Image Underst.*, 54(2):259–274.

Safaee-Rad, R., Tchoukanov, I., Smith, K. C., and Benhabib, B. (1992). Three-dimensional location estimation of circular features for machine vision. *IEEE Trans. Robot. Autom.*, 8(5):624–640.

Sagawa, R., Ota, Y., Yagi, Y., Furukawa, R., Asada, N., and Kawasaki, H. (2009). Dense 3D reconstruction method using a single pattern for fast moving object. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1779–1786.

Sakamoto, J. A., Barrett, H. H., and Goncharov, A. V. (2008). Inverse optical design of the human eye using likelihood methods and wavefront sensing. *Opt. Express*, 16(1):304–314.

Sakata, M., Yasumuro, Y., Imura, M., Manabe, Y., and Chihara, K. (2002). A location awareness system using wide-angle camera and active IR-tag. In *Proc. IAPR Workshop on Machine Vision Applications (MVA)*, pages 522–525.

Salvi, J., Fernandez, S., Pribanic, T., and Llado, X. (2010). A state of the art in structured light patterns for surface profilometry. *Pattern Recogn.*, 43(8):2666–2680.

Salvi, J., Pagès, J., and Batlle, J. (2004). Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849. Agent Based Computer Vision.

Scaramuzza, D., Criblez, N., Martinelli, A., and Siegwart, R. (2008). Robust feature extraction and matching for omnidirectional images. In *Field and Service Robotics*, volume 42 of *Springer Tracts in Advanced Robotics*, pages 71–81. Springer Berlin / Heidelberg.

Schindler, G. (2008). Photometric stereo via computer screen lighting for real-time surface reconstruction. In *Proc. International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, pages CD–ROM.

Schnieders, D., Fu, X., and Wong, K.-Y. K. (2010). Reconstruction of display and eyes from a single image. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1442–1449.

Schnipke, S. K. and Todd, M. W. (2000). Trials and tribulations of using an eye-tracking system. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)*, pages 273–274.

Seeing Machines Inc. (2011). faceLAB. http://www.seeingmachines.com/product/facelab/. Last accessed on January 9, 2011.

Semple, J. G. and Kneebone, G. T. (1952). *Algebraic Projective Geometry*. Oxford University Press.

SensoMotoric Instruments GmbH (SMI) (2011a). IVIEW X HED. http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/iview-x-hed.html. Last accessed on January 9, 2011.

SensoMotoric Instruments GmbH (SMI) (2011b). RED / RED250 / RED500. http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/red-red250-red-500.html. Last accessed on January 9, 2011.

Sezgin, M. and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging*, 13(1):146–168.

Shackel, B. (1967). *A Manual of Psychophysiological Methods*, chapter Eye movement recording by electro-oculography, pages 300–334. North-Holland Publishing Co., Amsterdam.

Shah, S. and Ross, A. (2009). Iris segmentation using geodesic active contours. *IEEE Trans. Inf. Forensics Security*, 4(4):824–836.

Sharma, R., Pavlovic, V. I., and Huang, T. S. (1998). Toward multimodal human-computer interface. *P. IEEE*, 86(5):853–869.

Shen, C.-H. and Chen, H. (2006). Robust focus measure for low-contrast images. In *Proc. IEEE International Conference on Consumer Electronics (ICCE)*, pages 69–70.

Shih, S.-W. and Liu, J. (2004). A novel approach to 3-D gaze tracking using stereo cameras. *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, 34(1):234–245.

Shih, S.-W., Wu, Y.-T., and Liu, J. (2000). A calibration-free gaze tracking technique. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pages 4201–4204.

Shiu, Y. C. and Ahmad, S. (1989). 3D location of circular and spherical features by monocular model-based vision. In *Proc. IEEE International Conference on Systems, Man and Cybernetics*, pages 576–581.

Simon, C. and Goldstein, I. (1935). A new scientific method of identification. *New York State J. Med.*, 35(18):901–906.

Smart Eye AB (2011a). Anti Sleep. http://www.smarteye.se/antisleep.aspx. Last accessed on January 9, 2011.

Smart Eye AB (2011b). Smart Eye Pro. http://www.smarteye.se/page1714564.aspx. Last accessed on January 9, 2011.

Smith, J. D., Vertegaal, R., and Sohn, C. (2005). Viewpointer: Lightweight calibration-free eye tracking for ubiquitous handsfree deixis. In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*, pages 53–61.

Snell, R. S. and Lemp, M. A. (1997). *Clinical Anatomy of the Eye*. Blackwell Publishing, Malden, 2nd edition.

Sorsby, A., Benjamin, B., Davey, J. B., Sheridan, M., and Tanner, J. M. (1957). Emmetropia and its aberrations. A study in the correlation of the optical components of the eye. *Medical Research Council Special Report Series*, 293.

SR Research Ltd. (2011a). EyeLink 1000. http://www.sr-research.com/EL_1000.html. Last accessed on January 9, 2011.

SR Research Ltd. (2011b). EyeLink II. http://www.sr-research.com/EL_II.html. Last accessed on January 9, 2011.

Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behav. Res. Meth. Instrum. Comput.*, 25(2):137–142.

Stenstrom, S. (1948). Investigation of the variation and the correlation of the optical elements of human eyes. *Am. J. Optom. Arch. Am. Acad. Optom.*, 25(7):340–350.

Stiefelhagen, R., Yang, J., and Waibel, A. (1996). A model-based gaze tracking system. In *Proc. IEEE International Joint Symposia on Intelligence and Systems (IJSIS)*, pages 304–310.

Stiefelhagen, R., Yang, J., and Waibel, A. (1997). Tracking eyes and monitoring eye gaze. In *Proc. Workshop on Perceptual User Interfaces (PUI)*, pages 98–100.

Sturm, P. and Bonfort, T. (2006). How to compute the pose of an object without a direct view? In *Proc. Asian Conference on Computer Vision (ACCV)*, pages 21–31.

Sugano, Y., Matsushita, Y., and Sato, Y. (2010). Calibration-free gaze sensing using saliency maps. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2667–2674.

Sugano, Y., Matsushita, Y., Sato, Y., and Koike, H. (2008). An incremental learning method for unconstrained gaze estimation. In *Proc. European Conference on Computer Vision (ECCV)*, pages 656–667.

Svoboda, T. and Pajdla, T. (2002). Epipolar geometry for central catadioptric cameras. *Int. J. Comput. Vision*, 49:23–37.

Swaminathan, R., Grossberg, M. D., and Nayar, S. K. (2006). Non-single viewpoint catadioptric cameras: Geometry and analysis. *Int. J. Comput. Vision*, 66(3):211–229.

Takemura, K., Kohashi, Y., Suenaga, T., Takamatsu, J., and Ogasawara, T. (2010). Estimating 3D point-of-regard and visualizing gaze trajectories under natural head movements. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 157–160.

Tarini, M., Lensch, H. P. A., Goesele, M., and Seidel, H.-P. (2005). 3D acquisition of mirroring objects using striped patterns. *Graph. Models*, 67(4):233–259.

Terr, D. (2010). "Resolvent Cubic." From MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/ResolventCubic.html. Last accessed on January 9, 2011.

Tobii Technology AB (2011a). Tobii Eye Tracking: An introduction to eye tracking and Tobii Eye Trackers. http://www.tobii.com/archive/files/20689/Tobii+EyeTracking+WhitePaper.pdf.aspx. Last accessed on January 9, 2011.

Tobii Technology AB (2011b). Tobii Glasses Eye Tracker Product Description. http://www.tobii.com/archive/files/20956/

TobiiGlassesProductDescription_080710_web.pdf.aspx. Last accessed on January 9, 2011.

Tobii Technology AB (2011c). Tobii T/X Series Eye Trackers Product Description. http://www.tobii.com/archive/files/17995/Tobii_TX_Series_Eye_Trackers_product_description.pdf.aspx. Last accessed on January 9, 2011.

Tsumura, N., Dang, M. N., Makino, T., and Miyake, Y. (2003). Estimating the directions to light sources using images of eye for reconstructing 3D human face. In *Proc. IS&T/SID Color Imaging Conference (CIC)*, pages 77–81.

Turk, M. (2004). Computer vision in the interface. *Commun. ACM*, 47(1):60–67.

Tuytelaars, T. and Mikolajczyk, K. (2008). Local invariant feature detectors: A survey. *Found. Trends. Comput. Graph. Vis.*, 3(3):177–280.

Tweed, D. and Vilis, T. (1990). Geometric relations of eye position and velocity vectors during saccades. *Vision Res.*, 30(1):111–127.

Vezhnevets, V. and Degtiareva, A. (2003). Robust and accurate eye contour extraction. In *Proc. Graphicon*, pages 81–84.

Villanueva, A. and Cabeza, R. (2007). Models for gaze tracking systems. *J. Image Video Process.*, 2007(3):1–16.

Villanueva, A. and Cabeza, R. (2008). A novel gaze estimation system with one calibration point. *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, 38(4):1123–1138.

Villanueva, A., Daunys, G., Hansen, D. W., B¥"ohme, M., Cabeza, R., Meyer, A., and Barth, E. (2009). A geometric approach to remote eye tracking. *Univers. Access Inf. Soc.*, 8(4):241–257.

Wagner, P., Bartl, K., Günthner, W., Schneider, E., Brandt, T., and Ulbrich, H. (2006). A pivotable head mounted camera system that is aligned by three-dimensional eye movements. In *Proc. ACM Symposium on Eye Tracking Research & Applications (ETRA)*, pages 117–124.

Wang, H., Lin, S., Liu, X., and Kang, S. B. (2005a). Separating reflections in human iris images for illumination estimation. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1691–1698.

Wang, H., Lin, S., Ye, X., and Gu, W. (2008). Separating corneal reflections for illumination estimation. *Neurocomputing*, 71(10–11):1788–1797.

Wang, J.-G. and Sung, E. (2001). Gaze determination via images of irises. *Image Vision Comput.*, 19(12):891–911.

Wang, J.-G. and Sung, E. (2002). Study on eye gaze estimation. *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, 32(3):332–350.

Wang, J.-G., Sung, E., and Venkateswarlu, R. (2003). Eye gaze estimation from a single image of one eye. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 136–143.

Wang, J.-G., Sung, E., and Venkateswarlu, R. (2005b). Estimating the eye gaze from one eye. *Comput. Vis. Image Underst.*, 98(1):83–103.

Waschbüsch, M., Würmlin, S., Cotting, D., Sadlo, F., and Gross, M. (2005). Scalable 3D video of dynamic scenes. *Visual Comput.*, 21(8-10):629–638.

Watson, A. B. (1986). *Handbook of Perception and Human Performance*, volume 1, chapter Temporal Sensitivity (Time dependent Human Perception of Visual Stimuli), pages 6–1–6–43. Wiley-Interscience New York.

Weisstein, E. W. (2011a). "Cubic Formula." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/CubicFormula.html. Last accessed on January 9, 2011.

Weisstein, E. W. (2011b). "Ellipse." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Ellipse.html. Last accessed on January 9, 2011.

Weisstein, E. W. (2011c). "Gray Code." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/GrayCode.html. Last accessed on January 9, 2011.

Weisstein, E. W. (2011d). "Quartic Equation." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/QuarticEquation.html. Last accessed on January 9, 2011.

Wildes, R. P. (1997). Iris recognition: An emerging biometric technology. *P. IEEE*, 85(9):1348–1363.

Wildes, R. P., Asmuth, J. C., Green, G. L., Hsu, S. C., Kolczynski, R. J., Matey, J. R., and McBride, S. E. (1996). A machine-vision system for iris recognition. *Mach. Vision Appl.*, 9(1):1–8.

Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 19(1):139–144.

Wu, H., Chen, Q., and Wada, T. (2005a). Estimating the visual direction with two-circle algorithm. In *Proc. International Workshop on Biometric Recognition Systems (IWBRS)*, pages 1–16.

Wu, H., Chen, Q., and Wada, T. (2005b). Visual direction estimation from a monocular image. *IEICE Trans. Inf. Syst.*, E88-D(10):2277–2285.

Wu, H., Kitagawa, Y., Wada, T., Kato, T., and Chen, Q. (2007). Tracking iris contour with a 3D eye-model for gaze estimation. In *Proc. Asian Conference on Computer Vision (ACCV)*, pages 688–697.

Wyatt, H. J. (1995). The form of the human pupil. *Vision Res.*, 35(14):2021–2036.

Xie, X., Sudhakar, R., and Zhuang, H. (1994). On improving eye feature extraction using deformable templates. *Pattern Recogn.*, 27(6):791–799.

Yagi, Y. (1999). Omnidirectional sensing and its applications. *IEICE Trans. Inf. Syst.*, E82-D(3):568–579.

Yamazoe, H., Utsumi, A., Yonezawa, T., and Abe, S. (2008). Remote and head-motion-free gaze tracking for real environments with automated head-eye model calibrations. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–6.

Yarbus, A. L. (1967). *Eye Movements and Vision*. Plenum Press, New York.

Yoo, D. H. and Chung, M. J. (2005). A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Comput. Vis. Image Underst.*, 98(1):25–51.

Yoo, D. H., Kim, J. H., Lee, B. R., and Chung, M. J. (2002). Non-contact eye gaze tracking system by mapping of corneal reflections. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition (AFGR)*, pages 94–99.

Young, D., Tunley, H., and Samuels, R. (1995). Specialised hough transform and active contour methods for real-time eye tracking. Technical Report 386, School of Cognitive and Computing Sciences, Univ. of Sussex.

Young, L. and Sheena, D. (1975). Survey of eye movement recording methods. *Behav. Res. Meth. Instrum.*, 7(5):397–429.

Yuille, A. L., Hallinan, P. W., and Cohen, D. S. (1992). Feature extraction from faces using deformable templates. *Int. J. Comput. Vision*, 8(2):99–111.

Zhang, L., Subramaniam, N., Lin, R., Raskar, R., and Nayar, S. (2008). Capturing images with sparse informational pixels using projected 3D tags. In *Proc. IEEE Virtual Reality Conference (VR)*, pages 11–18.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334.

Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458.

Zheng, Y. and Liu, Y. (2008). The projective equation of a circle and its application in camera calibration. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pages 1–4.

Zheng, Y., Ma, W., and Liu, Y. (2008). Another way of looking at monocular circle pose estimation. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 861–864.

Zhu, Z. and Ji, Q. (2007). Novel eye gaze tracking techniques under natural head movement. *IEEE Trans. Biomed. Eng.*, 54(12):2246–2260.

Zongker, D. E., Werner, D. M., Curless, B., and Salesin, D. H. (1999). Environment matting and compositing. In *Proc. ACM SIGGRAPH*, pages 205–214.