Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

PAUL SCHERRER INSTITUT

Elena Morandini

Eye Pose Estimation in Ocular Proton Therapy Using Anatomical Modeling and Refraction-Corrected Optical Imaging

Semester Project Report

Department of Information Techology and Electrical Engineering
Swiss Federal Institute of Technology (ETH) Zürich
Paul Scherrer Institut (PSI)

Supervisors

Prof. Dr. Anthony Lomax
Dr. Riccardo Via
Dr. Giovanni Fattori

June 2025

Abstract

Background: Ocular Proton Therapy (OPT) demands sub-millimeter precision in eye positioning to ensure accurate dose delivery. Traditional methods rely on surgically implanted fiducial markers and X-ray imaging, which, while accurate, are invasive and unsuitable for continuous monitoring. Emerging optical Eye Tracking Systems (ETS) offer non-invasive alternatives, but their effectiveness is limited by anatomical variability and light refraction in the anterior eye.

Methods: This study explores a refraction-aware, patient-specific modeling approach to eye pose estimation in OPT. A 3D anatomical model of the eye is derived from MRI data and refined by incorporating optimized corneal curvature and pupil center location through a ray-traced simulation framework based on Aguirre's model of the eye. Two strategies are compared against X-ray based ground truth: the first is based on an unaltered MRI-based model; the second uses an updated anatomically refined model.

Results: Refinement of corneal geometry and pupil center location significantly improved alignment between simulated and observed pupil contours, reducing optimization loss and increasing ellipse overlap across patients. Eye pose estimation using the updated model demonstrated improved accuracy, particularly in vertical (y) and beam-axis alignment (z) world coordinates, with clip and tumor centroid deviations showing statistically significant reductions compared to the original model.

Conclusion: While the results demonstrate the potential of non-invasive, clip-less eye localization in OPT using ETS and anatomically refined, refraction-aware models, limitations in anatomical modeling and pose validation must be carefully considered. Further validation, model refinement, and clinical evaluation are essential before such approaches can be translated into routine practice. These findings offer a promising direction, but should be interpreted with caution as part of ongoing efforts towards more accurate eye tracking in OPT.

Contents

Abstract **i**

Nomenclature **iv**

1 Introduction **1**

1.1 Background on Ocular Proton Therapy . . . 1

1.1.1 Treatment Workflow Overview . . . 1

1.1.2 X-ray imaging and the role of of implanted fiducials (clips) in current targeting strategies . . . 2

1.1.3 Potentials and Limitations of Eye Tracking Systems (ETS) . . . 2

1.2 The challenge of accurate eye pose estimation . . . 4

1.2.1 Impact of Light Refraction in the Anterior Chamber on Pupil Localization 4

1.3 Anatomical and Optical Modeling of the Eye . . . 4

1.3.1 Definition of a Patient-Specific 3D Eye Model from MRI . . . 4

1.3.2 Modeling Position and Orientation of the Eye (Fick parameters) . . . 4

1.4 Refraction-aware eye model refinement . . . 5

1.4.1 Overview of the ray-tracing model based on Aguirre et al . . . 5

1.4.2 Paradigm Shift: from *Direct* to *Inverse* Models for Eye Pose Estimation . . 6

1.5 Study Objectives . . . 7

2 Methods **8**

2.1 Patient Population and Imaging Data . . . 8

2.1.1 Patient Cohort Description . . . 8

2.1.2 Acquisition of ETS Images . . . 8

2.1.3 Calibration of Camera Systems . . . 9

2.1.4 Acquisition of MRI-Based Eye Model . . . 9

2.1.5 Integration of Anatomical and Optical Properties through Parameter Variations . . . 9

2.2 Optimization Framework . . . 10

2.2.1 Optimization Strategies . . . 10

2.2.2 Simulation of Pupil Contours using Ray Tracing . . . 11

2.3 Evaluation Strategy . . . 12

2.3.1 Retro-projection Error and Alignment Metrics . . . 12

2.3.2 Assessment of Eye Localization Accuracy vs MRI . . . 13

3 Results **14**

3.1 Varying Corneal Curvature . . . 14

3.1.1 Effect on simulated pupil contours and alignment with imaging data . . . 16

3.1.2 Application of Optimized Cornea Shape to MRI-based Eye Model 17
3.2 Eye Pose Estimation: Original vs Updated Eye Models 17
3.2.1 Relationship Between Estimated Eye Model and Image Ground Truth . . . 19
3.2.2 Accuracy of Pose Estimation Compared to Clip-Based Ground Truth 21

4 Discussion **24**
4.1 Interpretation of Results 24
4.1.1 Benefit of Incorporating Refraction and Anatomy into Pupil Localization . 24
4.1.2 Comparing Anatomical and Pose-Based Optimization Strategies 25
4.2 Limitations 26
4.2.1 Simplifications in Patient Specific Multi Modal Eye Model Definition 26
4.2.2 Lack of Anatomical Ground Truth for Eye Pose Validation 26
4.3 Future Work and Opportunities 27
4.3.1 What Can Be Achieved Without Clips 27
4.3.2 What is Needed to Overcome Clip Dependence 27
4.3.3 Timing Considerations 27

5 Conclusion **28**

A Appendix 1 - Raw Data **29**

Nomenclature

List of Acronyms and Abbreviations

OPT	Ocular proton therapy
ETS	Eye Tracking System
VOG	Video-Oculography
ROI	Region of Interest
L-BFGS-B	Limited–memory Broyden–Fletcher–Goldfarb–Shanno with Box constraints

Introduction

1.1 Background on Ocular Proton Therapy

1.1.1 Treatment Workflow Overview

Proton therapy is a highly precise radiation modality leveraging the physical properties of proton interactions with matter to distribute dose more conformally compared to conventional photon therapy, thus making it particularly apt to treat small lesions such as intraocular tumors. Ocular proton therapy (OPT) is a branch of proton therapy dedicated to the treatment of uveal melanomas, historically achieving tumor local control rates of well over 90% [1].

Before treatment, patients undergo a detailed ophthalmological evaluation including ultrasound and fundus photography imaging. The standard clinical workflow includes a surgical procedure where radiopaque fiducial markers (tantalum clips) are sutured onto the eye surface in convenient locations with respect to the tumor. These clips are then used as landmarks for the target volume definition as well as treatment geometry verification using X-rays [2, 3].

Treatment planning in OPT mainly consists of two parts: firstly, ophthalmological data, together with the clip positions, are used to define a geometrical eye model; then, the optimal eye orientation to achieve complete target dose coverage, while optimally sparing surrounding organs at risk, is defined [3]. As of today, the most popular treatment planning software is EyePlan, a non-commercial solution developed in the 1980s. Lately, however, a push for the integration of volumetric imaging such as MRI and CT in the treatment planning process, as well as a modernization effort, has led to the development of RayOcular (Ray-Search Laboratories, Stockholm, Sweden). This work includes eye models created using the latter planning system [4].

Finally, during treatment, patients are required to maintain a specified gaze orientation defined in the treatment plan by fixating on a dedicated light diode positioned at the optimal orientation. Orthogonal X-ray imaging is used to iteratively verify alignment by checking the position of the clips in space. Any necessary positional corrections can be made by translating the robotic chair where the patient is seated, or by performing adjustments in the light position, thus inducing rotational correction to the eye position. However, torsional eye movements (rotations about the visual axis) are more challenging to manage, often requiring adjustments to the collimator rather than the actual correction of the patient's position [1]. The setup of the treatment room for OPTIS 2, the machine utilized at the Center for Proton Therapy, can be seen in **Figure 1.1** below.

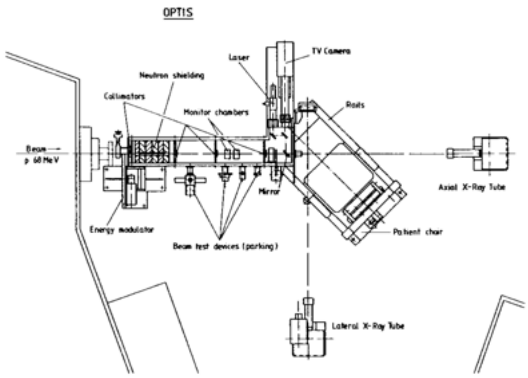


Figure 1.1: OPTIS Treatment Room Setup [5]

1.1.2 X-ray imaging and the role of of implanted fiducials (clips) in current targeting strategies

In current clinical practice for OPT, orthogonal radiographic imaging of implanted fiducial markers in the treatment room remains the gold standard for treatment localization and verification of the eye and tumor position. These markers provide an indirect geometrical reference for tumor localization and, assuming the body is rigid, for the whole eye. Orthogonal X-ray imaging of clips plays a role in OPT even before the eye model is created; in fact, X-ray images are acquired in the treatment room when the patient is seated in the treatment chair before treatment planning, during the so-called "planning image acquisition" phase. This couple of radiographic projections is then used to localize clips in the three-dimensional (3D) space and in turn construct the eye model incorporating tumor and clip positions [2, 1].

Before irradiation, orthogonal X-ray images are again used to align the apparent eye position with the treatment geometry. The alignment consists of comparing the clip positions seen in the X-rays with the nominal configuration corresponding to the treatment geometry [2, 3]. X-ray imaging is repeated iteratively while moving the treatment chair and light position until a satisfactory clips-to-clips alignment is achieved.

The clip-based method offers high accuracy and reliability and is well established in the clinic. However, it brings disadvantages as well: firstly, the invasive nature of the implantation procedure contributes to patient discomfort; secondly, X-ray imaging can only capture a snapshot of time, a less-than-ideal solution when monitoring the position of an ever-moving organ such as the eye. Finally, the iterative process lacks automation and heavily relies on the operator's experience.

More recently, optical eye tracking systems (ETS) are being explored as viable solutions for the aforementioned purposes. Optical tracking methods have shown promise as non-invasive surrogates for clip-based localization. However, their accuracy is still evolving and is influenced by factors such as eye moisture, lighting conditions, and modeling assumptions of the anterior part of the eye [6, 2, 1].

1.1.3 Potentials and Limitations of Eye Tracking Systems (ETS)

At the Center for Proton Therapy (CPT) at Paul Scherrer Institut, there is increasing interest towards non-invasive alternatives that would obviate the necessity of clip implantation and X-ray imaging while maintaining the utmost accuracy in patient positioning of OPT. Eye tracking systems (ETS) are currently being explored as non-invasive alternatives to clip-based alignment in OPT. These systems aim to estimate the 3D position and orientation of the eye in real time by capturing external ocular features, typically the pupil and corneal reflections, using a number of optical cameras.

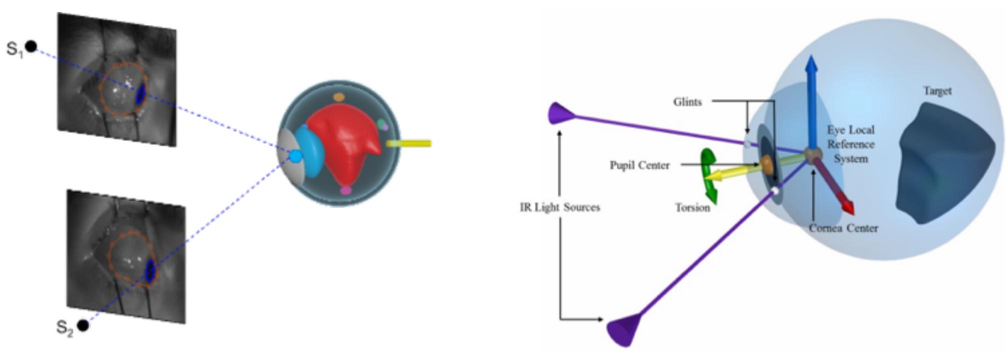
Potentials of ETS in clinical applications:

- Non-invasive localization: ETS eliminates the need for surgical implantation of fiducial markers. This makes treatment more accessible, particularly for patients unfit for surgery [3, 2].
- Real-time motion tracking: ETS can continuously monitor eye position throughout treatment, enabling detection of intra-fraction motion. On the contrary, X-rays are static and are only representative of a single snapshot in time [6, 1].
- Integration with MRI-based eye models: ETS measurements typically rely on a geometrical model of the subject's eye at varying levels of complexity. The use of MRI imaging to obtain an anatomically accurate description of the ocular anatomy of every patient can be easily integrated with eye tracking measurements [7].

On the other hand, ETS also faces various challenges, limiting its full application in the field of OPT. At CPT, several ETS technologies have been tested in the last few years:

- Pupil Tracking: It consists of measuring the 3D position of the pupil through geometric triangulation and the pinhole model. Assuming that the image of the pupil captured by the optical cameras corresponds to the subject's real anatomical pupil, geometric triangulation using a pair of calibrated cameras can provide a measurement of the 3D position of the patient's pupil. This method has proved to be reliable, as the pupil is the most prominent feature on ocular images, but not accurate enough when compared to X-ray imaging of clips. A major flaw of triangulation methods is the neglect of corneal refraction, which causes the apparent pupil center in 2D images to deviate from its actual 3D location. These errors result in nonlinear biases dependent on the corneal curvature and gaze angle, increasing with off-axis viewing [8].

- Eye Tracking using Video-Oculography (VOG): These systems estimate eye position based on corneal reflections and pupil segmentation, aiming to provide non-invasive localization of the eye and tumor. VOG uses multiple infrared light sources to illuminate the eye, generate corneal reflections, and adopts an eye model consisting of perfectly concentric spheres to model the cornea. While more sophisticated compared to pupil tracking, this method proved to be too unreliable due to the degradation of corneal reflections' visibility caused by the eyelid retractors worn by patients undergoing OPT, which is a clinical requirement for the treatment [2, 6, 1].



(a) Pupil Tracking using Triangulation (b) Pupil Tracking using Video-Oculography

Figure 1.2: Different methods of eye tracking: (a) triangulation methods between pupil image and camera and (b) VOG-based eye position estimation based on corneal reflections and pupil segmentations [9, 6].

While ETS has demonstrated promising sub-millimeter positional accuracy under ideal conditions, its repeatability and robustness under clinical conditions are inconsistent. Simplified ETS systems often treat the eye as a rigid sphere, neglecting optical distortions from the cornea and the complex curvature of anterior eye structures. Without correction, this leads to biased localization, particularly in estimating the true pupil center and gaze direction. To improve reliability, more research is currently being done regarding the use of MRI-based anatomical models for individualized ETS calibration, and integrating ray-tracing models for refraction-aware pupil localization [7, 10, 11]. This would enable the ETS to not only track eye movements but also estimate gaze direction. Aguirre's ray-traced eye model, for instance, simulates the appearance of the pupil under corneal refraction, enhancing the precision of model fitting for eye pose estimation [7, 8].

1.2 The challenge of accurate eye pose estimation

1.2.1 Impact of Light Refraction in the Anterior Chamber on Pupil Localization

Light refraction in the anterior structures of the eye such as the cornea plays a central role in defining the pupil appearance. While light travels through the curved layers of the eye, it is bent significantly due to the difference in refractive indices, distorting the apparent position and shape of the pupil [8].

The pupil is observed not as the anatomical aperture of the iris, but rather as the entrance pupil: a virtual image formed due to the refraction by corneal surfaces. The non-uniform and curved geometry of the cornea causes distortion of the pupil, making it appear larger and elliptical when viewed at an angle, shifting its center from its true geometric center, and altering the perceived depth of the pupil. This refraction must be modeled accurately to achieve accurate eye pose estimation [8].

Aguirre presents a ray-traced model of the human eye correctly simulating these effects, showing that accurate specification of the entrance pupil requires modeling of corneal refraction [8]. In OPT, mislocating the pupil by even a fraction of a millimeter can shift the estimated gaze vectors by several degrees, resulting in severe consequences for treatments, where dose gradients are steep and off-target delivery should be avoided. As studied by Lu et al., an offset in 0.5 mm error in pupil center estimation can translate to a gaze error of $> 3°$, which in the context of proton therapy could lead to significant misalignment of the therapeutic beam [12].

1.3 Anatomical and Optical Modeling of the Eye

1.3.1 Definition of a Patient-Specific 3D Eye Model from MRI

To overcome the aforementioned limitations, creating a patient-specific 3D eye model incorporating anatomical variability is a key step in improving the accuracy of eye positioning using ETS. Magnetic resonance imaging (MRI), with both T1 and T2 weighted sequences, has been recently explored for this purpose, as it provides information on several parts of the eye such as the optic nerve, the lens, and the globe contour. MRI can be used to visualize soft tissue boundaries and is critical for defining the shape and orientation of the eye [13].

MRI, however, does not provide perfect anatomical detail for all ocular structures. For instance, the fovea and retina cannot be clearly visualized, while the anterior chamber and lens are only partially resolved. Furthermore, ocular structures such as the ciliary body and iris remain challenging to visualize without ultra high resolution imaging [13].

Despite its limitations, MRI-based eye modeling could lay the groundwork for a clip-less workflow. An accurate model, together with a precise gaze direction obtained through ETS optical tracking, would allow the tumor and surrounding eye structures to be localized relative to the eye's anatomy alone, avoiding the need for surgical fiducials to be implanted [12, 8].

1.3.2 Modeling Position and Orientation of the Eye (Fick parameters)

In eye tracking applied to OPT, two key coordinate systems must be taken into consideration: the world coordinate system, aligned to the room isocenter, and the Fick coordinate system, centered at the center of the eyeball. The differences between these systems can be seen in **Figure 1.3** below. Eye pose in 3D space in OPT is defined using the Fick coordinates and can be described through a series of two angles and three translations $\mathbf{F}(\theta, \phi, \mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3})$. These parameters define the transformation between the anatomical eye model derived from MRI and its observed position in the treatment room [14].

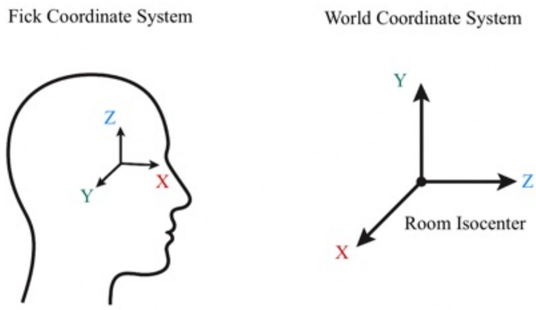


Figure 1.3: World and Fick coordinate systems as seen in the treatment room: the former centered at the room isocenter, the latter centered at the eyeball center of rotation.

The Fick rotations are represented by a gimbal sequence of horizontal, vertical and torsional rotations, described by the angles θ, ϕ and ψ respectively (see **Figure 1.4** below). This order is set to avoid gimbal lock and allows mapping between eye rotations and observable features. The eye position can therefore be completely characterized by a rotation about the vertical axis ($\mathbf{e_3}$) by θ, a rotation about the horizontal axis ($\mathbf{e_2}$) by ϕ and a rotation about the line of sight ($\mathbf{e_1}$) by ψ [14].

These parameters, however, are associated with some practical limitations. Firstly, they assume a single and fixed center of rotation, which was demonstrated by Aguirre to not be biologically accurate. On the other hand, torsional rotation is difficult to detect by optical tracking systems in the absence of implanted markers [8, 2].

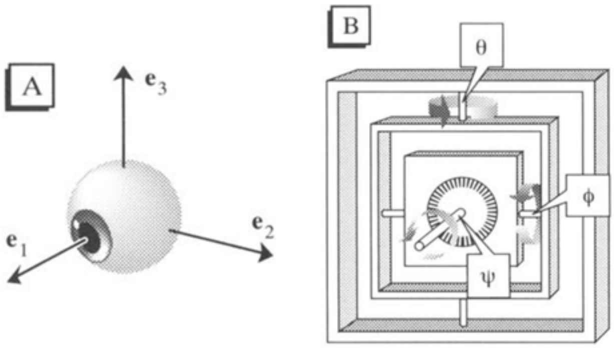


Figure 1.4: Fick translation and rotational parameters, adapted from [14].

1.4 Refraction-aware eye model refinement

1.4.1 Overview of the ray-tracing model based on Aguirre et al

The ray tracing eye model proposed by Aguirre presents an accurate method to model the appearance of the eye when seen by external cameras while considering different structures within the eye. This model is critical for refining eye pose estimation, especially when relying on visual cues like pupil shape in optical tracking systems. Rather than assuming straight-line projections from the pupil to the camera, this model simulates how light rays bend through the cornea and aqueous humor, producing a realistic representation of the entrance pupil. The main components of this model can be seen in **Figure 1.5** below.

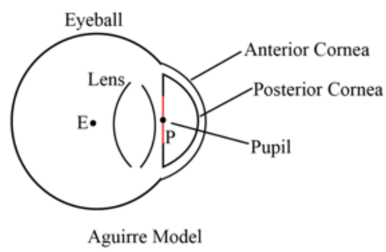


Figure 1.5: Schematic representation of the Aguirre eye model [12].

The key features of Aguirre's ray-tracing model include high anatomical accuracy, 3D ray tracing, and modeling of the entrance pupil. The model includes the corneal surfaces, an anterior chamber with a refractive index of ~ 1.336, an iris at realistic depth, and pupil diameter and shape as configurable parameters. The anterior and posterior corneal surfaces are modeled as triaxial ellipsoids separated by a fixed distance. Rays are traced from the external camera viewpoint into the eye, accounting for Snell's law at each surface interface, allowing for changes in ray direction due to curved boundaries as well as accounting for refractive index transitions between air - cornea and aqueous media. The resulting projection can accurately reproduce the position, size, shape, and orientation of the entrance pupil ellipse on images acquired with a calibrated optical camera. This can be compared to ETS images for pose estimation if at least two images pertaining to at least two optical cameras are available [8, 7].

Compared to simpler models, Aguirre's more complex model accounts for asymmetries in pupil distortion with off-axis gaze, and allows for different rotational centers for horizontal and vertical eye movements [7, 12]. This model can also be used to simulate eye appearance through ETS systems for a given pose, and allows comparing simulations to real-time camera views to estimate and verify gaze direction.

1.4.2 Paradigm Shift: from *Direct* to *Inverse* Models for Eye Pose Estimation

One of the main goals of this study is to be able to estimate eye pose accurately and non-invasively through optical tracking systems, rather than continuing to rely on surgically implanted fiducials for OPT. For this, it is hypothesized that having a complex and accurate model of the eye able to simulate the optical appearance of the entrance pupil based on anatomy and using ray-tracing is essential.

A complete model of eye optical properties would allow for a paradigm switch when measuring its position and orientation. Rather than relying on the direct approach, where 2D information on optical camera images is utilized to estimate the 3D position of features such as the pupil position or corneal reflections (see **Section 1.1.3**), we propose the inverse approach where a 3D eye model can move freely in space. Then, using a-priori knowledge on the eye optical properties and the camera calibration parameters, the eye appearance on the optical camerasâ images and design metrics can be estimated for comparison.

This simulation workflow would be structured as follows: after defining the patient-specific eye's anatomical geometry using MRI and other imaging modalities, a 3D eye pose is applied using Fick angles to rotate the model in space. Using ray tracing, light rays' propagation from the pupil through the cornea and into the air is simulated. The entrance pupil can be projected into the ETS camera coordinate system, producing a simulated 2D ellipse in each camera [7, 12] as can be seen in **Figure 1.6** below. The simulated 2D ellipse can then be directly compared with the appearance of the pupil in the ETS images.

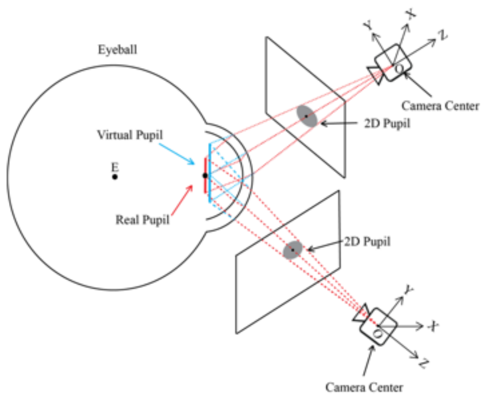


Figure 1.6: Re-projection of pupil contours onto ETS camera views [12].

1.5 Study Objectives

This report presents a thorough investigation into the estimation of eye pose in OPT using eye tracking systems in combination with an MRI-based model of the eye, with a focus on anatomical modeling and optical effects of the curvature of the anterior eye. Building on the limitations of geometric triangulation methods and the unreliability of corneal refractions and pupil appearance, this study aims to refine eye tracking accuracy.

The primary objective is to investigate whether refining and optimizing eye structures such as the cornea and entrance pupil in a ray-traced anatomical eye model improves the accuracy of 3D pupil center estimation.

To evaluate this, eye pose is estimated using two modeling strategies:

- A baseline model derived from the patient-specific MRI data
- An anatomically refined model that incorporates optimized corneal curvature and anterior eye geometry.

The performance of each approach is quantified by comparing the resulting eye poses to the ground truth provided by X-ray measurements of implanted fiducial clips.

Through this comparison, this study aims to examine the feasibility of clip-less eye positioning workflows in clinical settings. If accurate eye pose can be achieved through model-based simulation and surface tracking alone, this would represent a major step towards non-invasive and high-precision targeting in OPT.

Methods

2.1 Patient Population and Imaging Data

2.1.1 Patient Cohort Description

This study includes a small sample cohort of 12 patients undergoing ocular proton therapy at the Center for Proton Therapy at PSI in 2023, within the frame of the EKNZ 2019-01987 project approved by the ethics committee Ethikkommission Nordwest- und Zentralschweiz. All patients analyzed adhered to an experimental data acquisition campaign, agreeing to an MRI acquisition of the eye. The data acquired and utilized for this investigation originates from two specific treatment sessions: the planning image acquisition and the third treatment fraction. All patients had surgically implanted fiducial clips in the sclera prior to treatment. During X-ray image acquisition both in planning and in fraction, the patient is seated in the treatment chair and the head is immobilized using a custom-made thermoplastic mask and a bite block.

2.1.2 Acquisition of ETS Images

During the planning acquisition, the patient's pupil is roughly aligned to the beam axis, resulting in the eye x-axis in Fick coordinates being aligned to the z-axis of the world coordinate system. Several orthogonal X-ray acquisitions are taken at different gaze directions: straight into the beam, left or right based on tumor laterality, and one or more educated guess locations based on the operator's experience. Synchronously to these X-rays, ETS images are captured, one per X-ray acquisition, using a four-camera setup which can be seen in **Figure 2.1** below. The synchronous acquisition of the ETS images and X-rays enables the creation of a correlation between internal (clip-based) and external (pupil-based) eye positioning data.

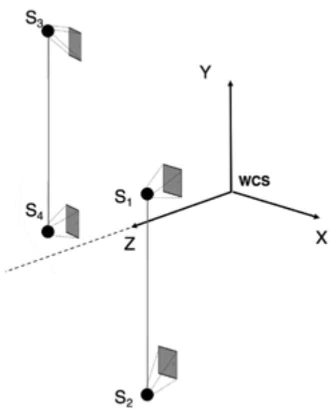


Figure 2.1: ETS four-camera setup (labeled S_1 - S_4) in the treatment room with two cameras placed on top of each other on either side of the patient.

During the fraction acquisitions, X-rays and ETS images are again acquired, this time in an iterative process used to guide patient alignment to match the clip configuration defined during the planning session. The iteration number varies based on patient cooperation and accuracy achieved; therefore, it can vary between patients. Again, ETS images are acquired simultaneously with X-ray imaging. 3D coordinates of the clips attached to the patient's sclera measured using

X-ray are saved in text files, while the ETS images are exported in binary files with proprietary format.

In this study, 12 patients were analyzed for a total of $n = 89$ acquisitions. For the planning image acquisition, a mean of 3 paired X-ray / ETS measurements were taken (min = 2, max = 4) whereas for the third fraction, a mean of 4 acquisitions were taken (min = 2, max = 7).

2.1.3 Calibration of Camera Systems

The ETS system in the room comprises four cameras: two placed on the right side of the beam axis, and two on the left. The selection of a pair of cameras for a given acquisition depends on the eye that contains the tumor. Each camera is calibrated using the pinhole camera model, which involves an intrinsic matrix K (describing focal length, principal point, and skew, set to 0), extrinsic parameters (rotation R and translation t defining the camera's pose), and distortion coefficients D (accounting for lens distortion). The complete camera projection from 3D world coordinates to 2D image coordinates is represented as $P = K[R|t]$ [15]. ETS images are acquired with a region of interest (ROI) smaller than the full optical camera sensor. Spatial correlation is maintained through a dedicated calibration procedure.

2.1.4 Acquisition of MRI-Based Eye Model

The ocular anatomy models used in this study were created using the RayOcular treatment planning system. To generate a patient-specific eye model based on the X-ray position of clips and the MRI of the patient, clip positions are first localized from the X-ray images in room coordinates. These X-rays are used to construct a virtual CT scan, which is then imported into the RayOcular software. The software performs point-based registration between clip locations in the virtual CT scan and in the MRI. In parallel, measurements provided by the ophthalmologists taken during clip implantation provide additional guidance to ensure accurate placement of the clips on the model's outer scleral surface.

After the rigid registration is applied between the MRI and virtual CT, manual adjustments can be made to better align the MRI image to the geometrical eye model. Furthermore, individual eye parameters can be adjusted to allow the model created using clips to better fit the anatomy seen on MRI scans.

The 3D MRI patient-specific model of the eye is then exported as a DICOM RTStruct file containing 3D point contours of different organs of the eye such as the sclera, lens, optic disc and nerve, tumor, and clips.

2.1.5 Integration of Anatomical and Optical Properties through Parameter Variations

The eye model utilized for optimization integrates both anatomical (geometric) and optical (refractive) parameters to capture variations between patients and correct for distortions caused by refraction in optical imaging. This combined approach allows to build a 3D representation of the front part of the eye, which is essential in order to accurately estimate the eye pose through optimization.

The front of the eye, particularly the corneal surface and the pupil, is modeled using a combination of geometrical and optical properties. As explained by Aguirre, the eye model's cornea is built through two separate spherical surfaces: a front (anterior) and a back (posterior) surface. Each is created as a unit sphere in 3D which is scaled to match the MRI-based eye model created using the RayOcular software [8].

2.2 Optimization Framework

Throughout the different optimization tasks performed in this work, the underlying optimization approach remains constant. Each task, in fact, optimizes a different subset of parameters using the same numerical method: the L-BFGS-B algorithm.

L-BFGS-B (Limited Memory Broyden-Fletcher-Goldfarb-Shanno with Box constraints) is a quasi-Newton optimization algorithm designed for optimizations with large numbers of variables and bound constraints. It approximates the inverse Hessian matrix to guide the optimization efficiently. The box constraint is needed in order to be able to specify upper and lower bounds for each parameter [16].

Moreover, this algorithm is particularly suited to the current investigation due to its robustness in smooth, moderately-sized, and computationally expensive landscapes such as those formulated in ray-traced models. The model parameters investigated in this study affect the simulation in continuous ways. Therefore, it is important that the algorithm can converge quickly and stably even in the presence of noisy losses [16]. In all objective functions to be minimized, the loss metric remains constant.

2.2.1 Optimization Strategies

In this study, two different optimizations are performed to refine the geometric and optical properties of the eye and subsequently predict pose parameters across the different acquisition segments accordingly.

2.2.1.1 Varying Eye Pose

The first optimization for eye pose utilizes the eye geometry coming from the MRI model and is carried out across all segments and fractions. The parameters optimized in these steps are the pupil radius and the five pose components: azimuth (θ) and elevation (ϕ) rotation angles and three translations in the x,y,z directions according to the Fick coordinate system. The bounds in this optimization are set to $\pm 25°$ for the former and ± 35 mm for the latter. The optimization is initialized with an educated pose guess derived from the conversion of the polar (φ) and azimuthal (θ) light source angles into eye pose rotations as well as triangulating the pupil location center recognized from the ETS images. While these estimates are often quite far from the truth, especially as they do not consider light refraction in the eye, they are still crucial in setting a starting point for the optimization.

The optimization strategy is subdivided as follows: first, the optimizer focuses on the pupil size parameter only, then on the eye pose parameters, and finally, a refinement is conducted by combining all variables.

2.2.1.2 Varying Corneal Curvature to Estimate Eye Pose

The second optimization applies the same aforementioned workflow to a slightly modified model of the eye. The updated eye model is created by optimizing three parameters: pupil radius, corneal curvature, and the offset of the pupil center along the eye optical axis (Δx in Fick coordinates). This optimization is conducted exclusively for the first segment of the planning acquisition, where the patient is looking forward into the beamline. This choice is based on the physiological assumption that the corneal geometry remains constant across fractions and segments for each patient; changes in curvature within days or even minutes are not biologically plausible [17].

The aim of this optimization is to find more suitable curvature and pupil parameters that will match the re-projected pupil onto the ETS plane with the pupil as seen on ETS images. To do this, the optimization is split into two main steps: first, the pupil size is optimized. The initial model features a fixed pupil diameter of 2.7 mm. However, when comparing the re-projected pupil to the one seen on ETS images, one can immediately appreciate the physiological variation of pupil size. For this reason, the pupil is optimized with bounds ranging from 2 mm to 4 mm.

After optimizing pupil size, the corneal surfaces and the pupil center offset are optimized simultaneously. The optimization takes the front corneal surface radii from the MRI model as input and iteratively calculates a constant scaling factor to apply to both the front and back surfaces of the cornea, in order to ensure both surfaces stay anatomically consistent throughout the optimization. Both radii optimizations are restricted to remain within $90 - 110\%$ of the original model radii to ensure that the model converges towards realistic limits. These bounds have been chosen to allow some flexibility in cornea curvature definition, as the cornea itself is not so clearly visible in the MRI and could therefore be wrongly contoured in the original model.

Simultaneously, the x coordinate of the pupil center is optimized (Δx), allowing a shift by up to ± 1.5 mm. This was included in the optimization to account for the iris thickness, which was not considered during ray-tracing. After this second optimization, a final refinement is conducted, taking all parameters, including pupil size, as input.

The resulting values for corneal radii and for Δx are used as fixed inputs for subsequent optimizations, while still allowing the pupil radius to vary. For this estimation in later segments, a grid search is performed over the defined bounds. This approach accounts for the fact that pupil size can vary significantly between gaze directions due to lighting conditions in the treatment room, patient-specific response, and imaging differences, and could potentially change with respect to the first acquisition.

Once the original model is updated through this optimization, a combined optimization for eye pose is carried out across all fractions and treatment segments, this time using the updated anatomical model with the optimized corneal curvature and Δx. The eye pose and pupil radius optimization strategy is unchanged compared to that in **Section 2.2.1.1**, which ensures that the estimated eye pose is consistent with the most accurate representation of the patient's eye anatomy.

In this way, the patient data is now expanded to comprise the following: X-ray ground truth coordinates of clip locations, an MRI-based model of the eye, and an updated model of the eye with modified pupil center location and corneal curvature. For all models, the optimized eye pose given the input parameters can be compared to the ground truth pose coming from the MRI-based eye model registered to the fiducial clips placed on the patients' eye.

2.2.2 Simulation of Pupil Contours using Ray Tracing

2.2.2.1 Loss Metric: Discrepancy between Observed and Simulated Pupil Contours

To evaluate the accuracy of the optimization, a loss function is iteratively computed between the retro-projected pupil with the candidate parameters being evaluated and the ground truth pupil contour extracted from the ETS image segmentation.

This comparison is performed first for each camera independently: the candidate and observed pupil are both modeled as ellipses, and the discrepancy between them is measured using a hybrid loss function composed of three terms:

- Algebraic Loss:
  This term evaluates how well the points generated on the candidate optimized ellipse satisfy the implicit equation of the target ellipse. It is computed as the mean squared error of the ellipse equation evaluated at the fitted points. This term is sensitive to parametrization and scale, which is why it is often paired with geometric loss.
- Geometric Loss:

This term geometrically measures the average Euclidean distance between corresponding points on the candidate and target ellipses. This is computed through the nearest neighbors of the candidate points in the target set and vice versa, allowing the metric to focus more on shape and position of the pupil.

- Area Penalty:
  To enforce consistency in overall scale, the area of both ellipses is iteratively compared and penalized according to the squared relative error between the candidate and target ellipse. This ensures that even when the shapes align well, the optimization does not over or under-estimate the pupil size.

Once this individual loss is calculated for both camera contributions, a final combined loss is computed as a weighted sum of the two according to 2.1 below

$$\text{Total Loss} = w_1 * L_1 + w_2 * L_2 \tag{2.1}$$

where:

- L_1 and L_2 are the hybrid losses per camera
- weights are calculated as follows: $w_1 = \frac{L_1}{L_1+L_2}$ and $w_2 = \frac{L_2}{L_1+L_2}$

This formulation ensures that the total loss is dominated by the camera where the model fits less well, guiding the optimizer to minimize the worst-case alignment between candidate and observed pupil ellipses.

2.3 Evaluation Strategy

To evaluate and compare the results of the different optimizations, this study aims to quantitatively assess the accuracy of the updated and original MRI-based eye models when used to estimate eye pose relative to the clip measurements from the X-ray acquisitions.

2.3.1 Retro-projection Error and Alignment Metrics

As previously mentioned, the metric guiding the optimization process is kept constant throughout all strategies. The loss between the retro-projected pupil ellipse obtained by tweaking certain parameters is iteratively compared to the pupil mask created from the two ETS images. The optimized parameters are then fed into the initial eye model to visualize and quantify differences and improvements. This is done through several key metrics:

- Clip Centroid Location:
  The location of clips from the optimized eye models are obtained by applying the optimized eye pose (for models both including the original cornea and the updated cornea) to the original eye model taken from RayOcular. The clip centroids are then computed from the respective clip contours.
- Tumor Metrics:
  Different tumor metrics are evaluated: the tumor center of mass, or centroid, computes the average position of all points making up the tumor contour. It is computed by taking the mean of the x, y, z coordinates of all points in the contour separately, to yield an average overall location.

- Clip Distance to the Sclera:
  To ensure anatomical plausibility, the distances between clips and the scleral surface are monitored to confirm that clips remain in a feasible location on top of the outer scleral surface despite the applied transformations. This is done by computing the Euclidean distance between sampled points around the clip fitted circle and the closest point on the outer layer of the sclera. The sign of the dot product of vectors from the clip to the sclera determines whether the clips lie on top, or below, the sclera.

- Pose Differences:
  The five pose parameters are compared across three models: the eye pose estimated from the X-ray clips, the original MRI model with optimized eye pose and the updated MRI model where both cornea and pupil are altered and and eye pose recalculated accordingly.

Altogether, these metrics provide a thorough assessment of how well the optimized parameters can improve alignment accuracy in eye pose while maintaining anatomical plausibility.

2.3.2 Assessment of Eye Localization Accuracy vs MRI

After quantitatively evaluating all metrics against their respective ground truth references, an additional qualitative verification was conducted using the RayOcular planning software and MRI data. Specifically, the estimated clip locations resulting from the optimized eye pose as well as the optimized corneal curvature and pupil center location were visualized on the MRI scans.

While metallic clips themselves are not directly visible in an MRI, their presence on the eye causes visible artifacts due to their material (tantalum), which cause signal voids in the images. When aligning the MRI-based eye model to the clip positions identified from the X-ray images, one can immediately notice that the artifacts are significantly larger than the actual clip size. This mismatch introduces uncertainty in the true clip location within the original MRI model.

While such uncertainty must be taken into consideration as it is carried forward in many other calculations, the artifact's size also provides a spatial margin, meaning that slight deviations of the optimized clips compared to the original ones may still fall within the visible void and therefore remain anatomically plausible. To perform this verification, the updated model's eye pose parameters were converted back into the original image space through a series of transformations which undid the measured eye pose and reverted from eye-coordinates to room-coordinates. The clips were created as spheres where the center pixel was the average of all clip locations across all segments for a particular patient, while the clip diameter was given by the maximum range in the three directions.

Results

3.1 Varying Corneal Curvature

Figure 3.1 below shows results for the optimization of the corneal radii executed on the first acquisition of the planning image phase across the patient population, while keeping the eye pose fixed. For the radii in the x and y Fick coordinates, the optimization process has converged at the allowed boundaries in around 50% of the cases, but this occurs much less in the z direction. As can be seen in **Table 3.1**, the x direction is the one that undergoes the most deviation from the original corneal values, tending to an average decrease of its radius across the tested patients.

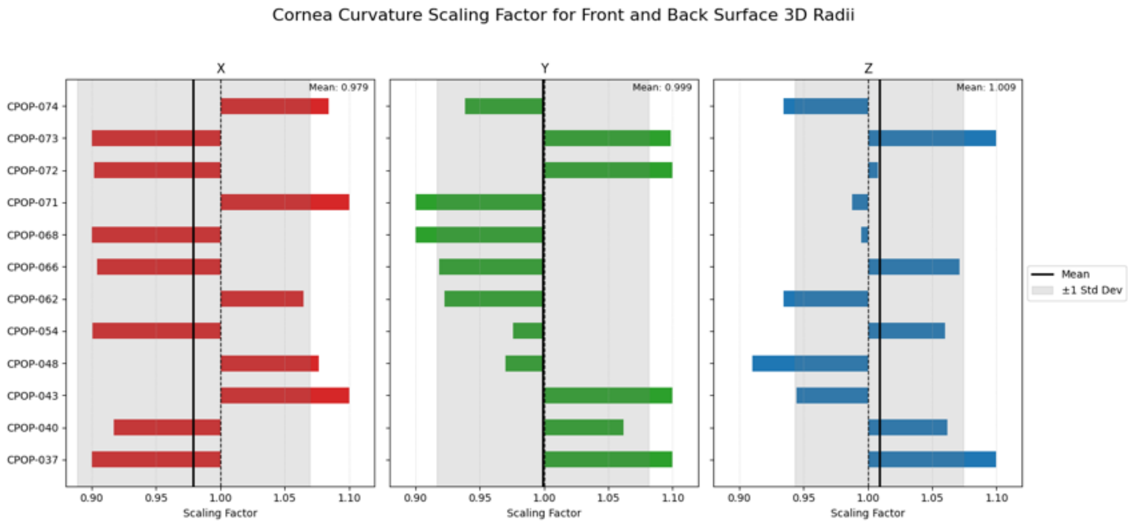


Figure 3.1: Optimized scaling factor deviation from original cornea curvature defined on the MRI model for all first segments of the Planning fraction across all patients. The same scaling factor is applied to front and back surface 3D radii to ensure consistency.

Table 3.1: Average and deviation scaling factors across all coordinates for all patients, shown above with a black line and shaded area respectively.

	X	**Y**	**Z**
Mean ± St. Dev	0.979 ± 0.090	0.999 ± 0.082	1.009 ± 0.066

During the iterative optimization of the corneal curvature, the horizontal position of the pupil center is allowed to shift along the optical axis by up to ±1.5 mm to facilitate the process. **Figure 3.2** shows the optimized values for this parameter, with only 25% of cases hitting the lower boundary, and with two patients for whom this deviation was not exploited by the optimizer. The resulting Δx values across patients yield an average of -0.894 ± 0.533 mm, therefore reflecting a tendency of pulling back the pupil around 1 mm on average. Despite the possibility of doing so, the optimizer never exploits the forward direction shift. This backward shift may be attributed to the fact that the iris thickness was assumed to be zero.

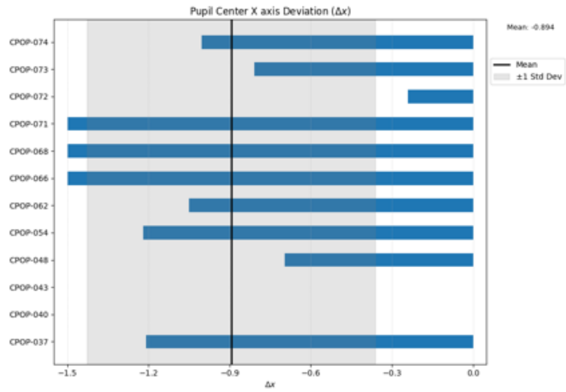


Figure 3.2: Optimized pupil center x axis deviation from original center of $(0, 0, 0)$, computed together with the cornea radii for the first acquisition of the Planning fraction, with the patient looking straight into the beam.

The hybrid loss (see **Section** 2.2.2.1) quantified in each optimization iteration was responsible for yielding the analyzed results. As can be seen in **Figure 3.3**, after the corneal curvature and pupil center offset optimization, this loss has yielded a reduction of at least 30% for 10 out of the 12 patients analyzed. Only for one subject (Patient CPOP-043), the optimization seems to have resulted in worse parameters compared to those given as input, reflecting how the optimizer has not worked successfully for this patient.

The decrease in mismatch between the two analyzed ellipses is reflected in a sharp increase in pupil % overlap after the optimization, with peaks of up to 84% for Patient CPOP-048. The failed optimization for Patient CPOP-043 is once again proven by the lowered overlap % after optimization.

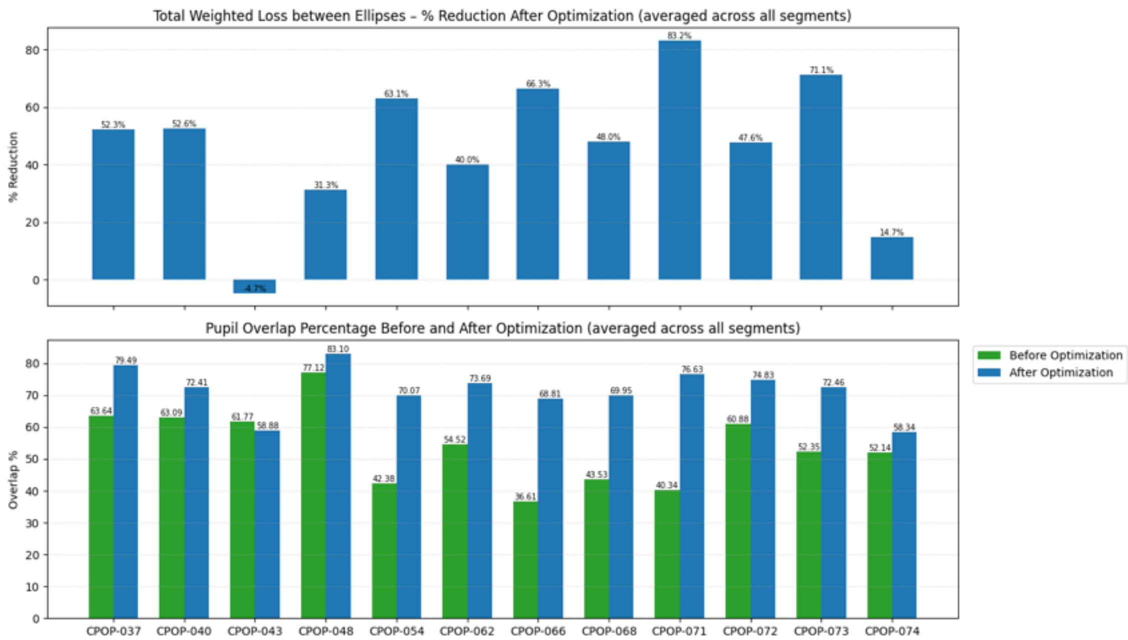


Figure 3.3: Average % loss reduction and average % pupil overlap before and after cornea curvature optimization for all patients.

3.1.1 Effect on simulated pupil contours and alignment with imaging data

The optimization results were visualized by overlaying the ETS images to the initial MRI-model retro-projected pupil (seen in **Figures 3.4** and **3.5** as a solid blue line, with the segmented ETS image ellipse in solid red, and the optimized pupil in dashed green).

Throughout the population, the optimizer was seen to behave differently, and therefore heavily relied on the initial MRI-based model. In **Figure 3.4** below, the optimizer visibly improves the retro-projection of the pupil by tweaking the cornea and pupil parameters in the first acquisition of the planning fraction. Not only, but when these parameters are applied to other fractions and segments, only allowing the pupil size to variate, the results also show a clear improvement (see **Figure 3.4**).

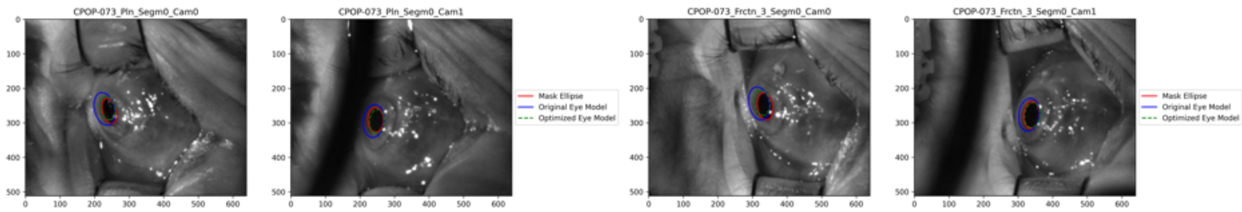


Figure 3.4: Patient CPOP-073 cornea and Δx pupil center optimization for first acquisition of the planning fraction and fraction 3 respectively. Scaling factor = [0.900,1.098,1.099], Δx = -0.809

This trend, however, is not stable across the population. **Figure 3.5** below is an example of a successful optimization of the cornea and pupil based on the first planning acquisition, but when these same parameters are applied to other acquisitions, the resulting pupil does not fit the mask perfectly.

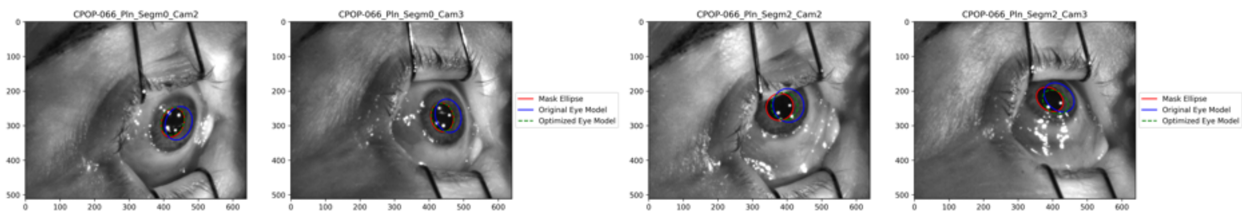


Figure 3.5: Patient CPOP-066 cornea and Δx pupil center optimization for first and third acquisition of the planning fraction. Scaling factor = $[0.904, 0.918, 1.071]$, Δx=-1.500

The individual results for the optimization of these two cases can be found in **Table 3.2** below.

Table 3.2: Optimization values for analyzed individual patients and fractions. Weighted loss calculated as a weighted sum of both camera contributions, summing the algebraic, geometric and area penalties.

Patient	**Weighted Loss % Reduction**	**Initial Pupil % Overlap**	**Final Pupil % Overlap**
CPOP-073 / Pln / 0	76.514	51.134	72.361
CPOP-073 / Fraction3 / 0	67.362	53.643	72.398
CPOP-066 / Pln / 0	89.080	50.878	91.924
CPOP-066 / Pln / 2	56.596	39.733	59.934

.

3.1.2 Application of Optimized Cornea Shape to MRI-based Eye Model

This first optimization has yielded visible improvements in results for the localization of the eye's pupil given the optimized measurements for corneal curvature and pupil center. To visually assess the anatomical feasibility of these structures, the new parameters were re-inserted in the RayOcular software and overlaid to the MRI images of each patient. **Figure 3.6** below shows the result of the optimized cornea for patients CPOP-054 and CPOP-062 respectively.

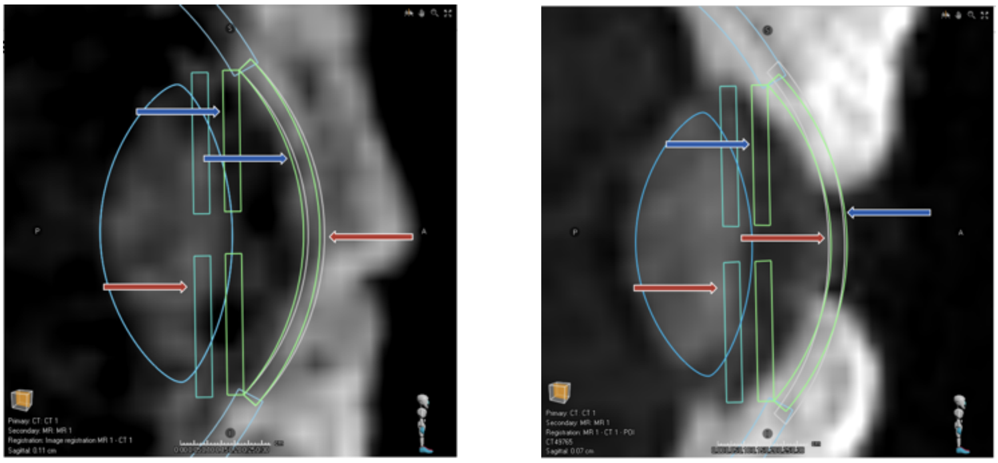
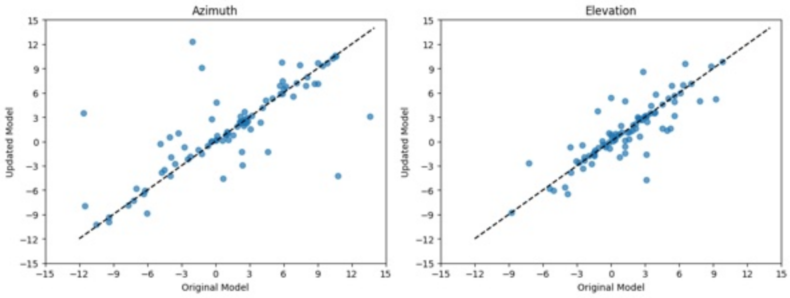


Figure 3.6: Original iris and cornea (blue arrows) overlaid to the patient's MRI compared to the optimized Δx and corneal curvature for patients CPOP-054 and CPOP-062 respectively (red arrows).

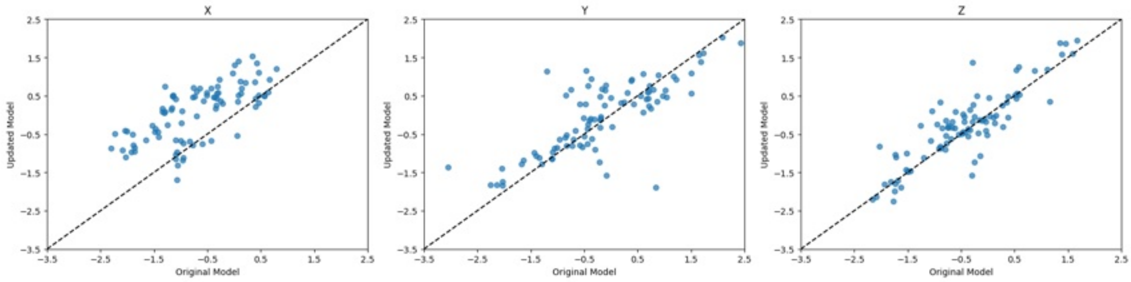
3.2 Eye Pose Estimation: Original vs Updated Eye Models

As previously mentioned, the eye pose, comprising of the eye's position and orientation in the treatment room, was estimated separately for the original MRI model and the updated model which included the optimized corneal curvatures and pupil offsets resulting from the optimization seen in **Section 3.1**.

The results were compared to the ground-truth estimated values obtained by registering the MRI model to the 3D position of the fiducials measured on X-ray images. **Figure 3.9** below shows the deviation compared to this ground truth for all five components of the pose: azimuth and elevation angles as well as translation in the x, y, and z Fick directions.



(a) Pose angles: Azimuth and Elevation



(b) Pose translations: Tx, Ty, Tz

Figure 3.7: Scatterplots showing differences in deviations for all patients for (a) eye pose angles and (b) translations compared to that estimated by aligning the X-ray fiducials to the MRI-based eye model.

Tables A.1 and **A.2** in **Appendix A** show that for most parameters, the mean deviation compared to the ground truth stays within the ± 1 *mm* or $\pm 2^\circ$ interval, demonstrating the robustness of the optimizer in converging towards a plausible eye pose for each patient and acquisition.

Between the two optimizations of eye pose, a Wilcoxon paired test was carried out for each of the five parameters. This test was chosen as it is suitable for patient paired data which is non-normally distributed. A test with $n = 89$ showed significant differences between original and updated pose methods for translations in the x and z directions, with p-values of $1.211 * 10^{-13}$ and 0.004 respectively. Non-significant differences were found for azimuth, elevation angles as well as translations in the y direction, with p-values of 0.122, 0.910 and 0.059 respectively.

These results can be visualized in **Figure 3.8** below, which shows the differences in deviations between the two models and the respective ground truth.

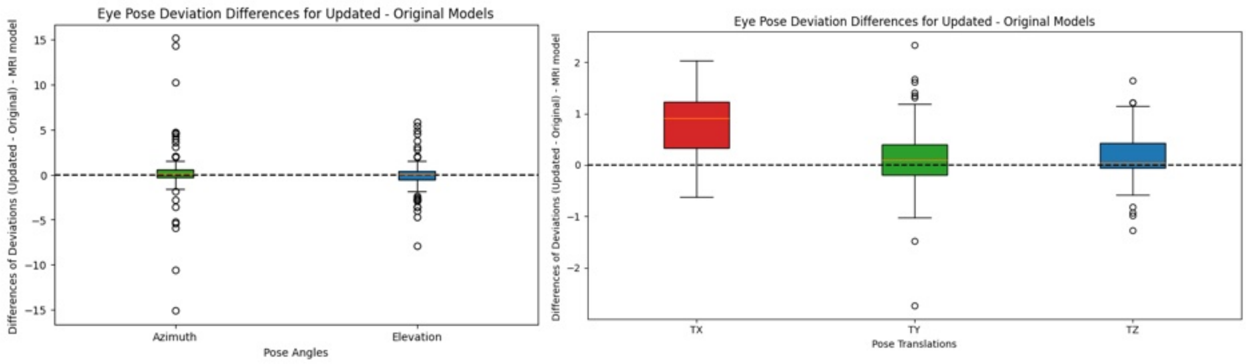


Figure 3.8: Difference in deviations between original and updated model eye pose for all parameters.

It is worthy to note that while the results for the eye pose estimation yield results very close to those obtained through the registration to the X-ray fiducials, the optimization process is strongly dependent on the initial guess given as input. The conversion of light position into eye pose angles, as well as the simple triangulation performed between pupil centers, is crucial in determining a reasonable starting point allowing the optimization process to converge correctly. Nevertheless, as can be seen in **Table 3.3** below, the optimization parameters on average differ substancially from those given as input.

Table 3.3: Differences in azimuth, elevation, and translation (T_x, T_y, T_z) between initial optimization input and optimized results (mean ± st. dev.).

Model	Azimuth (°)	Elevation (°)	T_x (mm)	T_y (mm)	T_z (mm)
Original	0.972 ± 5.715	-1.182 ± 2.969	-0.183 ± 0.672	-0.481 ± 3.119	-1.301 ± 2.244
Updated	1.276 ± 5.961	-1.277 ± 3.532	0.592 ± 0.889	-0.359 ± 2.849	-1.162 ± 1.914

3.2.1 Relationship Between Estimated Eye Model and Image Ground Truth

Compared to the results for the optimization of eye model parameters exclusively, the weighted loss reduction across all patients is larger when optimizing the eye pose: all results show at least a 50% reduction on average across all segments analyzed, with a corresponding steep increase in pupil % overlap for both models analyzed.

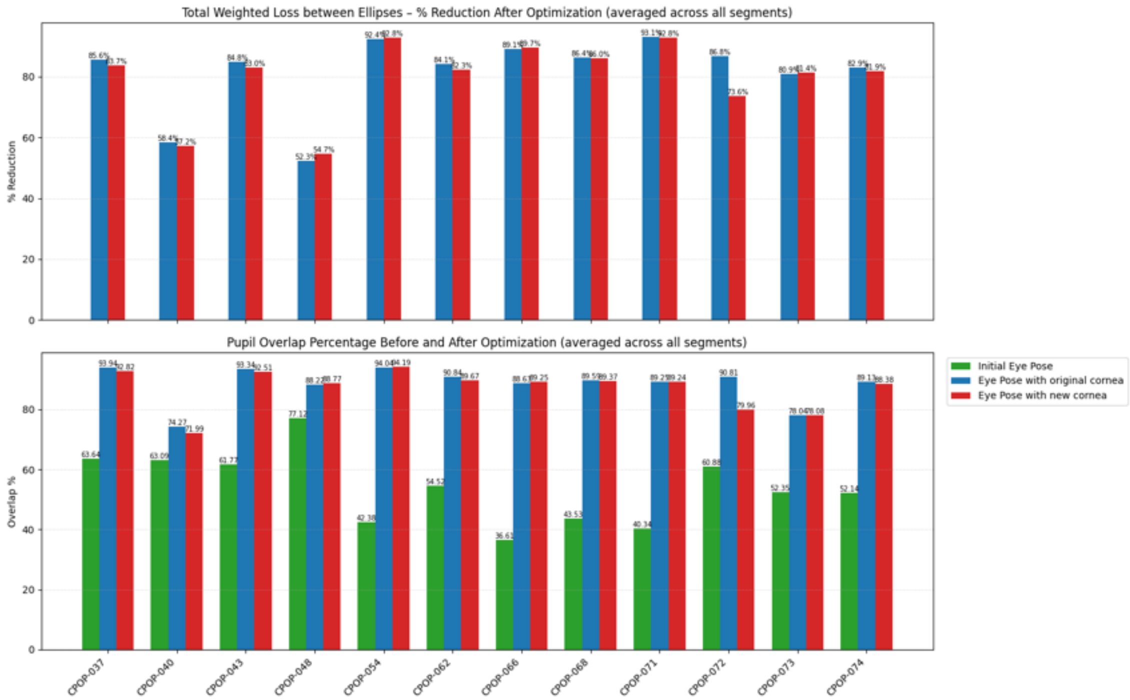


Figure 3.9: Average % loss reduction and average % pupil overlap before and after eye pose optimization for all patients for the two models: the original MRI model and the optimized model with new cornea curvature.

As for the first optimization, the results of the eye pose estimation were visualized by overlaying ETS images with the initial ellipse, mask ellipse from the image as well as the optimized ellipse with the new pose, while keeping all model parameters constant.

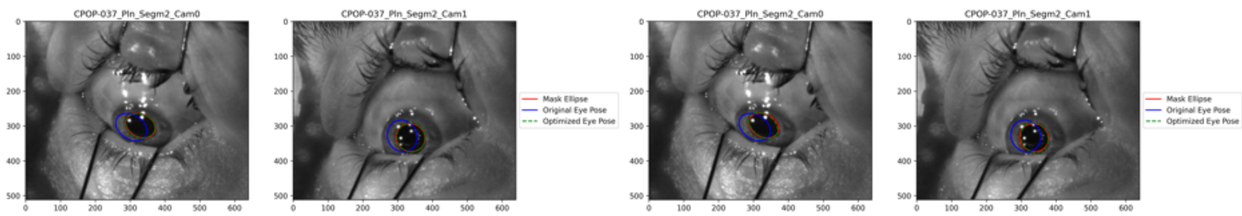


Figure 3.10: Patient CPOP-037 eye pose optimization for the original MRI model and the updated model respectively.

Of note, for patient CPOP-043, whose corneal curvature optimization yielded lower results compared to its initial values. The eye pose optimization successfully adjusted the pupil superposition, as can be clearly noted in **Figure 3.11** below.

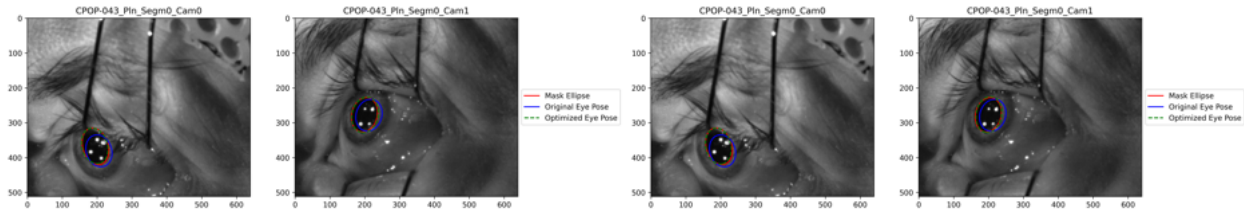


Figure 3.11: Patient CPOP-043 eye pose optimization for the original MRI model and the optimized corneal curvature model respectively. Both optimizations yielded better results in the separate strategy where pose and pupil size are handled aside.

The individual results of these optimizations can be seen in **Table 3.4** below.

Table 3.4: Optimization values for analyzed individual patients and fractions. Weighted loss calculated as a weighted sum of both camera contributions, summing the algebraic, geometric and area penalties.

Patient	**Model**	**Weighted Loss % Reduction**	**Initial Pupil % Overlap**	**Final Pupil % Overlap**
CPOP-037 / Pln / 2	Original	94.044	43.982	95.708
CPOP-037 / Pln / 2	Updated	89.282	43.982	91.016
CPOP-043 / Pln / 0	Original	83.482	63.902	93.261
CPOP-043 / Pln / 0	Updated	83.085	63.902	92.996

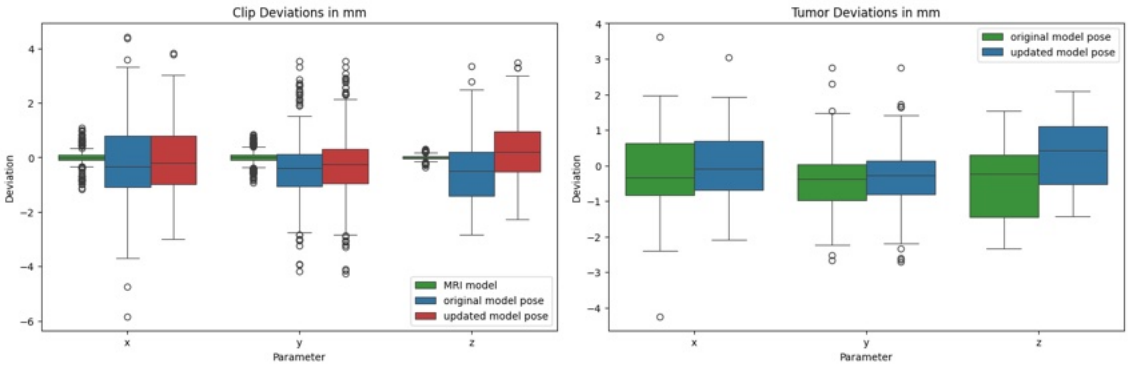
3.2.2 Accuracy of Pose Estimation Compared to Clip-Based Ground Truth

To assess the accuracy of the eye pose estimation using both eye models, the position of the clip centroids as well as the tumor center of mass were compared to those inferred by the X-ray measurements of the fiducial positions. While the MRI model registration is uncertain and prone to small errors and residuals, the X-ray clip positions are a fixed, secure measurement of the eye location.

Deviations for all patients, fractions, and acquisitions for both clip and tumor parameters (in x, y, and z axis for the world coordinate system) can be seen in **Figure 3.12** below. Once again, a paired Wilcoxon test was carried out between both models, yielding significant differences for translations in the y (vertical) and z (beam-axis) directions for both clip and tumor centroids, while the x (lateral) direction differences were not significant. The p-values resulted as follows:

- Clip Centroid Deviations: x = 0.140, y = 0.007, z = $4.502 * 10^{-12}$
- Tumor Center of Mass Deviations: x = 0.121, y = 0.005, z = $2.592 * 10^{-12}$

The overall statistics derived from these plots are summarized in **Tables A.3** and **A.4** in **Appendix A**.



(a) Clip Centroid Deviations (b) Tumor Center of Mass Deviations

Figure 3.12: Deviations for all patients, fractions and acquisitions of (a) clip centroids and (b) tumor center of mass compared to the original X-ray fiducial locations and the original MRI-based model tumor contour.

The correlation plots below show the agreement between tumor centroid coordinates resulting from eye pose estimation using the original MRI and updated eye models compared to the X-ray ground truth model (see **Figure 3.13**). Values for the correlation coefficients and sum of squared residuals are seen in **Table 3.5** below.

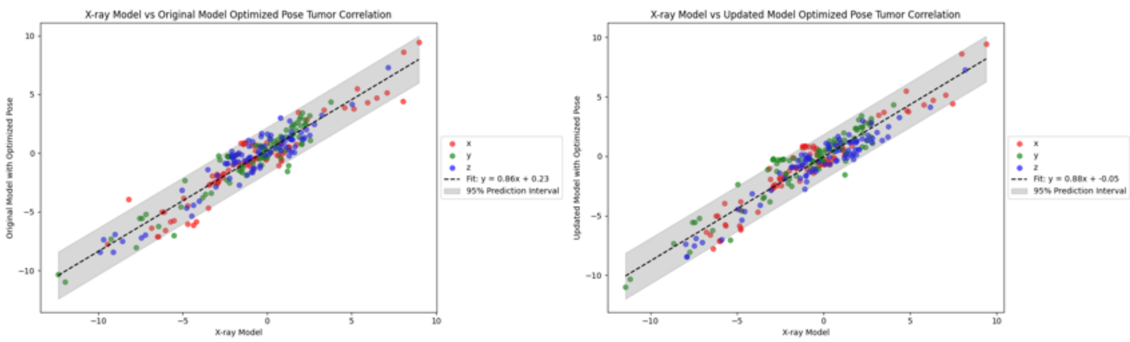


Figure 3.13: Correlation plots for all coordinates pertaining to tumor centroids from all patients, fractions and segments analyzed. On the x axis the ground truth values from the original MRI model; on the y axis the eye pose optimized models.

Table 3.5: Correlation Values for tumor centroid coordinates for each patient comparing the ground truth MRI model to both the optimized models

Model	Equation	R^2	SSE	X Mean ± St. Dev	Y Mean ± St. Dev	Z Mean ± St. Dev	Prediction Interval Width
Original Pose	$y = 0.86x + 0.23$	0.888	261.766	-0.193 ± 1.174	-0.462 ± 1.022	-0.452 ± 1.052	3.948
Updated Pose	$y = 0.88x - 0.05$	0.896	242.448	-0.071 ± 1.014	-0.351 ± 1.061	0.324 ± 0.915	3.799

To ensure the anatomical plausibility, the distances between the clip positions after eye position and orientation were calculated using the proposed optimization process, and the outer scleral surface aligned using X-ray, were plotted and calculated across all methods (as can be seen in **Figure 3.14**). In **Table 3.6**, both models, original and updated, show very similar results compared to the X-ray measurements, with a slight increase in mean deviation compensated by a decrease in the range of extremities.

Table 3.6: Clip – sclera distances (mm) averaged across all eye – pose optimized MRI models and eye – pose + cornea curvature optimized models.

Model	Mean	St. Dev	Minimum	Maximum
Original model	0.006	0.312	−0.725	0.712
Eye pose optimized model	0.028	0.307	−0.628	0.677
Eye pose + cornea optimized model	0.015	0.309	−0.654	0.676

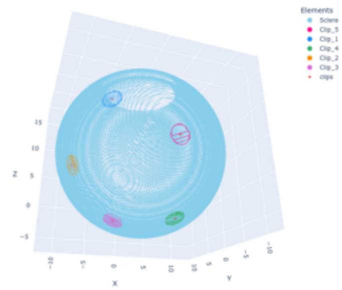


Figure 3.14: Patient CPOP-066 optimized eye pose + cornea curvature model clips for the first acquisition of the planning fraction.

A final qualitative check was performed in RayOcular, where clip centroids were plotted together with those coming from the original MRI model. Results for all averaged acquisitions for patient CPOP-054 can be seen in **Figure 3.15** below.

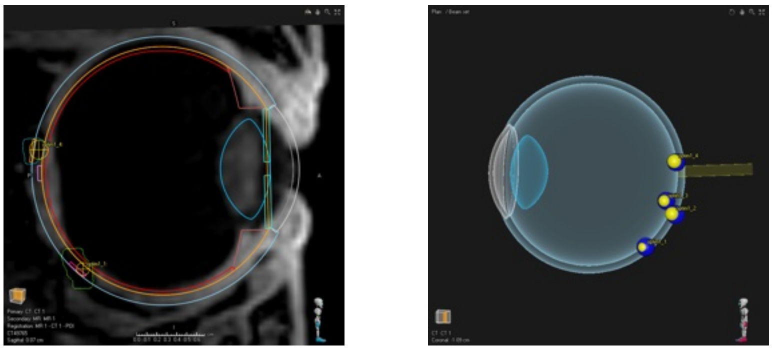


Figure 3.15: Clip positions seen in 2D and 3D for eye pose from the combined optimization using the new corneal curvature and pupil center location for patient CPOP-054. The clip centroid coordinates are computed through an average of all the segments for these patients, and the sphere diameter is defined by the largest range between segments in the 3 directions (x,y,z). On the right, the optimized clips are seen in yellow, while the original clips defined on the MRI model are seen in blue.

Discussion

4.1 Interpretation of Results

4.1.1 Benefit of Incorporating Refraction and Anatomy into Pupil Localization

The results presented clearly demonstrate that integrating refraction-aware modeling and anatomical information into the eye model significantly enhances pupil localization accuracy compared to simple triangulation methods. Across the patient cohort, the loss metric encompassing geometric, algebraic, and area-based terms showed consistent reduction following cornea curvature optimization, with almost all patients having improved alignment between pupil contours.

The initial model based on the MRI alone, while anatomically complete, still lacks fine-tuning of the anterior eye geometry, creating a clearly visible mismatch between the ray-traced pupil and its appearance when seen from the ETS cameras. The two parameters optimized in this study, the curvature of the corneal radii for front and back surfaces, as well as the Δx coordinate of the pupil center, show a first step towards more accurate, patient-specific modeling of the eye. While limiting their optimization bounds to maintain anatomically plausible results, small changes in these variables show a large impact on the improved localization of the entrance pupil. Across patients, the x-coordinate of the corneal radii was the one most exploited by the optimizer, as can be seen in **Table 3.1**, showing an increased or decreased convex-ness of the cornea in this direction.

The backward displacement of the pupil center, averaging at nearly 0.9 mm across patients, likely compensates for the unmodeled iris thickness. The entrance pupil is in fact assumed to stand at the front surface of the iris in the original model. The tendency of the optimizer to bring back the pupil center highlights the importance of subtle anatomical refinements.

The clear correlation between reduced mismatch between pupil ellipses as well as increased pupil overlap after optimization in this study emphasizes the clinical utility of incorporating accurate refractive modeling for enhancing eye pose estimation in clip-less workflows. These findings are consistent with the work by Aguirre, who emphasizes the need to simulate light refraction through curved corneal surfaces to correctly estimate the entrance pupil's true position and shape.

In the optimization of the corneal curvature and pupil center location, the parameters were kept constant across each patient, and therefore only optimized in the first planning acquisition, where the patient was looking straight into the beamline. This was done as studies do not report any significant changes in corneal curvature during brief periods of time. While a trend is seen in corneal flattening over a 14-year period by Scheiman et al., non-significant changes in corneal shape may occur in the short term as a result of the actions of extrinsic muscles and ciliary muscles [18, 17].

What was noted however by Kawamorita et al. is that changes in corneal curvature can occur when transitioning from a standing to a lying down position. This is particularly of interest because the MRI is taken in a supine position, while all treatments are performed with the patient in a sitting position. Studies have shown that curvature can increase in supine position, due to the different gravity and pressure forces acting on the eye [19]. As the cornea is defined on the supine MRI model, it is important that this finding is taken into consideration when comparing the entrance pupil derived from this model and the one seen through the ETS cameras when the patient is sitting in the treatment room.

The results seen in **Figure 3.5** could perhaps be explained by this phenomenon. A perfectly conformal pupil in the first acquisition, but a clear mismatch (despite the improvement) in a later segment could be justified by two factors. On one hand, a very slight change in curvature given sometime between the two acquisitions might have impacted this patient specifically. On the other hand, a mismatch in the gaze orientation during planning image acquisition might have been the cause of the error seen.

The convergence of the optimized parameters reaches the set boundaries several times in the analyzed patient population, showing a tendency of the optimizer to go further in curvature and offset compared to the limits given (see **Figures 3.1** and **3.2**). In this study, all ranges were empirically set a-priori to maintain anatomical feasibility. However, future studies might extend the range to include more curvatures and pupil centers, especially because the initial settings given by the MRI-based model created on RayOcular contain approximate locations for the iris location, especially.

It is also important to note that, contrary to the rest of the study, this optimization did rely on the eye pose obtained from registering the MRI model to the X-ray based locations of the clips. The pose was kept constant while different anatomies were explored by the optimizer. When moving towards a fully clip-less workflow, this step must be compensated by other measurements. For instance, an initial pose could be taken from the gaze optimization of the MRI model: this pose would be fed into the anatomical optimization whose parameters would in turn be utilized to fine-tune the pose again. In addition, recent efforts towards integrating scleral-cornea surface measurement to the ETS led to promising results [10]. Using such a system in combination with a refraction-aware model of the ocular anatomy of the patient could lead to further improvements in the results.

4.1.2 Comparing Anatomical and Pose-Based Optimization Strategies

When comparing optimization strategies, it becomes apparent that the greatest improvements stem from the combined optimization of the anatomical modeling of the cornea and pupil, as well as the pose optimization, compared to solely focusing on the latter. While the corneal curvature optimization seems to have improved the anatomical accuracy of the model, it was the subsequent estimation of eye pose parameters that led to the most substantial reduction in loss across all patients.

While the results of optimizing eye pose using the original MRI model also demonstrated a substantial improvement in pupil shape, orientation and size, the comparisons with the ground truth eye pose estimated from the X-ray fiducial clips yielded slightly worse results in deviations compared to the updated eye model. This suggests that the anatomical refinement provides a more complete structural basis for the pose optimization to converge towards the true patient gaze. It is clear from the MRI images that such structures cannot be clearly seen from the acquisition itself. Even though the optimized parameters do not seem to deviate greatly from the original values, a further validation could stem from using other forms of ophthalmologic imaging such as keratometry or optical coherence tomography.

The combined use of anatomical updates and eye pose estimation achieved lower deviations in clip and tumor centroid locations when compared to X-ray measurements, particularly in the z direction. Not only, but the agreement of the optimized tumor centroid coordinates compared to those coming from the original MRI model achieved a higher goodness of fit and means in all coordinates closer to 0 (demonstrating higher agreement with ground truth values, see **Table 3.5**).

Despite these advantages, it must be noted that pose optimization was highly sensitive to input conditions, relying heavily on approximate initial estimates derived from the light source position as well as the ETS based pupil location. This emphasizes the need for robust preprocessing and optimization strategies to enable correct estimation of the eye pose. As can be seen in **Table 3.3**, the optimizer adjusted the two angular parameters by an average of less than $\pm 2^\circ$ and translations by less than ± 2 mm, despite the much broader bounds being available during optimization. Considering that a shift in 3° corresponds to a displacement of around 0.5 mm, one can note that the differences in translation are of higher order compared to the angular displacement. This imbalance is further explained by the fact that translations and rotations were not normalized during optimization: angular parameters were kept in degrees, and translations in millimeters, resulting in inherently larger changes in the cost function due to translational adjustments. For these reasons, and given that multiple combinations of these five parameters can result in similar gaze projections, it is essential that the optimizer explores the parameter space more uniformly, to avoid converging to local minima that fit image data but do not reflect the true anatomical pose.

Furthermore, the significant differences and comparisons to the ground truth model must be analyzed carefully for two main reasons. First, the pose differences are computed against a ground truth which itself is the result of registering the MRI model to X-ray clip locations: this registration process carries its own residual errors and is therefore not a guaranteed ground truth. Clips appear as a void signal in MRI scans, typically much bigger than the volume of the clips themselves. This brings uncertainty in the registration process and, as a consequence, in the definition of the ocular anatomy with respect to these fixed fiducials. As such, what looks like a large error in clip localization using ETS images might just be a systematic component related to these uncertainties; a further analysis on this is definitely required.

Second, the five pose parameters are not independent of one another. In fact, if these parameters were to be combined into a single 5D vector and compared using the norm of their differences (rather than evaluating them individually), their difference could potentially not be statistically significant anymore. This would stand in contrast to the individual Wilcoxon tests carried out on each parameter, which did report significance for two out of the five variables. This discrepancy highlights the need to carefully consider whether to evaluate pose parameters individually or as a combined metric when judging the performance of these models.

4.2 Limitations

4.2.1 Simplifications in Patient Specific Multi Modal Eye Model Definition

A key limitation of the current approach lies in the simplifications made to the internal eye model. This can be seen in **Figure 3.6**, where the iris contours do not consider the pupil size optimized in this study. A wider pupil size would avoid the overlay between the iris and the lens as seen in both patients. Including lens parameters, as well as modeling the iris given a certain thickness rather than assuming the pupil lies on a single plane, would therefore allow for the creation of an even more precise eye model.

As noted previously, the patient-specific eye model was created based on the MRI acquisition, whose anatomical resolution is limited, especially in visualizing small structures such as the iris thickness, the lens position, and the anterior chamber depth or contour of the cornea. As a result, several anatomical parameters were estimated and had to be optimized. Moreover, the alignment between the MRI and the reconstructed CT clip positions carries with it a small registration residual, which can introduce a small variability that is carried through all the optimizations and calculations. All these factors limit the fidelity of the initial model and highlight once again the need for more integrated and higher resolution ophthalmological imaging data to better define the eye geometry.

4.2.2 Lack of Anatomical Ground Truth for Eye Pose Validation

A second limitation lies in the validation of eye pose estimation. The calculated offset between the optimized pose and the one computed by aligning the MRI model to the clip positions is constrained by the absence of a true anatomical ground truth. The fiducial clips from the X-ray imaging are a highly accurate reference, but they only provide rigid external markers for the eye and not a direct measurement of eye orientation and gaze.

Additionally to this, the MRI clip artifacts appear much larger than the clip sizes themselves;

therefore, it is not possible to create a direct link between the artifacts and the X-rays. While the larger size of the artifact seen on the MRI gives space for the optimization to converge to slightly different centroid locations, any deviations with respect to the MRI contoured clips must be interpreted with caution. The same applies to the measurement of tumor centroids and volume overlap: both metrics are, in fact, heavily dependent on the MRI model, and deviations could be misinterpreted when the ground-truth value deviates from the real truth.

4.3 Future Work and Opportunities

4.3.1 What Can Be Achieved Without Clips

This study provides quantitative evidence that accurate eye pose estimation could be feasible using non-invasive optical tracking in combination with anatomically and optically refined models. Nevertheless, the introduction of new technologies requires extreme caution because, however well-intentioned, changes must not have a negative effect on the quality of patient care. At the current state of development, non-invasive eye tracking cannot still guarantee the same level of accuracy and reliability provided by clips. At the same time, results are extremely encouraging, and the sources of uncertainties are being identified. Given the results described, this method could serve as a basis for real-time, non-invasive localization during treatment. This could be particularly feasible, initially, for patients who are susceptible to clip implantation, due to their age, eye condition, or surgical risk, and would therefore benefit from the exclusion of a surgical intervention.

4.3.2 What is Needed to Overcome Clip Dependence

To truly eliminate the reliance on implanted clips, further work is required in several key areas. First, improving the precision and automation of image registration, particularly between MRI and optical imaging, would be critical. There is a compelling need for more accurate eye models which not only encompass anatomical modeling variations in iris thickness, lens shape, corneal curvature, but also are able to accurately reproduce the true patient's anatomy from a given imaging modality. In the absence of clip measurements, other methods must be found to validate and monitor the consistency of pose derived from optical imaging to provide the necessary accuracy for clip-free alignment in treatment.

Furthermore, due to a divergence between the optical and visual axes, defining the eye pose solely on optical features such as the pupil center can lead to systematic misalignment [20]. In OPT, the eye is actively fixating during treatment and the beam is aligned relative to the patient's visual axis; therefore, all tracking and modeling systems must account for the angular offset between the former and the optical axis, an aspect which has not been considered in this study but has already been investigated in the past [20]. Ocular torsional rotation should also be accounted for. In the proposed method, torsion was modeled merely as a function of the azimuthal and elevation rotations of the eye. Moving towards a more patient-specific approach could potentially lead to improved results.

4.3.3 Timing Considerations

One final consideration which must be mentioned is the computational efficiency of the proposed workflow. The current optimization process, despite its consistent and accurate results, is time-consuming and heavily dependent on initial conditions, making its feasibility towards real-time applications challenging. Significant acceleration of the optimization processes, either through modeling, GPU-based systems, or refinement of the initial models, is therefore necessary. Future research should therefore prioritize the accuracy and reliability of models and imaging procedures, but also take their usability and clinical integration into consideration.

Conclusion

This study investigated the feasibility of improving eye pose estimation in OPT through anatomically refined and refraction-aware modeling of the eye. By optimizing the corneal curvature through its front and back surface radii and the pupil center position in an MRI-based ocular model, eye pose parameters using a ray-traced simulation framework based on Aguirre's model could be calculated. The extent to which non-invasive optical imaging can match the accuracy of traditional clip-based localization was therefore analyzed.

Results demonstrated that incorporating both anatomical and optical corrections substantially reduced re-projection error between simulated and observed pupil contours, achieving sub-millimeter alignment in most cases. While pose optimization yielded the largest improvements in pupil overlap, its combination with ocular anatomy refinement provided a more realistic and accurate eye model and measurement of eye position and orientation across treatment fractions , closely matching the ground truth derived from implanted fiducials, both in terms of clip and tumor centroid localization.

Despite this, limitations related to eye model simplification, computational efficiency, multi-modal alignment accuracy, and dependence on initial estimates must be taken into consideration before a fully clip-less workflow can be adapted to the clinical setting.

Overall, this work could represent a small step towards non-invasive and accurate eye localization in proton therapy, with the potential of improving patient comfort and reducing the burden of undergoing a non-curative surgical procedure.

Appendix 1 - Raw Data

Table A.1: Deviation statistics for all patient, fractions and segments of eye pose for the MRI-based model optimization.

Parameter	Mean	St. Dev	Minimum	Maximum
Azimuth	1.350	5.434	−11.627	13.619
Elevation	1.298	5.185	−8.774	9.830
$\mathbf{T_x}$	−0.708	0.790	−2.300	0.787
$\mathbf{T_y}$	−0.173	1.044	−3.045	2.433
$\mathbf{T_z}$	−0.442	0.895	−2.158	1.668

Table A.2: Deviation statistics for all patient, fractions and segments of eye pose for the new MRI-based model optimization with optimized corneal curvature.

Parameter	Mean	St. Dev	Minimum	Maximum
Azimuth	1.653	5.185	−10.279	12.307
Elevation	1.203	3.662	−8.773	9.880
$\mathbf{T_x}$	0.067	0.734	−1.696	1.541
$\mathbf{T_y}$	−0.051	0.935	−1.898	2.022
$\mathbf{T_z}$	−0.304	0.946	−2.258	1.954

Table A.3: Deviation statistics for all patient, fractions and segments of clip centroid positions (n clips = 54) for both optimized models.

Parameter	Model	Mean	St. Dev	Minimum	Maximum
X	X-ray	$2.039 * 10^{-7}$	0.283	−1.161	1.081
Y	X-ray	$6.592 * 10^{-8}$	0.268	−0.919	0.835
Z	X-ray	$-3.490 * 10^{-7}$	0.084	−0.373	0.308
X	original	−0.199	1.394	−5.841	4.428
Y	original	−0.438	1.161	−4.194	3.525
Z	original	−0.522	1.089	−2.839	3.347
X	updated	−0.103	1.240	−2.999	3.832
Y	updated	−0.313	1.203	−4.254	3.538
Z	updated	0.269	1.003	−2.262	3.474

Table A.4: Deviation statistics for all patient, fractions and segments of tumor center of mass location for both optimized models.

Parameter	Model	Mean	St. Dev	Minimum	Maximum
X	original	−0.193	1.174	−4.243	3.621
Y	original	−0.462	1.022	−2.655	2.747
Z	original	−0.452	1.052	−2.333	1.544
X	updated	−0.071	1.014	−2.094	3.045
Y	updated	−0.351	1.061	−2.710	2.760
Z	updated	0.324	0.915	−1.429	2.085

Bibliography

[1] Riccardo Via, Giovanni Fattori, Alessia Pica, Antony Lomax, Guido Baroni, Damien Charles Weber, and Jan Hrbacek. Eye tracking in ocular proton therapy. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA â20, page 1â3. ACM, June 2020. doi:10.1145/3379157.3391991.

[2] Riccardo Via, Fabian Hennings, Giovanni Fattori, Aurora Fassi, Alessia Pica, Antony Lomax, Damien Charles Weber, Guido Baroni, and Jan Hrbacek. Noninvasive eye localization in ocular proton therapy through optical eye tracking: A proof of concept. *Medical Physics*, 45 (5):2186â2194, March 2018. ISSN 2473-4209. doi:10.1002/mp.12841.

[3] Chiara Spaccapaniccia, Riccardo Via, Vincent Thominet, Amy Liffey, Guido Baroni, Alessia Pica, Damien Charles Weber, Antony John Lomax, and Jan Hrbacek. Non-invasive recognition of eye torsion through optical imaging of the iris pattern in ocular proton therapy. *Physics in Medicine amp; Biology*, 66(13):135014, July 2021. ISSN 1361-6560. doi:10.1088/1361-6560/ac0afb.

[4] Jörg Wulff, Benjamin Koska, Jens Heufelder, Martin Janson, Claus Maximilian Bäcker, Hilda Siregar, Carina Behrends, Christian Bäumer, Andreas Foerster, Nikolaos E. Bechrakis, and Beate Timmermann. Commissioning and validation of a novel commercial tps for ocular proton therapy. *Medical Physics*, 50(1):365â379, October 2022. ISSN 2473-4209. doi:10.1002/mp.16006. URL `http://dx.doi.org/10.1002/mp.16006`.

[5] Ch. Markovits, S. Jaccard, and Ch. Perret. The proton beam facility optis for the therapy of ocular tumours. In *THE PROTON BEAM FACILITY OPTIS FOR THE THERAPY OF OCULAR TUMOURS*, Proceedings of the Twelfth International Conference on Cyclotrons and their Applications, Berlin, Germany. ACM, 1989.

[6] Riccardo Via, Aurora Fassi, Giovanni Fattori, Giulia Fontana, Andrea Pella, Barbara Tagaste, Marco Riboldi, Mario Ciocca, Roberto Orecchia, and Guido Baroni. Optical eye tracking system for realâtime noninvasive tumor localization in external beam radiotherapy. *Medical Physics*, 42(5):2194â2202, May 2015. ISSN 2473-4209. doi:10.1118/1.4915921.

[7] Geoffrey K. Aguirre. A model of the appearance of the moving human eye. February 2021. doi:10.1101/2021.02.02.429411.

[8] Geoffrey K. Aguirre. A model of the entrance pupil of the human eye. *Scientific Reports*, 9 (1), June 2019. ISSN 2045-2322. doi:10.1038/s41598-019-45827-3.

[9] Riccardo Via. "imaging the eye, improving eye modeling for patient comfort. Paul Scherrer Institut, 2025.

[10] Riccardo Via, Katarina Bryjova, Alessia Pica, Guido Baroni, Antony Lomax, Damien Charles Weber, Giovanni Fattori, and Jan Hrbacek. Multi-camera optical tracking and fringe pattern analysis for eye surface profilometry in ocular proton therapy. *Physics and Imaging in Radiation Oncology*, 28:100517, October 2023. ISSN 2405-6316. doi:10.1016/j.phro.2023.100517.

[11] Mohamed Kilany Hassan, Emmanuelle Fleury, Denis Shamonin, Lorna Grech Fonk, Marina Marinkovic, Myriam G. Jaarsma-Coes, Gregorius P.M. Luyten, Andrew Webb, Jan-Willem Beenakker, and Berend Stoel. An automatic framework to create patient-specific eye models from 3d magnetic resonance images for treatment selection in patients with uveal melanoma. *Advances in Radiation Oncology*, 6(6):100697, November 2021. ISSN 2452-1094. doi:10.1016/j.adro.2021.100697.

[12] Conny Lu, Praneeth Chakravarthula, Kaihao Liu, Xixiang Liu, Siyuan Li, and Henry Fuchs. Neural 3d gaze: 3d pupil localization and gaze tracking based on anatomical eye model and neural refraction correction. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, page 375â383. IEEE, October 2022. doi:10.1109/ismar55827.2022.00053.

[13] Carlos Ciller, Sandro De Zanet, Konstantinos Kamnitsas, Philippe Maeder, Ben Glocker, Francis L. Munier, Daniel Rueckert, Jean-Philippe Thiran, Meritxell Bach Cuadra, and Raphael Sznitman. Multi-channel mri segmentation of eye structures and tumors using patient-specific features. *PLOS ONE*, 12(3):e0173900, March 2017. ISSN 1932-6203. doi:10.1371/journal.pone.0173900.

[14] Thomas Haslwanter. Mathematics of three-dimensional eye rotations. *Vision Research*, 35 (12):1727â1739, June 1995. ISSN 0042-6989. doi:10.1016/0042-6989(94)00257-m.

[15] Zhengyou Zhang. *Camera Parameters (Intrinsic, Extrinsic)*, page 135â140. Springer International Publishing, 2021. ISBN 9783030634162. doi:10.1007/978-3-030-63416-2_152. URL `http://dx.doi.org/10.1007/978-3-030-63416-2_152`.

[16] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550â560, December 1997. ISSN 1557-7295. doi:10.1145/279232.279236.

[17] Mitchell Scheiman, Jane Gwiazda, Qinghua Zhang, Li Deng, Karen Fern, Ruth E. Manny, Erik Weissberg, and Leslie Hyman. Longitudinal changes in corneal curvature and its relationship to axial length in the correction of myopia evaluation trial (comet) cohort. *Journal of Optometry*, 9(1):13â21, January 2016. ISSN 1888-4296. doi:10.1016/j.optom.2015.10.003.

[18] Ji C. He, Jane Gwiazda, Frank Thorn, Richard Held, and Walter Huang. Change in corneal shape and corneal wave-front aberrations with accommodation. *Journal of Vision*, 3(7):1, August 2003. ISSN 1534-7362. doi:10.1167/3.7.1.

[19] Takushi Kawamorita, Tomoya Handa, and Hiroshi Uozato. Changes of corneal aberrations in sitting and supine positions. *American Journal of Ophthalmology*, 141(2):412â414, February 2006. ISSN 0002-9394. doi:10.1016/j.ajo.2005.09.010.

[20] Daniel H. Chang. Determining ocular fixation for centering iols with purkinje images. *Cataract and Refractive Surgery Today*, July 2011.