

EYE GAZE METRICS FOR SKILL ASSESSMENT AND FEEDBACK IN KIDNEY STONE SURGERY

By

Yizhou Li

Dissertation

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements  
for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE

May 12, 2023

Nashville, Tennessee

Approved:

Jie Ying Wu, Ph.D.

Nicholas L Kavoussi, M.D.

## ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to those who have supported me in the completion of my master's thesis.

First and foremost, I would like to extend my sincere thanks to Dr. Wu and Dr. Kavoussi for their invaluable support and assistance during the experiment. Dr. Wu's guidance and encouragement have been instrumental in the progress of my project, and I am deeply grateful for her dedication and support. I also appreciate Dr. Nick's expertise and the way he brought our experiment closer to real-life surgical procedures.

Additionally, I would like to express my appreciation to Dr. Bodenheimer and the students in his lab for their patience and helpful answers when I encountered difficulties. Their support has been invaluable to me, and I am grateful for their contributions to my academic journey.

I am also thankful to my roommate, Xinyu Li, for his help and support in both my daily life and studies over the past two years. His friendship and assistance have been a source of comfort and motivation for me, and I am truly thankful for his presence in my life.

I would also like to acknowledge the support of Ph.D. student Daiwei Lu, who provided models for the project in this thesis. His contributions are greatly appreciated and have played a crucial role in the success of my research.

Most importantly, I would like to thank my family for their unwavering love and support throughout my academic journey. Their encouragement and belief in me have been a constant source of strength and motivation, and I am deeply grateful for all they have done for me.

In conclusion, I would like to thank all of these individuals for their unwavering support and guidance, without which this thesis would not have been possible. I am truly grateful for their contributions and will always remember their kindness.

## TABLE OF CONTENTS

	Page
<b>LIST OF TABLES</b> . . . . .	<b>iv</b>
<b>LIST OF FIGURES</b> . . . . .	<b>v</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
<b>2 Related Work</b> . . . . .	<b>2</b>
<b>3 Methods</b> . . . . .	<b>3</b>
3.1 Setting the Unity3D project to support eye-gaze recording . . . . .	3
3.2 Calibration Between Users' Eye-gaze and Hololens . . . . .	3
3.3 Creating a virtual plane . . . . .	4
3.4 Skill metrics . . . . .	5
3.5 Build a more realistic kidney phantom . . . . .	7
<b>4 Experiment</b> . . . . .	<b>9</b>
<b>5 Result</b> . . . . .	<b>11</b>
5.1 Qualitative Comparison of Gaze . . . . .	11
5.2 Statistical Analysis . . . . .	11
<b>6 Discussion and Conclusion</b> . . . . .	<b>14</b>
<b>References</b> . . . . .	<b>16</b>

## LIST OF TABLES

Table		Page
3.1	Coordinates of calibration squares . . . . .	4
3.2	Errors before and after five-squares calibration . . . . .	6
5.1	Features from the data . . . . .	12

## LIST OF FIGURES

Figure		Page
3.1	The proposed calibration procedure . . . . .	5
3.2	Virtual plane definition in the Hololens . . . . .	6
3.3	Workflow of building new kidney phantom . . . . .	8
4.1	User study to identify stones in kidney phantoms . . . . .	9
5.1	Heatmap of the eye movements . . . . .	11
5.2	Eye-gaze overlaid with monitor output . . . . .	12
5.3	Ratio of fixation to non fixation . . . . .	13

# CHAPTER 1

## Introduction

The assessment of surgical competency is essential for clinical training and safety purposes. Many of the current methods of surgical evaluation, however, are subjective. This is particularly true for endoscopic kidney stone surgery, where quantitative measures of kidney stone surgery performance are difficult, and there are no quantitative indicators to predict surgical outcomes. This results in almost no way to objectively assess competency. Flexible endoscopes give no kinematic data (unlike robotic surgery) and the review of endoscopic video recordings has not been able to distinguish endoscopic surgical expertise (1). Thus, validated, objective, real-time tools to evaluate expertise during endoscopic stone surgery are necessary. It is particularly important given the high incidence of kidney stone disease and an almost 30 percent risk of stone recurrence after index stone surgery due to residual stone fragments (2). While previous works have shown the effectiveness of eye gaze to distinguish novices from experts, most are retroactive metrics that can only be measured at the end of a task. We seek to develop a skill assessment algorithm based on eye-gaze data to provide temporally-resolved feedback for different segments during training procedures.

Eye tracking has shown potential in assessing surgical skills during laparoscopic/ robotic procedures. In the evaluation of real-time gaze data, metrics such as focal points, saccades, eye contact (3), and pupillary response (4) can distinguish between experts and trainees and give insight into the learning curve for different surgeries (5). Eye tracking, however, has yet to be applied to endoscopic kidney stone surgery. This is due partially to the fact that eye tracking is challenging in cases when a surgeon is concentrating on a restricted video display screen, leading to a relatively smaller area of eye focus. Additionally, surgeons may also move around to get tools and trade with other operators during endoscopic stone procedures which can impair eye-gaze tracking. This motivates head-mounted means to track eye gaze that will move with the surgeon.

Furthermore, different sub-tasks within endoscopic stone surgery have varying technical difficulties, making a summary statistic of eye-gaze data insufficient for accurately evaluating surgeons. Thus, eye tracking should be evaluated for each sub-task as different parts of the kidney are navigated for complete stone localization. We propose a method to integrate eye tracking during endoscopic kidney stone surgery with the endoscope monitor video feed using the Microsoft HoloLens 2 smart glasses - 64 GB (Microsoft Corporation, Redmond, WA, MFG.PART: NJX-00001 CDW PART: 6711261). We further evaluate the eye tracker to distinguish expert and novice surgeons during flexible ureteroscopy (fURS) in kidney phantoms.

## CHAPTER 2

### Related Work

Previous works have demonstrated that novices and expert surgeons have different eye-gaze patterns during both training and clinical tasks. Richstone et al. use a head-mounted tracker to measure pupil and gaze statistics and use them to train classifiers to identify surgeon skill levels (6). Law et al. found that experts tended to focus on target anatomy when during surgery, while novices tended to focus more on the surgical instruments (7). Likewise, Tien et al. showed that expert surgeons look at patient vital signs on the monitor more frequently, suggesting that expert surgeons have a higher level of surgical vigilance intraoperatively (8). Various works, moreover, show that eye-gaze patterns distinguishing between experts and novices hold over when viewing pre-recorded surgical videos (9; 10). This means that eye-gaze is a distinct measurement of skill from manual dexterity and (9) can potentially be used for surgical training.

Various works have used eye gaze for training in laparoscopic simulators. Vine et al. demonstrated that guiding novice training based on expert-gaze patterns resulted in better skill acquisition and retention (11). Liu et al. showed that the gaze-trained novices developed more efficient eye-gaze patterns faster (12). There are few equivalent studies for endoscopic procedures where training is less well-defined. Berges et al. measure eye gaze for an endoscopic sinus surgery task on a cadaver (13). They show that eye-gaze can also be used to distinguish skill level for endoscopic procedures. However, they use a room-mounted infrared camera, which cannot account for surgeon motion.

Eye-gaze-integrated augmented reality systems can be used to facilitate progression throughout a surgery. Lu et al. (14) integrated Pupil Core Eye trackers with the HoloLens 1 for surgical training. They detect when users reach a difficult point in the training task through eye-gaze and provide targeted instructions. A more thorough review of various applications of eye-tracking for surgical research can be found in (15).

While the previous methods can be used in fixed training scenarios, it fails in more complex operating room (OR) scenes where surgeons may be moving with respect to the camera. Since we use a head-mounted tracker that tracks a QR code on the monitor, our method is more robust to occlusions and surgeon movements. It also does not require additional hardware in the OR outside of the QR code.

## CHAPTER 3

### Methods

In order to measure task and eye-gaze-based metrics for skill assessment, we need to measure the eye-gaze of the surgeon with respect to the endoscope monitor. We utilized Unity3D to develop a mixed reality application for collecting eye gaze data. The project was built using the Mixed Reality Tool Kit(MRTK) supported by Microsoft, which provided a wide range of pre-built assets and tools for developing mixed reality applications. The primary goal of the project was to develop an application that could be deployed to Microsoft Hololens 2, allowing for the collection of eye gaze data during our user study. The following sections detail the basic setting for the Unity3D project to support eye-gaze recording (Sec. 3.1), the calibration from the eye-gaze to the Hololens (Sec. 3.2), the Hololens to the endoscope monitor (Sec. 3.3), the metrics we use for skill assessment (Sec. 3.4), and the new phantom we built for the user study (Sec. 3.5).

#### **3.1 Setting the Unity3D project to support eye-gaze recording**

To enable eye gaze data collection in our Unity3D project, we imported the Mixed Reality Toolkit package and activated the eye gaze data, gesture control data, and motion controller functionality. We then imported the default MRTK scene and configuration file that supported mixed reality into the Unity environment. By default, the configuration file calculates and displays the gaze position in real-time. However, when the world camera detects hand gesture input, the gaze position is fixed at a predetermined distance from the user and does not adjust with changes in distance. This can result in inaccurate gaze data collection. To address this issue, we disabled hand gesture control and hand model rendering. Additionally, we made the default gaze game object invisible to avoid interference with the user. As the gaze object is a prefab which is a predefined asset, we used a search algorithm that directly search the object name to locate the object representing the gaze position in the scene and dynamically updated its position in real-time. We also added the capability to record world camera videos by adding an empty game object to the scene, adding the video recording component, and recording the world camera throughout the data collection process. We upload the configuration files, C# scripts, and project settings to Github<sup>1</sup>.

#### **3.2 Calibration Between Users' Eye-gaze and Hololens**

Whenever a new user puts on the Hololens 2 for the first time, it automatically initiates a calibration protocol to measure user-specific eye characteristics for eye-gaze tracking. We found that after calibration, there was

---

<sup>1</sup><https://github.com/vu-maple-lab/Eye-Gaze-Metrics-Hololens2>

still a high error between the measured gaze and fiducials we place in the world. This motivated development of a custom calibration procedure.

Table 3.1: Coordinates of calibration squares

Square	Coordinate x (m)	Coordinate y (m)	Coordinate z (m)
Center	0	-0.2	2
Upper Left	-0.5	0.2	2
Upper Right	0.5	0.2	2
Lower Left	-0.5	-0.6	2
Lower Right	0.5	-0.6	2

In real kidney stone surgery, the surgeon will stand approximately 2 m away from the endoscope monitor. So we display 5 squares at a distance of 2 m in the scene Fig. 3.1, which allows us to achieve the best accuracy at this specific distance. More markers would increase calibration time without adding much coverage while fewer would leave areas of the field of view less calibrated. Each square is 1 cm<sup>2</sup>. The coordinate of each square is shown in Tab. 3.1. We ask the user to look at the center of each square. When the eye gaze sphere is detected within the square, we assume the user is attempting to look at the center. The app records the error between the center of the square and the gaze position. After four seconds of gaze has been detected within a square, the square disappears. Once the user finishes looking at all five squares, the calibration is completed. We turn off all visualizations to avoid introducing visual disturbances. If the deviation between the eye gaze and target never falls below our set threshold of 1 cm<sup>2</sup>, we ask the user to adjust the position of the Hololens, re-run the Hololens calibration task, and restart the proposed calibration task. It takes approximately 40 s to 50 s to complete Hololens' built-in calibration and 20 s to finish our calibration procedure.

During the calibration, we record the offset between the center of each square and the eye-gaze position. Since the eye gaze is a ray, its collision with the square is at the depth of the square and there is no offset on the z-axis. We show the average error on the x-axis and the y-axis in Tab.3.2. Our Hololens App saves those calibration errors. We correct the eye gaze position with saved errors during data replay.

### 3.3 Creating a virtual plane

In order to perform a task-based assessment of eye-gaze metrics, we map the gaze information to the endoscope monitor. We create a virtual plane in Unity to overlay onto the monitor based on a QR code placed in the environment, as shown in Fig. 3.2. To create this plane, we measure the physical size of the main display, which is 70 cm × 39 cm, and build a virtual plane of the same size with a collision model in Unity.

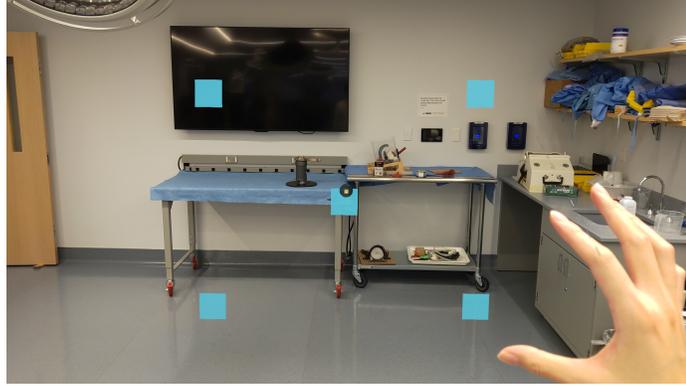


Figure 3.1: The proposed calibration procedure. Once calibration starts, all five squares will be displayed. The black sphere overlaid on the center square represents user’s eye gaze position. The user should stare at the center of each square for 4 s, at which point the active square disappears and the user should move on to the next square.

This plane is generated in the environment when the Hololens detects the QR code. The world camera continuously detects the QR code during the experiment and updates the position of the virtual plane relative to the user. When the user’s measured eye-gaze ray collides with the virtual plane, the app will record the collision position relative to the position of the center of the virtual plane.

Using the QR code frame gives more stable gaze measurements as the orientation sensor and displacement sensor of Hololens 2 accumulate errors, which cause offsets of the eye gaze during the replay. We turn off the spatial awareness of the Hololens. This feature builds the 3D collision model for the ambient environment as it detects the main monitor as several different planes.

### 3.4 Skill metrics

We define the following variables for skill assessment based on previous work. The overall time measures the time to complete each task. The overall distance records the displacement distance of the eye gaze on the virtual plane. Since the data is the coordinates of the eye gaze relative to the virtual plane, we define fixation as keeping the eye movement data within a circle with a diameter of 1.5 cm within about 300 ms. Since our scene is relatively static, with the endoscope motion causing the only change, we disregard eye-gaze patterns active in dynamic scenes. Thus, any eye-gaze pattern that is not a fixation is considered to be a non-fixation.

Fixation time is the total duration of fixation throughout a task, and the fixation to non-fixation ratio is defined by

$$\text{fixation to non - fixation ratio} = \frac{\text{fixation time}}{\text{total time} - \text{fixation time}} \quad (3.1)$$

We define the area of interest(AOI) as the part of the display showing the endoscope image.The actual

Table 3.2: Errors before and after five-squares calibration

User	x-axis (cm)	y-axis (cm)	After x-axis(cm)	After y-axis(cm)
Novice1	0.4	0.5	0	0
Novice2	0.1	0.3	0	0
Novice3	0.2	0.5	0	0
Novice4	0.06	0.5	0.09	0.24
Novice5	1.4	0.09	0	0
Novice6	0.01	0.29	0	0
Expert1	0.3	0.3	0	0
Expert2	0.1	0.2	0	0
Expert3	0.8	0.1	0.5	0
Expert4	0.04	0.13	0	0
Root mean square error	0.54	0.33	0.16	0.07
Standard deviation	0.41	0.15	0.14	0.07

size of the endoscope image is  $35.5 \text{ cm} \times 35.5 \text{ cm}$ . It is located at the center of the virtual plane. The gaze outside AOI denotes the number of times the gaze leaves this area. To calculate the total gaze area, we use Graham scan (16) to find the convex hull from the group of eye-gaze coordinates in AOI.

While most of these metrics have been shown to distinguish between novices and experts, we focus on using the fixation-to-non-fixation ratio for a more temporally resolved analysis. We hypothesize that the ratio changes throughout surgery and can indicate whether a particular segment is more challenging. Furthermore, we hypothesize that gaze patterns can be more indicative than total completion time as completion time may be affected by each surgeon becoming familiar with the novel task environment.



Figure 3.2: Virtual plane definition in the Hololens. The virtual plane is not rendered during the experiment to prevent visual interference.

### 3.5 Build a more realistic kidney phantom

To make our experimental setup more realistic, we create a new model of the human kidney. Fabian Adams et al. (17) constructed a human kidney model with a collecting system using 3D printing technology. However, their method required a 3D printer capable of using wax as a printing material. We propose an alternative method that eliminates the need for directly printing an internal mold made of wax. The inner model is obtained from multi-layer CT scan images through 3D reconstruction. Fig. 3.3 shows the whole workflow to build the phantom.

The inner model in Fig. 3.3a is then printed using Elastic 50A Resin (Formlabs Inc., Ohio, FLELCL01) material in a Formlabs3 3D printer (Formlabs Inc., Ohio, LASER PRODUCT, Form 3), which consume approximately 42.1 mL of Elastic 50A. The inner model made of Elastic 50A not only maintained its original shape but also had sufficient elasticity to facilitate demolding in subsequent steps. Next, we mix part A and part B of Ecoflex 00-20 silicone rubber (Smooth-On, Inc., Macungie, PA 18062, Ecoflex 00-20), submerge the inner model in it, an elastic platinum-catalyzed silicone, and leave it at room temperature for at least four hours to ensure complete solidification. Then, we cut the Ecoflex in half, with the inner model at the center, and remove the inner model to obtain a mold made of Ecoflex. The whole mold in Fig. 3.3b requires approximately 0.9 kg of Ecoflex. We spray a layer of acrylic lacquer evenly on the mold to prevent adhesion between the wax and the Ecoflex. The wax is then heated to 70°C and poured into the Ecoflex mold through a reserved hole. The heating temperature of the wax could not be too high; otherwise, the wax would shrink intensely during cooling, resulting in an inaccurate inner model. The wax is cooled at room temperature for at least an hour to ensure complete solidification. We carefully separate the Ecoflex mold to remove the wax without causing it to fracture. The inner mold is thus complete.

We use an open-source external model<sup>2</sup> and modify the internal structure of the ureter to fit our inner mold. We divide the external model into the left (Fig. 3.3c) and right parts (Fig. 3.3d), with the left part printed as a whole using black ABS material in a Stratasys 3D printer (Stratasys Ltd., Rehovot, Israel, Stratasys F170 3D printer) for about 30 hours, consuming approximately 52 cubic inches of ABS material. The right part is printed using a Formlabs 3D printer and divided into four small parts to fit the maximum printing size supported by Formlabs. The right part consumes approximately 370 mL of clear resin in a total of 34 hours. Finally, we assemble the seven parts of the left and right models to form the complete external model. We fix the inner model in the external model, mix Ecoflex 00-20, drop in about 1 mL of pink food coloring to simulate the color of a real kidney, degassed it for 10 to 15 minutes, and pour it into the assembled mold. The mixture is then left to solidify at room temperature for at least four hours. After the Ecoflex has completely solidified, we remove the mold and soak it in an ethanol solution at 70°C to dissolve the wax material inside,

<sup>2</sup><https://www.cgtrader.com/3d-models/science/medical/kidney-cross-section>

and finally rinse the inside and outside of the model thoroughly with water.

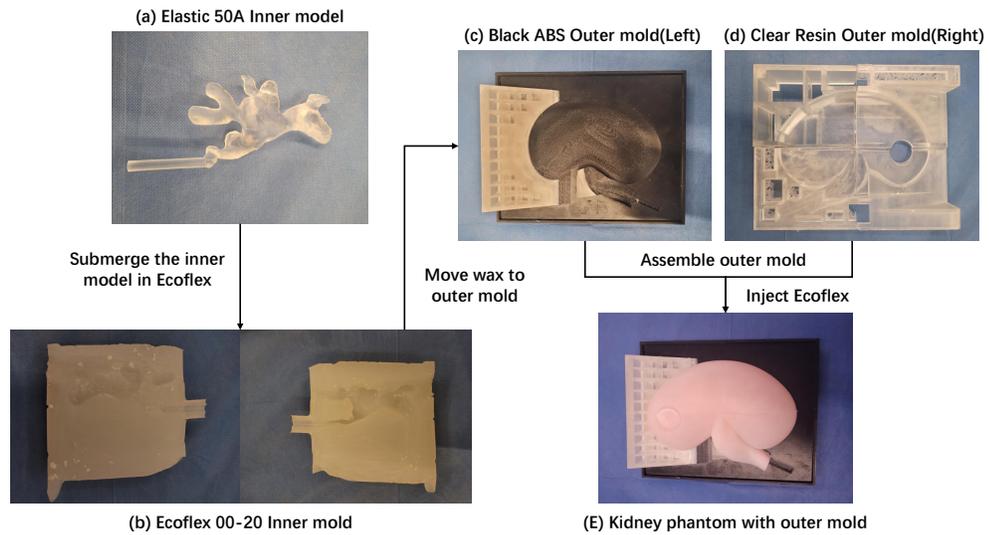


Figure 3.3: Workflow of building new kidney phantom. (a) The inner model is made by Elastic 50A printed from Formlab3 3D printer. (b) The left and right sides of the inner mold have the same shape inside as the inner model. (c) Left outer molds printed by Stratasys 3D printer. (d) Right outer molds printed by Formlab 3D printer. (e) Kidney phantom with the outer molds.

## CHAPTER 4

### Experiment

To test our hypothesis that we can use eye gaze to determine the skill level of surgeons in kidney stone procedures and track how difficulty changes throughout a task, we set up a user study to simulate the kidney stone identification task. The experiments are based on real kidney stone surgery. Two authors of the paper are surgeons and designed the experiment to mimic their clinical setup. The difference is that we save the process of the doctor entering the endoscope from the urinary tract into the body.



(a) Three kidney phantoms with needles inserted to simulate kidney stones. The needle tips are colored to prevent identification of the same fiducial.



(b) A surgeon is using the endoscope to identify the stone fiducials in kidney phantoms while wearing the HoloLens 2 to track her eye gaze

Figure 4.1: User study to identify stones in kidney phantoms.

We perform eye tracking of 10 surgeons performing fURS in kidney phantoms. The experimental setup is shown in Fig. 4.1a. The size of the surgical monitor is 75 cm by 40 cm and the size of the endoscope image is 35 cm by 35 cm. We used validated phantoms made from patient-specific data and obtained from Simagine Health<sup>1</sup>. Specifically, the renal collecting system is segmented from patient CT imaging and printed in a 3D, water-soluble resin. Models are made of silicone and made with traditional molding and casting techniques. A water-soluble resin is used and washed out to allow for a complex negative space within the silicone model, representing the renal collecting system. Surgeons are categorized as ‘expert’ (n=3, fellowship-trained endourologist, average case volume of >100 fURS per year) or novice (n=3, resident, <100 fURS per year). Each surgeon is tasked with fiducial localization in three separate kidney phantoms using a flexible ureteroscope. Three spinal needles are used as fiducials (representing stones) and placed percutaneously into upper, inter, and lower pole calyces in each phantom. Each needle is colored to indicate

<sup>1</sup><https://simaginehealth.com/index.html>

accurate localization by the surgeon and prevent reidentification of the same stone. Phantoms are silicone casts of 3D-printed collecting systems segmented from preoperative CT scans of three patients.

Each task involves entering the phantom with the endoscope, identifying the three needles, and exiting the phantom. Eye gaze is recorded using the Microsoft HoloLens 2 while each surgeon views the surgical monitor during fURS. We record the eye gaze at 30 fps and calculate metrics including total time for each task, distance traveled by eye gaze, number of gaze fixation points, gaze fixation dwell time and percentage, and area of eye gaze movement. We then compare these for experts and trainees.

## CHAPTER 5

### Result

#### 5.1 Qualitative Comparison of Gaze

We observe that experts and novices have distinct eye-gaze patterns. We plot the gaze data from one novice and one expert surgeon for one task in Fig. 5.1. The expert’s gaze is much more focused on the center of the surgical monitor than the novice’s. This aligns with previous studies’ findings that the expert has a preplanned search path they are following, whereas the novice performs a more random exploration (12).

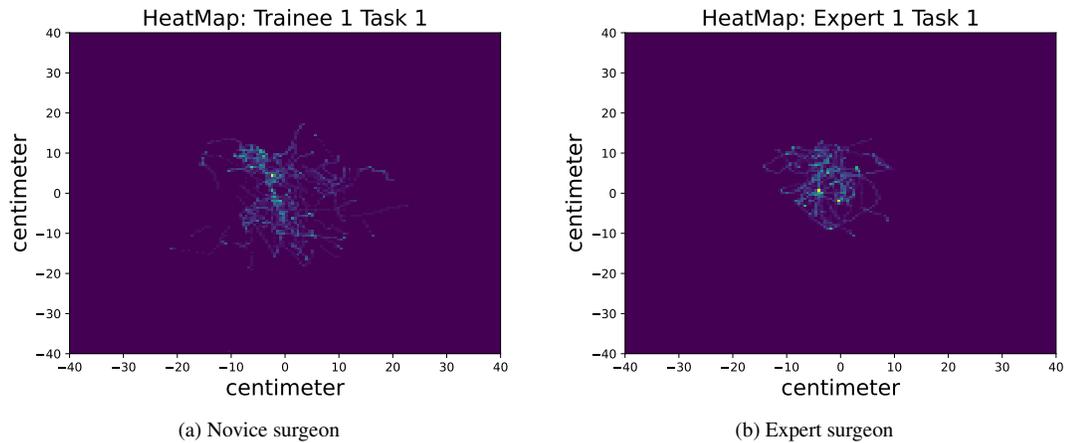


Figure 5.1: Heatmap showing the eye movements of a novice and an expert surgeon for one trial of phantom stone identification task.

We also overlay the gaze of another eye gaze of a novice and expert onto the surgical monitor output, shown in Fig. 5.2. The expert’s gaze is more concentrated at the center of the endoscope image. This is intuitively more efficient as moving the endoscope forward causes the new information to appear at the center of the image as opposed to the sides.

#### 5.2 Statistical Analysis

We report gaze metrics in Tab. 5.1. A two-sample t-test with two-tailed distribution was performed to compare gaze metrics in novices and experts. We find that experts had shorter task completion times as well as shorter gaze displacements than novices. The ratio of the total distance between the expert and the novice is smaller than the ratio of the total time. This supports our hypothesis that gaze may be a more differentiating factor in expertise than task completion time.

We find that a summary statistic of fixations is not a good metric between expert and novice. Experts had

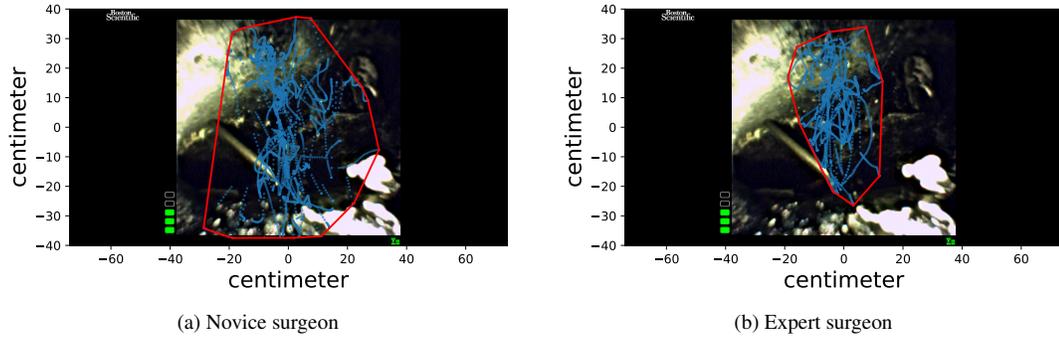


Figure 5.2: Eye-gaze path of a novice and an expert surgeon for one trial of phantom stone identification task overlaid with the surgical monitor output.

fewer fixations than novices, but the ratio between expert and novice is smaller for the number of fixations compared to the total time. The fewer fixations could simply have resulted from less time spent on the task. Furthermore, we find that the fixation to non-fixation ratio is the only nonsignificant metric for skill level among those we report.

Table 5.1: Features from the data

Features	Total mean (std)	Novice mean (std)	Expert mean (std)	$p$ -value
Total time (s)	84 (46)	318 (44)	52 (26)	** $p < .01$
Total distance (m)	14.88 (11.98)	19.92 (13.07)	7.32 (3.26)	** $p < .01$
Number of fixations	144 (85)	176 (85)	98 (65)	* $p < .05$
Fixation time (s)	33 (23)	44 (23)	18 (14)	** $p < .01$
Fixation to non-fixation	0.66 (0.4)	0.97 (0.5)	0.41 (0.28)	$p > 0.05$
Gaze outside AOI	1.7 (2.2)	2.4 (2.5)	0.67 (0.77)	* $p < .05$
Gaze area in AOI (cm <sup>2</sup> )	1283 (333)	1433 (236)	1056 (338)	** $p < .01$

While the fixation to non-fixation ratio was not significant as a summary metric, we also examined its utility in differentiating the difficulty of a task in the same surgeon over time. We separate each task into four equal-length segments and calculate the fixation to non-fixation ratio for each segment, shown in Fig. 5.3. We see that even for a static task of stone identification in a phantom, the ratio changes over the course of the task.

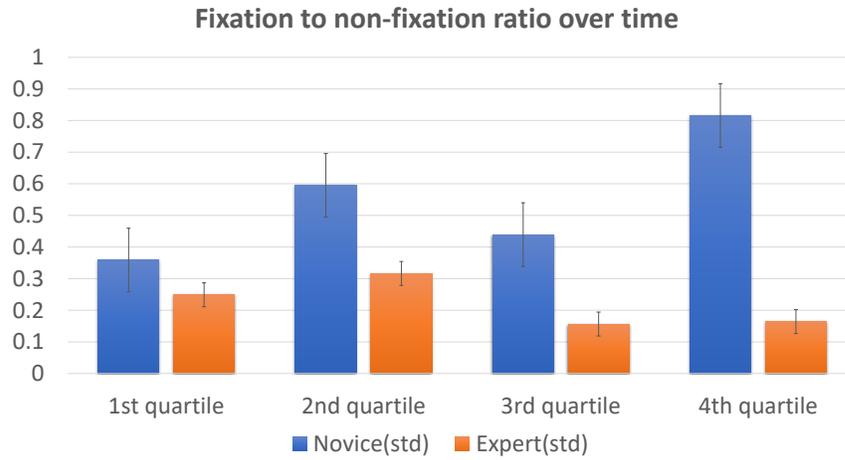


Figure 5.3: The ratio of fixation to non-fixation duration over time. Experts appear to learn to structure of the kidney fixate less over time whereas novices do not.

## CHAPTER 6

### Discussion and Conclusion

In this paper, we propose an eye-tracking method that accounts for the task space for kidney stone surgery. We implement a means of eye tracking during endoscopic surgery that is independent of head motion or surgeon movement. Furthermore, eye tracking was able to distinguish between experts and novices for multiple parameters, including gaze path, fixation number and time, gaze outside AOI, and gaze in AOI. Specifically, expert surgeons generally have a more focused gaze compared to novice surgeons. This suggests that experienced surgeons have more deliberate eye movements overall when exploring surgical anatomy, while novices have to visually examine more of the displayed anatomy to complete a task.

Currently, there are no objective, validated evaluation tools for endoscopic stone surgery. Additionally, kidney stone disease is very common and stone events often require endoscopic surgery. Though endoscopic surgery is a gold standard for the treatment of kidney stone disease, recurrence events are high due to incomplete treatment (2). This is partially due to variability in surgical competency among urologists (18). Thus, understanding the current learning curve and evaluating surgical competency is important for both patient safety and surgical outcomes. Though other groups have sought to evaluate this learning curve, eye tracking provides a non-invasive, automated, and objective assessment of surgical skill (19; 20). Our work uses this insight to provide temporally-resolved feedback that is based on the task space. As endoscopic kidney stone surgery has different levels of difficulty for each sub-task, we envision this feedback to be combined with the endoscope video recording to indicate which sub-tasks presented particularly difficult for a given surgeon for targeted training.

At present, this study has several limitations. First of all, our sample size is small, with only 4 experts and 6 novices. Each performed a phantom task, which does not take into account variations and dynamic changes present in real cases. Previous studies found that experienced surgeons would dedicate visual attention to checking on patient condition (8) but that gaze interaction is not captured in our study. Additionally, stone identification is not an interactive task. Treatment actions such as laser ablation could cause more adverse events, which have been shown to change eye-gaze patterns (10). Future work could consider monitoring the eye-gaze in real operations where these factors come into effect. Lastly, while we have shown that measuring eye-gaze metrics over time provides more information, our future work will focus on using this information for feedback during training.

By measuring which tasks presented the most difficulty for surgeons both during training tasks and in the operating room, we hope to build a platform to expedite training for endoscopic kidney stone procedures.

We envision tracking eye-gaze metrics with task data over different training sessions would allow us to plot granular learning curves for each sub-task and provide individualized feedback for novices.

## References

- [1] S. L. Conti, W. Brubaker, B. I. Chung, M. Sofer, R. S. Hsi, R. Shinghal, C. S. Elliott, T. Caruso, and J. T. Leppert, "Crowdsourced assessment of ureteroscopy with laser lithotripsy video feed does not correlate with trainee experience," *Journal of Endourology*, vol. 33, no. 1, pp. 42–49, 2019.
- [2] S. R. Khan, M. S. Pearle, W. G. Robertson, G. Gambaro, B. K. Canales, S. Doizi, O. Traxer, and H.-G. Tiselius, "Kidney stones," *Nature reviews Disease primers*, vol. 2, no. 1, pp. 1–23, 2016.
- [3] E. Chong, E. Clark-Whitney, A. Southerland, E. Stubbs, C. Miller, E. Ajodan, M. Silverman, C. Lord, A. Rozga, R. Jones, and J. Rehg, "Detection of eye contact with deep neural networks is as accurate as human experts," *Nature Communications*, vol. 11, 12 2020.
- [4] J. Klingner, R. Kumar, and P. Hanrahan, "Measuring the task-evoked pupillary response with a remote eye tracker," in *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ETRA '08, (New York, NY, USA), p. 69–72, Association for Computing Machinery, 2008.
- [5] B. Zheng, X. Jiang, R. Bednarik, and M. Atkins, "Action-related eye measures to assess surgical expertise," *BJS open*, vol. 5, 07 2021.
- [6] L. Richstone, M. J. Schwartz, C. Seideman, J. Cadeddu, S. Marshall, and L. R. Kavoussi, "Eye metrics as an objective assessment of surgical skill," *Annals of surgery*, vol. 252, no. 1, pp. 177–182, 2010.
- [7] B. Law, M. S. Atkins, A. E. Kirkpatrick, and A. J. Lomax, "Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment," in *Proceedings of the 2004 symposium on Eye tracking research & applications*, pp. 41–48, 2004.
- [8] G. Tien, B. Zheng, and M. S. Atkins, "Quantifying surgeons' vigilance during laparoscopic operations using eyegaze tracking," pp. 658–662, 2011.
- [9] R. S. Khan, G. Tien, M. S. Atkins, B. Zheng, O. N. Pantan, and A. T. Meneghetti, "Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation?," *Surgical endoscopy*, vol. 26, no. 12, pp. 3536–3540, 2012.
- [10] E. Fichtel, N. Lau, J. Park, S. Henrickson Parker, S. Ponnala, S. Fitzgibbons, and S. D. Safford, "Eye tracking in surgical education: gaze-based dynamic area of interest can discriminate adverse events and expertise," *Surgical endoscopy*, vol. 33, no. 7, pp. 2249–2256, 2019.
- [11] S. J. Vine, R. S. Masters, J. S. McGrath, E. Bright, and M. R. Wilson, "Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills," *Surgery*, vol. 152, no. 1, pp. 32–40, 2012.
- [12] S. Liu, R. Donaldson, A. Subramaniam, H. Palmer, C. D. Champion, M. L. Cox, and L. G. Appelbaum, "Developing expert gaze pattern in laparoscopic surgery requires more than behavioral training," *Journal of Eye Movement Research*, vol. 14, no. 2, 2021.
- [13] A. J. Berges, S. S. Vedula, A. Chara, G. D. Hager, M. Ishii, and A. Malpani, "Eye tracking and motion data predict endoscopic sinus surgery skill," *The Laryngoscope*, 2022.
- [14] S. Lu, Y. P. Sanchez Perdomo, X. Jiang, and B. Zheng, "Integrating eye-tracking to augmented reality system for surgical training," *Journal of Medical Systems*, vol. 44, no. 11, pp. 1–7, 2020.
- [15] A. M. Gil, S. Birdi, T. Kishibe, and T. P. Grantcharov, "Eye tracking use in surgical research: A systematic review," *Journal of Surgical Research*, vol. 279, pp. 774–787, 2022.
- [16] R. L. Graham, "An efficient algorithm for determining the convex hull of a finite planar set," *Info. Pro. Lett.*, vol. 1, pp. 132–133, 1972.

- [17] F. Adams, T. Qiu, A. Mark, B. Fritz, L. Kramer, D. Schlager, U. Wetterauer, A. Miernik, and P. Fischer, "Soft 3d-printed phantom of the human kidney with collecting system," *Annals of Biomedical Engineering*, vol. 45, pp. 963–972, 04 2017.
- [18] M. Goldenberg, M. Ordon, J. R. D. Honey, S. Andonian, and J. Y. Lee, "Objective assessment and standard setting for basic flexible ureterorenoscopy skills among urology trainees using simulation-based methods," *Journal of Endourology*, vol. 34, no. 4, pp. 495–501, 2020.
- [19] M.-T. Valovska, G. Gomez, R. Fineman, W. Woltmann, L. Stirling, and D. A. Wollin, "Analysis of flexible ureteroscopic motion and kinematic efficiency—a simulation-based pilot study," *Journal of Endourology*, no. ja, 2022.
- [20] C. S. Biyani, M. Kailavasan, N. Rukin, V. Palit, B. Somani, S. Jain, A. Myatt, G. Nabi, and J. Patterson, "Global assessment of urological endoscopic skills (gaues): development and validation of a novel assessment tool to evaluate endourological skills," *BJU international*, vol. 127, no. 6, pp. 665–675, 2021.