600.446: Advanced Computer Integrated Surgery

Final Project Report

Robo-ELF: Ergonomic Controller and Computer Vision Tools

Group 12

Members: *Jong Heun Kim, Tae Soo Kim, Steve Park*

Mentors: *Kevin Olds, Dr. Jeremy Richmon*

May 09, 2014

I. Background / Statement of Relevance

Traditional methods of laryngeal surgery involved disruption of the laryngeal framework which then led to motor and sensory deficits of the patients. Attempts to improve and possibly avoid such external invasive procedures led to advances in endoscopic laryngeal surgery that access the endolarynx through the mouth of the patient. Despite the enhancements to the endolaryngeal surgery triggered by improvements in microlaryngeal instrumentation and endoscopes, the current set up of endolaryngeal surgery continues to have disadvantages. The biggest shortcomings are the reduced depth perception from the monoscopic images, operator's distance from the surgical field and the crowded workspace for surgeons. [6]

To address these problems with the non-robotic endolaryngeal surgical procedures, robotic endolaryngeal flexible (Robo-ELF) scope surgical robot has been proposed. The current state of the robot enables the surgeon to manipulate the robot using a joystick controller attached to the surgical bed while a real-time endoscope images are displayed on a monitor in the operating room. An unmodified flexible laryngoscope is attached to the robot and the surgeon is able to manipulate the scope with only one hand using the joystick.

The Robo-ELF has undoubtedly improved the procedure of endolaryngeal surgery. However, there still exists shortcomings with the current design. Currently, in order to measure the level of subglottic stenosis, the otolaryngologist inserts increasing sizes of endotracheal tube into the stenotic segment to perform a leak test. Then the chosen endotracheal tube size is used to rank the severity of subglottic stenosis. Such an approach itself may alter the target that it is trying to measure and cylindrical endotracheal tubes are not completely apt for sizing stenotic segments.

Another noticeable drawback of the Robo-ELF is its user-unfriendly joystick controller. Initially, commercially available 3D space mouse is used to control the Robo-ELF because the mouse intuitively mimics three degrees of freedom of the scope: rotation, up/down, and advance/retract. However, the disadvantage of the 3D space mouse is that it’s very difficult to decouple user’s movements to reliably separate degrees of freedom. To overcome this shortcoming, a controller with two digital joystick replaced the mouse interface. Because each digital joystick controls different degrees of freedom, the isolation of degrees of freedom is very reliable. Even though the joystick has good isolation, the controller lacks intuitive interface as the two joysticks does not move in the same direction as the scope and gradation as the digital joystick has only on or off states. To manipulate Robo-ELF well, a new design of controller with intuitive interface movements and analog gradation control with haptic feedback is required.

To improve the current subglottic stenosis procedures and endoscopic surgery in general, our software will enable 3D distance calculations from the 2D endoscope images and produce a 3D reconstruction of the airway. To enhance the usability of the Robo-ELF, we will develop a more ergonomic and intuitive joystick controller.

II. Specific Aims

A. Controller

- Small and intuitive one-handed control interface
    - Overall design is compact (5’’x5’’x8’’) and requires only one hand to operate.
    - The motion of joystick is parallel to that of the scope system (intuitive interface).
- Self re-orientation and Haptic feedback
    - Gimbal system is implemented on each axis of rotation.
- Analog sensing instead of digital
    - Linear potentiometers are used to sense the degree of rotation.
- Redundant input for safety measures
    - A pair of potentiometer is mounted for each axis to ensure correct reading and abort the system if the readings do not match.

B. Computer Vision Tools

- System capable of providing quantitative endoscopic measurements.
  - Passive stereo vision principles are used to triangulate the depth from a pair of endoscope images.

III. Technical Approach

A. Controller

Our group took all considerations such as but not limited to ergonomics, isolation of motion, and size to come up with a new design for the controller. The main motivation for our controller design was to grant the surgeon the ability to operate intuitively enough to control without looking at the controller.

In order to achieve this, we implemented a velocity-control using linear potentiometers, haptic feedback and self-reorientation mechanism using the Gimbal system in addition to mimicking the motion of the robot relative to the display used to perform the endoscopy. Each degree of freedom as mentioned before and illustrated in Figure 1 has a different range and an isolated state of motion and with this knowledge, our controller design ensures precise control of navigation.

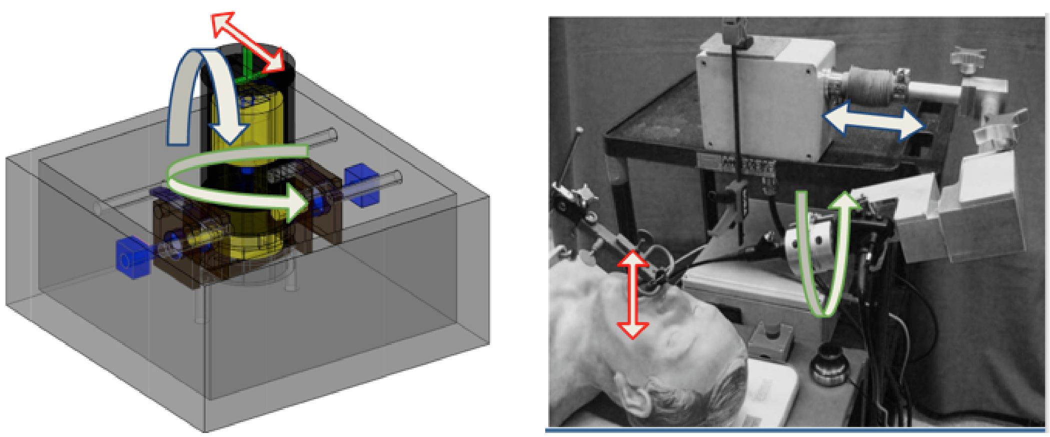

Figure 1. L) CAD design of the joystick. R) Image of the Robo-ELF with the flexible endoscope attached. The colored arrows represent each axis of motion. Red represents the up-down motion of the scope-tip. Green represents the rotation. Blue represents the front-back translation.

Fore-mostly, the main body parts consisting of the cylindrical handle, the inner shaft body for holding electrical parts, and the lever for scope-tip manipulation were all produced by

3D printing technology from the Department of Mechanical Engineering at Johns Hopkins University. The amount of time searching for optimal shape, size, and material of the main body parts was greatly reduced due to 3D printing and such saved time was directed to implementing more complex features including haptic feedback and self-reorientation mechanism. Many of other parts such as potentiometers, shaft coupling devices, sheet metal, and the encasing container for the controller were purchased and funded by the Department of Otolaryngology at Johns Hopkins University School of Medicine.

Linear potentiometers were chosen to mimic the motion of front-back translation and rotation as well as to precisely manipulate the different degrees of scope velocity by reading how much the potentiometers are rotated. Each axis of motion has two potentiometers for redundant sensing for the purpose of safety measures. Stock joystick potentiometers at which the 3D printed tip-lever is attached were used to emulate the up-down movement of the scope tip; redundant sensing measures were also implemented in this axis. Haptic feedback and self-reorienting mechanism were applied to each axis of motion as well. Using our knowledge of the Gimbal system, which utilizes retractive bending moment created against the user's force input, layers of sheet metal were applied in each axis to achieve this effect in the controller. All of the sensor signals are processed through an Arduino Uno which is responsible for data comparison within the controller and transmission to the computer. Finally, all parts are compactly encased in a stock container.

B. Computer Vision Tools

B.1 Passive Stereo Vision Principles

The main goal of the vision component of the Robo-ELF project was to yield a software that enables the user to measure 3D distance from 2D points on an endoscopic image. The software to produce such measurements require the following information: the focal length of the camera, the stereo baseline measurements and the disparity between the point of interest in stereo images. The block diagram below represents the means of obtaining relevant data and their relation to one another.

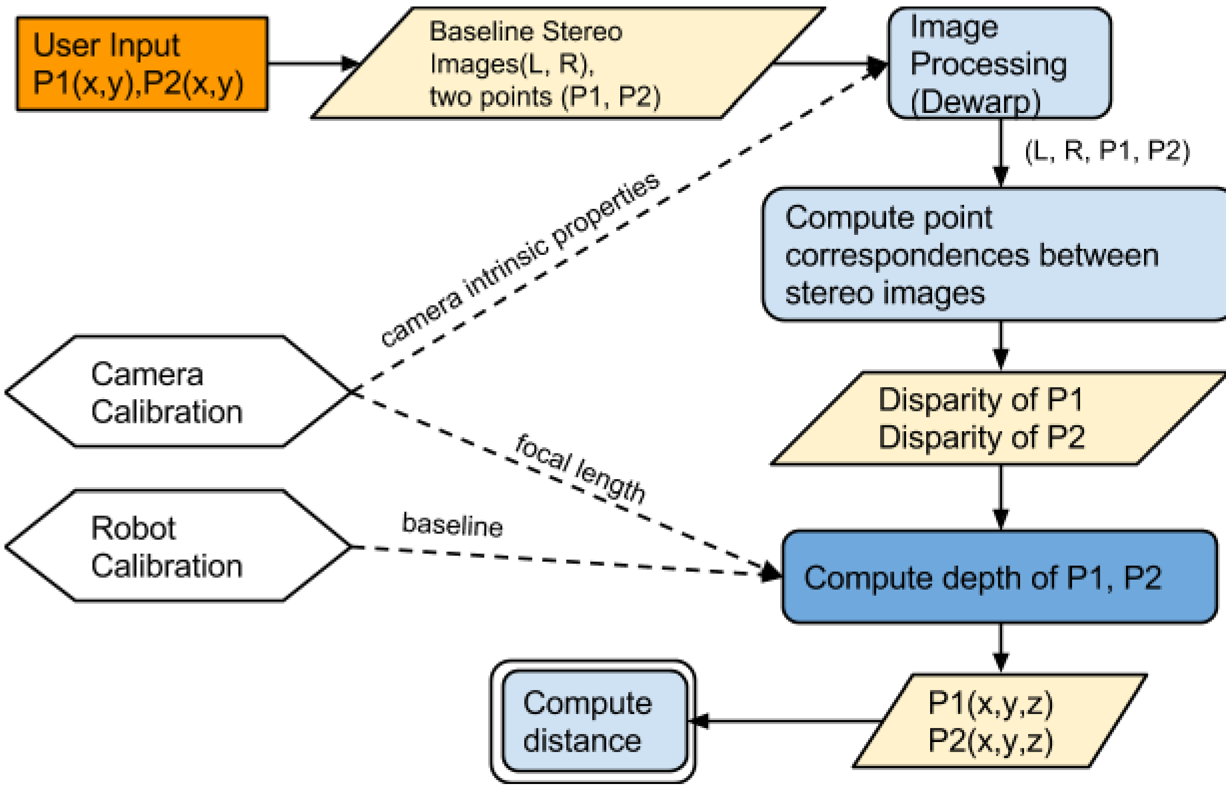

Figure 2. Flow diagram of stereo baseline 3D distance calculation from 2D image points.

The process specified above completes the proposed minimum computer vision requirement of the project which was to provide a software that measures distances of a known artificial scene. More specifically, given a point P1 in one of the stereo images (without loss of generality, assume the left image, L), the software must find the correct corresponding point P1' in the right image R. There are multiple suitable methods for identifying correspondences between stereo images. A leading idea is to use the SIFT (scale-invariant feature transform) feature descriptor for a specified window size around P1 in the first image and to find the corresponding point in the second image using the SIFT descriptor of P1. The correspondence solution will be found by picking a matching point P1' in the right image R that minimizes the cost function (in this case, the sum of squared differences between the SIFT descriptor values). Same procedure will be done for P2 as well. The current state of the software has an option to let the algorithm find the corresponding point using normalized cross correlation between the intensity of the neighboring pixels or let the user manually click on the corresponding point on the second image. Although the process is not automatic, the second choice simplifies the implementation complexity and is chosen to be the default setting for the program.

With corresponding point pair (P1, P1'), the algorithm can calculate the disparity between the point pair. Disparity is the measure of the pixel distance between the corresponding points on the image. Having found all necessary parameters (focal length, *f*, baseline, *B*, disparity, *d*), the algorithm can not only find the depth, *Z*, of a given point P1 in the real world but also find its exact 3D coordinates. The calculation involves simple algebra that involves similar triangles as shown below.

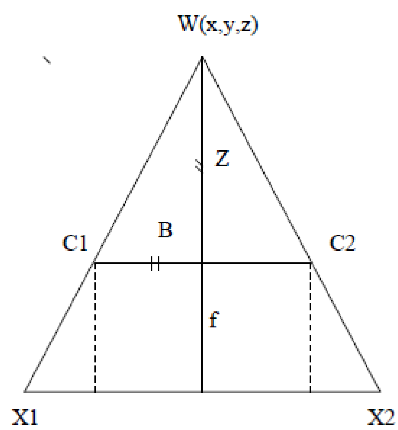

Figure 6.1: Stereo geometry.

Figure 3. Stereo baseline depth calculation problem set up.

In more detail, the depth can be calculated as follows:

$$Z = \frac{f * B}{x_1 + x_2}$$

Note that the disparity, *d,* is the sum of the pixel coordinates x_1 and x_2 in the respective images. The depth equation is inversely proportional to the magnitude of the disparity. In other words, the more the object moves in the stereo image pair, the closer it must be in the real environment. Since the disparity is found by the computer vision software as mentioned above and both the focal length and baseline are known from calibration, the depth value can be calculated accurately.

It is important to note that in order for the stereo vision algorithm to work correctly, the endoscope images must be first unwarped and rectified. Due to endoscopic camera characteristics, images are severely warped and the disparity values obtained from pre-processed images are meaningless. Images can be unwarped using camera calibration parameters obtained from the camera calibration software.

B.2 Experimental Setup

The procedure to acquire a 3D distance measurement from a 2D endoscope image uses Robo-ELF to manipulate the tip of the flexible endoscope. One of the biggest benefits of the software is that the normal operating set up of Robo-ELF need not be changed to measure distances. Under normal operating conditions, the graphical user interface of Robo-ELF now has an additional button to initiate the 3D measurement process. Upon pressing the button, the endoscopic image is captured at the current endoscope pose. Then, a signal is sent to one of the Robo-ELF motors that is responsible for manipulating the scope tip movement. The signal turns the motor exactly a fixed amount of encoder counts so that the scope tip movement is as consistent as possible over multiple trials. After the scope tip has moved, a second image is saved to complete the stereo image pair. Then, the user is prompted with the first endoscopic image so that the user can input two points on the image by simply clicking on the image. The two images are fed into the algorithm to compute the 3D distance between the given two points on the image. After the algorithm finishes running, the computed 3D distance value is displayed on the screen.

IV. Results and Discussion

A. Controller

Integrating and assembling all the parts into the final form only took several hours. However, this includes time making small modifications to some of the parts such as the 3D printed main body parts and mechanical parts like the shaft coupling set screws that were used to connect the potentiometers. Those minor modifications include abrading tight spaces between body parts using a filer to make more room for smoother operation and adjustments of the hole sizes on shaft coupling to fit the screws we were using and our intended design. Further, we had to create threaded profile on the inner surface drilled holes, so that a screw can be mounted to ensure firm connection between the potentiometers and the main body parts. In this regard, we suspect that time could have been greatly saved if the initial CAD design had reflected minuscule detail we implemented such as smoothing out touching surfaces for facilitated integration. Also, if we had been more experimental with mechanical parts by testing out and choosing the optimal material, the integration process can be faster and more precise.

B. Computer Vision Tools

To meet the minimum deliverable of the project, the software's capability to acquire measurements from a known artificial scene was tested. The chosen scene was a checker-board pattern as it is easy to accurately measure the ground truth value. The length and the width of each square of the checker-board pattern was measured with a straight ruler as a ground truth measurement.

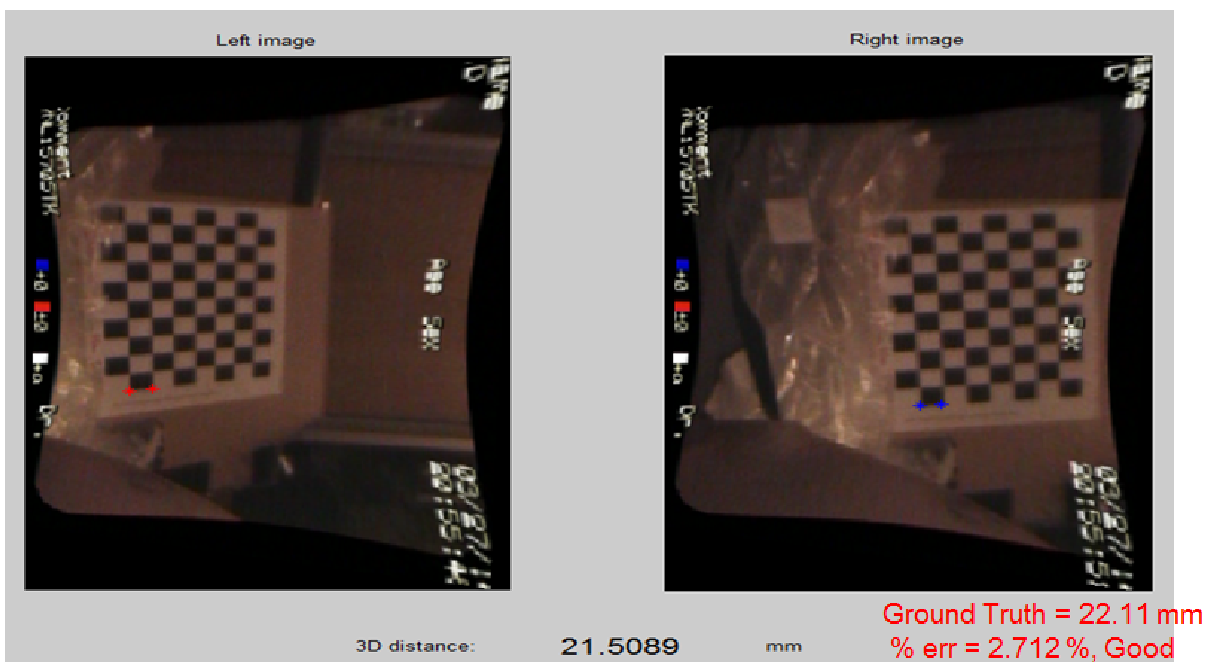

Figure 4. Measurement of the 3D distance of an artificial scene.

The figure above is a screenshot of what the user sees as a result of running the 3D measurement software. The user is prompted to input two points by clicking on the left picture. The physical distance in the 3D world of the two chosen points will be calculated by the program. The blue points on the right image represent the computed corresponding points of the user input points of the left image (in red in the left image). At the bottom of the screen is the final computed distance between the points. The ground truth measurement of the square in the checkerboard pattern was 22.11 mm and the computed distance of the algorithm was 21.5089 mm. The absolute error in the measurement was about 0.6 mm. The performance of the software for measuring the 3D distance of an artificial scene with the endoscope resting in free space yielded great results with error rates under 3%. The results were consistent because the amount of scope tip movement, the baseline, was reproducible over multiple trials. However, when the same algorithm and procedures were applied to a scene inside the throat, the 3D distance measurements were not only terrible but also inconsistent over trials.

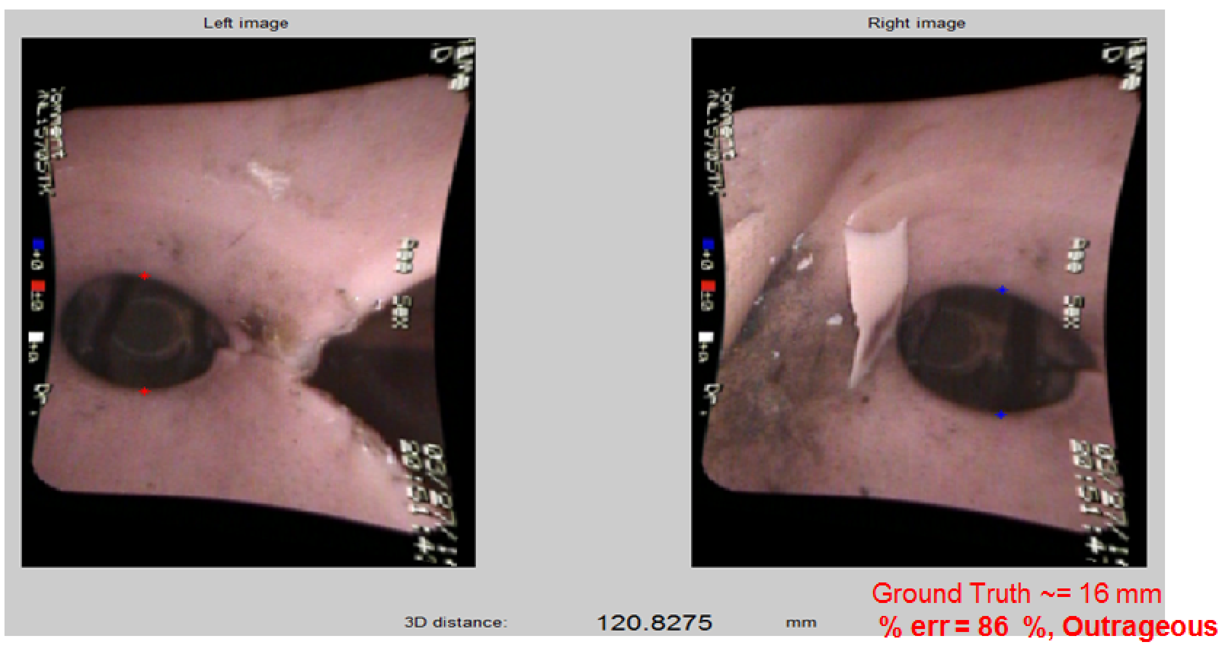

Figure 5. Measurement of the 3D distance inside the throat with an erroneous result.

As seen in the figure above, the software produced a 3D distance measurement of 120.8275 mm. The ground truth measurement of the diameter of the airway opening in the figure above was 16 mm. The outcome had a 86% error. The source of such huge error was the change in the scope tip movement which altered the baseline magnitude. In other words, compared to the scope tip movement in free space, the tip's movements were obstructed due to the tight configuration of the scope inside the throat. The flexible endoscope was unable to move as freely as it did for the artificial scene tests. Consequently, the change in baseline threw off the triangulation calculation and the depth values were erroneous which led to inaccurate 3D distance measurements.

To resolve the issue, the scope tip movement was recalibrated inside the throat. The recalibration procedure in fact showed that the scope tip moved much less inside the throat than it did in free space. With the new baseline values, the triangulated depth values and the resulting 3D distance values seemed reasonable. However, after many trials, the 3D distance measurements of an identical scene produced inconsistent results.

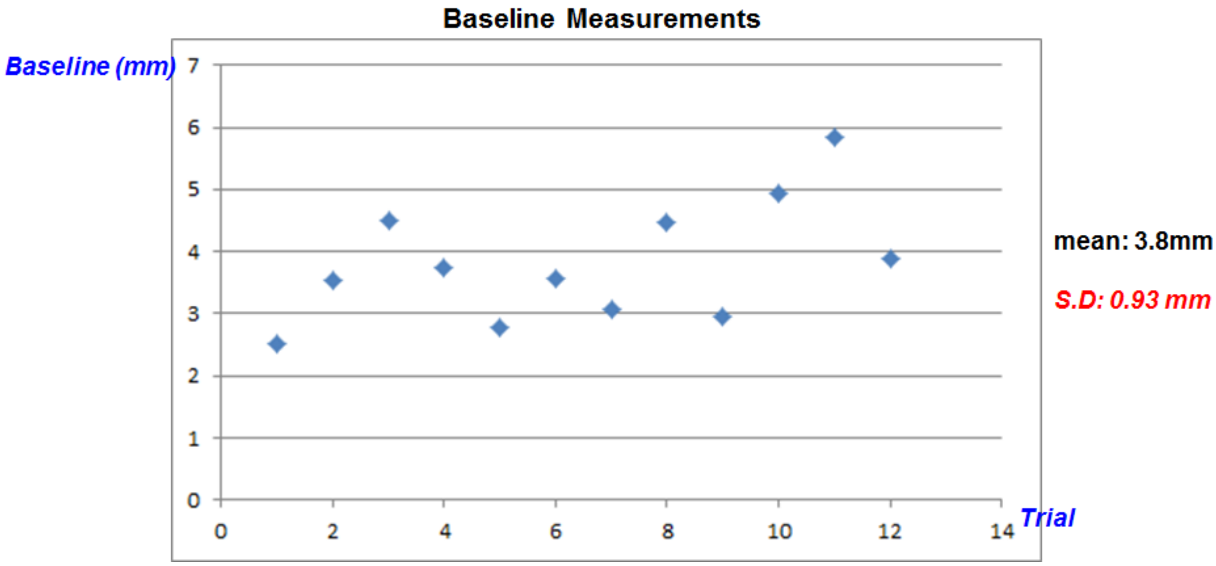

Figure 6. Inconsistent baseline measurements inside the throat.

As seen in the baseline measurement chart, the magnitude of the baselines differ from trial to trial with a standard deviation of 0.93 mm. Given the fact that the mean is 3.8 mm, the noise in the baseline readings was significant. The large discrepancy between trials was due to the fact that the configuration of the flexible endoscope was unpredictable inside the throat. The magnitude of the baseline depended heavily upon the configuration of the flexible endoscope inside the throat.

Consequently, the team has concluded that acquiring accurate and reliable 3D distance measurements relying solely on the scope tip movement was not clinically acceptable. To remove the uncertainty involved with the scope tip movement, the team chose to implement a novel approach to acquire 3D measurements using a system of endoscope and a laser. The new system uses active stereo vision principles to compute the depth of an object captured in an endoscope image.

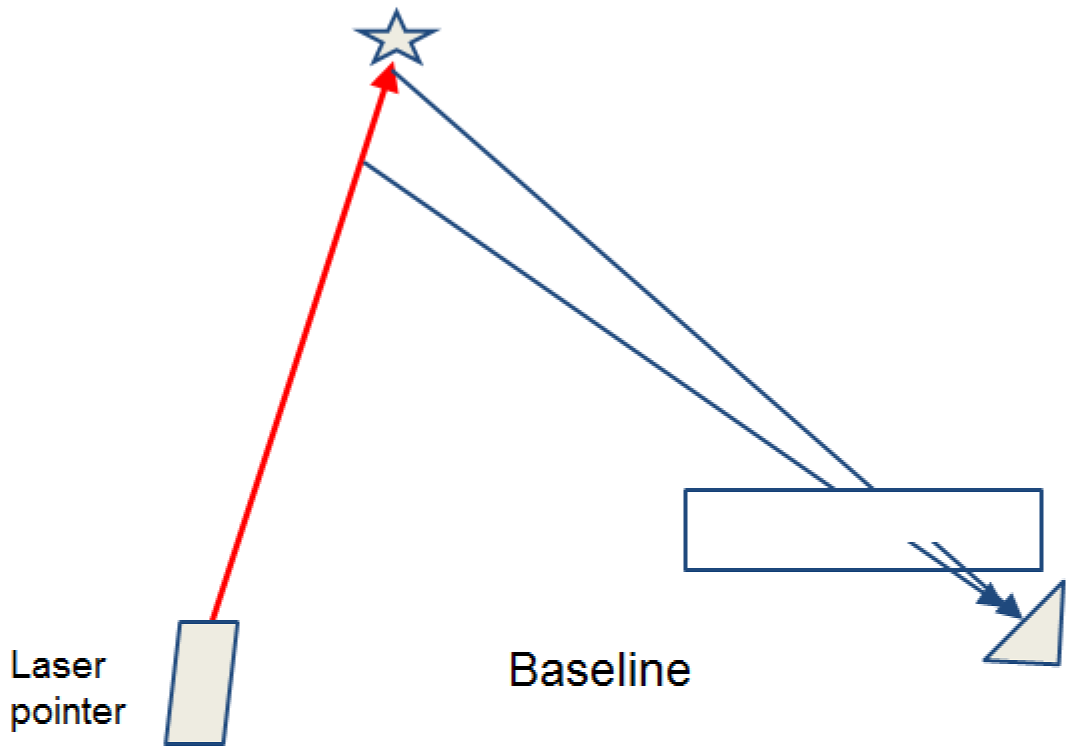

Figure 7. Active stereo vision geometry.

As shown in the figure above, active stereo triangulates the depth of an object by computing the intersection of a the ray that originates from the laser spot on the object (blue ray in the figure) and a laser beam that originates from the laser source (red ray in the figure). Since the laser source and the camera are stationary relative to each other, the baseline must be constant. If the laser beam is calibrated so that the exact location of the beam is known relative to the camera position, using a calibrated camera, an algorithm can triangulate the distance to the object. The team has started to implement the proposed endoscope+laser method. Further implementation of the software and the full 3D reconstruction algorithm will be implemented as a part of Tae Soo Kim's master's thesis project for the 2014-2015 school year if accepted by the review committee. The work on 3D reconstruction and distance measurements with an endoscope and a laser will resume on July of 2014. Below are some preliminary data showing a calibrated laser beam and a triangulated object point in real world coordinates.

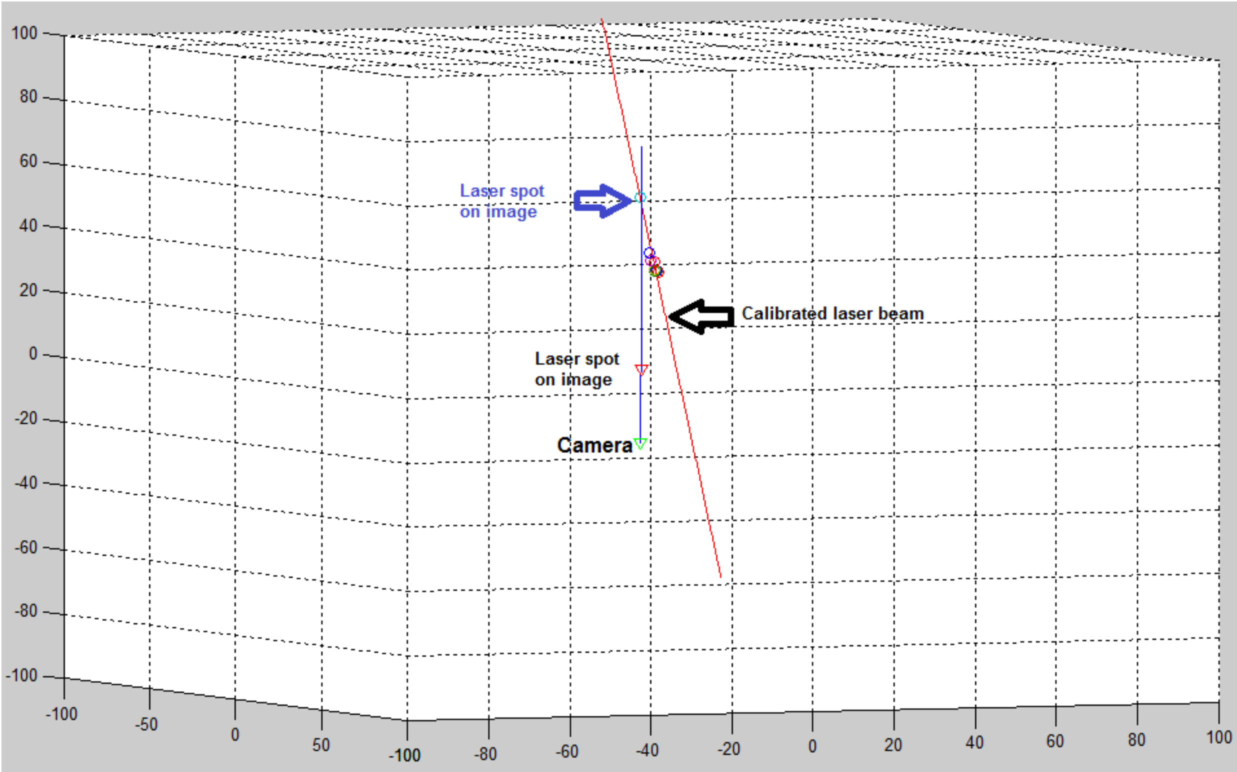

Figure 8. Preliminary results of the novel active stereo vision approach to calculate 3D distance of an object.

The green triangle above represents the camera position and the red triangle is the laser spot on the image plane of the camera. The triangulated position of the object is highlighted by the blue arrow in the above figure. The current state of the software requires a more precise calibration for the laser beam to generate more accurate distance measurements.

V. Deliverables

A full design of the new ergonomic controller is delivered. The design process has seen delays as the design of the controller had to be revised multiple times to produce the best possible result. Many optimizations were made to improve the ergonomics and the size of the controller. Another source of delay was the unforeseen delays in acquiring necessary parts for the controller. Many parts were ordered online and the delivery time of the parts were unpredictable. 3D printing of parts took longer than expected as the 3D printer had other tasks to perform as well. Despite the delays in the design and the manufacturing process, the team was able to successfully assemble a complete ergonomic controller.

A software to acquire 3D distances from a monoscopic endoscope image has been implemented. The original approach for measurement using passive stereo vision with scope tip baseline has produced results that were not reliable enough as mentioned in section IV above. As a result of inconsistent results, the team decided to implement a novel active stereo approach using a laser and an endoscope to acquire quantitative measurement. The complete software to acquire 3D measurements from a laser-endoscope system is still under development.

Due to unexpected complications in moving forward with human subject study of Robo-ELF, the clinical trials have been halted since the early stages of the project.

Delivered:

- A fully designed ergonomic controller for the Robo-ELF Scope manipulation.
- Documentation for the Robo-ELF Scope controller.
- Fully assembled ergonomic controller for the Robo-ELF Scope manipulation.
- Software to get real measurements from 2D scope images of a known artificial scene.
- Software to get real measurements from 2D scope images of the larynx. (Upgrade in progress with a novel approach)
- Documentation of the vision software.

To Be Delivered:

- Obtain more reliable and accurate 3D measurements by moving to active stereo approach using an endoscope+laser system. *(Already started)*
- Generate a dense and complete 3D reconstruction of the larynx using a novel endoscope+laser system. (*To be completed as a master's thesis project of Tae Soo Kim)*

VI. Management Summary

The project can be segmented into largely three components: human subjects study, creating an ergonomic controller and implementing 3D quantitative endoscope software. Steve Park took main responsibility in manufacturing and assembling the ergonomic controller. Jong Heun Kim was originally in charge of the human subjects study but as the team was unable to proceed with clinical trials, utilizing his design capabilities, Jong Heun Kim joined Steve Park in the design and fabrication of the controller. Tae Soo Kim focused on implementing the computer vision tools.

Kevin Olds mentored the design process of the controller and provided guidance on the principles behind the computer vision software.

VII. Future Work

The list below identifies future directions of the project.

- Gather feedback from surgeon and analyze pros and cons of the controller design.
- Comparison study between the current Robo-ELF joystick and the new ergonomic controller.
- If possible, help Kevin Olds and Dr.Jeremy Richmon with human subject clinical trial of the robot.
- Obtain more reliable and accurate 3D measurements by moving to active stereo approach using an endoscope+laser system. *(Already started)*
- Using a patterning lens or a diffraction grating with an endoscopic camera, generate a dense and complete 3D reconstruction of the larynx using a novel endoscope+laser system. (*To be completed as a master's thesis project of Tae Soo Kim)*

VIII. Lessons Learned

- Plan ahead for unexpected delays when ordering and printing parts.
- Better plan for unresolved/unforeseen dependencies.
- Endoscope images are severely warped.
- Accurate calibration is crucial for softwares that utilize cameras or optic principles.
- Frequent meetings with the mentor helps the team to stay on track and keep motivated.

IX. References

1. Darius Burschka, Ming Li, Russell Taylor, and Gregory D. Hager. Scale-Invariant Registration of Monocular Stereo Images to 3D Surface Models. In *Proceedings of IROS*, pages 2581-2586, 2004.
2. Darius Burschka and Gregory D. Hager. V-GPS – Image-Based Control for 3D Guidance Systems. In *Proc. of IROS, pages 1789*–1795, October 2003.
3. Darius Burschka and Gregory D. Hager. V-GPS(SLAM): – Vision-Based Inertial System for Mobile Robots. In *Proc. of ICRA, April* 2004.
4. Olof Enqvist, Fredrik Kahl, Carl Olsson. 2011. Non-Sequential Structure from Motion.

Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference. p. 264-271.

5. Daniel Mirota, Russell H. Taylor, Masaru Ishii, and Gregory D. Hager. Direct endoscopic video registration for sinus surgery. In *Medical Imaging 2009: Visualization, Image-guided Procedures and Modeling. Proceedings of the SPIE*, volume 7261, pages 72612K-1 - 72612K-8, February 2009.
6. H. Kawasaki, R. Furukawa, R. Sagawa, Y. Yagi. *Dynamic Scene Shape Reconstruction Using a Single Structured Light Pattern*. *Computer Vision and Pattern Recognition, 2008. CVPR 2008.* IEEE Conference on 1-8.
7. H. Kawasaki, R. Furukawa, R. Sagawa, Y. Yagi. *Shape from Grid Pattern Based on Coplanarity Constraints for One-shot Scanning*. *IPSJ Transactions on Computer Vision and Applications 2009.* Vol. 1, 139-157
8. Kevin Olds, Alexander T. Hillel, Elizabeth Cha, Martin Curry, Lee M. Akst, Russell H. Taylor, Jeremy D. Richmon. Robotic Endolaryngeal Flexible (Robo-ELF) Scope: A Preclinical Feasibility Study. *The Laryngoscope*; 2011; 121; 2371- 2374
9. Toan D. Pham, Rucker Ashmore, Lisa Rotelli. 2011. Proportional joystick with integral switch. U.S. Patent 7931101, filed October 13, 2006, and issued April 26, 2011.
10. Hanzi Wang, Daniel Mirota, and Gregory D. Hager. A generalized kernel consensus based robust estimator. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):178-184, 2010.
11. William M. Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima, Ron Kikinis. 1996. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*. v. 1. p 35–51.