

Computer Vision for navigation in robot-assisted rectal resection

Author
F.A. van der Kaaij, Bsc.

Supervisors
Dr. A.R. Wijsmuller
Prof. Dr. S. Misra
Dr. J. Sikorski

Department of surgical oncology, Universitair Medisch Centrum Groningen, the Netherlands

Clinical internship Technical Medicine

Master Medical Imaging & Interventions

University of Twente

M2-2 September 2021

Abstract

Background In surgical removal of locally advanced and recurrent rectal carcinomas, the resection margin rates remained tumor-positive in 15% of the patients with locally advanced rectal cancer and in even 50% of the patients with recurrent rectal cancer. Image-guided navigation techniques may improve safety and radicality of these robot-assisted surgical procedures. Two problems that arise here are the high mobility of the robotic arms and instruments and the sterile covers. A possible solution for these problems is the use of computer vision techniques in combination with stereotactic navigation. Computer vision techniques can be applied on the endoscopic stereo images to calculate the 3D position of the instruments relative to the stereo endoscope.

Goal Accuracy assessment of the use of computer stereo vision to calculate the 3D coordinates of manually selected markers in stereo images

Methods and materials After calibration, three experiments have been performed with different markers. The markers are manually selected in both the stereo images to obtain the pixel coordinates of the markers. The x-, y- and z-coordinate are calculated by the *triangulate* function. The accuracy of these calculations is evaluated by calculating the error between the expected coordinates and the calculated coordinates. The 3D coordinates are reprojected back to pixel coordinates with the function *worldToImage*. A sensitivity analysis is performed by varying the pixel coordinates and calculate the associated 3D coordinates with the *triangulate* function.

Results The results show an increasing overestimation of the calculation in every dimension, especially the z-dimension. The accuracy and precision in z-dimension decrease when the marker is further away from the camera. The reprojection errors in the left image are larger than in the right images. From the sensitivity analysis, the overall difference in z-coordinate is 19.87 mm when varying the pixel coordinates in a range of 10 pixels.

Conclusion The application of computer vision in robot-assisted laparoscopic surgery is accurate at 50 mm from the stereo endoscope. Nevertheless, the accuracy decreases with the distance from the camera. This can be explained by lens distortion and/or lack of rectification. Further research, such as standardization of manual clicking, image rectification and automatic feature detection, needs to be done to create a reliable navigation system based on computer stereo vision.

Keywords Computer stereo vision, depth calculation, robot-assisted colorectal surgery, image-guided navigation

I. Introduction

Globally, colorectal cancer covers 10.0% of the new cancer cases and is responsible for 9.4% of the cancer deaths in 2020. Overall, colorectal cancer ranks third in terms of incidence, but second in terms of mortality. When separating this group of patients by location of the carcinoma, rectal carcinomas cover 3.8% of the new cancer cases and they are responsible for 3.4% of the cancer deaths globally in 2020.¹ In the Netherlands, population screening is used to prevent or detect and treat colorectal carcinomas in an early stage. Early detection gives a greater chance of cure. Due to this screening, around 85% of the colorectal carcinomas are detected. Last year, 3300 patients in the Netherlands were diagnosed with a rectal carcinoma.²

The treatment options for rectal carcinomas depend on the stage of the disease. In general, 95% of the patients with colorectal cancer stage I to III in 2019 were treated surgically for curative treatment. Within this group, rectal carcinomas stage II and III were treated surgically in 76% of the patients.³ Nevertheless, surgical resection can be challenging, especially in patients with locally advanced primary and recurrent rectal cancer. The major challenge in surgical resection is the achievement of radical removal of the tumor, which implicates clear resection margins. But despite refinement of surgical techniques and improved neoadjuvant treatment, the resection margin rates remained tumor-positive in 15% of the patients with locally advanced rectal cancer and in even 50% of the patients with recurrent rectal cancer.⁴ These positive margin rates are associated with high local recurrence and low survival rates.⁵

Laparoscopic-assisted colorectal surgery (LACS) is preferred for surgical resection of locally advanced primary and recurrent rectal tumors compared to open surgery. This minimally invasive approach has several benefits for the patient, such as fast recovery of normal dietary intake and bowel function, reduced postoperative pain, and shorter hospitalization.⁶ However, laparoscopic resection of rectal tumors is more difficult and requires a prolonged learning curve to achieve equivalent outcomes compared to open surgery. To solve this problem, these laparoscopic resections are often assisted by the da Vinci[®] surgical robot (Intuitive Surgical, Sunnyvale, California). Robot-assisted colorectal surgery (RACS) has several advantages, such as better vision with a high definition three-dimensional view and enhanced accuracy of movement due to lack of human tremor and Endowrist technology with exceptional dexterity.⁶ Compared with LACS, RACS results in less blood loss, a lower bleeding rate and ileus rate and less chance of complications and mortality.⁷

In addition to RACS, image-guided navigation may improve the safety and effectiveness of surgical procedures for locally advanced and recurrent rectal cancer. In case of less mobile organs, such as the rectum attached to the bony pelvis, preoperative imaging with MRI or CT can be used for surgical navigation. Navigation provides real-time intra-operative feedback on the instruments' pose with respect to the intraoperative anatomy, which can be matched with the preoperative imaging.⁴ In robotic instruments, the robot joint encoder data can be used for navigation of the instruments. Nevertheless, the da Vinci[®] surgical robot is a cable driven system with multiple joints. Due to the cable driven system, errors can accumulate over time, which makes it hard to accurately determine the position of the instruments. Accurate position calculation also requires hand-eye calibration between the camera and the robot coordinates of the instruments, which is time-consuming and very difficult.⁸

At the University Medical Centre Groningen (UMCG), a possible application for surgical navigation is proposed by means of stereotactic navigation with the Brainlab[®] navigation system. This system uses a stereotactic camera which can detect and track infrared (IR) reflecting marker spheres attached to a surgical tool and the intraoperative anatomy. The position of the surgical tool with respect to the intraoperative anatomy can be calculated after calibration. This navigation system is often used in neurosurgery but was first applied during minimal-invasive rectal surgery in 2015. Nevertheless, two problems arise in the use of stereotactic navigation in combination with the da Vinci[®] Surgical robot. The first problem is that the IR reflecting markers should always be in sight of the camera. But due to

continuous movement of the robotic arms and the sterile covers around these arms, the IR markers are not always visible for the camera. The second problem is the mechanism of multiple joints and the Endowrist technology, which cause the instruments to move relative to the IR marker on the robotic arm. Therefore it is difficult to determine the location of the tip of a surgical instrument based on the IR marker.⁹

A possible solution for these problems is the use of computer vision techniques applied on the endoscopic stereo images to calculate the 3D position of the instruments relative to the stereo endoscope. With use of the Brainlab[®] navigation system and IR markers on the da Vinci[®] stereo endoscope and the anatomy of the patient, the position of the stereo endoscope can be determined relative to the intraoperative anatomy. A schematic representation of this set up is shown in Figure 1. By combining these two coordinate systems, the position of the instruments based on the endoscopic stereo images can be matched to the intraoperative anatomy. This could enable image-guided navigation based on 3D MRI topography with the Brainlab[®] navigation system during laparoscopic rectum resections assisted by the da Vinci[®] surgical robot. This may lead to improvement in radicality of rectum resection because of better recognition of anatomical landmarks, dissection planes and margins towards the tumor. A visual representation of the proposed design is shown in Appendix A.

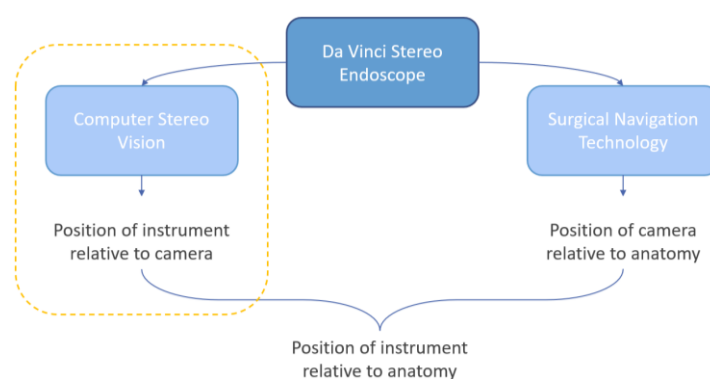


Figure 1 Schematic representation of the proposed design. This report will be focused on Computer Stereo Vision to determine the position of the instruments relative to the camera

1.1 Objectives

The objective of this study is to validate the use of computer stereo vision and geometry to calculate the 3D coordinates of manually selected markers in stereo images. This will be accomplished by evaluating the accuracy of the calculations compared to the real coordinates.

1.2 Hypothesis

Regarding the accuracy of coordinate calculations, the results are expected to not be fully accurate, especially at larger depths. In previous research from Mussabayev et al., the dependence of accuracy on the distance from the camera is measured.¹⁰ The results show an increased inaccuracy when the distance from the stereo camera increases. This can be explained by geometry, in which a small error in pixel coordinates of the images has more impact on the calculation of the z-dimension of the selected point. Furthermore, the calculations are based on manual selection of features in the stereo images. Therefore, the calculated coordinates may deviate from the actual coordinates of the selected features relative to the stereo endoscope.

II. Materials and Methods

2.1 Stereo camera calibration

For application of computer vision techniques, the Computer Vision Toolbox in MATLAB R2020a is used.¹¹ The first step is calibration of the stereo endoscope, which is done with use of the Stereo Camera Calibrator app in MATLAB R2020a. This app requires multiple stereo images of a checkerboard image to estimate the intrinsic and extrinsic camera parameters of both cameras in the stereo camera pair. These parameters are necessary for calculation of the 3D coordinates of a specific point. The used checkerboard needs to have the size of an actual surgical instrument, which resulted in a checkerboard squares size of 4.7 mm.

2.2 Experiment 1 – Test image

After calibration, an image pair is made from an optical table, which has M6 tapped holes on 25 mm spacing. To compensate the lens distortion, the function *undistortImage* is used for correction of the images. With the function *ginput*, six holes in the same row are selected manually in both images to get the pixel coordinates of these holes. To determine the x-, y-, and z-coordinate of the manually selected holes, the camera parameters and the pixel coordinates of the left and right images are used in the function *triangulate*. The output of this function is the x-, y- and z-coordinate of the manually selected holes. The difference in z-coordinates is calculated to determine the calculated distance between the holes. The error in this calculation is defined by the difference in the actual distance and calculated distance between the holes. The overall error is calculated by taking the root-mean-square (RMS) of the errors.

2.3 Experiment 2 – 3D printed markers

For further validation of computer vision and the *triangulate* function, the procedure will be repeated with 3D printed spherical white markers. These markers are designed in *SolidWorks 2020*, along with a stand for the stereo endoscope to keep the camera stable. The experimental set up of the experiment is shown in Figure 2. The markers have a height of respectively 100 mm, 150 mm, and 200 mm. The position of the markers with respect to the camera are predetermined and vary in depth between 175 and 625 mm. The markers are manually selected in both the stereo images to obtain the pixel coordinates of the markers. The x-, y- and z-coordinate are calculated by the *triangulate* function. The accuracy of this calculation is evaluated by calculating the error between the expected coordinates and the calculated coordinates. The overall accuracy is calculated by taking the root-mean-square (RMS) of the errors. The precision is calculated by calculating the standard deviation (STD). By repeating the manually selecting of the markers, six calculated coordinates are obtained from the same marker position. This is performed on the second and third marker.



Figure 2 Experimental set up of the Da Vinci endoscope and the 3D printed markers

2.4 Experiment 3 – Adjustable laser cut marker

To vary the position of the marker independent from the holes in the table, a new marker is made by laser cutting. This marker has two engraved scales which can be used to vary the x- and y-coordinate of the manually selected points. The marker can be slide in x-direction and can be positioned at a specific point on the scale. The scale on the marker is used to click on different heights in processing of the stereo images to vary the y-coordinate. The holes in the table are used to vary the z-coordinate of the marker. This enables to vary the position with smaller intervals and on multiple heights. An image of the laser cut marker is shown in Figure 3. Stereo images are taken at $x = -25$ mm, $x = -15$ mm, $x = 0$ mm, $x = 15$ mm, and $x = 25$ mm at depths of $z = 50$ mm, $z = 100$ mm, $z = 150$ mm, $z = 200$ mm, and $z = 250$ mm. The y-coordinate is manually selected in the stereo images at $y = -25$ mm, $y = -20$ mm, $y = -10$ mm, $y = 0$ mm, and $y = 10$ mm. The pixel coordinates obtained by clicking these y-coordinates are used to calculate the x-, y- and z-coordinate by the *triangulate* function. The accuracy of the calculation is evaluated by calculating the error between the expected coordinates and the calculated coordinates. The overall accuracy is calculated by taking the root-mean-square (RMS) of the errors. The function *worldToImage* is used to project the calculated 3D coordinates back to the pixel coordinates in both images. The difference between the selected pixels and calculated pixels in calculated in x- and y-direction for both images. The overall error in x- and y-direction is calculated by taking the RMS of the errors.

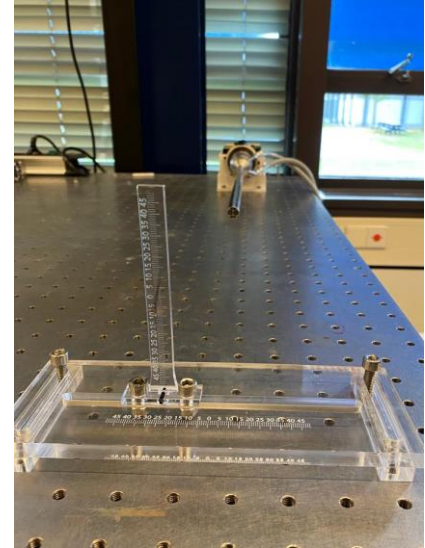


Figure 3 Experimental set up of the laser cut marker with scales in front of the Da Vinci endoscope

2.4.1 Sensitivity analysis

On the selected points from experiment 3, a sensitivity analysis is performed to evaluate the influence of manual error in the clicking process. The results can also be used to determine the desired accuracy of marker selection. The manually selected pixel coordinates are adjusted over a range of 10 pixels in both directions in x- and y-direction with steps of 1 pixel. This results in 441 pixel coordinates (21×21 pixels), which are used for calculation of the 3D coordinates. The resulting 3D coordinates are shown in a surface plot to visualize the influence of error in manual marker selection. For every individual point, the minimum and maximum z-coordinate is taken and the difference between these depths is calculated. This will provide insight in the range of depths when varying the pixel coordinates of the manually selected points.

III. Results

In this section the results from the described method are elaborated. The results of the calibration are elaborated in Appendix B.

3.1 Experiment 1 – Test image

The stereo images of the optic table are visualized in Figure 4. The manually selected points are visualized by yellow dots in the left image (green) and blue dots in the right image (purple). The error in calculated distance between the selected holes is shown in Table 1. The overall error in this calculation was 8.6 mm.

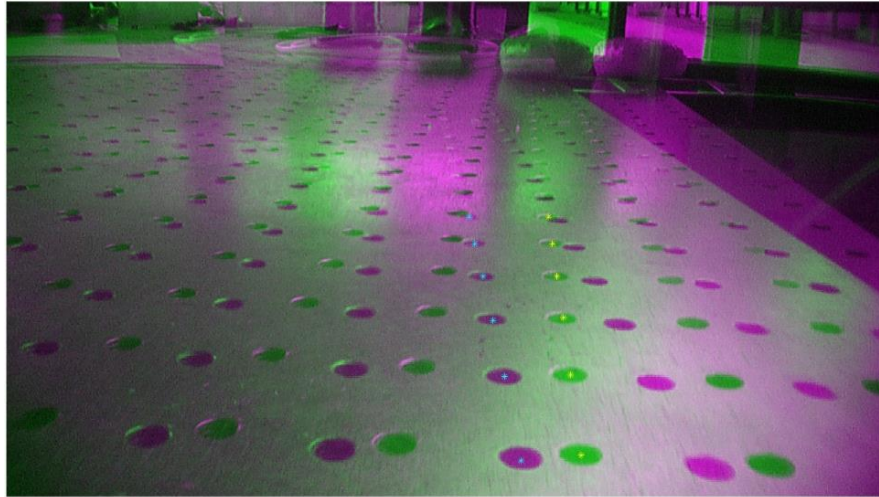


Figure 4 Overlay image of the left (green) and right (purple) stereo images of the optic table. The manually selected holes in the left image are shown by yellow dots, the ones in the right image by blue dots.

Table 1 The error between the expected distance and calculated distance between the holes.

Expected distance (mm)	Calculated distance (mm)	Error (mm)
25.00	19.4213	-5.5787
25.00	21.7787	-3.2213
25.00	14.4543	-10.5457
25.00	31.3514	6.3513
25.00	11.6790	-13.3210
Overall Error (mm)		8.6081

3.2 Experiment 2 – 3D printed markers

The stereo images of the 3D printed markers are visualized in Figure 5. The manually selected points are visualized by yellow dots in the left image (green) and blue dots in the right image (purple). A plot of the expected coordinates (blue dots) and calculated coordinates (red dots) of the first markers is shown in Figure 6. The coordinates are related to each other by a number. The overall error in x-, y- and z-coordinate, along with a distinction between near and far positions of the first marker, are shown in Table 2. The precision expressed in the standard deviation in x-, y- and z-coordinate for marker 2 and 3 is shown in Table 3.

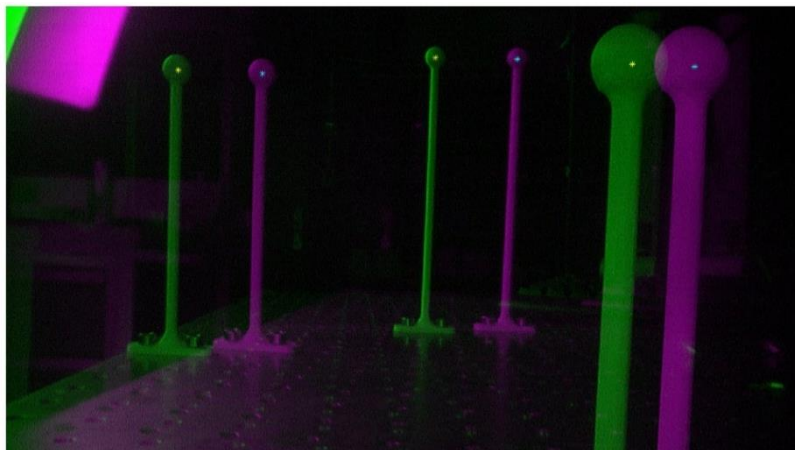


Figure 5 Overlay image of the left (green) and right (purple) stereo images of the 3D printed markers. The manually selected markers in the left image are shown by yellow dots, the ones in the right image by blue dots.

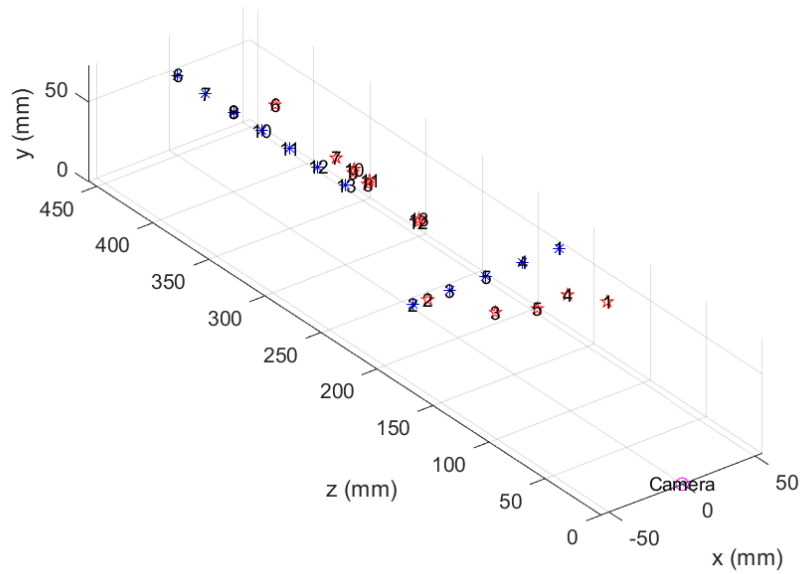


Figure 6 3D plot of the expected coordinates (blue dots) and calculated coordinates (red dots) of the first marker

Table 2 The overall error in x-, y- and z-coordinate (marker 1, 2 and 3); Error in marker 1 near by the camera; Error in marker 1 far away from the camera

	x (mm)	y (mm)	z (mm)
Overall error	9.6588	13.4799	101.2439
Error marker 1 near	3.9162	6.4492	37.7692
Error marker 1 far	1.2090	11.5233	89.0467

Table 3 Standard deviation in calculated coordinates of marker 2 and 3 in x-, y- and z-dimension

	x (mm)	y (mm)	z (mm)
STD marker 2	3.6554	4.9550	14.3540
STD marker 3	2.4510	9.3738	26.4110

3.3 Experiment 3 – Adjustable laser cut marker

The stereo images of the adjustable laser cut marker are visualized in Figure 7. The manually selected points on the scale are visualized by yellow dots in the left image (green) and blue dots in the right image (purple). The overall error in x-, y- and z-coordinate at the different depths are shown in Table 4. The plots of the expected coordinates (blue dots) and calculated coordinates (red dots) at the different depths are shown in Figure 9. The error between the manually selected pixel coordinates and the pixel coordinates calculated from the 3D coordinates was larger in the left image, specifically in the x-coordinates. Therefore, the error in x-coordinates is visualized in Figure 8 for both the left and right image. The RMS of the errors in x- and y-dimension are shown in Table 5 for the left and right image.

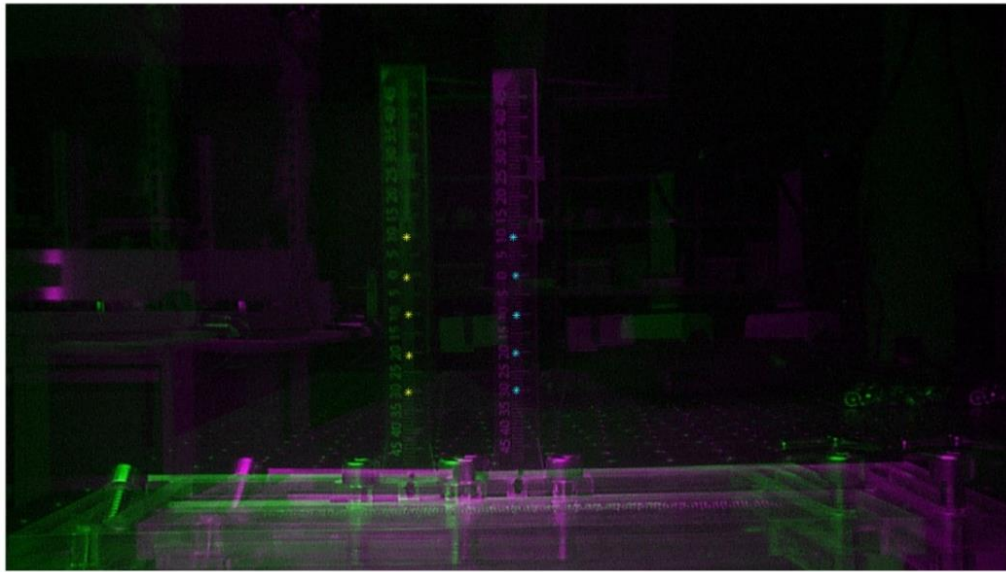


Figure 7 Overlay image of the left (green) and right (purple) stereo images of the laser cut marker. The manually selected points on the scale in the left image are shown by yellow dots, the ones in the right image by blue dots.

Table 4 The overall error in x-, y- and z-coordinate at different depths, i.e., $z = 50$ mm, $z = 100$ mm, $z = 150$ mm, $z = 200$ mm, and $z = 250$ mm.

z-coordinate (mm)	x (mm)	y (mm)	z (mm)
50	3.4998	1.1802	6.0998
100	9.3896	8.1321	40.2027
150	17.2553	22.1045	118.0980
200	33.6112	51.3828	291.7728
250	60.9624	99.1747	704.4984

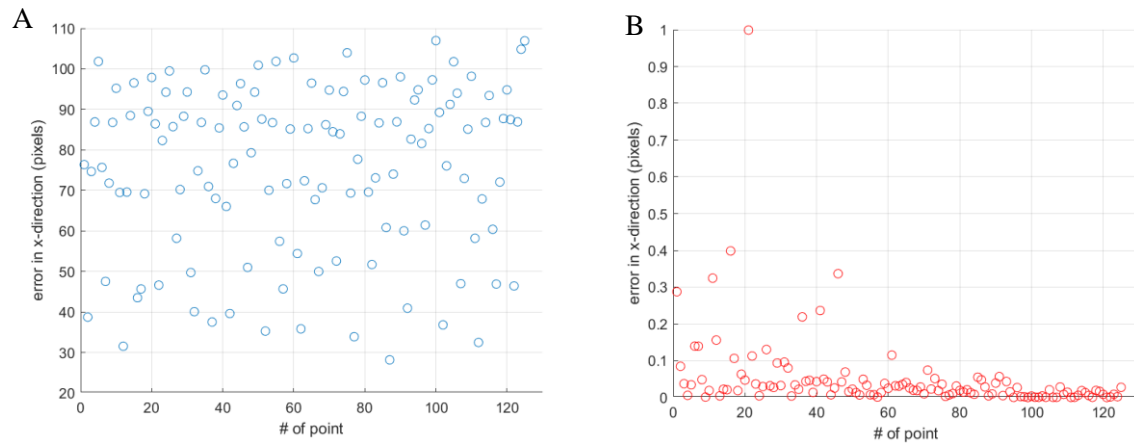


Figure 8 Plots of the error between the manually selected pixels coordinates and the calculated pixel coordinates from the 3D coordinates. A) The error in x-direction per point in the left image. B) the error in x-direction per point in the right image.

Table 5 Overall errors in x- and y-direction between the manually selected pixels coordinates and the calculated pixel coordinates from the 3D coordinates for the left image and right image

	x (pixels)	y (pixels)
Left image	78.3186	7.2652
Right image	0.1198	2.3284

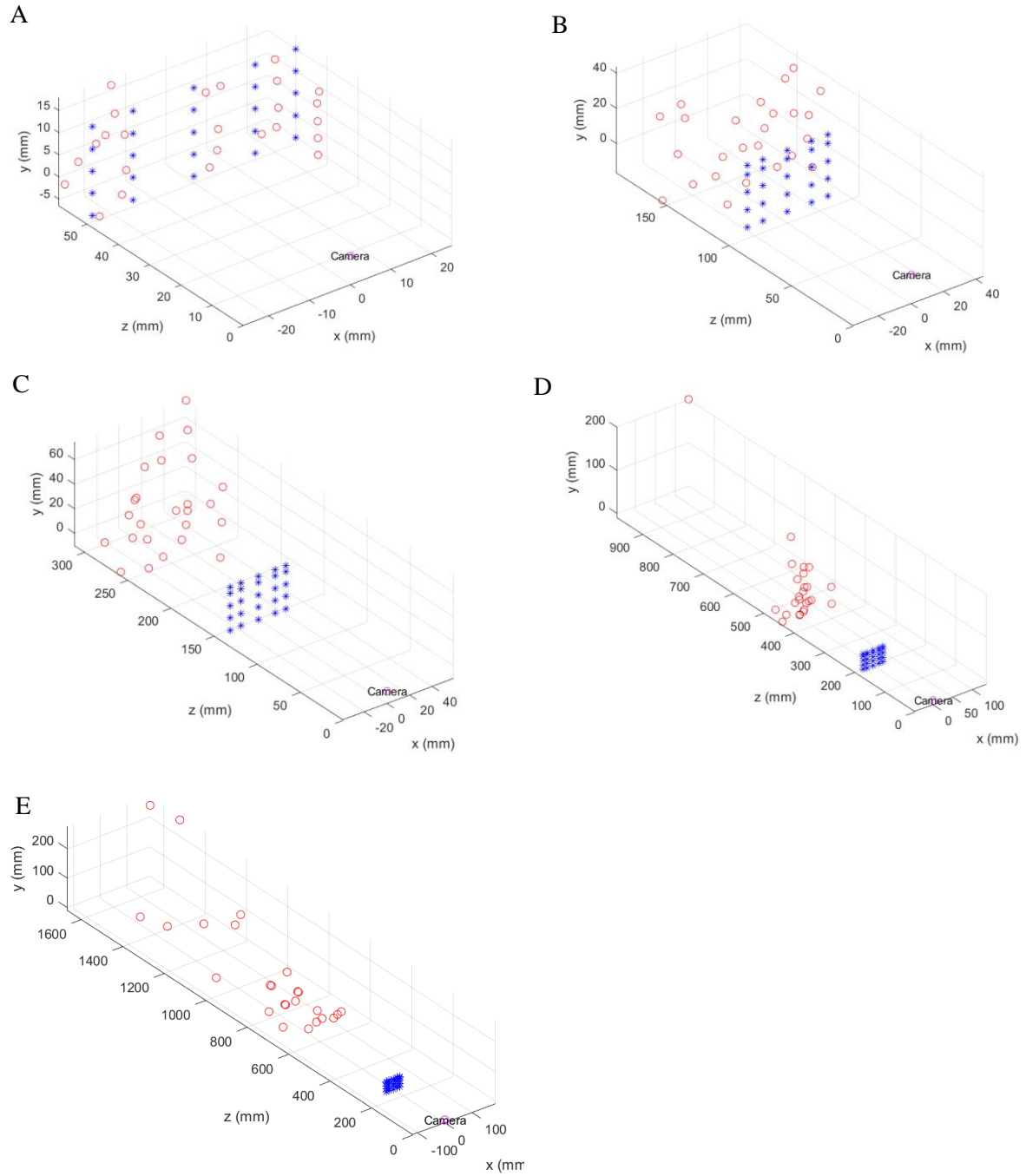


Figure 9 3D plot of the expected coordinates (blue dots) and calculated coordinates (red dots) at different depths. A) Coordinates of markers at $z = 50$ mm; B) Coordinates of markers at $z = 100$ mm; C) Coordinates of markers at $z = 150$ mm; D) Coordinates of markers at $z = 200$ mm; E) Coordinates of markers at $z = 250$ mm.

3.3.1 Sensitivity analysis

The resulting surface plots of the 3D coordinates calculated from the adjusted pixel coordinates are shown in Figure 10. At larger depths, the surface planes tend to bend to the right sight. Because the z-coordinate (i.e. the depth) is influenced most by the variation of pixel coordinates, the difference between the minimum and maximum z-coordinate is calculated per point. The maximum difference in one point was 117.34 mm. The RMS of all difference in z-coordinate is 19.87 mm.

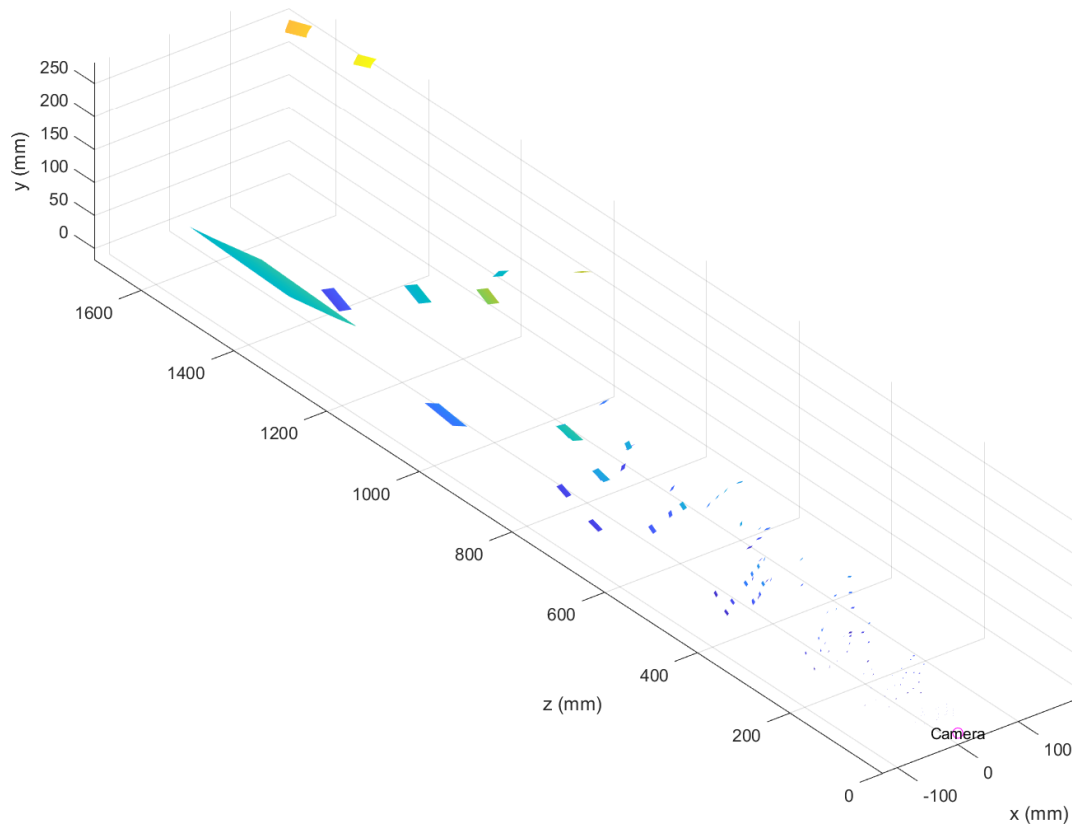


Figure 10 The surface plots of the 3D coordinates calculated from the adjusted pixel coordinates from different planes. A) yz-plane, B) xy-plane, C) xz-plane

IV. Discussion

In this report the accuracy of 3D coordinate calculation based on computer stereo vision is evaluated. In this section the results will be discussed, along with the limitations and recommendations for further research.

Based on Experiment 2, the accuracy and precision in z-dimension decrease in case the marker is further away from the camera. Nevertheless, the accuracy and precision in x- and y-dimension remain in the same range and do not show a relation to the depth of the marker. Therefore, distinction had been made in the depth (z-dimension) in Experiment 3. The results from Experiment 3 confirm an increase of the inaccuracy based on the calculation error. This error mainly increases in the z-dimension, but also in the x- and y-dimension. This can be explained by the small baseline between the two cameras. Because both cameras need to fit in the endoscope, the baseline is only 5 mm. Mussabayev et al. showed that this base distance determines the accuracy of the calculations. The farther the object from the camera, the longer the base distance must be. Therefore, the distance over which the coordinates can be calculated accurately is restricted by the small baseline.¹⁰

The calculation errors can also be explained by manual clicking. Due to geometry, these errors have more impact on the calculation when the marker is further away from the camera. In Experiment 3, this manual selection of markers is performed using a scale without one specific point in the center to click on. To make this process more reproducible, the use of a marker with dots would be more accurate. This would standardize the procedure and may lead to more accurate pixel coordinates and with that 3D coordinates.

To evaluate the influence of manual errors, the sensitivity analysis is performed. The results show that adjustment of the pixel coordinates influence the z-coordinate of the calculated 3D point. Nevertheless, the large difference between the calculated 3D coordinate and the actual 3D coordinates of the marker are not declared by the difference in z-coordinate. The calculated 3D coordinates at larger depths ($z = 20$ cm or $z = 25$ cm) have z-coordinates of around 1000 mm. The largest difference in z-coordinate after adjusting the pixel coordinates was 117.34 mm.

Another cause of the large error in z-coordinate can be the lack of image rectification or lens distortion. This is also shown in the results from experiment 3, where the calculated pixel coordinates from the 3D coordinates deviate significantly in the left image. In the right image, the error between the calculated pixel coordinates and the selected pixel coordinates is almost zero. This can be explained by a lack of rectification, which result in corresponding points in both images to have the same row coordinates. In further research, it may be valuable to investigate this finding and the use of image rectification as a possible solution.

Calculation errors can also be caused by errors in the calibration. Calculation of the 3D coordinates with the *triangulate* function is also based on the camera parameters which result from this calibration. After rejection of stereo pairs with high reprojection errors, 31 stereo pairs were used with a mean reprojection error of 1.07 pixels. Generally, an acceptable mean projection error needs to be less than one pixel.¹² The calibration might have been more accurate when more high-error images were removed. Nevertheless, the calibration improves by using more images. Therefore, a trade-off is to be made between removing high-error images and keeping as many images as possible. This is done by only removing the three images with the highest reprojection error.

The problems of manual selection of markers can potentially be solved using keypoint- and edge-based methods. This enables accurate and standardized detection of markers. Tekin et al. showed that these automatic detection methods require textured objects and high resolution images.¹³ In the used endoscope, the image quality of the left image is worse compared to the right image. The left image might need to be edited to avoid a difference in keypoint detection between both images. In the

context of automatic keypoint detection, it might be valuable to study the combination between different kind of markers and the use of different detection algorithms.

The use of automatic keypoint detection in clinical settings requires attachment of markers to the surgical instruments. A problem that might arise here is the visibility of these markers. Due to movement of the instruments, the markers can easily disappear out of the camera's sight. The shape of the marker can also deform if the marker has an angle relative to the camera. This makes it harder to detect the marker and determine the exact location of the center of the marker. These two limitations complicate the ability of real-time tracking of the instruments.

Instead of automatic keypoint detection, the detection of surgical instrument can also be performed using machine learning. In previous research by ten Brink, a DoubleU-Net was able to accurately segment a Da Vinci surgical instrument from a set of 2D monocular images.¹⁴ Important limitations from this study were the use of ex vivo images, segmentation of only one instrument, non-real time performance and the use of mono-images instead of stereo images.

V. Conclusion

The application of computer vision in robot-assisted laparoscopic surgery is accurate at 50 mm from the stereo endoscope. Nevertheless, the accuracy decreases with the distance from the camera. Therefore, it would not be possible to rely on the calculated positions in case of surgical instruments relative to the patient in preoperative scans based on these experiments. Further research, such as standardization of manual clicking, image rectification and automatic feature detection, needs to be done to create a reliable navigation system based on computer stereo vision.

VI. Bibliography

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249. doi:10.3322/CAAC.21660
2. Integraal Kankercentrum Nederland. Incidentie darmkanker. Published 2021. Accessed September 2, 2021. <https://iknl.nl/kankersoorten/darmkanker/registratie/incidentie>
3. Integraal Kankercentrum Nederland. Behandeling darmkanker. Published 2021. Accessed September 2, 2021. <https://iknl.nl/kankersoorten/darmkanker/registratie/behandeling>
4. Kok END, Veen R van, Groen HC, et al. Association of Image-Guided Navigation With Complete Resection Rate in Patients With Locally Advanced Primary and Recurrent Rectal Cancer: A Nonrandomized Controlled Trial. *JAMA Netw Open.* 2020;3(7):e208522-e208522. doi:10.1001/JAMANETWORKOPEN.2020.8522
5. Huang C-M, Huang C-W, Ma C-J, et al. Outcomes of neoadjuvant chemoradiotherapy followed by radical resection for T4 colorectal cancer. *World J Gastrointest Oncol.* 2020;12(12):1428. doi:10.4251/WJGO.V12.I12.1428
6. Cheong C, Kim N. Minimally Invasive Surgery for Rectal Cancer: Current Status and Future Perspectives. *Indian J Surg Oncol.* 2017;8(4):591-599. doi:10.1007/S13193-017-0624-7
7. Sheng S, Zhao T, Wang X. Comparison of robot-assisted surgery, laparoscopic-assisted surgery, and open surgery for the treatment of colorectal cancer A network meta-analysis. Published online 2018. doi:10.1097/MD.00000000000011817
8. Du X, Kurmann T, Chang P, et al. Articulated Multi-Instrument 2-D Pose Estimation Using Fully Convolutional Networks. *IEEE Trans Med Imaging.* 2018;37(5):1276-1287. doi:10.1109/TMI.2017.2787672
9. Atallah S, Parra-Davila E, Melani A, Romagnolo L, Larach S, Marescaux J. Robotic-assisted stereotactic real-time navigation: initial clinical experience and feasibility for rectal cancer surgery. *Tech Coloproctol.* 2019;23(1):53-63. doi:10.1007/S10151-018-1914-Y
10. Mussabayev RR, Kalimoldayev MN, Amirgaliyev YN, Tairova AT, Mussabayev TR. Calculation of 3D Coordinates of a Point on the Basis of a Stereoscopic System. *Open Eng.* 2018;8(1):109-117. doi:10.1515/ENG-2018-0016
11. The MathWorks Inc. Computer Vision Toolbox Documentation - MathWorks Benelux. Accessed September 28, 2021. <https://nl.mathworks.com/help/vision/>
12. The MathWorks Inc. Using the Stereo Camera Calibrator App - MATLAB & Simulink - MathWorks Benelux. Accessed November 11, 2021. <https://nl.mathworks.com/help/vision/ug/using-the-stereo-camera-calibrator-app.html>
13. Tekin B, Sinha SN, Fua P. Real-Time Seamless Single Shot 6D Object Pose Prediction. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* Published online December 14, 2018:292-301. doi:10.1109/CVPR.2018.00038
14. ten Brink R. *Automatic Segmentation Using the DoubleU-Net and 3D Visualisation of Non-Rigid Da Vinci Surgical Instruments.*; 2021.

Appendix A – Visual representation of design

in Figure 11, a visual representation is shown of the proposed design for image-guided navigation in robot-assisted rectal surgery. On the left, the Brainlab navigation systems detects the IR markers on the Da Vinci stereo camera and the anatomy of the patient. In this way, the position of the camera relative to the anatomy can be determined. On the right, the image of a single stereo image is shown with sight on the instruments with detectable markers. With computer stereo vision, the position of the instruments can be determined relative to the camera. By combining both coordinate systems, the position of the instruments can theoretically be determined relative to the anatomy of the patient.

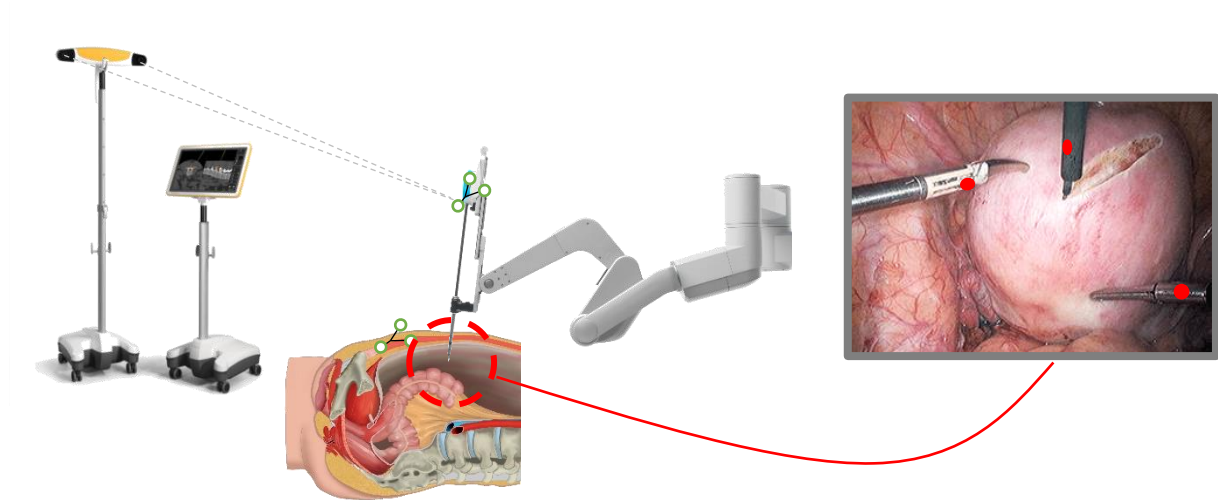


Figure 11 A visual representation of the design. Left: Brainlab navigation. Middle: Da Vinci arm with IR markers in patient's anatomy. Right: Single stereo image with markers on instruments.

Appendix B – Camera Calibration

For calibration in the Stereo Camera Calibrator app in MATLAB R2020a, checkerboard images are used with a checkerboard squares size of 4.7 mm. In total, 43 paired stereo images were made, of which 9 pairs were rejected by the Stereo Camera Calibrator app. An example of a stereo image pair is shown in Figure 12. From the 34 remaining stereo pairs, 3 images were manually discarded because their reprojection error was higher than 2 pixels. The reprojection error is the distance in pixels between the detected and reprojected points. This results in 31 stereo pairs remaining for calibration. The radial distortion was set to 3 coefficients and computation of skew or tangential distortion had no effect on the reprojection error. The reprojection errors of the remaining stereo pairs are shown in Figure 13. The visualization of the extrinsic camera parameters in Figure 14 shows the position and orientation of the checkerboard image relative to the position of the cameras of the endoscope. The checkerboard image is held in different angles and at different distances from the cameras.

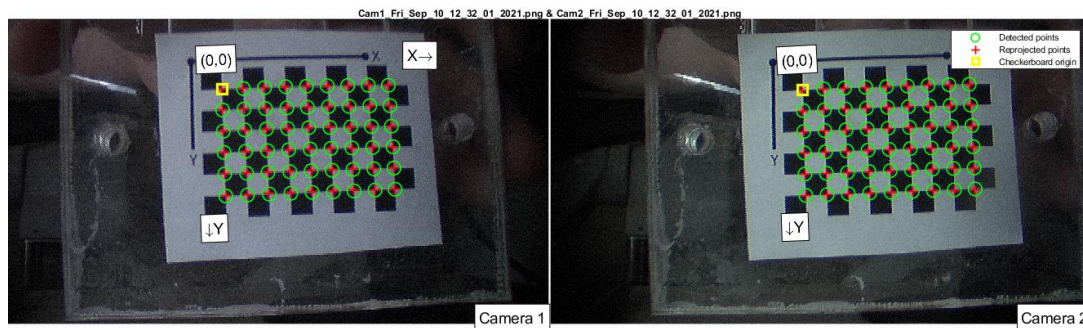


Figure 12 A stereo image pair of the checkerboard image with the detected x- and y-axis. The yellow square represents the checkerboard origin, the green circles the detected points and the red crosses the reprojected points.

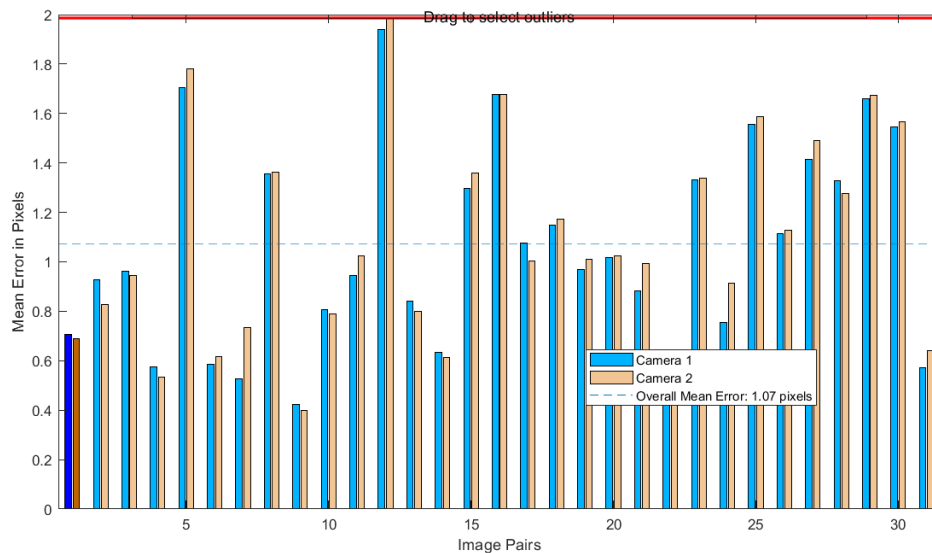


Figure 13 Bar graphs of the mean reprojection error per camera and per image pair. The dashed line represents the overall mean error, which is 1.07 pixels.

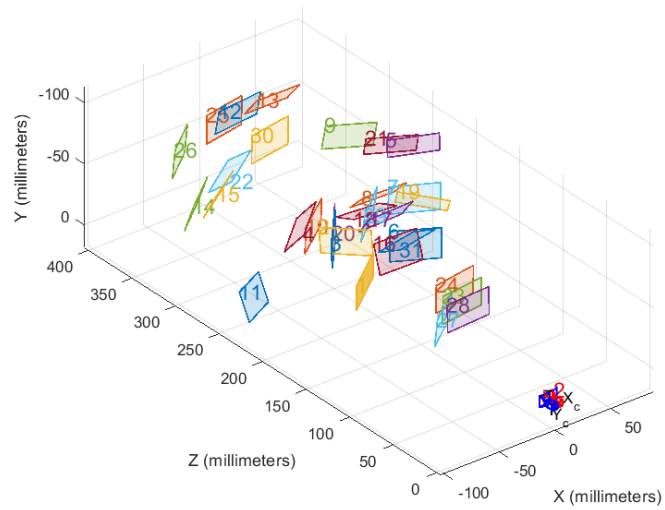


Figure 14 A visualization of the extrinsic camera parameters. The positions of the checkerboard image are numbered and shown in different colors. The X_c and Y_c represent the position of the stereocamera.