Depth Cues in Augmented Reality for Training of Robot-Assisted Minimally Invasive Surgery

Full Paper

Nicklas H. Christensen
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
nhch14@student.aau.dk

Oliver G. Hjermitslev
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
ohjerm14@student.aau.dk

Frederik Falk
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
ffalk13@student.aau.dk

Marco B. Madsen
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
mma14@student.aau.dk

Frederik H. Østergaard
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
fhas14@student.aau.dk

Martin Kibsgaard
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
kibsgaard@create.aau.dk

Martin Kraus
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
martin@create.aau.dk

Johan Poulsen
Aalborg University Hospital
Reberbansgade 15
9000 Aalborg, Denmark
jop@rn.dk

Jane Petersson
Aalborg University Hospital
Reberbansgade 15
9000 Aalborg, Denmark
jane.petersson@rn.dk

ABSTRACT

Training of robot-assisted minimally invasive surgery often includes supervised practice with a robotic surgical system. In this case, augmented reality can improve the communication between instructor and trainee, for example, by allowing the instructor to demonstrate skills with virtual surgical instruments that are shown to the trainee by means of augmented reality. However, virtual instruments are more difficult to handle than the real instruments—partly due to the lack of depth cues. In order to improve the usability of virtual surgical instruments, we compared five depth cues. Results showed both a preference for the artificial highlight cue and an aversion to transparency and depth of field. The highlight cue was therefore reviewed by experienced surgery instructors. These experts agreed that the highlight cue was beneficial and that the prototype could be used to some extent already and fully upon further development.

CCS CONCEPTS

•**Human-centered computing →Mixed / augmented reality;** *Collaborative interaction;* •**Computing methodologies →Mixed / augmented reality; Perception;**

KEYWORDS

minimally invasive surgery, robot-assisted surgery, teleoperation, telesurgery, telementoring, training, mixed reality, augmented reality, depth perception

ACM Reference format:
Nicklas H. Christensen, Oliver G. Hjermitslev, Frederik Falk, Marco B. Madsen, Frederik H. Østergaard, Martin Kibsgaard, Martin Kraus, Johan Poulsen, and Jane Petersson. 2017. Depth Cues in Augmented Reality for Training of Robot-Assisted Minimally Invasive Surgery. In *Proceedings of AcademicMindtrek'17, Tampere, Finland, September 20–21, 2017,* 7 pages.
DOI: 10.1145/3131085.3131123

1 INTRODUCTION

Since its introduction in 1970, Minimally Invasive Surgery (MIS) has become increasingly popular due to the reduced damage to human tissue [14]. It is a type of surgery where operations are performed through small openings or incisions in the patient. The technique has been improved with several modern advancements, most notably the introduction of video cameras. Unlike open surgery, the surgical team only has vision of the surgery site using an endoscopic camera and a monitor. The advancements in cameras and displays is one of the key factors that allowed for the recent implementation of robots such as the da Vinci Surgical System, which is shown in Figure 1. This proved successful as the amount of robot-assisted surgeries tripled between 2007 and 2010 [13].

In robot-assisted minimally invasive surgery (RAMIS), the surgeon sits at a console and controls a robot with small instruments that are inserted into the patient through small incisions. The most wide-spread general purpose robotic surgical systems are the da Vinci systems from Intuitive Surgical. In these systems, the console

AcademicMindtrek'17, Tampere, Finland

DOI: 10.1145/3131085.3131123

Figure 1: The da Vinci S Surgical System with a surgeon controlling the robot from the console on the left and an assistant communicating with the surgeon by drawing on the view of the surgical field. Copyright 2017 Intuitive Surgical, Inc.

provides the surgeon with a stereoscopic view of the surgery site, while the rest of the surgical team rely on monoscopic monitors.

During supervised training for RAMIS, the trainee's vision is limited to that of the robot and the immersive interface of the console makes visual communication with the instructor difficult. Augmented reality can improve the visual communication between instructor and trainee by allowing the instructor to point at parts of the operating field and to demonstrate the use of a virtual surgical instrument [4, 7]. However, virtual surgical instruments are more difficult to handle than real surgical instruments due to the lack of depth cues and other differences between the virtual instruments and their real counterparts. Therefore, we designed and evaluated a selection of five depth cues in an augmented reality environment, namely an overlay for consoles and monitors of da Vinci Surgical Systems. The best performing version was then evaluated in an expert review.

We review previous work in Section 2 and summarize the theory of relevant depth cues in Section 3. Section 4 describes the experimental comparison of our implementations of selected depth cues, while Section 5 reports on the expert evaluation of the depth cue that performed best in the experiment. Section 6 discusses our results and Section 7 concludes this paper.

2 PREVIOUS WORK

Several researchers have observed the need for improvements to the visual communication in RAMIS [6, 7, 15].

In particular, previous work has been published on improving or creating alternatives to the standard 2D telestration of the da Vinci Surgical Systems. The telestration feature allows instructors to make line drawings on a touchscreen and show the line drawings in the console. However, when the line drawings are shown on the stereoscopic display, they are offset and placed at a fixed depth causing double vision and making precise pointing impossible [6].

Ali et al. [1] implemented a system that is able to show 2D telestration on a 3D display by processing the right and left images and finding matching features to offset the drawings from one image to the other. They compared their system to 2D telestration and found no significant difference in error rates between the two versions. However, the overall error rate was 9.1% (9 errors in 99 trials) which indicates that neither system is reliable for precise pointing. They evaluated the systems using an experienced surgical instructor and three trainees of varying surgical experience. The mentor pointed to a pin in clusters of pins, which the trainee had to touch with the instruments, and measured the errors made and time taken to reach the pin from a starting point.

To eliminate the errors caused by showing the drawing at a fixed depth or translating the drawings, we proposed a solution that shows the line drawings (and a cursor) to only one eye. This makes the drawings precise and showed improvements of the communication between trainee and instructor. This is also in line with the findings of Teather and Stuerzlinger [17] who showed that one-eyed cursors improve performance compared to other screen cursors on a stereoscopic output.

Jarc et al. [4] found that the use of 3D instruments to instruct trainees was preferred over the standard 2D telestration. The virtual instruments were controlled by the instructors using a Razer Hydra game controller and they could view the augmented video stream on a passive 3D TV. In their work each test group consisted of one instructor and one trainee pairing up to perform four different dry-lab exercises. Each pair completed the tests in random order. Both instructor and trainee then had to score eight different questions on a standardized questionnaire. The results showed that trainees are capable of advancing their learning curve faster with the use of virtual 3D instruments as guidance. However, transferring from dry-lab exercises to wet-lab training and actual surgeries has not yet been tested and therefore the effect of this is not known. We have previously implemented similar overlay of virtual instruments [6] with the addition of simple thread simulation that made it possible for instructors to demonstrate knot tying. However, due to difficulties perceiving the depth of the virtual instruments, and especially the depth of the thread, it was difficult for the instructors to grab and interact with the thread.

The system used to overlay the video streams of the robot with computer graphics in previous work has usually resulted in delaying the original video signal more than 100 ms [1, 4, 18]. In [6] we present a system that is able to overlay the video signals in less than 1 ms [8] and it is this system that we use for the expert evaluation described in Section 5.1. The drawback of this approach is that the video never enters system memory and is not accessible for processing on the computer, e.g. to perform depth matching like Ali et al. [1]. However, the systems that delay the original video signal significantly have shown to be unusable during real training [11].

3 DEPTH CUES IN AUGMENTED REALITY

According to Matatko et al. [9], four groups of depth cues are used by humans: binocular disparity, pictorial cues, oculomotor cues, and motion-produced cues. Binocular disparity is the assessment of difference in stereoscopic vision and therefore already exists in the stereoscopic display of da Vinci robots. However, it is not available on the monitors that the assistants and instructors use to view the operational field. Pictorial cues include relative height (in the image), relative size, occlusion, shading, and texturing.

Occlusion is especially noteworthy as instructors have discussed transparent virtual tools in order to mitigate obstruction of vision, which to some degree negates the occlusion depth cue. According to Epstein [2], occlusion is widely used to estimate relative depth at all distances. Removing it may cause more issues with perception than anticipated. Oculomotor cues are induced by eye movement, such as the convergence of the eyes to fixate on a point. Unfortunately, the endoscopic cameras cannot be influenced by the system. However, it is possible to blur the non-focused objects by using depth of field. This cue is only secondary according to Kelle [5]. Nonetheless, it is still relevant to test whether depth of field is a useful cue; in particular, in the absence of other cues. This can be simulated using image blur, which, according to Mather [10], should equal stereopsis as depth cue at the right distances. Motion-produced cues include deletion and accretion, the disappearance and appearance of moving objects behind closer ones. As with occlusion, it is already a standard part of most graphics engines and does not require additional implementation. However, when overlaying graphics on stereoscopic video, where the depth of the real objects is unknown, it is difficult to have the real objects occlude the virtual. In this work, we make sure that the virtual objects are always in front of the real objects to avoid any conflicts between occlusion and other depth cues.

Motion parallax is the effect of close objects moving "faster" than far objects as the viewer or the environment moves [16]. It occurs due to the relative angular change in the retina between near and far objects and represents another possible implemented depth cue. There are concerns that motion parallax, as the distance increases, becomes a perception of motion rather than a perception of depth [3, 12]; however, this should not be the case in this setup where the objects are close to the camera.

Out of all depth cues discussed, some are automatically implemented in the system while some cannot be implemented due to restrictions imposed by the cameras. We were able to implement five depth cues; shadows, depth of field, parallax, transparency and a cue that we call "highlight." These are shown in Figure 2 with the exception of motion parallax due to its moving nature. Shadows were tested even though the final system is an overlay and thus cannot have a (virtual) background. Should shadows prove to significantly improve depth perception, it would require additional work to correctly implement it as it would require finding the depth of the real objects (e.g. using stereo depth mapping). The "shadows" cue enabled shadows in the scene using artificial light sources pointed directly forward from the virtual camera, similar to the light setup of the da Vinci systems.

Depth of field was implemented using a real-time post-processing image effect with focal length equal to the distance from the camera to the instrument closest to the virtual needle in the scene. As the real camera is rarely moving, motion parallax was implemented in the virtual scene by continuously moving the camera left and right a small distance (the game engine's equivalent of 10 cm). This ensured test participants were subject to the depth cue at all times. Transparency was implemented by setting the alpha value of all available objects in the scene, namely the instruments and the needle, to 0.5. This means that at any point, whether behind or in front,

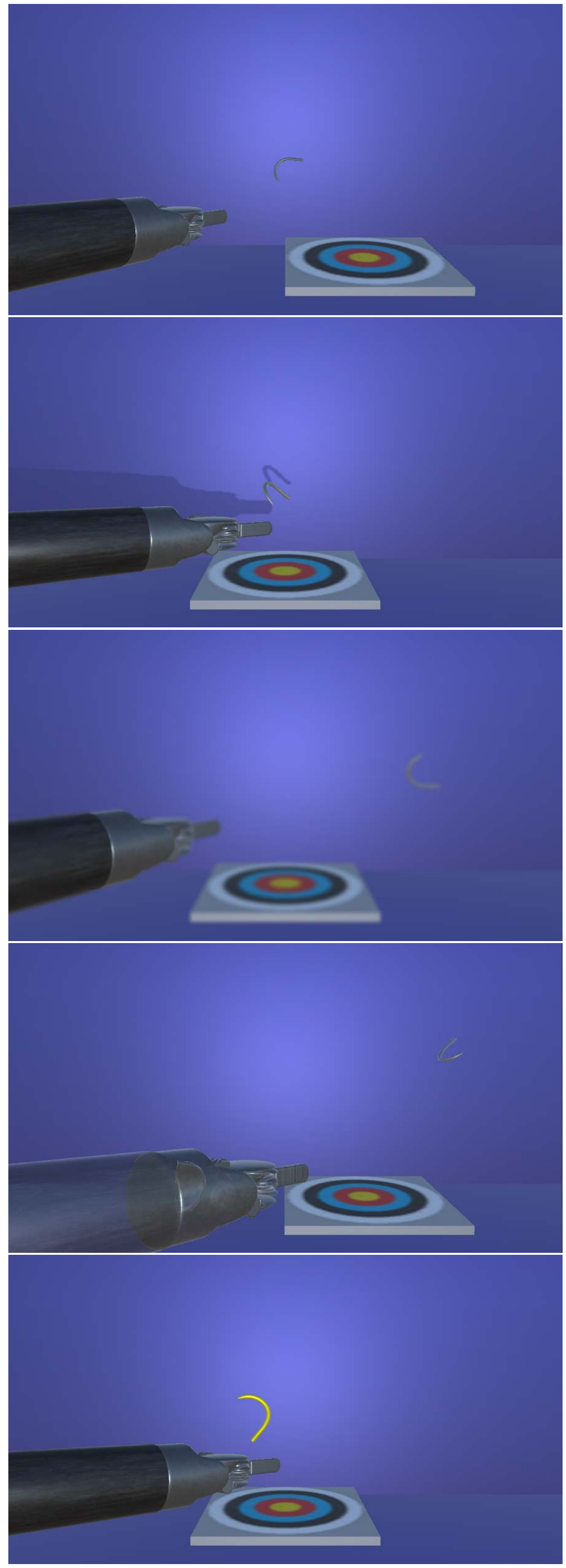


Figure 2: From top to bottom: The Normal, Shadow, Depth of Field, Transparency, and Highlight cues.

the objects should add equally to the output image. The artificial highlight cue was designed to color the needle in the scene red, yellow, or green depending on the distance to the closest tool. Green indicates that the instrument is close enough to grab, yellow means just out of reach and red is when the instrument is far from the needle.

4 QUANTITATIVE EXPERIMENT

A set of tests was conducted to assess the effect of the designed depth cues. The tests were designed such that participants with no former knowledge of RAMIS could perform them. The participants were all students from Aalborg University. In the tests, the participants were tasked with moving a needle to a target location in a virtual environment 30 times with different depth cues in a pseudo-random order (exactly 5 repetitions per depth cue and as a control). To move the needle, the participants used a Razer Hydra to control two virtual instruments resembling the instruments of the da Vinci robot. When the needle was placed on the target, the scene reloaded and the needle and target would spawn pseudo-randomly to ensure that the distance was kept relatively uniform while not allowing participants the benefit of a static starting position. To account for the learning curve all participants were given as much time as needed to get accustomed to the controls before starting the test.

Upon completion of each instance, the program logged the time taken to grab the needle, the total time of the instance, the distance to the center of the target from the point of the needle, errors, correct grabs, an ID, and which cue was used in that instance.

The aim of the experiment was to test the following hypotheses:

H_A: *"There is a difference in performance between the different depth cues."*

H_0: *"There is no difference in performance between the different depth cues."*

Performance in this context is defined as a combination of time, accuracy, and errors. A difference in performance can be in any individual or combination of these values. However, we did not evaluate their relative importance as this presumably depends on the individual instructor of RAMIS.

4.1 Method and Materials

The test was performed at Aalborg University with 30 participants and consisted of them grabbing and placing a needle inside of a designated target with a random depth cue applied to the scene. The procedure was evaluated beforehand by conducting a pilot test.

The following procedure list was created:

- The participant is given an introduction to the test
- The participant learns the controls of the system
- Once ready, the test starts and the first depth cue is applied
- The test is repeated until each of the six depth cues has been tested five times
- The participant is given a questionnaire regarding their experience

The purpose of the questionnaire was to assess the experience of the controls and the realism of the 3D tools.

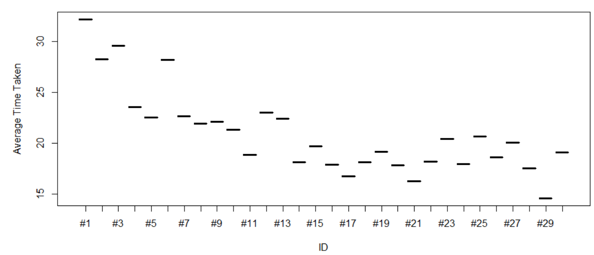


Figure 3: The average completion time for each ID.

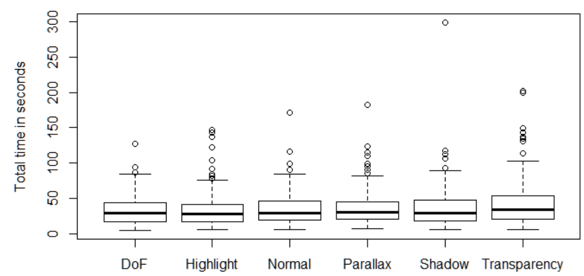


Figure 4: Boxplot of total time taken on each cue.

4.2 Results

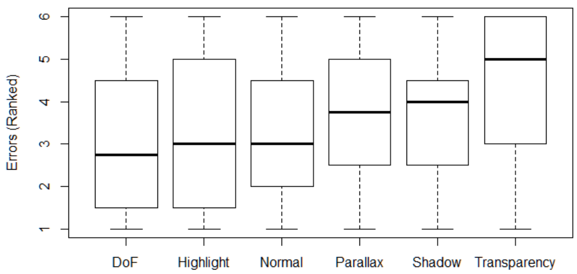
Each of the 30 participants contributed 30 observations of seven variables. Each observation was given an ID from 1 to 30 (first observation of each participant having ID 1, second ID 2, etc.) The total amount of observations was 896 instead of 900 because four observations had to be skipped. Figure 3 shows the average completion time for all ID's. Apparently, there is a tendency that participants spent more time in the first few tasks — independently of the depth cues.

As can be seen in Figure 4, plotting the raw time measurements does not show an effect of the introduced cues on user performance. The plots for errors and accuracy are very similar and have thus been omitted.

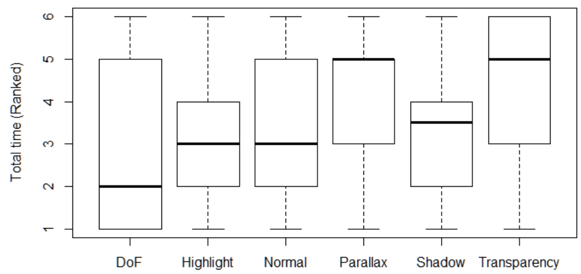
To take advantage of the within-subject design, we averaged and ranked the measurements from each participant individually. This has the benefit of reducing the effect of difference in skill between subjects. As can be seen in Figure 5, box plots of the ranked data show a tendency for parallax and transparency to have a negative effect on user performance.

The non-parametric Friedman's test was used to analyze four response variables: errors, total time, grab time, and accuracy. The results of the test are summarized in Table 1, which shows that the test gave a significant result for the total time and grab time metrics.

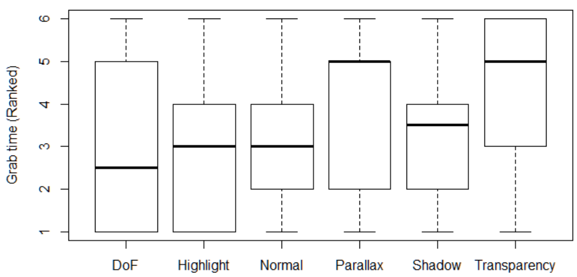
To find out which cue is significant in relation to total time and grab time, we did pairwise comparisons of average errors using paired Wilcoxon signed rank tests with Holm-Bonferroni corrected p-values. As shown in Table 2, the only significant difference is between transparency and highlight; in favor of the highlight cue. The pairwise comparison of the grab time measurements showed a



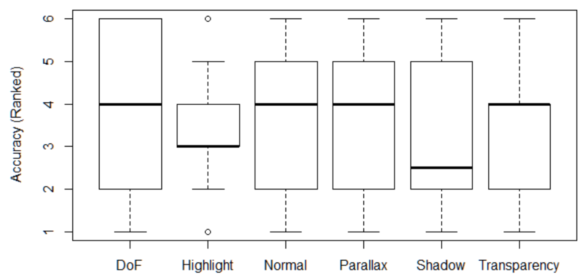
(a) Errors



(b) Total time



(c) Grab time



(d) Accuracy

Figure 5: Box plots of the four different performance metrics. Values have been averaged and ranked per participant.

Metric	Friedman chi^2	df	p-value
Errors	8.9	5	0.115
Total Time	13.7	5	**0.017**
Grab Time	12.0	5	**0.035**
Accuracy	2.8	5	0.731

Table 1: The results from the Friedman's tests. Bold text indicates a significant result (p < 0.05).

Cue	DoF	Highlight	Normal	Parallax	Shadow
Highlight	1.000	-	-	-	-
Normal	1.000	1.000	-	-	-
Parallax	0.523	1.000	1.000	-	-
Shadow	1.000	1.000	1.000	1.000	-
Transp.	0.106	**0.045**	0.056	0.469	0.265

Table 2: Pairwise comparisons of total time using paired Wilcoxon signed rank tests with Holm-Bonferroni corrected p-values. Bold text indicates a significant result (p < 0.05).

similar tendency in disfavor of transparency and to a smaller degree in disfavor of parallax.

The experiment also included a questionnaire that each participant filled out after testing. Together with observations from the test and participants' comments during the test, these assessed the user experience of the system. Several participants commented on the difficulty of estimating the depth in the scene and the majority of participants used occlusion as the main cue both when grabbing and placing the needle. This meant that if the tools were near the needle, the participant would typically continue until the tools clipped through it. This technique was also used when placing the needle.

Several participants commented on their preference of the highlight cue. When the colour of the needle changes to green, the participants assume that they can grab on to the needle, however this is not always the case as the colour is dependent on the distance rather than the participant's potential to grab it. Therefore, the highlight cue does not live up to the users' perceived affordance. Some commented that it would be nice with a change of colour when the needle was successfully grabbed.

The depth of field cue was found annoying by several participants and some even reported that they got a slight headache or felt discomfort in their eyes when trying to focus on the needle. The transparency cue made grabbing the needle difficult for some participants. This is to be expected since the majority relied on the occlusion cue as described above.

The participants would primarily use one of two strategies when placing the needle. The first strategy was to drag the needle into the target with little focus on using the tip of the needle. The second strategy resembles the way surgeons handle a needle: transferring the needle between the tools to angle it correctly before moving it to the target location. This allows for precision with a trade-off in completion time.

Most participants found the controls intuitive giving it a mean score of 6 on a scale from one to seven. In total, 20% of the participants gave a score 5 or below.

5 EXPERT EVALUATION

We conducted an expert review to determine whether the solution is usable in an operational environment. It was evaluated during a training session at Minimal Invasiv Udviklings Center in Aalborg. The evaluations followed a simple structure of demonstrating the system followed by an open interview between observations. The demonstrations included the experts trying out the solution. Four trainees and three instructors were present during the evaluation other than Johan Poulsen and Jane Petersson. Two of the trainees were experienced and two were relatively new with the da Vinci Surgical System. All participants had a chance to observe and comment on the system in use, however, only the instructors had a chance to control the virtual instruments.

5.1 Prototype in Operational Environment

We developed a prototype to be used in a real training environment with the overlaying system from [6]. The system was altered to include the findings from Section 4.2. This means that only the highlight cue was available. The buttons on the Razer Hydra were mapped to key functions, such as clutch, reset position, and disable overlay. The depth cue was also toggle-able. The system was integrated with the overlay system presented in [6] and included two virtual instruments and a needle.

5.2 Findings

A common observation during the evaluation was the speed with which the experts learned to use the controls. This was generally faster than the participants in the depth cue evaluation. This indicates that the controls are similar to the da Vinci console and is supported by the interviews in which they stated that the controls were intuitive to them. The RAMIS surgeons agreed that the solution is good at visualizing specific orientations and actions of the tools and would reduce the amount of interruptions during a training session. Johan Poulsen stated that the highlight cue was beneficial when estimating when to grab. Both trainees and instructors saw potential in the solution and agreed that the specific visual assistance would be helpful in training sessions, both real and simulated. This extends to human surgery where new RAMIS surgeons require supervision in procedures that were not taught in training sessions.

The solution was easy to use and depth was easy to perceive even on 2D monitors, however, the latency meant that the needle was increasingly difficult to work with. This means that the RAMIS surgeons would prefer to use the system without the needle in its current form.

To determine the motion-to-photon latency of the virtual instruments, we extended the system described in [8]. We used a microcontroller and an optocoupler to trigger a button on the Razer Hydra and used a phototransistor to detect a resulting change on different monitors of the system. The latency of the virtual instruments were 249 ms on the telestration monitor, 193 ms on the console's displays, and 77 ms on the regular computer monitor, which only displays the virtual instruments.

6 DISCUSSION

The quantitative experiment defines performance as a combination of three categories; time, errors, and accuracy. Only one of these were found to be significantly different between depth cues when compared to each other. One reason that the accuracy data is not significant could be due to the way the system calculated distance to the target. The target in the test environment is two dimensional and the participants were asked to place the tip of the needle in the middle of the target. However, for the test to continue, the program had to register a collision between the target surface and the needle's tip. This resulted in the participant having to clip through the surface, effectively decreasing accuracy, on purpose. Unfortunately, this was not detected in the pilot tests. A more accurate solution would be to project the tip of the needle to the target surface before calculating the offset distance.

A factor that might affect the validity of the findings is that 76.6% of the participants were students of media technology. This could have an impact on the performance since they were more familiar with VR tools (3.68 vs. 3.41 familiarity with VR tools on a self-reported 1-5 scale).

The expert evaluation is based on the opinions of three experts and the comments from trainees. This means that results are subjective and not entirely reliable. An observation of the system in use during a number of training sessions would have been preferable to ascertain their experiences. A larger sample of experts across similar contexts would quantify the results, improving reliability. The training sessions at Minimal Invasiv Udviklings Center have tight schedules for the trainees and instructors, meaning they have had a short amount of time to learn the controls of the solution. Testing in a more relaxed environment might have given a more in-depth evaluation.

From the expert evaluation we found that the motion-to-photon latency was too high when viewed on the standard monitors of the da Vinci system. The measurements showed that the latency is significantly smaller when using a regular PC monitor to view the virtual instruments. However, as the video signals never enter system memory in the current setup, it is not possible for the surgeon to e.g. point using only the PC monitor. Newer generations of the video grabber cards that we use in the system are able to simultaneously overlay and input the signals (e.g. Blackmagic Design DeckLink Quad 2), which allows for showing the augmented video stream with minimal latency on a regular PC monitor.

7 CONCLUSION AND FUTURE WORK

Robot-assisted minimally invasive surgery is an increasingly popular method of operating with many advantages compared to open surgery. One of the disadvantages is the cost and duration of training new surgeons capable of operating the da Vinci Surgical System. This work studied possibilities of improving training by overlaying the video streams with virtual instruments and introducing additional depth cues to assist instructors in perceiving the depth of the virtual instruments.

The test showed a visible learning curve in performance, specifically in the mean total task time. This indicates that practice rather than depth cues is most important for user performance. Although training efficiency is the priority, we cannot avoid the need for additional practice with this solution.

The experiment showed a significant difference in the total task time between the highlight and transparency cues. Additionally, it showed a tendency for transparency and to some degree parallax to have a negative impact on user performance. The participants expressed preference for the highlight cue.

The expert evaluation showed a fast learning curve for the instructors when using the solution for the first time, indicating that the controls are similar to those of the robotic surgical system. The experts stated that the highlight cue was beneficial in conjunction with the intrinsic depth cues. The expert evaluation also indicated that transparency was not currently a necessity and that the virtual tools are not interrupting the workflow.

In future work, the highlight cue should be made to fade between colors to provide the instructors with more gradual depth information. The parallax cue showed a tendency to affect user performance negatively and it might be relevant to look into using head tracking to test a more natural version of the cue. In future evaluations, the training period where participants familiarize themselves with the controls and virtual instruments should be longer to counter the significant increase in performance over the first 10-15 measurements.

ACKNOWLEDGMENTS

We thank Minimal Invasiv Udviklings Center and Aalborg University Hospital for their continuous support and collaboration in our shared work on improving training for robot-assisted minimally invasive surgery.

The authors would also like to thank the anonymous reviewers for their valuable comments and critique.

REFERENCES

[1] Mohamed R. Ali, Jamie P. Loggins, William D. Fuller, Brian E. Miller, Christopher J. Hasser, Peter Yellowlees, Tamas J. Vidovszky, Jason J. Rasmussen, and Jonathan Pierce. 2008. 3-D Telestration: A Teaching Tool for Robotic Surgery. *Journal of Laparoendoscopic & Advanced Surgical Techniques* 18, 1 (Feb. 2008), 107–112. https://doi.org/10.1089/lap.2007.0051

[2] William Epstein and Sheena Rogers. 1995. *Perception of Space and Motion.* Elsevier Science, Saint Louis. http://public.eblib.com/choice/PublicFullRecord.aspx?p=4732014 OCLC: 973191204.

[3] Eleanor J. Gibson, James J. Gibson, Olin W. Smith, and Howard Flock. 1959. Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology* 58, 1 (1959), 40–51. https://doi.org/10.1037/h0043883

[4] Anthony M. Jarc, Swar H. Shah, Troy Adebar, Eric Hwang, Monish Aron, Inderbir S. Gill, and Andrew J. Hung. 2016. Beyond 2D telestration: an evaluation of novel proctoring tools for robot-assisted minimally invasive surgery. *Journal of Robotic Surgery* 10, 2 (June 2016), 103–109. https://doi.org/10.1007/s11701-016-0564-1

[5] Olaf Kelle. 1995. *Dynamische Tiefenwahrnehmungskriterien in computergenerierten interaktiven Szenen und virtuellen Simulationsumgebungen* (als ms. gedr ed.). Number 356 in Fortschritt-Berichte VDI Reihe 10, Informatik, Kommunikationstechnik. VDI-Verl, Dsseldorf. OCLC: 75607158.

[6] Martin Kibsgaard and Martin Kraus. 2015. Real-Time Augmented Reality for Robotic-Assisted Surgery. In *The 3rd AAU Workshop on Robotics: Proceedings.* Aalborg Universitetsforlag, 19–23.

[7] Martin Kibsgaard and Martin Kraus. 2016. Pointing with a one-eyed cursor for supervised training in minimally invasive robotic surgery. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9515. Springer, Cham, 12–21. https://doi.org/10.1007/978-3-319-29965-5_2

[8] Martin Kibsgaard and Martin Kraus. 2017. Measuring the Latency of an Augmented Reality System for Robot-Assisted Minimally Invasive Surgery. In *12th International Conference on Computer Graphics Theory and Applications.* SCITEPRESS.

[9] Anja Matatko, Jrgen Bollmann, and Andreas Mller. 2011. Depth Perception in Virtual Reality. In *Advances in 3D Geo-Information Sciences*, Thomas H. Kolbe, Gerhard Knig, and Claus Nagel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 115–129. http://link.springer.com/10.1007/978-3-642-12670-3_7

[10] George Mather. 1997. The Use of Image Blur as a Depth Cue. *Perception* 26, 9 (1997), 1147–1158. https://doi.org/10.1068/p261147 arXiv:http://dx.doi.org/10.1068/p261147 PMID: 9509149.

[11] Florin Octavian Matu, Mikkel Thøgersen, Bo Galsgaard, Martin Møller Jensen, and Martin Kraus. 2014. Stereoscopic augmented reality system for supervised training on minimal invasive surgery robots. In *Proceedings of the 2014 Virtual Reality International Conference on - VRIC '14.* ACM Press, New York, New York, USA, 1–4. https://doi.org/10.1145/2617841.2620722

[12] M. E. Ono, J. Rivest, and H. Ono. 1986. Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology. Human Perception and Performance* 12, 3 (Aug. 1986), 331–337.

[13] Ontario, Ministry of Health and Long-Term Care, and Medical Advisory Secretariat. 2011. *Robotic-assisted minimally invasive surgery for gynecologic and urologic oncology: an evidence-based analysis.* Medical Advisory Secretariat, Ontario Ministry of Health and Long-Term Care, Toronto, Ont. http://www.deslibris.ca/ID/227830 OCLC: 732604098.

[14] Branka Radojcifj, Radoica Jokifj, Slobodan Grebeldinger, Igor Meljnikov, and Nikola Radojifj. 2009. [History of minimally invasive surgery]. *Medicinski Pregled* 62, 11-12 (Dec. 2009), 597–602.

[15] Rebecca Randell, Stephanie Honey, Natasha Alvarado, Alan Pearman, Joanne Greenhalgh, Andrew Long, Peter Gardner, Arron Gill, David Jayne, and Dawn Dowding. 2016. Embedding robotic surgery into routine practice and impacts on communication and decision making: a review of the experience of surgical teams. *Cognition, Technology & Work* 18, 2 (2016), 423–437. https://doi.org/10.1007/s10111-016-0368-0

[16] Brian Rogers and Maureen Graham. 1979. Motion parallax as an independent cue for depth perception. *Perception* 8, 2 (1979), 125–134. https://doi.org/10.1068/p080125

[17] Robert J. Teather and Wolfgang Stuerzlinger. 2013. Pointing at 3D Target Projections with One-eyed and Stereo Cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13).* ACM, New York, NY, USA, 159–168. https://doi.org/10.1145/2470654.2470677

[18] Francesco Volonté, Nicolas C Buchs, François Pugin, Joël Spaltenstein, Boris Schiltz, Minoa Jung, Monika Hagen, Osman Ratib, and Philippe Morel. 2013. Augmented reality to the rescue of the minimally invasive surgeon. The usefulness of the interposition of stereoscopic images in the Da Vinciffl robotic console. *The International Journal of Medical Robotics and Computer Assisted Surgery* 9, 3 (2013), e34–e38.