

# Teleoperating Robots from Arbitrary Viewpoints in Surgical Contexts

Mark Draelos, Brenton Keller, Cynthia Toth, MD,  
Anthony Kuo, MD, Kris Hauser, PhD, and Joseph Izatt, PhD

**Abstract**—Intraoperative 3D imaging has great potential for enhancing surgical visualization. This is especially so in ophthalmic surgery where live volumetric imaging from optical coherence tomography systems recently incorporated into surgical microscopes has freed surgeons from the otherwise universal top-down viewpoint. New viewpoints, however, disorient surgeons when directions of their hand motions and viewed tool motions do not align. We propose introducing a robotic surgery system to decouple surgeons’ hands from their tools and ensure that viewed tool motions align in arbitrary viewpoints. We present a framework entitled Arbitrary Viewpoint Robotic Manipulation (AVRM) which governs how hand and tool motions should interact to minimize disorientation and thereby enable operations from desirable but previously untenable viewpoints. A crossover study in which 20 subjects completed mock surgical scenarios with an AVR-M testbed system demonstrated that arbitrary viewpoints do not improve task performance unless automatic hand-tool misalignment correction is provided. When provided together with arbitrary viewpoints, automatic hand-tool misalignment correction reduces task completion time by 50% on average compared to a fixed top-down viewpoint.

**Index Terms**—Teleoperation, medical robotics, compensation

## I. INTRODUCTION

Visualization is key in surgery. Adequate tissue exposure promotes safety and efficiency whereas visual obscuration leads to long and difficult cases. In ophthalmic surgery especially, the universal top-down view of modern surgical microscopes deprives surgeons of visualization in key areas such as the peripheral retina, layered retinal tissues, and tissue immediately posterior to the iris (Fig. 1A). Real-time 3D imaging modalities such as optical coherence tomography (OCT), 3D ultrasound, and, possibly, fast MRI/CT have great potential for enhancing surgical field visualization. The 3D images they capture enable the surgeon to “see” from otherwise physically impossible yet advantageous perspectives. This has already been demonstrated in ophthalmic surgery with intraoperative OCT guidance (Fig. 1B) [1]–[5]. Moreover, real-time 3D imaging permits rapid and seamless viewpoint changes during procedures.

Operating from new viewpoints, while desirable, has the potential to disorient surgeons if changes in viewpoint are not matched by similar changes in their motor frame (Fig. 2A–B) [6], [7]. (The surgeon’s motor frame is their mental

representation of their body’s configuration in space.) These new potential viewpoints introduce misalignment between surgeons’ hands and their view such that *routine hand motions cause non-intuitive viewed tool motions*. Surgeons must consciously correct for these non-intuitive tool motions, losing the speed and efficiency they have acquired from years of training, especially when large mental rotations are involved [8]. Automatic compensation to align surgeons’ visual and motor frames is thus desirable to minimize disorientation when capitalizing on new viewpoints (Fig. 2C). Otherwise, “manual” correction severely reduces the utility of intraoperative 3D imaging.

Alignment of hand and tool motions avoids disorienting surgeons because mental rotations are eliminated. This approach, however, requires that a surgeon’s hands be decoupled from their tools, because hand motions in a given direction may require tools to move in yet another direction, depending upon the viewpoint. We propose introducing a robotic surgery system between the surgeon and their tools that performs this decoupling and automatically compensates for hand-tool misalignments (Fig. 3). Such a system would record hand motions via haptic input devices and transform them into robotic tool motions in the surgical frame of reference, thereby avoiding disorientation and restoring surgeon efficiency. Rather than directly manipulating surgical tools, the surgeon would use a virtual “wand” that moved with the viewpoint to represent their haptic input for each hand

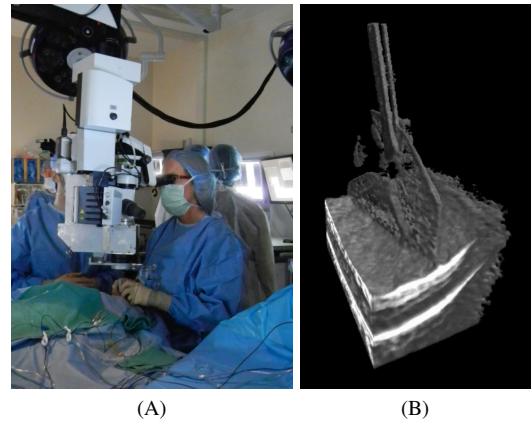


Fig. 1. (A) Traditional surgical microscope configuration for ophthalmic surgery. The microscope is suspended over the patient’s eyeball and can only provide an overhead view. (B) Example microscope-integrated OCT surgical field visualization of an epiretinal membrane peel with forceps. This side rendering reveals key tool-tissue interactions at the peeling site which are difficult if not impossible to visualize from above.

\*This research is partially supported by NIH R01-EY023039, NIH F30-EY027280, and NIH T32-GM007171.

M. Draelos, B. Keller, and J. Izatt are with the Department of Biomedical Engineering, Duke University, Durham, NC, USA. K. Hauser is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. C. Toth and A. Kuo are with the Department of Ophthalmology, Duke University Medical Center, Durham, NC, USA. mark.draelos@duke.edu

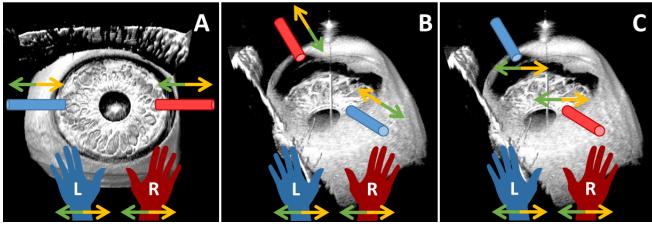


Fig. 2. Conceptual illustration of a view compensation usage case in the anterior segment of the eye with two tools (blue and red cylinders) controlled with the surgeon's matching hands (blue and red hands). (A) Traditional surgical configuration where left/right hand motions (green/yellow arrows) induce left/right tool motions. The surgeon remains oriented because hand and tool motions are intuitively connected. (B) After OCT view rotation to enhance visualization, the tools have become misaligned such that left/right hand motions induce non-intuitive angled tool motions. Additionally, the left hand now controls the rightmost tool and vice versa due to a rotation beyond  $180^\circ$  about the vertical axis. The surgeon's efficiency is consequently decreased. (C) The view compensation system has corrected tool motions and, after notifying the surgeon, swapped the tools such that left/right hand motions (controlling haptic input devices) induce left/right tool motions (effected by robotic manipulators) of the appropriate tool. The surgeon continues the procedure from the new visualization perspective without interruption or awkward motions.

in a 3D visualization. Under this scheme, surgeon control input would originate in a coordinate system that follows the dynamic view rather than the stationary operative field. The surgeon would thus be free to change viewpoints as ergonomics or the case demands. We term this approach Arbitrary Viewpoint Robotic Manipulation (AVRM).

AVRM is an extensible framework for governing how surgeon hand motion (or more generally, any sufficiently dexterous input) should influence tools to minimize disorientation. This framework is most notable in that it allows surgeons to operate from radically unconventional viewpoints (e.g., upside-down) with comparative ease during robotically-assisted procedures guided by intraoperative 3D imaging. A crossover study in which 20 subjects completed mock surgical scenarios from a fixed top-down view and again with an unrestricted view demonstrated that arbitrary viewpoints do not improve task performance *unless hand-tool misalignment is automatically corrected*.

## II. RELATED WORK

Several groups have investigated techniques for reducing mental workload when teleoperating robots. DeJong et al. described minimizing mental rotation in a task where users teleoperated a robot arm using multiple camera views [8]. They found that arranging monitor views with the user positioned along the monitor centerlines improved task efficiency, provided that the input device and robot end-effector coordinate systems were aligned. Similarly, Hiatt et al. defined three specific classes of coordinate systems, including a viewpoint-based one, for a teleoperation task [9]. Their experiments showed that although users considered the viewpoint-based coordinate system to require the lowest mental effort, this coordinate system had the largest task completion times. Both of these works use stationary cameras and/or display alignment rather than arbitrary viewpoints.

Several groups have explored synthesizing new viewpoints of surgical fields. Korrea et al. described an image transformation system that displays an existing endoscopic scene from an alternate viewpoint [10]. Their technique relied upon estimating a surface from stereoscopic endoscopic images and then projecting that surface from a different perspective. Bichlmeier et al. described an augmented reality system in which a surgeon interacts with anatomical data through a virtual mirror [11]. By moving the physical handle, the surgeon could view anatomical data in the augmented reality world through the mirror, peering around corners. Unfortunately, the transformation of control input into the new viewpoint frame is beyond the scope of these works.

Intuitive Surgical's da Vinci Surgical System implements a form of camera-relative control using the "substantially connected" concept [12]. An input device and a robotic tool are considered substantially connected if their directions of incremental motion are aligned within a given tolerance for a given viewpoint [13]. The da Vinci surgeon console, however, incompletely performs hand-tool decoupling because the console masters force the surgeon's hands to follow tool orientations, especially for scissors, forceps, and other graspers (e.g., needle drivers). This limits achievable tool orientations to possible (and from a practical standpoint, ergonomic) hand orientations so that operating from viewpoints at large angular differences from a given tool is difficult. Complete hand-tool decoupling with a virtual wand eliminates this difficulty by allowing the surgeon to control tools from any viewpoint without exceeding their hands' ergonomic workspace. Furthermore, obtaining arbitrary viewpoints with the da Vinci patient cart requires arbitrary positioning of the laparoscopic camera, which is impossible for extreme viewpoints given the constraints on physical camera placement. Reliance on 3D imaging overcomes this restriction because changing viewpoints only requires modifications to the rendering modelview and projection matrices; the physical imaging system need not move necessarily. Operating from physically impossible viewpoints (from an image capture standpoint) thus becomes possible.

Many robotic surgery systems presently available or under development can incorporate the AVRM concept. Such systems need hand-tool decoupling and sufficient dexterity to articulate the surgical tool. For the eye alone, there are several such systems. Examples include the PRECEYES company [14] and the ARMA group at Vanderbilt [15] for vitreoretinal surgery. Bourla et al. demonstrated that the da Vinci Si patient cart has sufficient precision for human anterior segment surgery [16] despite previously reported challenges with intraocular surgery [17]. Notably, the requirement for hand-tool decoupling excludes strictly cooperative or handheld robotic surgery systems. Such systems cannot perform large-scale motions independent of the surgeon's hands without compromising their operation.

## III. FRAMEWORK

Our Arbitrary Viewpoint Robot Manipulation (AVRM) framework decomposes the view compensation problem into

interactions between three concepts: cameras, wands, and tools. Each concept embodies a rigid-body transformation, which will be expressed as a homogeneous transformation matrix, that varies as a function of time  $t$ . The framework specifies how motion of each concept influences the others to avoid hand-tool misalignment.

To keep the surgeon oriented, the AVR-M framework must be updated at minimum before each visualization frame render, because camera motion, the primary source of disorientation, occurs between frames. In practice, the AVR-M framework is updated much faster than the visualization framerate because haptic input devices and robotic arms typically refresh faster than 100 Hz. The sections that follow describe the rules for how camera, wand, and tool motions relate.

#### A. Camera

The camera  $C$  describes the viewpoint (or vantage pose) from where the surgeon views the operating field. Commonly, this is the inverse modelview matrix with which the intraoperative 3D imaging is rendered. At the start of a camera viewpoint change, the initial camera vantage pose is recorded at  $t_0$ . The misalignment at a subsequent time  $t$  is

$$D(t) = C(t)C(t_0)^{-1}, \quad (1)$$

which represents the relative camera motion.

Frequently, it is desirable to constrain relative camera motion along or about specific translational or rotational axes, respectively. For example, the surgeon may wish to only compensate camera motion about the vertical axis. In such cases,  $\tilde{D}(t)$  is the constrained version of  $D(t)$  which effects a zero translation or rotation for the desired axes. For translation constraints, the translation component of  $D(t)$  is extracted and the appropriate displacement is zeroed. For rotation constraints, the orientation component of  $D(t)$  is extracted as a quaternion and the appropriate quaternion

unit coefficients are zeroed. After computing constraints, the translation and orientation components are re-composed to form  $\tilde{D}(t)$ .

#### B. Wand

A wand is a virtual representation of a haptic's stylus that exists only in visualization. We consider hand motions captured by the haptic as direct control inputs to the wand rather than the surgical tool, thus decoupling hand and tool motions. This provides a natural way to maintain surgeon orientation during viewpoint changes because the wand moves only in response to hand motions. Viewpoint changes will not affect the wand position because the surgeon's hands remain stationary. The wand concept is fundamental in that it provides an intuitively-connected and viewpoint-invariant link between the surgeon's hands and the visualization.

Each wand is described by its tip pose  $W$  (position and orientation) and its pivot  $P$ , both in the operating field coordinate system, where  $P$  is a translation-only pose representing the wand pivot point during camera viewpoint changes. Conceptually, the wand embodies the view transformation  $V$  of its haptic input pose  $H$  such that

$$W = VH. \quad (2)$$

Thus, the view transformation directly relates haptic stylus motion to wand motion. During camera motion, the wand records its initial pose  $W(t_0)$  and initial haptic input  $H(t_0)$  and updates  $V$ .

$$V(t) = P\tilde{D}(t)P^{-1}V(t_0) = P\tilde{D}(t)P^{-1}W(t_0)H(t_0)^{-1} \quad (3)$$

$$W(t) = V(t)H(t) = [P\tilde{D}(t)P^{-1}]W(t_0)[H(t_0)^{-1}H(t)] \quad (4)$$

The net effect is that the initial wand pose  $W(t_0)$  is rotated according to the relative camera motion  $\tilde{D}$  about the axis-aligned pivot pose  $P$ . Subsequent haptic input  $H(t)$  influences the wand relative to  $W(t_0)$ . Thus, the wand moves with the camera about the pivot while responding to haptic input. This corrects hand-tool misalignment.

Although strictly arbitrary,  $P$  is most usefully chosen as the camera center point (for an orbit-style camera) or the wand tip. If  $P$  is chosen as the camera center point and the camera motion is about  $P$ ,  $P\tilde{D}(t)P^{-1}W(t_0)$  is the initial wand pose pivoted about  $P$ . Thus, subject to haptic motion, the wand will appear stationary in the surgeon's view as desired. Similarly, if  $P$  is chosen as the wand tip,  $P\tilde{D}(t)P^{-1}W(t_0)$  holds the wand tip stationary in the operating field but maintains the wand's orientation as viewed by the surgeon.

Other relationships between wand and camera motion are possible to correct hand-tool misalignment. For example, camera motion could induce haptic force feedback to move the surgeon's hand with the view. In this paper, we discuss only the above relationship for clarity.

#### C. Tool

A tool is an abstraction of an object, physical or virtual, that the surgeon manipulates via the wand. Physical tools are typically robotically-manipulated surgical tools, whereas virtual tools can take many forms, such as representations for

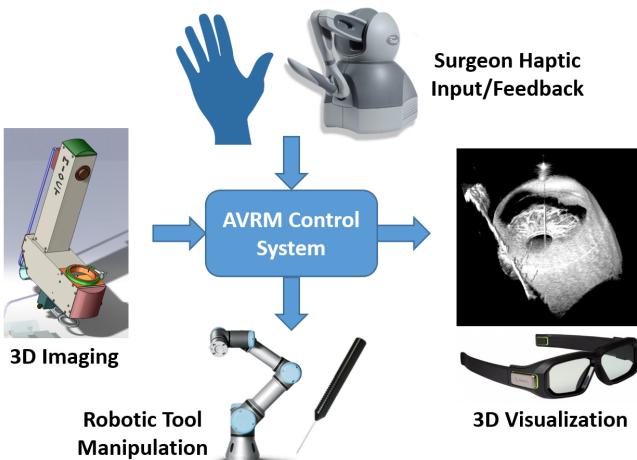


Fig. 3. Concept diagram of AVR-M. The surgeon's handheld tools and surgical microscope view are replaced with haptic input devices and a 3D operative field visualization. The surgical tools are held and manipulated by robots. The AVR-M control system translates the surgeon's intent into robotic tool motion.

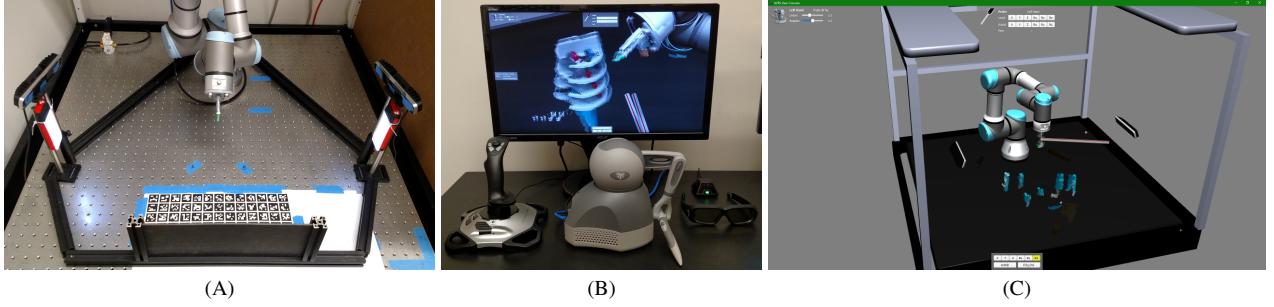


Fig. 4. The surgical system testbed robot workcell (A), surgeon console (B), and example surgeon visualization (C). The visualization includes physically accurate robot and environment models and point cloud display from the RGB-D cameras. The surgeon views the surgical field view using a 3D monitor and controls the surgical tool with the haptic input device. The RGB-D cameras are registered to the world coordinate system using the ArUco grid target.

virtual fixtures and so forth. In the case of physical tools, the tool pose ultimately drives the setpoint for that tool’s robotic manipulator.

Each tool is described by its tip pose  $T$  in the operating field coordinate system. Conceptually, the tool embodies the grab transformation  $G$  of its wand input pose  $W$  such that

$$T = WG. \quad (5)$$

The tool tip thus acts as an extension of the wand. Notably, the tool is not directly affected by camera motion like the wand; camera motion influences the tool pose through wand input. During a mount operation (Section III-D), the tool records its initial pose  $T(t_0)$  and initial wand input  $W(t_0)$  and updates  $G$ .

$$G = W(t_0)^{-1}T(t_0) \quad (6)$$

$$T(t) = W(t)G = [W(t)W(t_0)^{-1}]T(t_0) \quad (7)$$

The net effect is that the initial tool pose  $T(t_0)$  is affected by the wand’s motion from the grab site  $W(t_0)$  in both position and orientation. Once again, there exist other useful relationships between tool and wand motion, such as enforcing a grab site at the tool tip, which we omit for clarity.

Tools with internal degrees of freedom may require additional control input beyond a pose for full operation. For simple tools, such as forceps or scissors, extra control inputs from sliders or buttons on haptic styluses are readily mapped to appropriate joints. More sophisticated tools, such as those with articulated wrists, may require specialized control strategies or external input devices (e.g., foot pedals). In these cases the wand serves a secondary role of selecting the tool to which additional control input is routed.

#### D. Mounts

Mounting is the process by which a particular wand is associated with a particular tool. For an AVR-M system with  $n$  wands ( $W_i$ ’s) and  $m$  tools ( $T_j$ ’s and  $G_j$ ’s), we define the  $n \times m$  mount matrix  $M$  with element  $m_{ij}$  such that

$$T_j = \left( \sum_{i=1}^n m_{ij} W_i \right) G_j \quad (8)$$

with  $m_{ij} \in \{0, 1\}$  and  $\sum_{i=1}^n m_{ij} \leq 1$ . The mount matrix  $M$  thus governs which wands control which tools because the

$j$ th grab  $G_j$  affects the  $i$ th tool  $T_i$  only when  $m_{ij} = 1$ . The constraints enforce that at most one wand controls a given tool at any time. When the surgeon requests that wand  $W_i$  control tool  $T_j$ , the mount matrix is updated such that  $m_{ik} = \delta_{kj}$  for  $k = 1, \dots, n$  and  $m_{kj} = \delta_{ki}$  for  $k = 1, \dots, m$ . Control of tool  $T_j$  is removed from all wands except wand  $W_i$ . When this change occurs, tool  $T_j$  records its new grab transformation.

#### IV. EXPERIMENTAL SETUP

To assess the AVR-M framework, we assembled an AVR-M-enabled surgical system testbed with a robot workcell and operator console for conducting mock surgical procedures (Fig. 4). The workcell included a UR3 robot arm (Universal Robots; Odense, Denmark) which held a soft-tipped compressed air nozzle as the primary tool. Two opposed RealSense SR300 RGB-D cameras (Intel; Santa Clara, CA) registered with an ArUco grid target [18], [19] provided live 3D surface imaging of the workcell’s operative field portion at 30Hz. The console exposed tool and camera control using a six degree of freedom Touch haptic input device (Geomagic; Morrisville, NC) and a three axis Extreme 3D Pro joystick (Logitech; Lausanne, Switzerland), respectively. The joystick also exposed buttons for clutching the haptic stylus position when a workspace boundary was reached. The operator viewed the surgical field with a 3D Vision (NVIDIA; Santa Clara, CA) infrared shutter glasses-based stereoscopic monitor (ASUS; Taipei, Taiwan) refreshing at 60Hz per eye. The control software processed haptic input poses and force feedback at 1kHz and generated tool position-velocity-acceleration commands subject to velocity limits and collision avoidance at 125Hz.

We designed the operator console to replicate the standard operating room setup where the surgeon is seated while holding surgical tools and looking into the surgical microscope (Fig. 4B). The console replaced the surgical tools and microscope with haptic input devices and a stereoscopic display, respectively. The surgeon received the same content as that available in the usual overhead view but now had the freedom to change their viewpoint as needed.

We created two mock surgical scenarios to evaluate the testbed. The “Tubes” scenario consisted of nine angled tubes into which the operator guided the tool tip (Fig. 5A). This

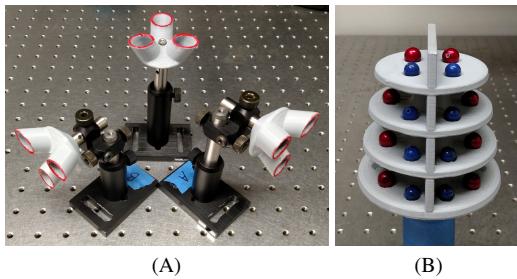


Fig. 5. (A) The “Tubes” scenario rig with nine angled 16 mm projections in sets of three and red highlighted apertures. (B) The multilayered “Beads” scenario rig with eight red (12 mm) and eight blue (10 mm) beads resting in 3 mm deep wells.

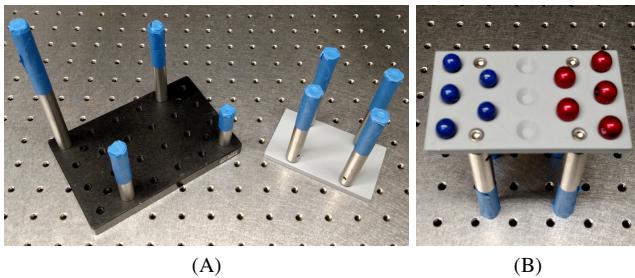


Fig. 6. (A) Structures with targets at varying heights for tool control training. (B) Well plate with beads for compressed air training.

modeled the frequent task of passing instruments through ports or incisions in surgery. The tube apertures were approximately twice the tool diameter. The “Beads” scenario required the operator to dislodge sixteen beads from a multilayered structure using compressed air at the tool tip (Fig. 5B). This exercised the precise tool control required in surgical practice. We designed both scenarios to motivate substantial viewpoint changes using obvious access directions (e.g., tube angles) and vertically overlapping structures.

In addition, we created two training structures that allowed operators to practice tool control and compressed air use. The first structure presented targets at varying heights for the operator to practice touching with the tool tip (Fig. 6A). The second structure presented beads for the operator to practice dislodging with compressed air (Fig. 6B). These structures facilitated operator training without compromising the diagnostic utility of the mock surgical scenarios.

## V. METHODS

We designed a two-group crossover user study to assess the efficacy of our AVR framework’s hand-tool misalignment correction. Subjects were evenly split into two groups labeled “N-UR” and “N-CR”. In the N-UR group, subjects completed scenarios in the “N” and “UR” view configurations. Similarly, in the N-CR group, subjects completed scenarios in the “N” and “CR” view configurations. In the N configuration, subjects were given an initial top-down view and could only pan and zoom. This modeled top-down microscope visualization without intraoperative 3D imaging. In the UR and CR configurations, subjects were

allowed to tilt and rotate their view; however, only the CR configuration provided automatic hand-tool misalignment correction. Thus, the UR configuration modeled procedures guided by intraoperative 3D imaging without hand-tool decoupling. Other than these differences in view limitations and hand-tool misalignment correction, the framework operated identically in each configuration. For the purposes of this study, we limited misalignment correction to rotations about the vertical axis. We randomized subjects to complete the N or UR/CR configuration first.

Subjects completed the study in two sessions separated by at least one day. This washout period was designed to mitigate the effects of learning between each of the two sessions. During each session, subjects practiced with the training structures, attempted the Tubes and Beads scenarios in that order, and answered survey questions. The training portion applied the UR or CR configurations for the N-UR and N-CR groups, respectively, even if the scenario portion applied the N configuration. We included two time-limited competency checks in the training portion to ensure that subjects had gained a baseline proficiency with the operator console. Subjects who failed to complete each competency check within 75 seconds after multiple attempts were excused from the study. At the end of each session, subjects rated the difficulty of both scenarios on a scale from 1 (easy) to 10 (difficult).

For the Tubes scenario, subjects were instructed to insert the tool tip into each tube and pulse the compressed air. For the Beads scenario, subjects were instructed to displace the beads from their wells without touching them and to finish the eight red beads before starting the eight blue beads. In both cases, subjects were asked to work as fast as possible without rushing. We recorded the elapsed time from the scenario start at which each task (tool tip insertion or bead displacement) was completed. Each individual scenario attempt continued until all tasks were complete or 10 minutes had elapsed. We computed the Mean Task Time (MTT) as the elapsed time of the last completed task divided by the number of completed tasks. Thus, subjects were not penalized for reaching the time limit.

To evaluate the utility of arbitrary viewpoints with and without automatic hand-tool misalignment correction, we compared the mean paired change of MTT and difficulty rating from N to UR/CR configuration for each subject regardless of configuration testing order. We assessed for statistical significance to the  $p < 0.05$  level using the two-tailed paired  $t$ -test for MTT and the Wilcoxon signed-rank test for difficulty rating. This analysis methodology was designed to normalize against subjects’ prior experience (e.g., video games or 3D modeling).

## VI. RESULTS

We recruited 20 subjects from Duke University under IRB protocol approval. Only one subject had prior surgical experience. All subjects passed the two competency checks and attempted both scenarios. Figures 7 and 8 show the paired change in Mean Task Time (MTT) and difficulty rating from

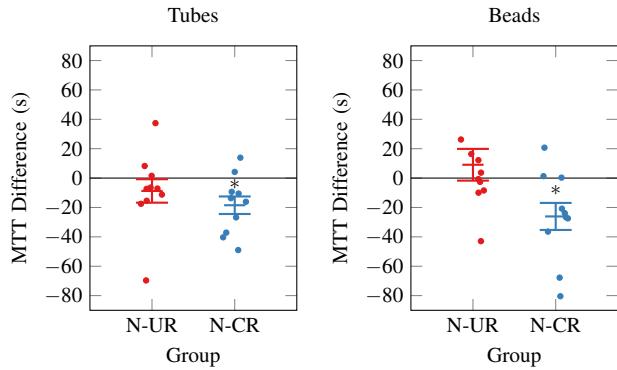


Fig. 7. Mean change in mean task time (MTT) from N to UR/CR configurations across all subjects for each scenario. Improved performance is negative change. Each dot represents the performance change of one subject. Error bars indicate standard error of the mean. Asterisks denote results statistically different from zero.

TABLE I

MEAN PAIRED DIFFERENCE AND SIGNIFICANCE FOR MEAN TASK TIME (MTT) AND DIFFICULTY (D) BY GROUP AND SCENARIO

Group	Metric	Tubes			Beads		
		$\Delta$	$\Delta\%$	$p$	$\Delta$	$\Delta\%$	$p$
N-UR	MTT	-8.8 s	-20 %	0.33	9.1 s	29 %	0.45
	D	-1.2	-18 %	0.23	-1.5	-19 %	0.12
N-CR	MTT	-18.5 s	-50 %	0.017	-26.1 s	-52 %	0.024
	D	-2.3	-37 %	0.028	-3.2	-35 %	0.0078

N to UR/CR configuration for all subjects. For the Tubes scenario, there was a statistically significant reduction in MTT ( $p = 0.017$ ) and difficulty rating ( $p = 0.028$ ) in the N-CR group. Similarly, for the Beads scenario, there was a statistically significant reduction in MTT ( $p = 0.024$ ) and in difficulty rating ( $p = 0.0078$ ) in the N-CR group. Neither scenario produced a statistically significant MTT or difficulty rating change in the N-UR group. Table I lists the mean paired difference, mean paired difference as percent of mean baseline, and significance for each group and scenario.

## VII. DISCUSSION

The data show positive results for AVRMs utility in both scenarios. Most notably, Mean Task Time (MTT) exhibited a statistically significant reduction in the N-CR group but not in the N-UR group. This indicates that the ability to change viewpoints improved average performance *only in combination with automatic hand-tool misalignment correction*. Moreover, the magnitude of this performance improvement is substantial: arbitrary viewpoints with automatic hand-tool misalignment correction (CR) yielded an approximately 50% reduction in MTT compared to the overhead fixed viewpoint (N) on average.

Subjects in the N-UR group did not exhibit this effect with enablement of new viewpoints, however. The distributions in Fig. 7 suggest possible MTT reduction and increase for the Tubes and Beads scenarios, respectively, yet these trends

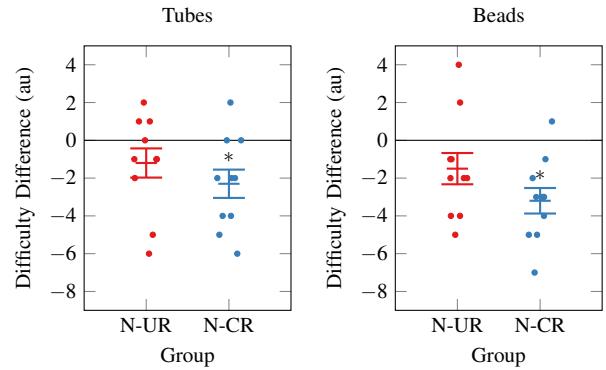


Fig. 8. Mean change in difficulty rating from N to UR/CR configurations across all subjects for each scenario. Improved rating is negative change. Each dot represents the rating change of one subject. Error bars indicate standard error of the mean. Asterisks denote results statistically different from zero.

were not statistically significant. We suspect that subjects in the N-UR group likely resorted to minor viewpoint changes near the initial overhead viewpoint to avoid disorienting tool motions. This would most readily account for the similar performance between the N and UR configurations.

From a subjective difficulty standpoint, subjects once again found scenarios easier with arbitrary viewpoints only in combination with automatic hand-tool misalignment correction. We attribute this to two effects. First, automatic correction reduced the mental transformations necessary to move the tool in the desired direction from a new viewpoint. This is consistent with the previously described increases in cognitive workload as mental rotation angle increases [8]. A mental rotation angle of zero due to automatic correction thus minimizes cognitive workload. Second, the fixed overhead viewpoint of the N configuration challenged subjects to complete scenarios that were designed to work best with varying viewpoints. Arbitrary viewpoints eliminated this difficulty, but the UR configuration added extra cognitive burden due to mental transformations. (The N configuration has no cognitive burden associated with mental transformations because hand-tool motions are initially aligned and cannot become misaligned.) Thus, subjects in the N-UR group did not report a statistically significant reduction in difficulty likely because cognitive burden offset improved viewpoints.

## VIII. CONCLUSION

We have demonstrated the necessity of AVRMs for improving performance of surgically-relevant tasks guided by 3D imaging from arbitrary viewpoints. We hope that incorporation of AVRMs into robotic surgery systems will shorten operative times, reduce surgeons' mental workloads, and potentially improve patient outcomes in those surgical contexts with intraoperative 3D imaging.

## REFERENCES

- [1] Y. K. Tao, J. P. Ehlers, C. A. Toth, and J. A. Izatt, "Intraoperative spectral domain optical coherence tomography for vitreoretinal surgery," *Optics Letters*, vol. 35, no. 20, pp. 3315–7, 2010.

- [2] O. M. Carrasco-Zevallos, B. Keller, C. Viehland, L. Shen, M. I. Seider, J. A. Izatt, and C. A. Toth, "Optical coherence tomography for retinal surgery: Perioperative analysis to real-time four-dimensional image-guided surgery," *Investigative Ophthalmology and Visual Science*, vol. 57, no. 9, pp. Oct37–50, 2016.
- [3] O. M. Carrasco-Zevallos, B. Keller, C. Viehland, L. Shen, G. Waterman, B. Todorich, C. Shieh, P. Hahn, S. Farsiu, A. N. Kuo, C. A. Toth, and J. A. Izatt, "Live volumetric (4D) visualization and guidance of in vivo human ophthalmic surgery with intraoperative optical coherence tomography," *Scientific Reports*, vol. 6, p. 31689, 2016.
- [4] L. Shen, O. Carrasco-Zevallos, B. Keller, C. Viehland, G. Waterman, P. S. Hahn, A. N. Kuo, C. A. Toth, and J. A. Izatt, "Novel microscope-integrated stereoscopic heads-up display for intrasurgical optical coherence tomography," *Biomedical Optics Express*, vol. 7, no. 5, pp. 1711–26, 2016.
- [5] C. Viehland, B. Keller, O. M. Carrasco-Zevallos, D. Nankivil, L. Shen, S. Mangalesh, T. Viet du, A. N. Kuo, C. A. Toth, and J. A. Izatt, "Enhanced volumetric visualization for real time 4d intraoperative ophthalmic swept-source OCT," *Biomedical Optics Express*, vol. 7, no. 5, pp. 1815–29, 2016.
- [6] B. P. DeJong, J. E. Colgate, and M. A. Peshkin, *Mental Transformations in Human-Robot Interaction*, ser. Intelligent Systems, Control and Automation: Science and Engineering. Springer Netherlands, 2011, vol. 1010, book section 3, pp. 35–51. [Online]. Available: [http://dx.doi.org/10.1007/978-94-007-0582-1\\_3](http://dx.doi.org/10.1007/978-94-007-0582-1_3)
- [7] M. A. Menchaca-Brandan, A. M. Liu, C. M. Oman, and A. Natapoff, "Influence of perspective-taking and mental rotation abilities in space teleoperation," in *Proceedings of the ACM/IEEE International Conference on Human-robot interaction*. ACM, 2007, Conference Proceedings, pp. 271–278.
- [8] B. P. DeJong, J. E. Colgate, and M. Peshkin, "Improving teleoperation: reducing mental rotations and translations," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 4. IEEE, 2004, Conference Proceedings, pp. 3708–3714.
- [9] L. M. Hiatt and R. Simmons, "Coordinate frames in robotic teleoperation," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE, 2006, Conference Proceedings, pp. 1712–1719.
- [10] Y. Koreeda, S. Obata, Y. Nishio, S. Miura, Y. Kobayashi, K. Kawamura, R. Souzaki, S. Ieiri, M. Hashizume, and M. G. Fujie, "Development and testing of an endoscopic pseudo-viewpoint alternating system," *Int J Comput Assist Radiol Surg*, vol. 10, no. 5, pp. 619–628, May 2015.
- [11] C. Bichlmeier, S. Heining, M. Feuerstein, and N. Navab, "The virtual mirror: A new interaction paradigm for augmented reality environments," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 9, pp. 1498–1510, Sept 2009.
- [12] G. S. Guthart and J. K. Salisbury, "The Intuitive(TM) telesurgery system: overview and application," in *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 1, 2000, pp. 618–621 vol.1.
- [13] G. Niemeyer, G. Guthart, W. Nowlin, N. Swarup, G. Toth, and R. Younge, "Camera referenced control in a minimally invasive surgical apparatus," Dec. 30 2003, US Patent 6,671,581. [Online]. Available: <http://www.google.com/patents/US6671581>
- [14] M. D. de Smet, S. H. Popma, G. Naus, T. H. C. M. Meenink, M. J. Beelen, and M. Steinbuch, "Robot-assisted choroidotomy and sub-retinal bleb creation," *Investigative Ophthalmology and Visual Science*, vol. 55, no. 13, pp. 2324–2324, 2014. [Online]. Available: <http://dx.doi.org/>
- [15] H. Yu, J.-H. Shen, K. M. Joos, and N. Simaan, "Design, calibration and preliminary testing of a robotic telemانipulator for OCT guided retinal surgery," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, Conference Proceedings, pp. 225–231.
- [16] T. Bourcier, J. Chammas, P.-H. Becmeur, J. Danan, A. Sauer, D. Gaucher, P. Liverneaux, and D. Mutter, "Robotically assisted pterygium surgery: First human case," *Cornea*, vol. 34, no. 10, pp. 1329–1330, 2015.
- [17] D. H. Bourla, J. P. Hubschman, M. Culjat, A. Tsirbas, A. Gupta, and S. D. Schwartz, "Feasibility study of intraocular robotic surgery with the da vinci surgical system," *Retina*, vol. 28, no. 1, pp. 154–158, 2008.
- [18] S. Garrido-Jurado, R. M. noz Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280 – 2292, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320314000235>
- [19] S. Garrido-Jurado, R. M. noz Salinas, F. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern Recognition*, vol. 51, pp. 481 – 491, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320315003544>