



Published in final edited form as:

J Robot Surg. 2015 September ; 9(3): 223–233. doi:10.1007/s11701-015-0520-5.

Augmented reality and cone beam CT guidance for transoral robotic surgery

Wen P. Liu¹, Jeremy D. Richmon², Jonathan M. Sorger³, Mahdi Azizian³, and Russell H. Taylor¹

Wen P. Liu: wen.p.liu@gmail.com

¹Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

²Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins Hospital, Baltimore, MD, USA

³Intuitive Surgical Inc., Sunnyvale, CA, USA

Abstract

In transoral robotic surgery preoperative image data do not reflect large deformations of the operative workspace from perioperative setup. To address this challenge, in this study we explore image guidance with cone beam computed tomographic angiography to guide the dissection of critical vascular landmarks and resection of base-of-tongue neoplasms with adequate margins for transoral robotic surgery. We identify critical vascular landmarks from perioperative c-arm imaging to augment the stereoscopic view of a *da Vinci si* robot in addition to incorporating visual feedback from relative tool positions. Experiments resecting base-of-tongue mock tumors were conducted on a series of ex vivo and in vivo animal models comparing the proposed workflow for video augmentation to standard non-augmented practice and alternative, fluoroscopy-based image guidance. Accurate identification of registered augmented critical anatomy during controlled arterial dissection and en bloc mock tumor resection was possible with the augmented reality system. The proposed image-guided robotic system also achieved improved resection ratios of mock tumor margins (1.00) when compared to control scenarios (0.0) and alternative methods of image guidance (0.58). The experimental results show the feasibility of the proposed workflow and advantages of cone beam computed tomography image guidance through video augmentation of the primary stereo endoscopy as compared to control and alternative navigation methods.

Keywords

Transoral robotic surgery; Video augmentation; *da Vinci*; Cone beam computed tomography; Image-guided robotic surgery

Correspondence to: Wen P. Liu, wen.p.liu@gmail.com.

Compliance with ethical standards

Ethical studies All institutional and national guidelines for the care and use of laboratory animals were followed.

Informed consent Statement of informed consent was not applicable since the manuscript does not contain any patient data.

Conflict of interest Russell H. Taylor presents no conflict of interest.

Introduction

The rising incidence of oropharyngeal cancer related to the human papilloma virus has become a significant health care concern. Both surgical and non-surgical treatment modalities have been advocated [1–3]. Surgical strategy and navigational approaches to excise a tumor with adequate margins are derived from preoperative volumetric data [i.e., computed tomography (CT) and magnetic resonance (MR)]. However, currently the integration of preoperative planning to the surgical scene is conducted as a mental exercise. Furthermore, perioperative positioning for a transoral robotic procedure requires an extended neck, open mouth, and tongue pulled anteriorly, presenting a surgical workspace highly deformed from that of preoperative acquisitions.

In orthopedics [4], laparoscopy [5–7], and other head and neck interventions [8–10] researchers have sought to overcome some of the above limitations by integrating information from medical images through augmented reality. Direct overlay of anatomical information onto existing video sources [7, 11, 12] for surgery have a minimal footprint and natural integration with primary visual displays. Presenting supplementary navigational information to the surgeon directly within the primary means of visualization (i.e., the endoscopic video) has shown to be advantageous in skull base studies [13, 14, 9, 10]. Similar to these efforts, stereoscopic augmented reality has been realized in operating microscopes [15] and transoral robotic surgery [16]. Falk et al. [17] used augmented reality in endoscopic coronary bypass grafting, the fused information was displayed directly within the visual field of the *da Vinci's* surgeon side console. In general, either manual methods [16] or external tracking systems [18, 19] are employed to update augmented reality in image-guided robotic surgery.

Related studies not only demonstrate the potential application of augmented reality [20] from projected virtual scenes, but also the integral role of enhanced depth perception. The role of depth was explored with multi-axial views of preoperative CT [7] and further emphasized when Herrell et al. [21] resected embedded targets from gel phantoms using the *da Vinci* robotic system. Using registered CT augmented with the location of a tracked tool tip researchers achieved a resection ratio closer to the ideal, compared to the resection ratio from procedures without depth-enhanced image guidance.

Image guidance derived from both CT (vasculature) and MR (tumor) preoperative data [5–7, 12] can combine the advantages of different modalities, while intraoperative imaging [11, 22] can capture real-time patient positioning and tissue deformation during surgery. In previous simple target localization experiments [23], an intensity-based algorithm developed by Reaungamornrat et al. [24] is used to deformably register preoperative CT to the perioperative cone beam CT (CBCT), and the deformation field is then used to update the graphical models of the anatomic structures.

The goal of this paper is to explore different methods of augmented reality for in vivo tumor resection, with adequate margin, for transoral *da Vinci*-assisted (Intuitive Surgical, Inc., Sunnyvale, CA) surgery. Critical anatomical structures, including tumor and lingual arteries, are directly segmented from perioperative CBCT angiography (CBCTA). In addition to

stereo video augmentation, we create a novel orthogonal view of the virtual scene in order to explicitly enhance depth perception. Furthermore, colors of tumor margin boundaries dynamically changed to reflect tool position. A series of experiments resecting base-of-tongue mock tumors were conducted on ex vivo and in vivo animal models comparing the proposed workflow for video augmentation to simulated standard of practice and fluoroscopy-based image guidance.

Materials and methods

System overview and workflow

The ideal proposed clinical workflow has the patient positioned in a standard perioperative position. After placement of surface registration fiducials, contrast material is injected to enable visualization of critical vascular oropharyngeal structures while a CBCTA image is obtained. The acquired volumetric data not only captures the deformation of the oral workspace (tongue, neck, mandible) in the operative position, but can serve as the anchor to register multimodal preoperative images and plans as proposed in [23]. The lingual artery, and tumor if visible, are then segmented from the CBCTA using ITK-Snap [25] (www.itk-snap.org) by manual initialization and refined with intensity-based, region growing techniques. Registration fiducials are also manually colocated both in the CBCTA and stereo video camera of the *da Vinci* to establish point-pair correspondence which resolves the initial Euclidean transformation, registering the image volume to the robot. Detailed preoperative planning, including localization of the tumor and adequate margins based on standard diagnostic CT/MR, can be created prior to the operation and registered using similar methods described in [23].

Our research visualization system provides guidance during base-of-tongue resection by overlaying segmented targets (tumor/margins) and the lingual artery directly onto the endoscopic video. This system is implemented by extending the cisst/Surgical Assistance Workstation open-source toolkit [26, 27], developed at the Engineering Research Center for Computer Integrated Surgery (Johns Hopkins University). The augmentation, after initial registration, follows camera kinematics, provided by the *da Vinci* application programming interface with software components as described in [23]. In this work, our contributions in image guidance consist of a new method for following rigid resection motion using intraoperative tracking of custom fiducials and providing orthogonal views of tracked tools relative to the critical data in order to supplement the surgeon's stereo perspective in depth, (i.e., parallel to the camera axis). Both methods are discussed in Sect. 2.3 Image Guidance below.

Phantom models

Ex vivo (EV) porcine tongue phantoms—Ex vivo (EV) excised porcine tongues were used as one type of phantom models. Each specimen was embedded with an 8-mm diameter nitrile sphere as synthetic mock tumor (green in Fig. 1d). Five to eight 3.2-mm diameter nylon spheres (green in Fig. 1c) were affixed to the tongue surface as registration and landmark fiducials. A CBCT (109 kVp, 290 mA, $0.48 \times 0.48 \times 0.48$ mm³ voxel size) was then acquired with the tongue secured onto a flat foam template.

These ex vivo models were used in several experimental scenarios. First, to simulate current standard of practice (control) EV phantoms were used in mock tumor resection without integrated image guidance (i.e., the surgeon was given CBCT to be viewed in offline displays). In a second scenario, EV phantoms were used with video augmentation of the tumor target and tool tracking in order to customize features and settings (i.e., determining color and opacity values for augmented structures and thresholds for tool tracking) for the user interface prior to in vivo experiments.

In vivo (IV) porcine animal phantoms—Additionally, in vivo (IV) porcine animal phantoms were also used in our experiments. To set up, a live pig was placed supine on an operating table (Fig. 1d), anesthetized, catheterized and a tracheostomy tube was placed. The specimen's jaw was opened with a bite block. The animal's tongue was pulled anteriorly with sutures, similar to the positioning of a human patient undergoing robotic base-of-tongue resection.

An angiography was acquired after injection of 40 ml of iodine during a volumetric CBCT scan (90 kVp, 290 mA, $0.48 \times 0.48 \times 0.48$ mm³ voxel size). Two mock tumors (urethane, medium durometer spherical medical balloons, 10 mm diameter) were placed anterior/superior to bilateral lingual arteries (using a radiopaque FEP I.V. catheter (Abbotcath®-T 14G \times 140 mm) in the base-of-tongue (Fig. 1a, b). Balloons were injected with a mixture of 0.5 ml rigid polyurethane foam (FOAM-IT®) and 0.25 ml iodine to retain shape and provide tomographic contrast, respectively. Acrylic paint (0.25 ml) was also added to the filling mixture to provide visual feedback. For the first day of experiments eight 3.2-mm diameter nylon spheres (green surface fiducials in Fig. 1c) were placed on the tongue surface as registration fiducials; during a second set of experiments these were replaced by a custom resection fiducial (Fig. 2 triangular green lattice with inset white, black and yellow spheres).

Image guidance

Video augmentation: critical structures and additional orthogonal view—

During mock tumor resection our proposed guidance with video augmentation superimposed mesh models of critical data segmented from perioperative CBCTA onto to the stereographic projection viewed through the *da Vinci Si* surgeon side console. Mesh models structures were saved as Visualization Toolkit [28] objects and loaded into an OpenGL 3D scene, managed by a *cisst* software component for vision. This component supports stereo endoscopic video capture through a high-end graphics card (Nvidia Quadro SDI) which allowed us to overlay the virtual objects through alpha-blending. Augmentation for EV models included the synthetic tumor and surface fiducials while IV models also included segmented lingual arteries.

To provide navigational information in the axis orthogonal to the camera plane, i.e., depth, we created supplemental camera views of tracked tools within the virtual scene (model meshes of critical segmented landmarks and CBCT slices and volumes). This additional camera perspective, rendered picture-in-picture (Fig. 2, lower inset left) can be dynamically changed, to reflect the surgeon's preference, but was observed to be most useful in the lateral, left-to-right sagittal plane, orthogonal to the primary stereo endoscopy view axis. We

implemented the picture-in-picture display by extending the OpenIGTLink module for Slicer 3D [28] (<https://www.slicer.org>) as a bidirectional socket-based communication interface with the image guidance program. Registration and tool transformations were streamed to the Slicer module, which rendered the tools, CBCT data, and tumor models in the picture-in-picture display.

Margin resection guidance—A sphere (Fig. 2, green sphere, radius = 10 mm) concentric with the mock tumor was included in the augmentation to provide the surgeon with the overlay of an ideal margin for resection. Additional depth information relative to the primary *da Vinci* tool was communicated by both a chromatic changes and explicit distance information. We dynamically altered the color of the margin to indicate the proximity of the tracked tool tip with respect to the ideal resection. A default blue hue changed to green when the tool tip of the primary instrument (5 mm monopolar cautery) was 0–2 mm outside of the margin, then yellow and red when the tip moved within the margin by –2 and –4 mm, respectively. In addition to these chromatic cues, a numeric label on the wrist of the instrument displayed a real-time update of the explicit distance of the tool to the ideal margin boundary.

Tool tracking—Motion of the *da Vinci* tools can be derived using forward kinematics from instrument joint encoders provided by the application programming interface (API). Although the *da Vinci* arms provide very precise incremental motion, they are not designed for stereotactic accuracy. Positions reported by API of the robotic arms have been found to produce as much as 25 mm of error at the tool tip [29]. To correct for this offset we tested two methods:

1. Compute a Euclidean transformation as an initial correction. This is accomplished by recording tool tip locations in stereo video and their corresponding positions as given by the API. The rigid transformation is solved using point correspondence from several tool poses.
2. Continuously derive setup joint corrections using a vision-based solution. A proprietary technique developed by Intuitive Surgical Inc., tracks custom grayscale markers, attached to the shaft of instruments, through the stereo endoscope. For our research we modified this software, which only supported 8 mm instruments, to accommodate 5 mm instruments, used for base-of-tongue procedures and incorporated the correction algorithm within our modular image guidance system architecture.

Resection volume tracking—Superimposed virtual structures were initially rigidly registered by identifying point-based correspondence with spherical fiducials, glued onto tongue surfaces (Fig. 1c) and visible in stereo video and CBCTA. This workflow was tested with simple target localization tasks in previous experiments using ex vivo pig tongue phantoms [30].

We updated the orientation of our resection volume by locating and continuously tracking a custom, rigid, 3D, surface fiducial to address motion created during the surgical resection. Using sutures, the 3D fiducial was attached directly above the resection target (Fig. 2) with

an assumption of a constant spatial relationship between the fiducial and the resected volume of tissue. We updated the overlay of the tumor and margin mesh with the rigid transformation from point-based tracking of colored spheres embedded within the customized fiducial.

This fiducial was fabricated on a 3D printer and designed as a planar right isosceles triangular lattice with a hypotenuse of 10 mm in length. Each corner of the symmetric triangle was connected by a ring with an inner radius of 1.5 mm. The triangular frame (1 mm in width) was painted green, and white, yellow and black 1.6 mm (radius). Teflon spheres were each inserted into the corner rings. Using color thresholds, the green framework of the fiducial was first located as an initial region of interest. Corner rings of the green frame created circular negatives that were segmented using contour detection and matched by their average color to the nylon spheres.

Additional modifications were further explored to increase robustness of this tracking method. Chromatic thresholds, used to locate the surface fiducials, were updated based on the average of successful segmentation and pose recovery. This feature was implemented in order to be dynamically adaptive and robust to fiducial color changes due to cautery artifact.

Fluoroscopy-based augmentation—For comparison with our proposed video-based augmentation, we conducted several tumor resection experiments with fluoroscopy-based image guidance using the Siemen's Syngo workstation. During a second BOT resection during IV experiments the C-Arm was placed laterally to capture sagittal X-rays of IV experiments after docking the robotic arms to the operating table. During these intraoperative fluoroscopic-guided experiments the surgeon side console was located in a radiation-shielded workspace with access to manually activate X-ray on request. Using proprietary Siemens guidance software, the live fluoroscopic images and its overlay onto the CBCTA of the head of the porcine specimens was rendered in 2D in the bottom left and right corners of the surgeon side console through TilePro.

Experiments

The robotic experimental studies conducted were exempt from Institutional Review Board approval at Johns Hopkins Medical Institutions and University. Two individual sets of experiments were conducted on a research *da Vinci* Si console with variations of the proposed image guidance using both EV and IV phantoms. Each set of experiments (variable scenarios summarized in Table 1) included four resections as follows:

- Control with EV phantom (CEV).
- Video augmentation with EV phantom (VEV).
- Fluoroscopy-based augmentation with IV phantom (FIV).
- Video augmentation with IV phantom (VIV).

Two EV specimens, CEV1 and CEV2, were used as control (i.e., preoperative images were available offline and not integrated to the robotic system) in order to simulate current standards of practice. The clinician was given access to view preoperative CBCTs with

visible tumors and surface landmark fiducials on offline monitors displaying the reconstructed volumes in Multi-Planar Reconstruction views. Scenarios VEV1 and VEV2 served to gauge user experience and feedback on proposed features of the video augmentation software on simple EV specimens prior to testing on comprehensive IV models. Experiments comparing video to fluoroscopic augmentation were conducted on IV specimens, which provided a realistic operative workspace. FIV1 and FIV2 both used fluoroscopic augmentation, but differ with the capability of our X-ray system to enlarge up to 4x regions of interest for FIV2.

Video augmentation for experimental set 1 differed from set 2 as follows. For tool tracking, to calibrate for the inherent offset at the remote center of motion, set 1 (VEV1, VIV1) used an initial point-based calibration for corrections while a vision-based technique to track artificial markers was employed for set 2 (VEV2, VIV2). These techniques I and II, used for set 1 and 2, respectively, are as described in 2.3 Image Guidance Tool Tracking. In addition, set 2 also tested our initial implementation for guided margin resection. Augmented overlays of critical data included an ideal margin, updated during intraoperative tracking of a custom resection fiducial.

For evaluation purposes, sub-volumes (mm^3) and ratios of the resected tissues, as specified in Table 2, were measured using thresholds and label maps in Slicer 3D as follows:

Specimen

The volume of the whole resected specimen. Specimen is calculated by measuring the volume of the entire resection.

Margin

The volume of the margin achieved. Margin is calculated by taking the intersection of the ideal margin with the resected tissue concentric with the tumor.

Tumor

The volume of the synthetic tumor. Tumor is calculated by measuring the volume after filtered with an intensity threshold based on the contrast-enhancement of the tumors.

Excess

The volume of excess resection. We measured the depth of the margin as the closest point on the superior surface of the oral tongue to the closest point on the ideal margin. Excess is calculated by subtracting from the specimen the volume of the tumor, the volume of the margin and the ideal cylindrical access [18] required to arrive at the depth of the margin.

Resection ratio

$$\frac{\text{volume of margin achieved}}{\text{volume of ideal spherical margin}}$$

Excess ratio

$$\frac{\text{volume of excess resection}}{\text{volume of the whole resected specimen}}$$

Results

A head and neck surgeon (JDR) proficient in robotic surgery performed all the resections with the goal of achieving a 10 mm margin around the tumor while avoiding and/or controlling the lingual artery.

For both control scenarios (CEV1, CEV2, i.e., tumor resection without integrated image guidance on an EV tongue), the resected specimen did not contain the target mock tumor. All remaining experiments with integrated image guidance led to successful resections of the whole tumor. In the live animal lab cases accuracy of the lingual artery overlays was visually confirmed (Fig. 3, video augmented overlay of exposed lingual dissection) along with successful arterial dissection and control (i.e., no inadvertent hemorrhage) in both video and fluoroscopic augmentation.

Measurements of the specimen resected from all eight robotic experiments are summarized in Table 2. Corresponding postoperative slices/volumes of the resected tumors (blue in ‘Segmentation’ column) and their intersection with an ideal margin (yellow in ‘Segmentation’ column) is demonstrated in Fig. 4.

Resection ratios in order from high to low was achieved with VIV2 (1.00), VEV2 (0.87), VIV1 (0.81), VEV1 (0.71), FIV2 (0.58) and FIV1 (0.44), respectively. Improvements achieved by VIV2 compared to VIV1 (similarly from VEV2 to VEV1) can be attributed to the addition of margin overlay and intraoperative tracking of the resected volume. Scenarios that utilized fluoroscopic overlays (FIV1, FIV2) had the advantage of live, intraoperative X-ray projections of the operative workspace, but updates were restricted to a single (2D) plane.

The **excess ratio** estimates the amount of extraneous tissue resected (i.e., tissue aside from the tumor, margin and estimated required access). This ratio in order from low to high was achieved with VEV2 (0.17), FIV2 (0.29), VEV1 (0.42), VIV1 (0.50), FIV1 (0.54), VIV2 (0.74), CEV1 (1.00), and CEV2 (1.00), respectively. Ratios for CEV1, CEV2 account for the entire volume having not resected the tumor.

In addition to the two forms of resection ratios discussed above, two measurements of accuracy are of interest here: (1) Projection Distance Error—the 2D pixel distance between projected overlay and the true image location of the object; (2) Tool Tracking Error—the 3D position [mm] of the tool tip compared. Mean projection distance error, from point-based manual registration, has been previously established at 2 mm using an anthropomorphic skull phantom [23]. During video-based image guidance for Set 2 visual estimates of tool tracking error (distance of virtual to true tool tip in video) for VIV2, was measured (retrospectively using virtual rulers) to be 5 mm (mean), 10 mm (max).

Discussion

This is a proof-of-concept study that assessed the value of multiple techniques augmenting the surgeon's endoscopic view with CBCTA data with the goal of improving surgical accuracy and optimizing margins for transoral robotic base-of-tongue surgery.

These experiments demonstrate that video-based augmentation (VEV1, VIV1, VEV2, VIV2) achieved superior resection ratio when compared to fluoroscopy-based guidance (FIV1, FIV2) and the control scenario without guidance (CEV1, CEV2). EV phantoms presented an abnormally challenging environment consisting of a featureless tongue volume which led to failed tumor resections in the both the control scenarios. In contrast to a featureless EV phantom, dental, oropharyngeal and neurovascular anatomies serve as landmarks in the IV model. However, using the proposed video-augmentation guidance the surgeon was able to successfully resect both mock tumors despite the challenging environment posed by EV phantoms.

Superior resection ratios achieved by experiments with video augmentation, compared to fluoroscopy-based augmentation, indicates the significance of the method of integration between guidance information and the primary visual field. The fluoroscopic overlays were rendered through TilePro and thus shown below the native stereo endoscopy in a separate window. Informal surveys and similar work for monocular video augmentation in skull base surgery [14] have suggested advantages of guidance through augmentation [31] of the primary, "natural" window rather than having the surgeon to shift his focus between difference sources of information. Improvements in resection ratio on the second set of the experiments using fluoroscopy-based guidance suggest the significance of scale of the augmentation used for image guidance. Set 2 featured an enlarged region of interest at 4 \times , taking advantage of sub-millimeter resolution of 2D X-rays, and is closer to the scale of the endoscopic video.

In addition to resection ratio, we also measured excess ratio. Excess ratio for VIV2 is large since the surface fiducial, used for intraoperative volume tracking was required to be removed with the resection and was attached with a lateral offset with respect to the tumor. Results for excess ratio did not discriminate between our video-based proposal and alternative methods of image guidance. The experimental protocol in this paper did not penalize excess tissue, though the incorporation of this rule, especially in resections that cross the midline of the oral tongue could be included for future experiments. The difficulty lies in experiment setup for in vivo phantoms which requires consistent placement of the tumor since it would be a significant factor in the access required for adequate tumor and margin resection.

Despite encouraging results achieved by the proposed video augmentation system issues of robustness and accuracy remain. During our experiments with intraoperative resection volume tracking vision-based detection of the custom fiducial was susceptible to failure when the fiducial surface was not orthogonal to the endoscope and upon occlusion by tools and debris. To address inaccurate detections due to occlusion we are exploring algorithms, such as Kalman filters, in order to numerically estimate tracking locations based on

observation. This approach is built upon the assumption of a constant spatial relationship between the surface fiducial and resected volume. This limitation, along with the requirement for surface fiducials as part of the workflow, though empirically acceptable for the purposes of our preclinical research experiments, could be replaced with a vision-based reconstruction algorithm.

The 5 mm (mean) tool tracking error is not acceptable for clinical use and can be improved through intraoperative fluoroscopy, using 2D–3D registration to correct for kinematic inaccuracies, tissue deformation and external forces. Comprehensive phantom studies quantifying tracking error and techniques for improvement are currently underway.

Conclusions and future work

Experimental results show the feasibility and advantages of our proposed guidance through video augmentation of the primary stereo endoscopy as compared to control and fluoroscopy-based guidance for transoral robotic surgery. Augmentation from CBCTA can include critical vascular structures and incorporate resection target information (i.e., tumor and desired margin) from preoperative image data. Additionally, guidance provided by this system used augmented reality fusing not only virtual image information derived from volumetric image acquisitions, but also tool localization and other virtual reconstructions in order to present novel enhanced depth information to the surgeon.

Future work will focus on clinical engineering efforts in order ready materials and methods introduced for clinical assessment. For example, we need to determine the appropriate replacement of the wooden blocks and sutures eliminated the need for stainless steel clinical mouth and tongue retractors. Towards this effort, we will continue experiments with 3D printed, radiolucent retractor component replacements, and review advanced reconstruction algorithms [32] with metal noise reduction. Surface fiducials and their placement, though acceptable for our preclinical research protocol and possibly clinical workflows as they were considered similar to vascular clips, could be replaced with a vision-based reconstruction algorithm. During our two sets of experiments we noted that most IV specimens in general required longer dissections and resulted in larger excess ratios as compared to EV due to volumes removed for arterial control and workspace limitations of the transoral access. Evaluation of such costs in time for all steps, and radiation must be substantiated with further experimentation. The small number of experiments presented is an obvious limitation. However, we will conduct further extensive experiments to optimize the workflow such that more TORS surgeons can be included, a more realistic tongue/tumor model will be incorporated, and increased iterations will improve the validity of our model.

Acknowledgments

The authors extend sincere thanks to support provided by Intuitive Surgical Inc., Johns Hopkins, NIH-R01-CA-127444, and the Swirnow Family Foundation. The SAW software infrastructure used in this work was developed under NSF grants EEC9731748, EEC0646678, MRI0722943, and NRI1208540 and under Johns Hopkins University internal funds.

Wen P. Liu, PhD, is a student fellow sponsored by Intuitive Surgical, Inc., Jeremy D. Richmon, MD, is a proctor for Intuitive Surgical, Inc., Jonathan M. Sorger, PhD, is an employee of Intuitive Surgical, Inc., Mahdi Azizian, PhD, is an employee of Intuitive Surgical, Inc.,

References

1. Zhen W, Karnell LH, Hoffman HT, Funk GF, Buatti JM, Menck HR. The National Cancer Data Base report on squamous cell carcinoma of the base of tongue. *Head Neck*. 2004; 26:660–674. [PubMed: 15287033]
2. Weinstein GS, O'Malley BW Jr, Magnuson JS, Carroll WR, Olsen KD, Daio L, Moore EJ, Holsinger FC. Transoral robotic surgery: a multicenter study to assess feasibility, safety, and surgical margins. *Laryngoscope*. 2012; 122:1701–1707. [PubMed: 22752997]
3. Weinstein GS, Quon H, Newman HJ, Chalian JA, Malloy K, Lin A, Desai A, Livolsi VA, Montone KT, Cohen KR, O'Malley BW. Transoral robotic surgery alone for oropharyngeal cancer: an analysis of local control. *Arch otolaryngol-head and neck surg*. 2012; 138:628–634. [PubMed: 22801885]
4. Van de Kelft E, Costa F, Van der Planken D, Schils F. A prospective multicenter registry on the accuracy of pedicle screw placement in the thoracic, lumbar, and sacral levels with the use of the O-arm imaging system and StealthStation Navigation. *Spine*. 2012; 37:E1580–E1587. [PubMed: 23196967]
5. Su LM, Vagvolgyi BP, Agarwal R, Reiley CE, Taylor RH, Hager GD. Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration. *Urology*. 2009; 73:896–900. [PubMed: 19193404]
6. Hughes-Hallett A, Mayer EK, Marcus HJ, Cundy TP, Pratt PJ, Darzi AW, Vale JA. Augmented reality partial nephrectomy: examining the current status and future perspectives. *Urology*. 2013; 83:266–273. [PubMed: 24149104]
7. Volonte F, Buchs NC, Pugin F, Spaltenstein J, Jung M, Ratib O, Morel P. Stereoscopic augmented reality for da Vinci robotic biliary surgery. *Int J Surg case rep*. 2013; 4:365–367. [PubMed: 23466685]
8. Cabriolo I, Sarrafzadeh A, Bijlenga P, Landis BN, Schaller K. Augmented reality-assisted skull base surgery. *Neurochirurgie*. 2014; 60(6):304–306. [PubMed: 25245926]
9. Caversaccio M, Garcia Giraldez J, Thoranaghatte R, Zheng G, Egli P, Nolte LP, Gonzalez Ballester MA. Augmented reality endoscopic system (ARES): preliminary results. *Rhinology*. 2008; 46(2):156–158. [PubMed: 18575020]
10. Caversaccio M, Langlotz F, Nolte LP, Häusler R. Impact of a self-developed planning and self-constructed navigation system on skull base surgery: 10 years experience. *Acta Otolaryngol*. 2007; 127(4):403–407. [PubMed: 17453461]
11. Chen X, Wang L, Fallavollita P, Navab N. Precise X-ray and video overlay for augmented reality fluoroscopy. *Int J Comput Assist Radiol Surg*. 2013; 8:29–38. [PubMed: 22592259]
12. Volonte F, Buchs NC, Pugin F, Spaltenstein J, Schiltz B, Jung M, Hagen M, Ratib O, Morel P. Augmented reality to the rescue of the minimally invasive surgeon. The usefulness of the interposition of stereoscopic images in the da Vinci robotic console. *Int J Med Robotics + Comp Assisted Surg : MRCAS*. 2013; 9:e34–e38.
13. Mirotta, DJ.; Uneri, A.; Schafer, S.; Nithiananthan, S.; Reh, DD.; Gallia, GL.; Taylor, RH.; Hager, GD.; Siewerdsen, JH. High-accuracy 3D image-based registration of endoscopic video to C-arm cone-beam CT for image-guided skull base surgery. In: Wong, KH.; Holmes Iii, DR., editors. *SPIE Medical Imaging*. Vol. 7964. Lake Buena Vista, FL: SPIE; 2011. p. J-79640-J-79610.
14. Liu, WP.; Mirotta, DJ.; Uneri, A.; Otake, Y.; Hager, GD.; Reh, DD.; Ishii, ML.; Siewerdsen, JH. *SPIE Medical Imaging 2012. Image-Guided Procedures, Robotic Interventions, and Modeling*; 2012. A clinical pilot study of a modular video-CT augmentation system for image-guided skull base surgery; p. 8316-8112.
15. Pratt P, Edwards E, Arora A, Tolley N, Darzi AW, Yang G-Z. Image-guided transoral robotic surgery for the treatment of oropharyngeal cancer. *Hamlyn Symposium*. 2012
16. Falk V, Mourgues F, Vieville T, Jacobs S, Holzhey D, Walther T, Mohr FW, Coste-Maniere E. Augmented reality for intraoperative guidance in endoscopic coronary artery bypass grafting. *Surgical technology international*. 2005; 14:231–235. [PubMed: 16525977]

17. Pietrabissa A, Morelli L, Ferrari M, Peri A, Ferrari V, Moglia A, Pugliese L, Guarracino F, Mosca F. Mixed reality for robotic treatment of a splenic artery aneurysm. *Surg Endosc*. 2010; 24:1204. [PubMed: 19826869]
18. Suzuki N, Hattori A, Suzuki S, Otake Y. Development of a surgical robot system for endovascular surgery with augmented reality function. *Stud Health Technol Inform*. 2007; 125:460–463. [PubMed: 17377326]
19. Herrell SD, Kwartowitz DM, Milhoua PM, Galloway RL. Toward image guided robotic surgery: system validation. *J Urol*. 2009; 181:783–789. discussion 789–790. [PubMed: 19091336]
20. Edwards PJ, King AP, Hawkes DJ, Fleig O, Maurer CR Jr, Hill DL, Fenlon MR, de Cunha DA, Gaston RP, Chandra S, Mannss J, Strong AJ, Gleeson MJ, Cox TC. Stereo augmented reality in the surgical microscope. *Stud Health Technol Inform*. 1999; 62:102–108. [PubMed: 10538337]
21. Liu WP, Reaungamornrat S, A D, Sorger JM, Siewerdsen JH, Richmon JD, Taylor RH. Surgery Toward Intraoperative Image-Guided Transoral Robotic. *Robotic Surgery*. 2013; 7:217–225.
22. Badani KK, Shapiro EY, Berg WT, Kaufman S, Bergman A, Wambi C, Roychoudhury A, Patel T. A Pilot Study of Laparoscopic Doppler Ultrasound Probe to Map Arterial Vascular Flow within the Neurovascular Bundle during Robot-Assisted Radical Prostatectomy. *Prostate cancer*. 2013; 2013:810715. [PubMed: 23862066]
23. Liu WP, Reaungamornrat SAD, Sorger JM, Siewerdsen JH, Richmon JD, Taylor RH. Toward intraoperative image-guided transoral robotic surgery. *Robotic Surg*. 2013; 7:217–225.
24. Reaungamornrat S, Liu WP, Wang AS, Otake Y, Nithiananthan S, Uneri A, Schafer S, Tryggestad E, Richmon J, Sorger JM, Siewerdsen JH, Taylor RH. Deformable image registration for cone-beam CT guided transoral robotic base-of-tongue surgery. *Phys Med Biol*. 2013; 58:4951–4979. [PubMed: 23807549]
25. Yushkevich, Paul A.; Piven, Joseph; Hazlett, Heather Cody; Smith, Rachel Gimpel; Ho, Sean; Gee, James C.; Gerig, Guido. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*. 2006; 31(3):1116–1128. [PubMed: 16545965]
26. Deguet A, Kumar R, Taylor RH, Kazanzides P. The cisst libraries for computer assisted intervention systems. *MICCAI Workshop*. 2008 <https://traclcsrjhuedu/cisst/>.
27. Jung MY, Balicki M, Deguet A, Taylor RH, Kazanzides P. Lessons learned from the development of component-based medical robot systems. *software engineering for robotics*. 2014; 5(2):25–41.
28. Pieper S, Lorenson B, Schroeder W, Kikinis R. The NA-MIC kit: ITK, VTK, pipelines, grids, and 3D Slicer as an open platform for the medical image computing community. *Proc. IEEE Intl. Symp. Biomed. Imag*. 2006:698–701.
29. Reiter A, Allen PK, Zhao T. Feature classification for tracking articulated surgical tools. *Med Image Comput Comput Assist Interv*. 2012; 15:592–600. [PubMed: 23286097]
30. Liu WP, Reaungamornrat S, Sorger JM, Siewerdsen JH, Taylor RH, Richmon JD. Intraoperative image-guided transoral robotic surgery: pre-clinical studies. *Int J Med Robot*. 2014; 11(2):256–267. [PubMed: 25069602]
31. Rieger A, Blum T, Navab N, Friess H, Martignoni ME. Augmented reality: merge of reality and virtuality in medicine. *Dtsch Med Wochenschr*. 2011; 136:2427–2433. [PubMed: 22094972]
32. Stayman JW, Otake Y, Prince JL, Khanna AJ, Siewerdsen JH. Model-Based tomographic reconstruction of objects containing known components. *IEEE Trans Med Imaging*. 2012; 31(10): 1837–1848. [PubMed: 22614574]

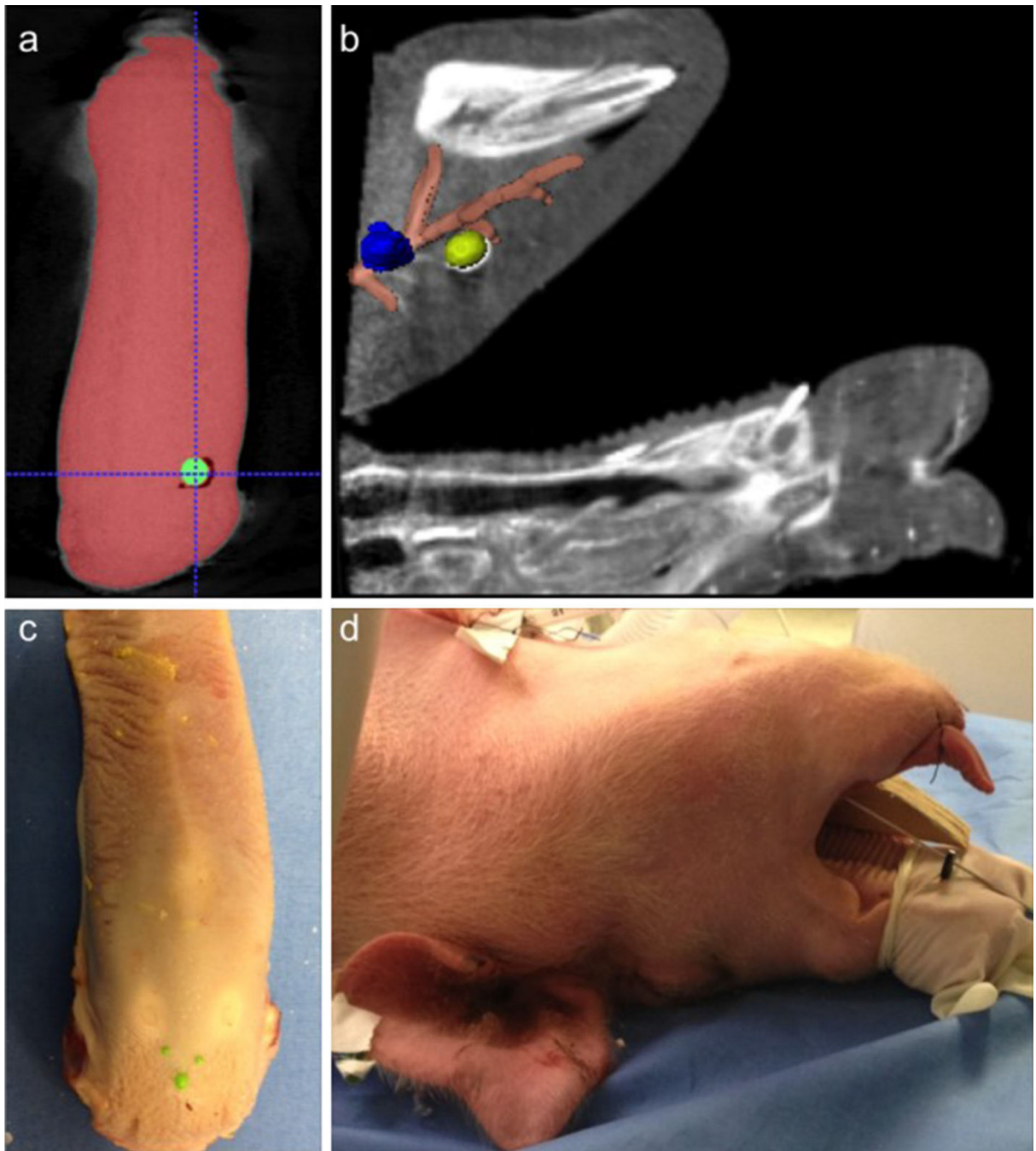


Fig. 1.
a Single axial slice from CBCT of an ex vivo pig tongue phantom with embedded tumor (green). **b** Single sagittal slice CBCTA of an in vivo pig phantom with segmented models of the right lingual artery (orange), and two base-of-tongue tumors (right in yellow, left in blue). **c** Photograph of an ex vivo pig tongue phantom affixed with green registration fiducials. **d** Photograph of an in vivo pig phantom supine and readied for tumor placement

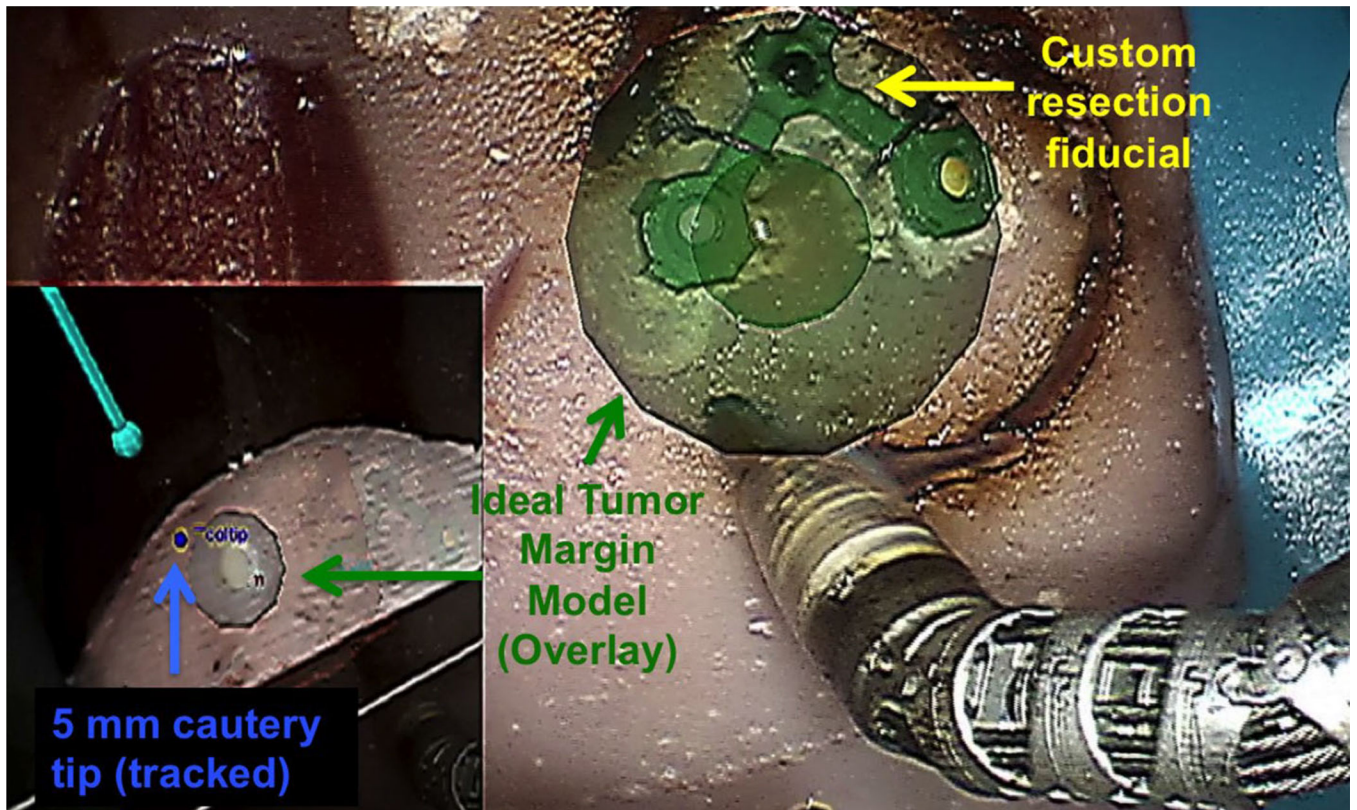


Fig. 2. Screen capture of an ex vivo phantom experiment using video augmentation of margins (the spherical representation of an ideal margin is *green* since tool tip is within +2 mm proximity) and tool tracking in a novel view (*lower left* picture-in-picture) for image guidance

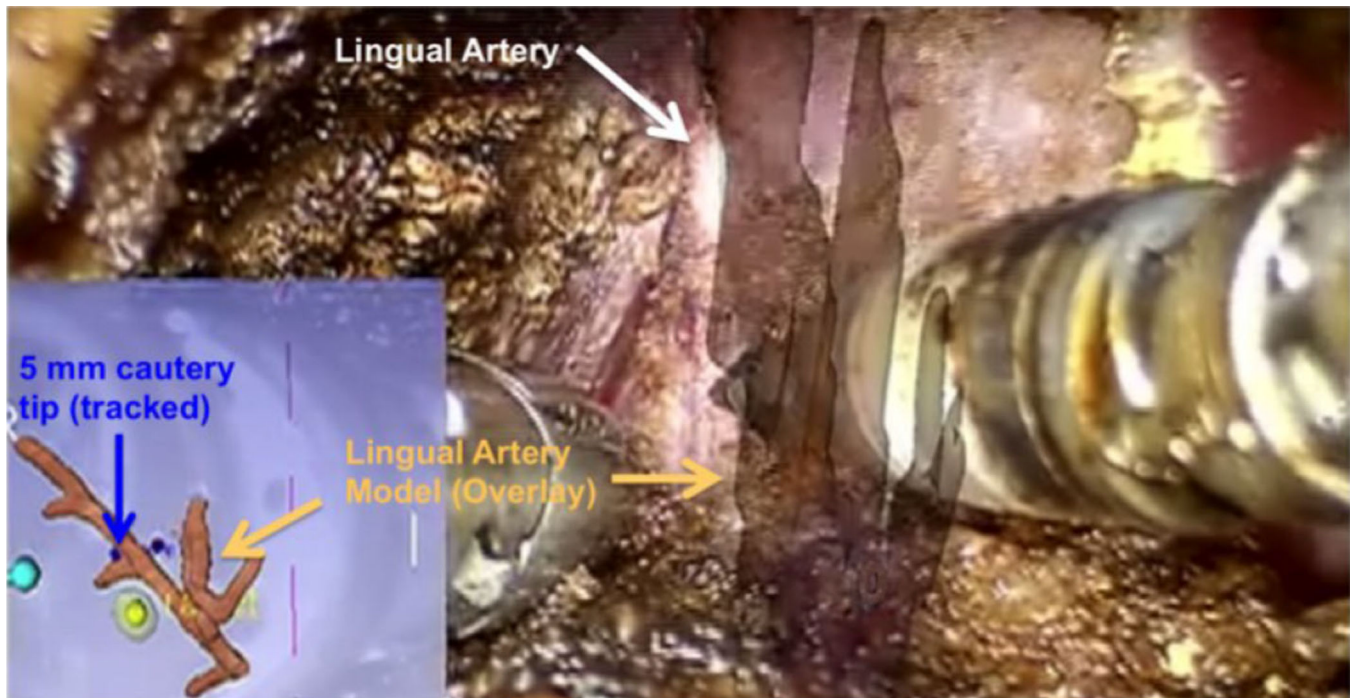


Fig. 3. Screen capture of a lingual artery (*white*) dissection during an in vivo porcine lab experiment using video augmentation as image guidance. The segmented model of the lingual artery is overlaid onto the primary visual field while a lateral virtual view is shown in the *lower left* (picture-in-picture) with localization of the tracked tool tip (*blue*) and tumor/margin model (*yellow*)

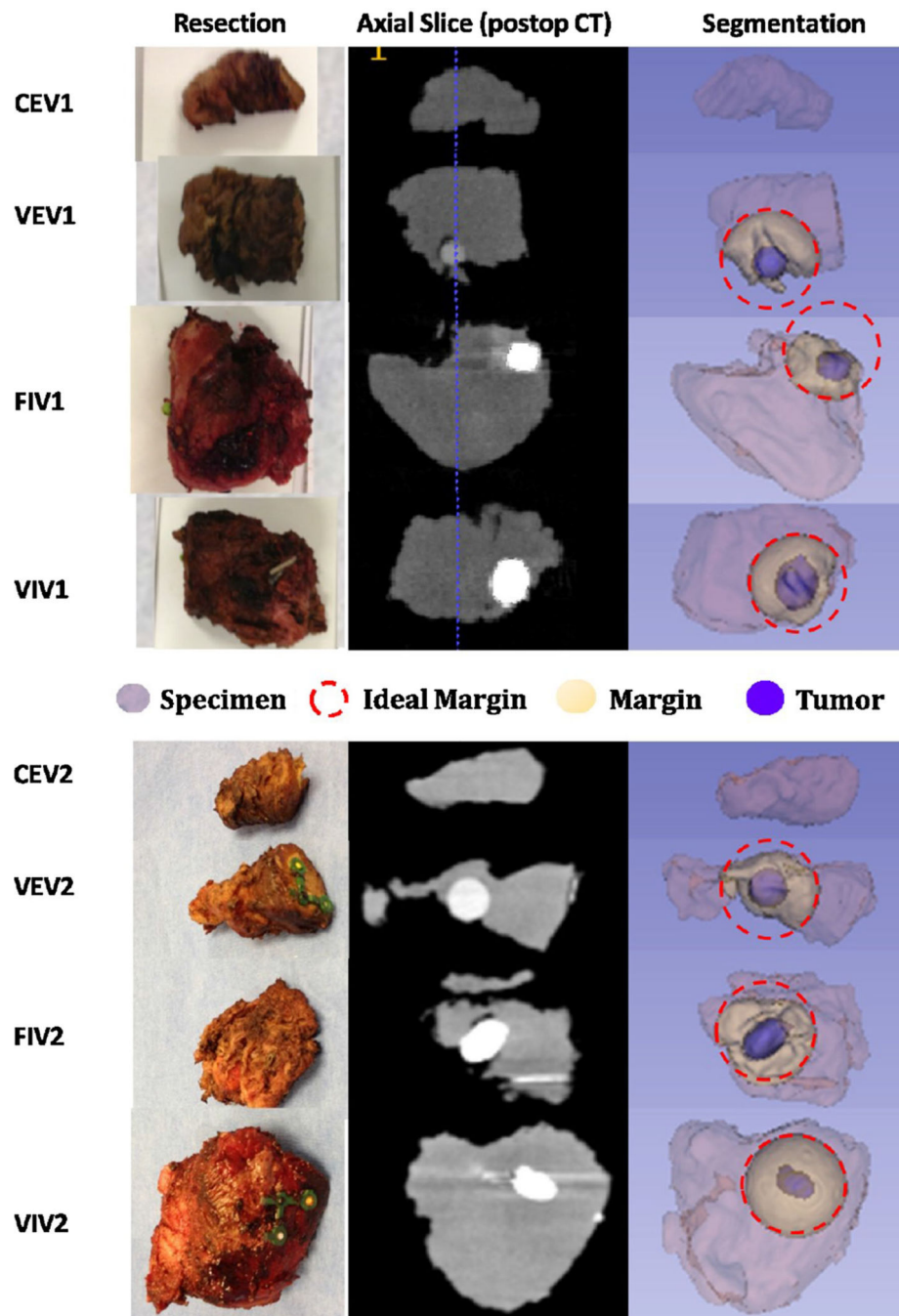


Fig. 4. Resected tumor and margins imaged with postoperative CBCT. Column ‘Resection’ shows a photograph of the resected specimen for each experiment. Column ‘Axial Slice (postop CT)’ is a *grayscale* slice of the CBCT volume with the tumor in white due to added contrast. Column ‘Segmentation’ labels different regions of the specimen as follows: tumor in *blue*, intersection with ideal margin in *yellow*, ideal spherical margin dash-outlined in *red*, and the specimen is *light purple*

Table 1

Variation of image guidance in animal phantom experiments

Experiment model	Image data	Guidance	Margin	Image guidance	
Set 1	CEV1 EV	Periop CBCT	None	No	The control for the first set, simulated current practice with preoperative CBCT shown in offline 2D display, i.e., separate from surgeon's console
	VEV1 EV	Periop CBCT	Video	No	Represented the basic proposed image guidance workflow where stereo video endoscopy was superimposed segmented tumor and fiducials from CBCT along with tracked tools in novel virtual views, picture-in-picture
	FIV1 IV	Periop CBCTA	Fluoro	No	Superimposed fluoroscopy on CBCTA used for guidance in tumor resection and lingual dissection
	VIV1 IV	Periop CBCTA	Video	No	Used video augmentation guidance similar to VEV1, however, with in vivo porcine lab requiring lingual dissection
Set 2	CEV2 EV	Periop CBCT	None	No	The control for the second lab, with the same guidance given as CEV1
	VEV2 EV	Periop CBCT	Video	Yes	Included all features implemented for VEV1, but extended image guidance for margins with resected specimen tracking using a custom fiducial
	FIV2 IV	Periop CBCTA	Fluoro	No	Fluoroscopic guidance, as given in FIV1, with extended capabilities to zoom up to 4× on region of interest
	VIV2 IV	Periop CBCTA	Video	Yes	The same guidance and scenario as VIV1, but extended image guidance for margins with resected specimen tracking using a custom fiducial

Table 2

Results from animal phantom experiments

Experiment	Specimen mm ³	Margin mm ³	Tumor mm ³	Excess mm ³	Margin/ specimen	Excess ratio	Resection ratio	Tumor/ margin	Tumor/ideal tumor
Set 1	CEV1	2959.82	0.00	2959.82	N/A	1.00	N/A	N/A	0.00
	VEV1	8059.67	2960.45	3365.42	0.37	0.42	0.71	0.09	0.49
	FIV1	19602.55	1832.28	10613.50	0.09	0.54	0.44	0.15	0.54
	VIV1	13450.55	3410.59	6756.12	0.25	0.50	0.81	0.17	1.11
Set 2	CEV2	2474.34	0.00	2474.34	N/A	1.00	N/A	N/A	N/A
	VEV2	5485.52	3646.92	943.32	0.66	0.17	0.87	0.08	0.59
	FIV2	6656.81	2429.50	1926.02	0.36	0.29	0.58	0.13	0.62
	VIV2	22287.22	4199.81	16422.70	0.19	0.74	1.00	0.05	0.42

Ideal margin: a sphere of radius = 10 mm

Ideal tumor: a sphere of radius = 5 mm

Excess: specimen - (tumor + margin + cylinder at margin depth)