The International Journal of **Medical Robotics** and **Computer Assisted Surgery**

THE INTERNATIONAL JOURNAL OF MEDICAL ROBOTICS AND COMPUTER ASSISTED SURGERY
Int J Med Robotics Comput Assist Surg (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/rcs.1602

ORIGINAL ARTICLE

Intraoperative image-guided transoral robotic surgery: pre-clinical studies

Wen P. Liu[1]*
Sureerat Reaugamornrat[1]
Jonathan M. Sorger[2]
Jeffrey H. Siewerdsen[1,3]
Russell H. Taylor[1]
Jeremy D. Richmon[4]

[1]*Department of Computer Science, Johns Hopkins University, Baltimore, USA*

[2]*Intuitive Surgical Inc., Sunnyvale, USA*

[3]*Department of Biomedical Engineering, Johns Hopkins University, Baltimore, USA*

[4]*Department of Otolaryngology - Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, USA*

*Correspondence to:
Wen P. Liu, Department of Computer Science, Johns Hopkins University, 112 Hackerman Hall, 3400 N. Charles Street, Baltimore, MD 21218, USA.
E-mail: wen.p.liu@gmail.com

Accepted: 5 June 2014

Abstract

Background Adequate resection of oropharyngeal neoplasms with transoral robotic surgery (TORS) poses multiple challenges, including difficulty with access, inability to palpate the tumor, loss of landmarks, and intraoperative patient positioning with mouth retractor and tongue extended creating significant tissue distortion from preoperative imaging.

Methods This study evaluates a workflow integrating intraoperative cone beam computed tomography (CBCT) for image-guided TORS through robotic experimentation locating 8–10 embedded targets in five porcine tongues and a cadaveric head phantom, conducted under various modes of visualization and integration of preoperative/intraoperative imaging.

Results A statistically significant improvement in mean target localization error was achieved for both the porcine tongue ((9.8 ± 4.0) mm vs. (5.3 ± 1.3) mm, *P*-value = 0.0151) and cadaver ((11.2 ± 5.0) mm vs. (5.8 ± 2.5) mm *P*-value = 0.0189) in experiments comparing scenarios simulating current standard-of-care practice and the proposed image guidance system.

Conclusion Intraoperative image guidance with augmentation of critical surgical structures has the potential to improve target localization for TORS.

Keywords transoral robotic surgery; intraoperative image-guidance; cone-beam CT; stereo video augmentation; base of tongue resection

Introduction

Transoral robotic surgery (TORS) is a minimally invasive surgical approach to resect benign and malignant lesions of the oropharynx. TORS has become a progressively more viable option in addressing a trend of increasing oropharyngeal cancer (1,2). Despite multiple studies demonstrating the safety and feasibility of TORS (3) with excellent functional and quality-of-life outcomes (4) there remain multiple challenges to the resection of selected oropharyngeal neoplasms. This is particularly apparent in base of tongue tumors that may be deeply infiltrative with a poorly defined submucosal margin. Poor exposure of the tumor, a bloody field, inability to palpate the tumor, loss of landmarks and perspective

may all contribute to a challenging and ultimately unsuccessful oncologic resection.

The resulting challenge to reliably delineate tumor margins is compounded by the fact that preoperative imaging of the tongue in repose may not accurately reflect the tumor position when the tongue is extended and deformed with patient positioning during surgery. Such distortion can greatly limit the surgeon's ability to safely obtain negative margins in the highly vascular base of tongue with limited landmarks to guide the dissection. Expert surgeons rely on experience to remain oriented with respect to critical anatomy, even after tissue deformation, by intuitively mapping preoperative data and physical examination findings to the (highly deformed) surgical field. Such practice leaves considerable room for improvement and may lead to compromised margins and ultimately recurrent disease.

We proposed the use of intraoperative cone-beam computed tomography (CBCT) to deformably register key anatomical structures delineated from preoperative diagnostic computed tomography (CT) or magnetic resonance imaging (MRI). These critical structures would include targets of interest (i.e. oropharyngeal neoplasms + negative margins) and functional anatomy (i.e. carotid and lingual arteries). We anticipate that increased accuracy in localization of these key anatomies throughout the surgery would lead to improved surgical performance. Our study evaluates image guidance through video augmentation overlaying these select structures on a standard TORS endoscopic view on two different tongue models. This paper presents the initial development and evaluation of the proposed image guidance system for transoral robotic surgery.

Materials and methods

The study was exempt from Institutional Review Board approval at Johns Hopkins Medical Institutions and University Robotic experiments involved locating 'mock tumors' or targets embedded beneath the surface (i.e. not visible) within two different tongue models. The objective of each task was to place a pin as close to a particular target as possible with various levels of image-guidance and post-experiment analysis measured the accuracy of pin placement. Robotic experiments were conducted on a da Vinci 'S' system in the Computer Integrated Interventional Systems Laboratory in the Department of Computer Science, CBCT image data were acquired at the Imaging for Surgery Therapy and Radiology Lab in the Department of Biomedical Engineering and CT images were obtained from the Department of Radiology at the Johns Hopkins University.

Model porcine (Model P) tongue phantoms

Experiments were first conducted on five fresh porcine tongues. Initially, three porcine tongues were each embedded with eight frozen peas simulating soft tissue targets. In the remaining two tongues the soft-tissue-simulating spheres were replaced with ten 1.6 mm diameter Teflon spheres, as this resulted in reduced collateral tissue trauma and improved accuracy of target analysis. Between eight and ten 3.2 mm diameter nylon spheres were affixed to the tongue surface to serve as registration fiducials. The number of utilizable embedded targets varied due to random placement and different tongue specimen sizes, which also affected the number of required surface fiducials. The tongue specimens were placed on one of two interchangeable foam templates retained on a custom fabricated frame. The first template maintained the tongue in a flat preoperative (PO) position similar to that of a patient lying supine, while the intra-operative (IO) tongue template placed the tongue in an extended, curved position, simulating that of the human tongue retracted during a TORS base of tongue resection. The tongue phantom, including the template and frame, were imaged positions (PO and IO) by a C-arm CBCT system with a $(15 \times 15 \times 15)$ cm^3 field of view as detailed below (Figure 1).

Model cadaver (Model C) head phantom

A fresh adult cadaver head from the Maryland State Anatomy Board was lightly preserved in a phenol-glycerin solution to maintain joint and tissue flexibility, allowing for a range of motion in the neck, mandible and tongue. In the preoperative pose, simulating that of a patient in a CT scanner, the cadaver's mouth was closed with tongue in repose. In the intraoperative position, replicating that of a patient positioned for base of tongue TORS, the neck was extended, mouth opened, and the tongue pulled anteriorly with sutures along a custom radiolucent frame. These positions are illustrated in Figure 2. Six 3.2 mm diameter nylon sphere fiducials were glued to the surface of the tongue while ten 1.6 mm diameter Teflon spheres were implanted within the tongue to serve as targets. Clinical mouth and tongue retractors, such as the Feyh Kastenbauer (FK) retractor (Gyrus ACMI/Explorent GmbH, Tuttlingen Germany), were not used as these stainless steel instruments would cause metal artifacts in CBCT. Alternatively, custom radiolucent instruments or advanced reconstruction algorithms with metal artifact reduction (5) could address this challenge but are topics beyond the scope of this paper.

DOI: 10.1002/rcs

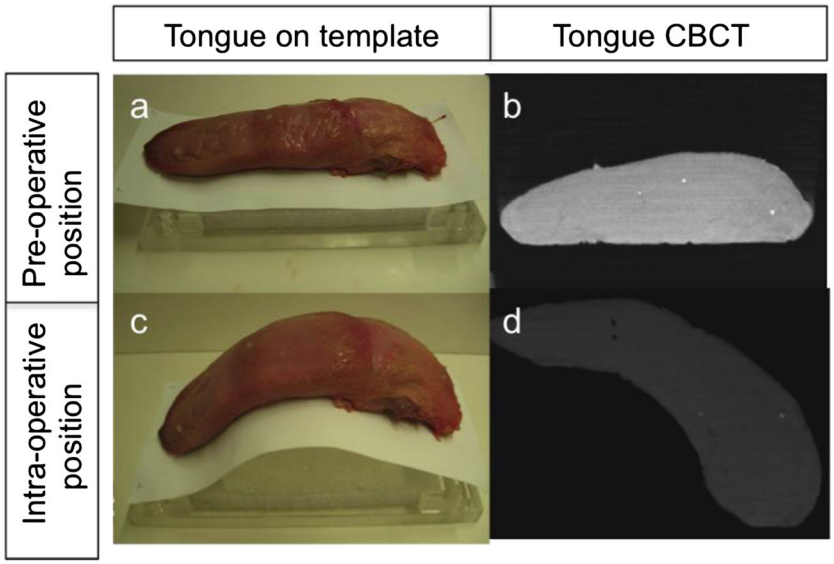

Figure 1. Photographs and CBCT images of a porcine tongue positioned by (a,b) a flat, preoperative template and (c,d) a curved, extended intraoperative position. The bright punctuated lesions in the scans represent the Teflon targets

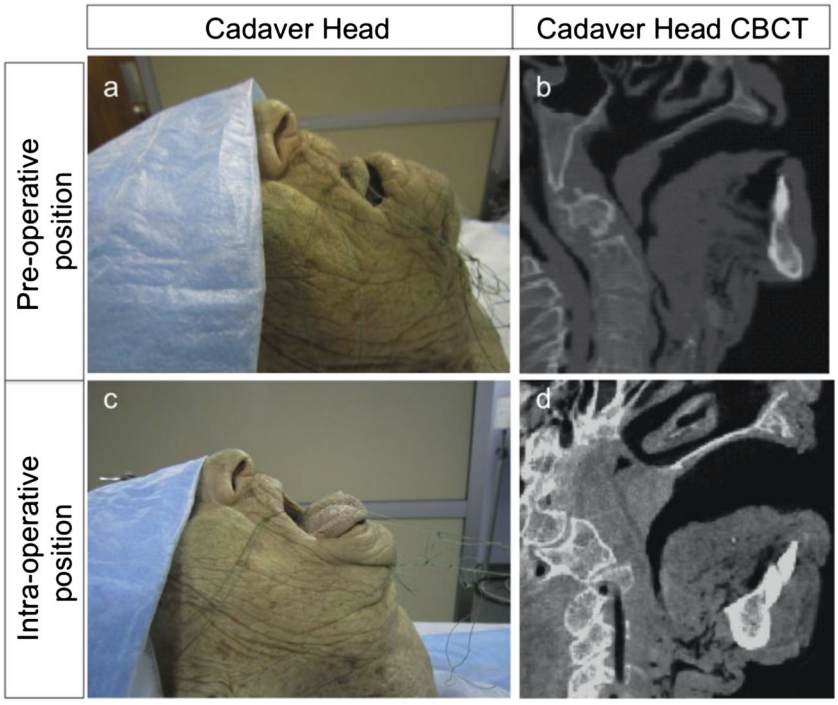

Figure 2. Cadaver head positioned in (a) a preoperative pose and (b) imaged in CT. The same cadaver was positioned in (c) an intraoperative pose [tongue sutured and extended] and (d) imaged using C-arm CBCT

Image guidance workflow

The main steps of the proposed workflow to integrate image guidance with TORS are illustrated in Figure 3.

Step 1: Surgical planning

The workflow begins by segmenting TORS anatomy, including oncologic (i.e. oropharyngeal lesions) and key functional structures (i.e. carotid and lingual arteries) from standard diagnostic CT. In general, segmentation in such volumetric images refers to designating each voxel in the image (or some subset of voxels) by a label specifying the anatomy or tissue type to which the voxel belongs. Research on this topic encompasses a broad and active field (6,7) beyond the scope of this paper, and within the current work we applied relatively simple, semi-automatic

DOI: 10.1002/rcs

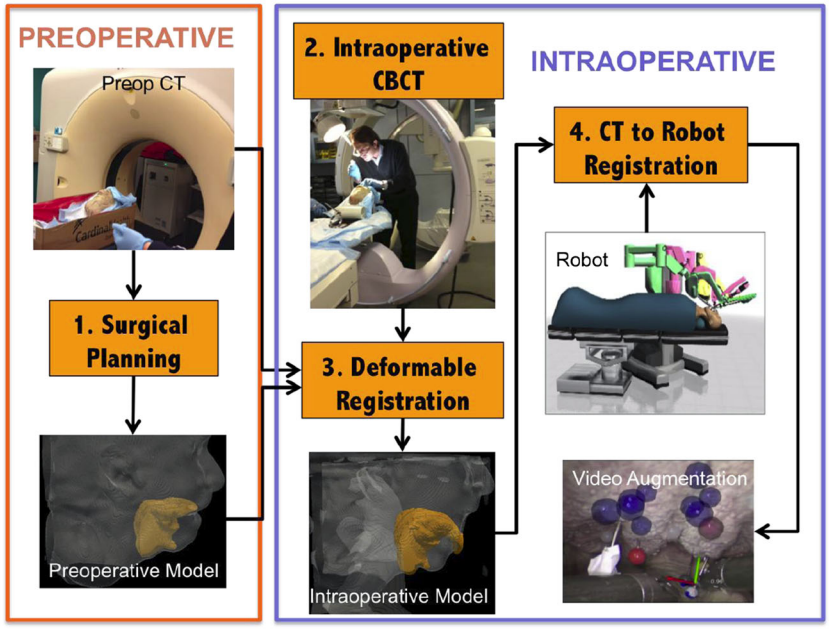

Figure 3. Proposed image guidance workflow to deformably register surgical planning from preoperative CT to robotic stereo video using intraoperative CBCT

segmentation techniques initialized with intensity-based thresholds and edge-based snake methods (8) using ITK-snap (7) and 3DSlicer (9) (Brigham & Women's Hospital, Cambridge MA). The segmentation of structures of interest constitutes a *preoperative model,* which in our experiments included the tongue targets (centroid and boundary), registration fiducials, oral tongue, and tongue base volume. For the Model P specimens, surgical planning in the PO position was performed directly in CBCT (detailed below). For Model C, surgical planning in the PO position was performed in preoperative CT (Philips Brilliance CT, Head Protocol, 120 kVp, 277 mAs, $0.7 \times 0.7 \times 1.0\,\text{mm}^3$ voxel size, and 'Standard' reconstruction kernel).

Step 2: Intraoperative imaging

The workflow, shown in Figure 3, incorporates intraoperative CBCT acquired using a mobile, fixed-room, or robotic C-arm (10–12) with the patient in the standard IO position, but prior to docking of the robot. In the IO position the patient should be readied for transoral robotic intervention with the neck flexed, mouth retracted with the tongue deformed and held in place by the tongue blade of the retractor. The surface fiducials establish a reference coordinate system for point-based rigid registration. The acquisition of a CBCT after these steps captures the intraoperative deformation of the tongue in the IO position with artificial registration fiducials. In these experiments, we employed a mobile C-arm prototype for CBCT reported in previous work (11,13,14) with scan protocols for Model P and Model C appropriate to CBCT of the head and neck (15): 100 kVp, 230 mAs, $0.6 \times 0.6 \times 0.6\,\text{mm}^3$ voxels, and a 'Standard' reconstruction kernel. The IO position of Model P specimens was as illustrated in Figure 1(c), (d), and that of Model C was that in Figure 2(c), (d), with the mouth open and tongue retracted.

Step 3: Deformable registration

Following imaging and segmentation of structures in the IO position, a deformable 3D image registration is performed to geometrically align the preoperative CT (the 'moving' image) with the intraoperative CBCT (the 'fixed' image) (16). This registration step transforms the preoperative plan to an accurately deformed *intraoperative model* whose components are to be overlaid onto the stereo endoscopy for guidance. Intraoperative CBCT provides an accurate image of anatomy in the IO position but is limited with respect to soft tissue visibility. Therefore, rather than visualizing the target neoplasm directly in CBCT (which exhibits low contrast relative to surrounding muscle and lingual tonsils), the CBCT image is used to deformably register the preoperative CT in which preoperative planning was conducted in the context of well visualized anatomic detail.

Step 4: CBCT-to-robot registration

Following 3D image registration, the robotic endoscopic video is registered with the *intraoperative model* via a manual process in which registration fiducials segmented in CBCT are identified through the da Vinci console's

DOI: 10.1002/rcs

endoscopic view. These corresponding data provide a rigid point-based transformation between the CBCT data and the robotic endoscope. When the surgeon moves the endoscope, this transformation is updated using the real-time joint positions of the robot as provided by the Application Programming Interface (API) (Intuitive Surgical Inc., Sunnyvale, CA) to maintain registration of overlaid objects in the stereo endoscopic scene.

Step 5: Video augmentation

The *intraoperative model* is visualized through video augmentation of the real-time stereo endoscopy in which the segmented target and critical structures are overlaid directly within the surgeon's field of view in the da Vinci console. Figure 4 illustrates the difference between current simulated practice and the proposed image guided workflow. Stereoscopic video augmentation was achieved with a modular architecture by extending the *cisst/ SURGICAL ASSISTANCE WORKSTATION (SAW)* (17) as previously reported (18). During robotic experimentation to place pins in the centroid of the targets, the *daVinci S* console showed the embedded synthetic targets (extended as a virtual sphere with a 4 mm radius) and a surface mesh of the tongue. The degree of transparency, color, and visibility of all objects could be set according to user preference and adjusted directly using the surgeon's console.

In addition to projective overlay, preliminary experiments (18) concluded a need for subsurface information along the camera axis, i.e. a depth gauge, to improve overall target localization. This required tracking of the pin tip, held by a *da Vinci* needle driver, with respect to the virtual targets. The *da Vinci S* API reports the tip position of any tool in the coordinate system of the endoscope. However, an unknown intrinsic transformation of the robotic arm offsets the true tool position. To calibrate this offset, we positioned a needle driver (inserted into arm 2 of the patient side cart – i.e. the surgeon's right handle robotic arm) in several known positions and computed a rigid transformation between the reported position from the API and the observed video coordinates. To estimate the tip of the pins held by the needle driver during experimentation, we assumed a simple translation (i.e. length of the pin, as measured preoperatively) from the center of the calipers of the tracked needle driver. These steps allowed tracking of the tip of each pin held by the right hand needle driver within the endoscopic view and overlaid objects. Thus, in addition to the original overlay, as a depth cue we rendered a transparent sphere at the tracked pin tip that changed from green, to yellow, to red when the estimated distance from the pin tip to the closest target was within 4 mm, 2 mm, and 1 mm, respectively, along with a quantitative marker (depth gauge) displaying the absolute distance.

Robotic experimental protocol

A head and neck surgeon experienced in TORS (JDR) was asked to use the research da Vinci S console with variations of the image guidance system described above. The goal of each experiment was to place 9.5 mm pins using a right-handed needle driver as close as possible to the center of each target embedded in the Model P and Model C phantoms. Experiments on both models included a control, viewing preoperative image data independently from robotic visualization, as well as one or more image-guided scenarios. Experiments conducted on Model P (P1–P5), tested variations of image guidance such as the influence

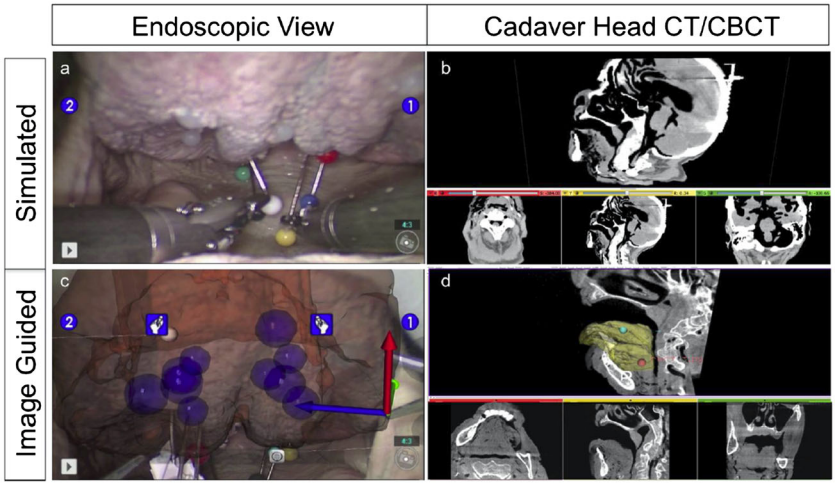

Figure 4. Simulated current practice in which the surgeon has access to only (a) the raw endoscopy (b) preoperative CT to help guide the surgical navigation. Our proposed image guidance workflow showing (c) a 3D image overlay of the targets on the tongue. derived from (d) the preoperative plan registered to intraoperative CBCT

DOI: 10.1002/rcs

of deformable registration and depth information. In contrast, experiments on Model C (C1–C2) used a more realistic workspace with a cadaveric head to compare simulated current workflow and the fully proposed image guidance system. Variations in the experiments are summarized in Table 1.

Model porcine

Experiment P1 simulated current clinical practice where preoperative images are viewed separately from the robotic system. The preoperative image data were available offline in standard 3D triplanar views on a laptop next to the surgeon's console. The surgeon had free access to the PO images prior to placement of each needle in its designated target. In scenario P2, the IO image data that accurately represents the deformed tongue in the intraoperative set-up replaced the PO data. The inclusion of scenario P2 allowed direct comparison with the case where IO imaging is available but is not integrated, i.e. viewed separately from the stereo endoscopy. Scenario P3 deformably registered intraoperative models onto the stereoscopic view. This allowed the surgeon to view a 3D overlay of the target images on top of the standard endoscopic view, allowing for the evaluation of the influence of image guidance with overlay and the integration of deformable registration. Although not envisioned in the final clinical workflow, Scenario P4 allowed assessment of the deformable registration step. Planning structures were defined directly in the IO images, rather than deformably registered as in P3. Experiment P5 extended the image guidance scenario from P4 with an additional depth gauge to evaluate the impact of explicit stereo information.

Model cadaver

C1 (Figure 4(a), (b)), similar to P1, simulated current clinical practice where standard 3D triplanar views of preoperative CT were available offline for pin placement. Image guidance for experiment C2 (Figure 4(c), (d)) tested the most realistic phantom with the full proposed workflow, using video augmentation with additional depth information of deformably registered preoperative CT data.

Analysis

Following each of the above experimental protocols and after all pins were placed, the Model P and Model C specimens were imaged with the CBCT C-arm. The pin tips, pin axis, and targets were manually segmented in CBCT using ITK-Snap and Slicer3D. Target Localization Error (TLE) was measured as the distance between the pin tip and target. We decomposed TLE into four types: Edge, Center, Projection, and Depth, as defined in Figure 5. The TLE_{Edge} is the distance between the needle tip and

Table 1. Experimental scenarios. Experiments simulating current practice are in italics. Experiments using the proposed image guidance system are in bold

Phantom		Image Data	Overlay	Depth	Image Guidance
	P1	*Preop*	*No*	*No*	*Preop image is displayed in 2D, offline. This simulates current practice where preop data is viewed separately from the surgeon's console.*
Model Porcine	P2	Intraop	No	No	Intrao CBCT is displayed in 2D, offline. This simulates availability of imaging from intraop C-arm without integration with stereo endoscopy.
	P3	Preop (Registered)	Yes	No	Preop segmented plan is deformly registered to stereo endoscopy. This represented the basline proposed image guidance workflow with video augmentation only.
	P4	Intraop	Yes	No	Intraop segmented plan is overlayed onto stereo endoscopy. This simulates the ideal scenario where TORS critical structures are sufficiently visible and can be segemented in intraoperative CBCT, thus excluding the need for deformable registration between preop and intraop imaging*.
	P5	**Intraop**	**Yes**	**Yes**	**Intraop segmented plan and expicit depth gauge is overlayed onto stereo endoscopy. This extends image guidance used in P4 with augmented depth information (i.e. tracking the tip of the pin with respect to targets).**
Model Cadaveric	*C2*	*Preop*	*No*	*No*	*Preop image is displayed in 2D, offline. This simulates current practice, in a cadaveric oropharyngeal workspace, where preop CT is viewed separately from the surgeon's console.*
	C2	**Preop (Registered)**	**Yes**	**Yes**	**Preop segmented plan and explicit depth gauge is deformly registered to stereo endoscopy. This tests our entire proposed image guidance system (augmentation + depth) in a cadaveric oropharyngeal workspace.**

*This is an unrealistic situation as preop CTA and MR are expected to better delineate soft tissue oropharyngeal structures that may not be visible in intraop CBCT. However, this serves as a basis of comparison where image-based registration errors do not contribute to overall accuracy.

DOI: 10.1002/rcs

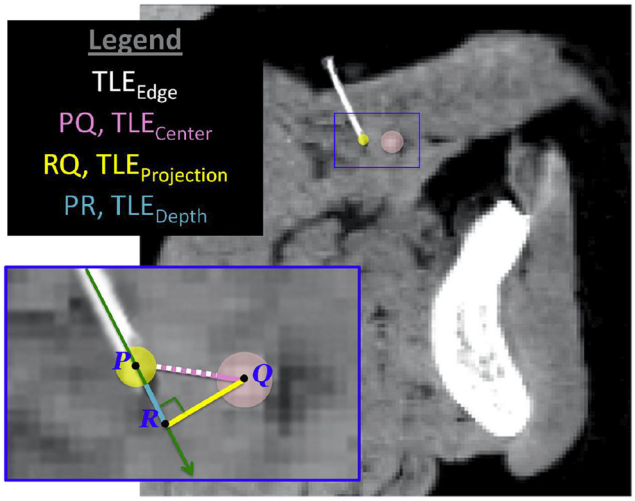

Figure 5. Sagittal slice of post experiment CBCT from the Model C phantom (cadaver head). Target Localization Error (TLE) is deconstructed into four types: TLE (Edge, Center, Projection, and Depth)

the closest edge of the target (needles placed on or inside the target assigned a TLE_{Edge} value of 0 mm). The TLE_{Center} best captures the given task of needle placement at the centroid, while $TLE_{Projection}$, or reprojection distance, has been used in previous work to evaluate accuracy of video augmentation (19). If the needle tip, the center of the target, the projection of the target onto the needle axis is labeled *P, Q, R*, respectively, then TLE_{Center} is the line segment *PQ* (Figure 5). We deconstruct *PQ* into segments along the needle, *PR*, and orthogonal to the needle, *RQ*, as TLE_{Depth} and $TLE_{Projection}$, respectively.

Results

Model porcine phantom experiments

Measurements of TLE from the Model P experiments with the TORS robot are summarized in Table 2, with *P*-values shown in Table 3. The mean TLE_{Edge} [mm], improving progressively in each image guidance scenario, was evaluated at 4.8 ± 4.0, 3.9 ± 2.9, 3.2 ± 3.6, 2.2 ± 1.9, 1.3 ± 1.2 for P1–P5, respectively. Similarly, the mean TLE_{Center} [mm] improved from 9.8 ± 4.0, 8.9 ± 2.9, 7.1 ± 2.8, 6.7 ± 2.8, 5.3 ± 1.3 for P1–P5, respectively while achieving a *P*-value of 0.0151 between P1 and P5. Comparable results were measured for the deformable workflow (P3) and the direct intraoperative overlay (P4), and the lack of statistical significance for all TLE measures between these two cases (e.g. P-value = 0.7036 for TLE_{Center}) suggested that the proposed deformable image registration system approaches the ideal scenario of planning directly in intraoperative CBCT (Table 4).

TLE_{Center} was further deconstructed into its projection and depth components. Experiments on porcine phantoms show $TLE_{Projection}$ improving by ~3 mm between no overlay ((6.4 ± 3.3) mm for P1, and (6.12 ± 1.7) mm for P2) and scenarios with overlay ((3.2 ± 1.6) mm for P3, and (3.2 ± 1.6) mm for P4). However, TLE_{Depth} showed no improvement comparing the same two scenarios with and without overlay (Figure 6 box plot of TLE_{Edge} in the upper right). This points to intraoperative imaging as the source for the improvement in depth error as opposed to an effect from video augmentation, whose influence clearly plateaus. Stereoscopic overlay localizes targets within the camera image plane well, but it does not provide clear depth localization along the camera axis, i.e. orthogonal to the camera plane. With additional augmentation of explicit depth information (i.e. experiment P5), TLE_{Depth} is reduced to (3.5 ± 2.0) mm. This demonstrated the usefulness of enhancing stereo perception with information along the camera axis. Using stereoscopic video augmentation and a depth gauge, the surgeon was able to place needles with a hit ratio of 25% in experiment P5, compared with 0% for P1 and P2. Improvements between P1 and P2 indicate the positive influence of intraoperative imaging, and the larger improvement between P2 and P3 reinforces the value of visualizing guidance information directly in the surgeon's natural endoscopic window.

Table 2. Target localization error from cadaver and porcine phantom experiments. Results show improved TLE between simulated current clinical workflow (P1 and C1) and the proposed image-guidance system (P5 and C2)

	TLE_{Center} [mm]			TLE_{Edge} [mm]			$TLE_{Projection}$ [mm]			TLE_{Depth} [mm]			
Phantom	Mean	STDDev	Max	Mean	STDDev	Max	Mean	STDDev	Max	Mean	STDDev	Max	Hit ratio
P1	9.8	4.0	18.5	4.8	4.0	13.5	6.4	3.3	11.3	6.6	4.0	14.6	0/8
P2	8.9	2.9	15.4	3.9	2.9	10.4	6.2	1.7	8.7	5.5	4.1	14.5	0/8
P3	7.1	3.7	14.1	3.2	3.6	10.1	3.2	1.6	5.8	5.9	4.3	12.8	1/8
P4	6.7	2.8	9.8	2.2	1.9	4.7	3.3	2.0	6.0	5.3	3.3	9.3	2/8
P5	5.3	1.3	7.0	1.3	1.2	3.0	3.5	1.1	4.7	3.5	2.0	6.5	2/8
C1	11.2	5.0	19.0	7.4	4.7	15.0	8.1	5.6	16.4	8.1	2.3	10.4	1/10
C2	5.8	2.5	10.1	2.0	2.3	6.1	3.5	1.3	5.5	2.3	1.1	3.2	3/10

DOI: 10.1002/rcs

Table 3. Statistical significance (*P*-values) in the measured differences in TLE between the various modes of operation in the porcine phantom experiments. Statistically significant results (*P*-value < 0.05) are in underlined bold

Porcine experiments (*P*-value)					
TLE (Center)					
Phantom	P1	P2	P3	P4	P5
P1	-	0.6313	0.0657	**0.0139**	**0.0151**
P2	-	-	0.1418	0.2569	**0.0199**
P3	-	-	-	0.7036	0.2596
P4	-	-	-	-	0.2214
P5	-	-	-	-	-
TLE (Edge)					
Phantom	P1	P2	P3	P4	P5
P1	-	0.6313	0.0831	**0.0159**	**0.0170**
P2	-	-	0.1642	0.2734	**0.0207**
P3	-	-	-	0.8183	0.1647
P4	-	-	-	-	0.2109
P5	-	-	-	-	-
TLE (Projection)					
Phantom	P1	P2	P3	P4	P5
P1	-	0.8580	0.0588	**0.0258**	0.0439
P2	-	-	**0.0054**	**0.0159**	**0.0124**
P3	-	-	-	0.9313	0.6498
P4	-	-	-	-	0.5160
P5	-	-	-	-	-
TLE (Depth)					
Phantom	P1	P2	P3	P4	P5
P1	-	0.6179	0.4015	0.6471	0.0887
P2	-	-	0.8948	0.8221	0.2702
P3	-	-	-	0.6957	0.3360
P4	-	-	-	-	0.2156
P5	-	-	-	-	-

Table 4. Statistical significance (*P*-values < 0.05) were achieved in measurements of all TLEs in the cadaver specimen between the conventional mode of operation (C1) and the proposed workflow (C2) integrating intraoperative imaging and endoscopic overlay

Cadaveric experiments (*P*-value)				
TLE	Center	Edge	Projection	Depth
Phantom	C2	C2	C2	C2
C1	**0.0189**	**0.0136**	**0.0406**	**0.0111**

Model cadaver experiments

Similar to the results detailed above for the porcine phantom, the Model C (cadaver head) data using proposed image guidance system (experiment C2) achieved improved TLE in all four categories, as shown in the boxplots of Figure 7 and by comparisons of experiments C1 and C2 in Table 2. TLE_{Center} improved from (11.2 ± 5.0) mm for scenario C1 to (5.8 ± 2.5) mm for scenario C2 ($P = 0.0189$). This demonstrates a statistically significant improvement between simulated current practice (C1 – unregistered preoperative imaging) and the proposed image guidance process (C2 – deformable registration of planning data via intraoperative CBCT and overlay in endoscopic video). Comparing experiment P1 (porcine model) with C1 (cadaver), it is clear that the baseline level of difficulty was higher for the latter. This is likely a product of increased difficulty in visualization and maneuverability due to a realistic robotic setup for TORS in the cadaver oral cavity. The endoscopic camera (0 degree endoscope) was obliquely positioned in relation to the tongue surface when passed through the mouth. When viewed through the console, this presented crowded targets with more overlap and created a more challenging space for localization both visually and spatially. However, the comparable results achieved in experiment C2 (compared with P5) shows that the proposed image guidance system overcomes the additional challenges presented by the realistic setup in Model C.

Discussion

Transoral robotic surgery has become an increasingly prevalent technique for treatment of oropharyngeal cancer. It offers a minimally invasive intervention with excellent oncologic, functional, and long-term quality of life (20). However, despite growing enthusiasm for TORS culminating in the first randomized, prospective surgical trials (ECOG 3311 and RTOG 1221) there remain barriers to its

DOI: 10.1002/rcs

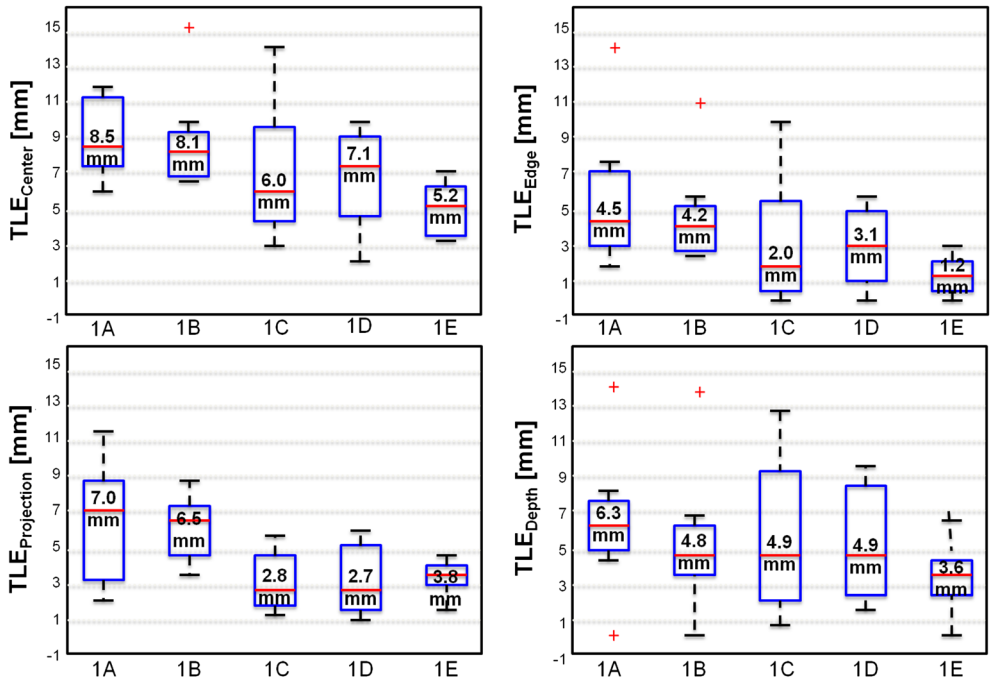

Figure 6. Box plots of TLE (Center, Edge, Depth, Projection) [Clockwise from upper left] for porcine experiments

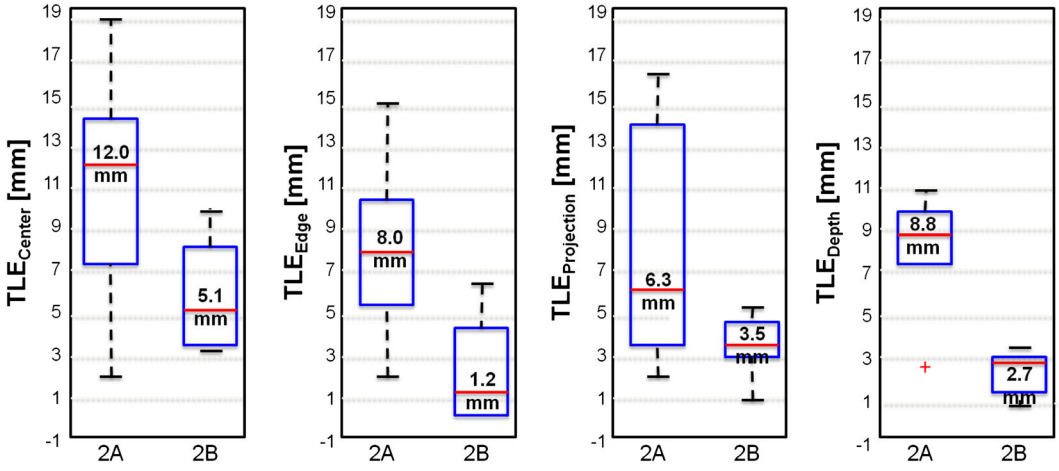

Figure 7. Box plots of TLE (Center, Edge, Projection, Depth) [From left to right] for cadaver experiments

adoption. Infiltrative, submucosal base of tongue cancers present a particular challenge. The surgeon must rely on preoperative imaging of a tumor that may deform substantially with the patient in the operative position with mouth open and tongue extended. The base of tongue also lacks conspicuous landmarks to help guide the surgeon safely around the tumor while maintaining an adequate oncologic margin. This can result in an undesirable situation where the actual tumor depth and depth of dissection are not fully appreciated resulting in a challenging procedure and potentially incomplete resection. Ultimately, this may limit a surgeon's confidence to resect larger tumors. Although studies have demonstrated that TORS is associated with a low positive margin rate (between 0 and 7% (3,21,22)), such an occurrence is associated with significantly more toxic adjuvant therapy and negatively impacts survival rates. We therefore sought to develop an image guidance platform for TORS to overcome such limitations in base of tongue resections.

Endoscopic and laparoscopic surgical interventions integrating various imaging modalities, such as ultrasound (23) and CBCT (24,25), have been applied to similarly high-risk and complex, minimally invasive procedures. Previously reported work in the context of image-guided sinus

DOI: 10.1002/rcs

and skull-base surgery demonstrates operative tools tracked within preoperative images (26,27). Visualization and feedback is often accomplished through rendering of virtual (28) or reconstructed environments (29). To date, intraoperative imaging with visualization directly in the primary stereo endoscopic view has had limited applicability in robotic surgery, including TORS (30), where it was shown to assist in the resection of a mass in the parapharyngeal space.

The work detailed in this paper demonstrates the feasibility of intraoperative image-guided TORS using a research *da Vinci S* robot in porcine tongues and a human cadaver. The proposed workflow uses a surgical plan defined in preoperative CT and/or MRI that is subsequently registered deformably to intraoperative CBCT. This captures the large deformation associated with patient setup in the operating room while directly integrating rich, multi-modal preoperative data delineating the cancer target as well as critical functional anatomy within the surgeon's natural field of view. The experimental results demonstrate statistically significant improvements in TLE when comparing intraoperative image-guided TORS with simulated current practice.

Resection of tumors via TORS is achieved in an en bloc fashion. Surgeons rely on pre-operative imaging, physical examination, visual feedback, the input of the bedside assistant, and personal experience to the approach of a tumor, which is often individualized in each case. There is a steep learning curve to the resection of deep base of tongue cancers given the lack of anatomic landmarks, the potential for significant bleeding, and the unfamiliar orientation requiring dissection in a plane beyond the tumor margin as the tongue curves towards the vallecula with limited visualization of the true depth of dissection. This may result in either breach of the tumor margin or excessive resection of normal tissue and injury to critical neurovascular structures. Our approach seeks to provide real time feedback to the surgeon through augmented visual cues to optimize resection and avoid some of the drawbacks listed above. These results beg the question: what level of accuracy is needed in TORS base of tongue resections? While the ultimate goal is to safely excise the tumor with a clear margin of normal tissue, the definition of 'clear' and 'close' margins has varied in the literature. A close margin has been defined in a TORS context (31) as a tumor within 2 to 5 mm of the cut surface edge. Therefore, we speculate that an improvement in accuracy in tongue base resections on this scale would result in a decreased close and positive margin rate. Our results demonstrate that image-guidance overlay with depth cues was able to decrease the TLE by an average of 5.4 mm in the cadaver tongue model, which is consistent with this desired level of improved accuracy.

For resection in soft tissue, real-time, high-fidelity deformable intraoperative tracking is the ultimate objective. However, we believe that an assumption of a locally rigid resection volume, within a globally deformable environment, would be a sufficient starting point. Using this assumption, next steps include resection volume tracking via rigid surface fiducials, possibly composed from surgical pins/clips, using optics and 3D localization with dual-projection fluoroscopy. The proposed workflow currently requires an additional low-dose CBCT scan (24). The incremental radiation, especially with many patients requiring postoperative radiotherapy, is minimal. Future efforts will include integration with digital tomosynthesis and ultrasound, as an alternative non-irradiative modality, to better address deformable updates with reconstruction. This significant limitation of the current system must be addressed prior to clinical deployment.

Ongoing and future work will improve the human–computer interface of the research system, including visualization and rendering of depth perception and camera-axis information. The significance of subsurface information in image guidance systems has been investigated for similar robotic interventions, as in urology, for example, where image guidance has been shown to resolve questions of depth, decrease the standard deviation of target margin resection, and shorten overall procedure time (32). The registration of CBCT to robotic coordinates may also be refined with additional methods, such as registration of 3D structures recovered from disparity and motion (33,34). Future work will also include a more comprehensive analysis of the navigation system in order to understand error sources and propagation to enable comprehensive systematic refinement. Further preclinical studies should include additional cadaveric and *in vivo* robotic experiments, emphasizing intraoperative deformation and updates, with the ultimate goal of bringing this technology to the clinical arena.

Acknowledgements

This research was supported in part by Intuitive Surgical Inc., the Department of Otolaryngology – Head and Neck Surgery at Johns Hopkins School of Medicine, and other Johns Hopkins University internal funds. The authors extend sincere thanks to Dr Iulian Iordachita and Woo Jin Kim (Department of Computer Science, Johns Hopkins University) for assistance with tongue phantom fabrication and to Dr Eric Ford (Department of Radiation Oncology, Johns Hopkins University) for acquisition of preoperative CT images. The authors also extend their gratitude to Dr Adam Wang and Dr Sebastian Schafer (Department of Biomedical Engineering, Johns Hopkins

DOI: 10.1002/rcs

University) for acquisition of C-arm CBCT images and to Mr Ronn Wade (University of Maryland, State Anatomy Board) for assistance with the cadaver specimen. The C-arm CBCT prototype was developed in partnership with Siemens XP (Erlangen, Germany) with thanks to Dr Rainer Graumann, Dr Gerhard Kleinszig, and Dr Sebastian Vogt. Infrastructure support was provided from NIH-R01-CA-127444, NSF EEC9731748, and the Swirnow Family Foundation.

Conflicts of interest

Jeremy D. Richmon, MD is a proctor for Intuitive Surgical, Inc.
Jonathan M. Sorger, PhD is an employee of Intuitive Surgical, Inc.
Wen P. Liu, M.S. is a student fellow sponsored by Intuitive Surgical, Inc.
Jeffrey H. Siewerdsen has a research collaboration with Siemens Healthcare.

Funding

Department of Otolaryngology – Head and Neck Surgery Johns Hopkins Hospital
Department of Computer Science, Johns Hopkins University
Intuitive Surgical, Inc.
Siemens Healthcare

References

1. Weinstein GS, Quon H, Newman HJ, *et al.* Transoral robotic surgery alone for oropharyngeal cancer: an analysis of local control. *Arch Otolaryngol Head Neck Surg* 2012; **138**(7): 628–634.
2. Westra WH. The changing face of head and neck cancer in the 21st century: the impact of HPV on the epidemiology and pathology of oral cancer. *Head Neck Pathol* 2009; **3**(1): 78–81.
3. Weinstein GS, O'Malley BW, Jr, Magnuson JS, *et al.* Transoral robotic surgery: a multicenter study to assess feasibility, safety, and surgical margins. *Laryngoscope* 2012; **122**(8): 1701–1707.
4. Weinstein GS, O'Malley BW, Jr, Desai SC, Quon H. Transoral robotic surgery: does the ends justify the means? *Curr Opin Otolaryngol Head Neck Surg* 2009; **17**(2): 126–131.
5. Stayman JW, Otake Y, Prince JL, *et al.* Model-based tomographic reconstruction of objects containing known components. *IEEE Trans Med Imaging* 2012; **31**(10): 1837–1848.
6. Xu C, Pham D, Prince J. Medical image segmentation using deformable models. In Handbook of Medical Imaging, Vol 2: Medical Image Processing and Analysis. SPIE Press: Bellingham, WA, USA, 2000.
7. Yushkevich PA, Piven J, Hazlett HC, *et al.* User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 2006; **31**(3): 1116–1128.
8. Yushkevich JP, Cody H, Ho S, *et al.* User-guided level set segmentation of anatomical structures with ITK-SNAP. *Insight J* 2005; **1**: 1116–1128.
9. Fedorov A, Beichel R, Kalpathy-Cramer J, *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012; **30**(9): 1323–1341.
10. Chan Y, Siewerdsen JH, Rafferty MA, *et al.* Cone-beam computed tomography on a mobile c-arm: novel intraoperative imaging technology for guidance of head and neck surgery. *J Otolaryngol Head Neck Surg* 2008; **37**(1): 81–90.
11. Siewerdsen JH. Cone-beam CT with a flat-panel detector: from image science to image-guided surgery. *Nucl Instrum Methods Phys Res A* 2010; **648**(1): S241–S250.
12. Siemens' Artis zeego brings surgery and industry together. *Cardiovasc J Afr* 2009; **20**(4): 258.
13. Siewerdsen JH. Cone-beam ct with a flat-panel detector: from image science to image-guided surgery. *Nucl Instrum Methods Phys Res A* 2011; **648**(S1): S241–S250.
14. Siewerdsen JH, Chan Y, Rafferty MA, *et al.* Cone-beam CT with a flat-panel detector on a mobile C-arm: preclinical investigation in image-guided surgery of the head and neck. In Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display Conference, San Diego, CA, 2005.
15. Daly MJ, Chan H, Nithiananthan S, *et al.* Clinical implementation of intraoperative cone-beam CT in head and neck surgery. In SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling Conference, Lake Buena Vista, Florida, USA, 2011.
16. Reaungamornrat S, Liu WP, Schafer S, *et al*. A Gaussian mixture + demons deformable registration method for cone-beam ct-guided robotic transoral base-of-tongue surgery. Paper presented at: SPIE Medical Imaging, Orlando, 2013.
17. Deguet A, Kumar R, Taylor RH, Kazanzides P. The cisst libraries for computer assisted intervention systems. *MICCAI Workshop*, 2008, https://trac.lcsr.jhu.edu/cisst/.
18. Liu WP, Reaungamornrat S, Deguet A, *et al*. Toward intraoperative image-guided transoral robotic surgery. Paper presented at: Hamlyn Symposium, London, UK, 2012.
19. Mirota D, Wang H, Taylor R, *et al*. Toward video-based navigation for endoscopic endonasal skull base surgery. MICCAI, Berlin, Germany, 2009.
20. Leonhardt FD, Quon H, Abrahao M, *et al.* Transoral robotic surgery for oropharyngeal carcinoma and its impact on patient-reported quality of life and function. *Head Neck* 2012; **34**(2): 146–154.
21. Hurtuk A, Agrawal A, Old M, *et al.* Outcomes of transoral robotic surgery: a preliminary clinical experience. *Otolaryngol Head Neck Surg* 2011; **145**(2): 248–253.
22. Moore EJ, Olsen KD, Kasperbauer JL. Transoral robotic surgery for oropharyngeal squamous cell carcinoma: a prospective study of feasibility and functional outcomes. *Laryngoscope* 2009; **119**(11): 2156–2164.
23. Klein T, Hansson M, Navab N. Modeling of multi-view 3D free-hand radio frequency ultrasound. Paper presented at: Medical Image Computing and Computer-Assisted Interventions (MICCAI-2012), France, 2012.
24. Daly MJ, Siewerdsen JH, Moseley DJ, *et al.* Intraoperative cone-beam CT for guidance of head and neck surgery: assessment of dose and image quality using a C-arm prototype. *Med Phys* 2006; **33**(10): 3767–3780.
25. Siewerdsen JH, J. DM, Chan H, *et al*. High performance intraoperative cone-beam CT on a mobile C-arm: an integrated system for guidance of head and neck surgery. Paper presented at: SPIE Medical Imaging, Orlando, 2009.
26. Balachandran R, Fitzpatrick JM, Labadie RF. Accuracy of image-guided surgical systems at the lateral skull base as clinically assessed using bone-anchored hearing aid posts as surgical targets. *Otol Neurotol* 2008; **29**(8): 1050–1055.
27. Liu WP, Mirota DJ, Uneri A, *et al*. A clinical pilot study of a modular video-CT augmentation system for image-guided skull base surgery. Paper presented at: SPIE Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling, 2012.
28. Schulze F, Buhler K, Neubauer A, *et al.* Intra-operative virtual endoscopy for image guided endonasal transsphenoidal pituitary surgery. *Int J Comput Assist Radiol Surg* 2010; **5**(2): 143–154.

DOI: 10.1002/rcs

29. Stoyanov D. Stereoscopic scene flow for robotic assisted minimally invasive surgery. Paper presented at Medical Image Computing and Computer-Assisted Interventions (MICCAI-2012), France, 2012.
30. Desai SC, Sung CK, Genden EM. Transoral robotic surgery using an image guidance system. *Laryngoscope* 2008; **118**(11): 2003–2005.
31. Nason RW, Binahmed A, Pathak KA, *et al.* What is the adequate margin of surgical resection in oral cancer? *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2009; **107**(5): 625–629.
32. Herrell SD, Kwartowitz DM, Milhoua PM, Galloway RL. Toward image guided robotic surgery: system validation. *J Urol* 2009; **181**(2): 783–789; discussion 789–790.
33. Zhang L, Curless B, Seitz S. Spacetime stereo: shape recovery for dynamic scenes. Paper presented at Computer Vision and Pattern Recognition, 2003.
34. Tombari F, Stefano L, Mattoccia S, Mainetti A. A 3D reconstruction system based on improved spacetime stereo. Paper presented at International Conference Control Automation, Robotics and Vision, Singapore, 2010.

DOI: 10.1002/rcs