Published in final edited form as:

IEEE Trans Med Imaging. 2013 July ; 32(7): 1215–1226. doi:10.1109/TMI.2013.2243464.

Evaluation of a System for High-Accuracy 3D Image-Based Registration of Endoscopic Video to C-Arm Cone-Beam CT for Image-Guided Skull Base Surgery

Daniel J. Mirota [Member, IEEE],

Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218 USA

Ali Uneri,

Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218 USA

Sebastian Schafer,

Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218 USA

Sajendra Nithiananthan,

Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218 USA

Douglas D. Reh,

Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, MD 21218 USA

Masaru Ishii,

Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, MD 21218 USA

Gary L. Gallia,

Department of Neurosurgery and Oncology, Johns Hopkins Medical Institutions, Baltimore, MD 21218 USA

Russell H. Taylor [Fellow, IEEE],

Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218 USA

Gregory D. Hager [Fellow, IEEE], and

Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218 USA

Jeffrey H. Siewerdsen*

Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218 USA

Daniel J. Mirota: dan@cs.jhu.edu; Ali Uneri: ali.uneri@jhu.edu; Sebastian Schafer: sebastian.schafer@jhu.edu; Sajendra Nithiananthan: sajendra@jhu.edu; Douglas D. Reh: dreh1@jhmi.edu; Masaru Ishii: mishii3@jhmi.edu; Gary L. Gallia: ggallia1@jhmi.edu; Russell H. Taylor: rht@jhu.edu; Gregory D. Hager: hager@cs.jhu.edu

Abstract

© 2013 IEEE

*jeff.siewerdsen@jhu.edu.

A preliminary version of this work was presented at SPIE Medical Imaging 2011.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

The safety of endoscopic skull base surgery can be enhanced by accurate navigation in preoperative computed tomography (CT) or, more recently, intraoperative cone-beam CT (CBCT). The ability to register real-time endoscopic video with CBCT offers an additional advantage by rendering information directly within the visual scene to account for intraoperative anatomical change. However, tracker localization error ($\sim 1\text{--}2$ mm) limits the accuracy with which video and tomographic images can be registered. This paper reports the first implementation of image-based video-CBCT registration, conducts a detailed quantitation of the dependence of registration accuracy on system parameters, and demonstrates improvement in registration accuracy achieved by the image-based approach. Performance was evaluated as a function of parameters intrinsic to the image-based approach, including system geometry, CBCT image quality, and computational runtime. Overall system performance was evaluated in a cadaver study simulating transsphenoidal skull base tumor excision. Results demonstrated significant improvement ($p < 0.001$) in registration accuracy with a mean reprojection distance error of 1.28 mm for the image-based approach versus 1.82 mm for the conventional tracker-based method. Image-based registration was highly robust against CBCT image quality factors of noise and resolution, permitting integration with low-dose intraoperative CBCT.

Index Terms

Cone-beam computed tomography (CBCT); endoscopy; image-guided treatment; registration; surgical guidance/navigation

I. Introduction

Minimally invasive surgical access to the skull base is becoming an increasingly popular approach over traditional open access (e.g., lateral rhinotomy). Endonasal skull base surgery (ESBS) offers a minimally invasive approach to remove lesions of the skull base transnasally and has been shown to provide reduced morbidity in comparison to open approaches [1]. The technique is also used to treat pituitary lesions, non-neoplastic skull base lesions, and other tumors of the skull base and nasal cavity. Such lesions are particularly challenging to access endonasally due to the proximity of critical neurovasculature, including the carotid arteries and cranial nerves [2]. Due to the complexity of these surgeries, navigation systems are often used to assist the surgeon [3], and studies have shown that navigation helps reduce the morbidity of the endonasal approach [4], [5]. Active research continues to improve the accuracy of the navigation systems [6]–[8] and reports that current less invasive registration methods, e.g., surface tracing, are not sufficient for all skull base procedures. Furthermore, current surgical technique requires the careful identification of key anatomic landmarks in both CT and endoscopically because of the limited accuracy of today's registration methods [9], though there is no substitute for anatomic knowledge. Most of these systems rely on navigation in the context of preoperative data—e.g., pre-operative CT—that does not convey intraoperative anatomical change nor allow for intraoperative assessment of the resection.

Recently, systems capable of high-quality intraoperative 3D imaging have become available based on cone-beam CT (CBCT) [10]–[15]. For example, CBCT on a mobile C-arm has demonstrated high-quality images with sub-millimeter spatial resolution and soft-tissue

visibility [11], [16]. Previous work used a conventional tracking system to track the endoscope and navigate within preoperative or intraoperative data [17]–[20], but did not use the video data directly in the registration process. Other related work registers endoscopic video using image-based methods, but registered to preoperative data [21], [22].

We build on initial studies [23] and extend such methods by directly using the endoscopic video data to register to intraoperative CBCT, which enables high-precision registration of video endoscopy with the most up-to-date tomographic data during surgery. The system detailed below reconstructs 3D point clouds from the endoscopic video using structure from motion (SfM) [24] and registers the resulting 3D point cloud directly to intraoperative CBCT. In this way, the system extends and improves previous methods that only include registration to preoperative data [25] or only use a tracking system for estimating the endoscope position [26]. Furthermore, the system extends beyond [23] by investigating the effects of relevant system parameters and further experimentation.

II. Methods

A. System Setup

The proposed system was first tested and evaluated using the benchtop arrangement shown in Fig. 1, which allowed for precise, reproducible control of the system geometry. The benchtop used two phantoms illustrated in Fig. 1: a polycarbonate “red skull” phantom [27] provided a rigid context approximating the anatomy of the sinus and oral cavity; the black anthropomorphic phantom included a natural human skeleton within tissue-equivalent plastic, modified to allow an endoscopic approach to the sphenoid, nasopharynx, and oropharynx. The red skull was used in studies of fundamental geometric registration accuracy and the anthropomorphic phantom for studies simulating the CBCT image quality properties of the head. The phantoms were rigidly secured to a rotary table (Velmex, Inc., Bloomfield, NY, USA) fastened to the optical table as in Fig. 1. The benchtop also included two linear stages (Velmex, Inc., Bloomfield, NY, USA) for precise positioning of the endoscope (Karl Storz GmbH & Co. KG, Tuttlingen, Germany), which was held rigidly by a clamp attached to a passive articulated arm (NOGA Ltd., Shlomi, Israel).

Tracking of surgical tools (a rigid pointer and endoscope) was performed using the Vicra infrared tracker (Northern Digital Inc., Waterloo, ON, Canada). A simple pointer tool and pivot calibration provided initial registration to the tracking system. The position of the phantom was tracked using a reference marker rigidly attached to the skull. A rigid body with four infrared markers was attached to the endoscope as shown in Fig. 1, allowing the endoscope to be tracked in real-time, with methods for registration and calibration detailed below.

B. Registration Methods

We compared two registration methods: 1) a conventional tracker-based method in which video and CBCT were registered based on the endoscope pose estimate provided by the tracker, and 2) an image-based method in which video (i.e., 3D SfM point clouds) and 3D CBCT were directly registered. The tracker-based method provides the initialization for the image-based method, thereby yielding an integrated system that is potentially both robust

(coarse pose initialization by the tracker) and precise (fine registration using video and CBCT images directly). It is worth noting that the accuracy of the tracker used does affect the final registration accuracy, as shown in [25]. Fig. 2 shows the relationships and transformations between the pertinent frames of reference.

1) Tracker-Based Registration—In the tracker-based method, the position of the endoscope is determined solely by the tracking system. To compute the position of the endoscope, a pointer tool was first calibrated and used to record fiducial locations on the exterior surface of the rigid phantom. A rigid transformation between these fiducial points localized by the tracking system and the same points segmented in CBCT was computed, providing the transformation ${}^{\text{patient}}T_{\text{reference}}$ and ${}^{\text{CBCT}}T_{\text{patient}}$. A camera calibration was then performed using a checkerboard grid attached to the reference marker. The intrinsic camera parameters—including the first-, third-, and fifth-order radial distortion parameters—and the extrinsic transformation ${}^{\text{endoscope}}T_{\text{camera}}$ were computed using the German Aerospace Agency (DLR) camera calibration toolbox [28]. The optical camera center of the endoscope (O) in Fig. 2 was then computed as shown in (1). As illustrated in Fig. 2 and shown algebraically in (1), the tracker-based method requires a long series of transformations to compute the transformation from the camera to the CBCT. It includes passing through the tracking system, the reference frame of the patient and the endoscope rigid-body. Each of these transformations may add error to the estimated location of the optical center of the endoscope

$${}^{\text{CBCT}}T_{\text{camera}} = ({}^{\text{CBCT}}T_{\text{patient}}) ({}^{\text{patient}}T_{\text{reference}}) \cdot ({}^{\text{tracker}}T_{\text{reference}})^{-1} ({}^{\text{tracker}}T_{\text{endoscope}}) \cdot ({}^{\text{endoscope}}T_{\text{camera}}). \quad (1)$$

2) Image-Based Registration—The image-based method directly uses the video image data itself to register the endoscope images to the CBCT volume. To do so, both the CBCT and the endoscope images require processing. For the endoscopic video, processing involved the pipeline shown in Fig. 3. A pair of images first underwent SIFT feature [29] extraction. The resulting features were matched using SVD-SIFT Match [30] to create candidate correspondences from which the motion between pairs of images was estimated using Adaptive Scale Kernel Consensus (ASKC) [31]. Following triangulation, the feature point cloud was registered to CBCT using the trimmed least-squares method described in [25].

The CBCT images—the volumetric data—are segmented at the air/tissue boundary using a simple intensity-based threshold. Marching cubes [32] is applied and provides a 3D surface at the air/tissue boundary. The CBCT is optionally smoothed before surface extraction to reduce noise in the segmentation, treated in detail in Section II-E2 below. Image-based registration therefore involves a match of the 3D SfM point cloud derived from endoscopic video to the air/tissue surface derived from CBCT.

The top row of Fig. 3 illustrates the conventional tracker-based approach, which provides initialization for the image-based approach illustrated in the bottom row. The diagram highlights in gray the pertinent parameters of each step. The distance from the endoscope optical center to the target (on the CBCT surface) is denoted d . The parameter d was first

measured using tracking software to compute the distance between the registered camera location and to a segmented target. Once an initial measurement was recorded (e.g., 10 mm), changes to the distance to target were made using the linear stage directly. The parameter b is the baseline distance between endoscopic video image pairs and was measured and changed from the linear stage directly. The SIFT detector [29] is applied to the image pair, parameterized by the number of features, N_f . The features detected on the pair of images are matched with SVD-SIFT Match [30]. In this step, we add two additional constraints on correspondences—the distance between features, d_f , and the correlation of the features, c . After initial feature correspondences are established, the motion between the image pair is estimated using ASKC [31], which is similar to Random Sample Consensus (RANSAC) [33] and is parameterized by the number of samples, s , used to achieve consensus. Based on the motion between video image pairs, the features are triangulated to form a 3D point cloud that is registered to the segmented CBCT surface. The entire surface is used in the registration process because the correspondence between the point cloud and the surface is unknown. The registration is a reweighted least squares method characterized by the number of iterations, N_i . Among the image-based method pipeline steps, the parameters of SVD-SIFT match, robust motion estimation and registration were previously investigated in [25].

The image-based registration is potentially affected by parameters governing CBCT image quality—specifically, the radiation dose, D (which affects the level of quantum noise in CBCT image as described in [34]) and optional smoothing, σ , representing the width of a uniform 3D Gaussian filter applied to the CBCT reconstruction to reduce quantum noise at the cost of spatial resolution. The segmentation of the air/tissue CBCT surface is parameterized by the threshold value ($t \sim -500$ HU) and percentage of decimation d_c .

The dependence of image-based video-CBCT registration accuracy was evaluated as a function of each of the aforementioned parameters to elucidate the factors governing the performance of the system. Such measurements provide understanding of the robustness of the proposed system to variation in any particular parameter and offer a guide for future development aimed at improving registration accuracy and computational speed.

C. Analysis of Registration Accuracy

The geometric accuracy of each registration method was assessed in terms related to the target registration error (TRE) [35], which describes the root mean squared (rms) distance between target points (i.e., anatomical points not included in the registration process) transformed by the estimated registration and their corresponding fixed locations. TRE metrics suitable for projective geometry are illustrated in Fig. 4 and described in [36].

We consider two rays emanating from the camera optical center—one containing the target point plane ($\mathbf{t}_{im}$) in the 2D video image and the other containing the target point ($\mathbf{t}_{3D}$) in the 3D CBCT image. From these can be defined the projection distance (PD, in pixels), the angular error (AE, in degrees), and the reprojection distance (RPD, in millimeters) similar to discussion in [37] for X-ray projections. However, unlike X-ray projection imaging, the camera imaging geometry in Fig. 4 describes the imaging plane as a virtual image. In a complete pinhole camera model the image forms upside-down behind the optical center of the camera on the imaging sensor where the pixel size is fixed. The virtual image is

magnified in front of the camera and the pixels are also magnified to match the virtual image size. For this reason, there is no magnification of PD, unlike that in X-ray projection geometry, which is magnified by the source-detector distance (SDD). In place of SDD, we have the focal length of the camera, measured as the distance from the optical center to the image sensor, which was fixed throughout all studies reported below. The registration error can be alternatively described in terms of the AE as shown in Fig. 4, describing the angle between the two rays emanating from the optical center and containing the target points in 2D and 3D.

The metric primarily used below is the RPD, which is the perpendicular distance from the 3D target point ($\mathbf{t}_{3D}$) to the ray $\mathbf{r}$ extending from the camera center ($\mathbf{o}$) through the point on the image plane ($\mathbf{t}_{im}$)

$$RPD = \left\| \mathbf{t}_{3D} - \left(\mathbf{o} + \mathbf{r} \left(\frac{\mathbf{r} \cdot (\mathbf{t}_{3D} - \mathbf{o})}{\mathbf{r} \cdot \mathbf{r}} \right) \right) \right\|. \quad (2)$$

This form of reprojection distance is zero when the registration is perfect and is similar to the RPD defined in [37]. However, unlike RPD under X-ray projection geometry, the RPD for camera geometry is magnified by the distance (d) between the optical center of the camera ($\mathbf{o}$) and the target in 3D ($\mathbf{t}_{3D}$).

D. Evaluation Methodology

Given the number of parameters of the system (nine investigated in detail below), a systematic approach was undertaken to evaluate each parameter first using a rigid phantom to best isolate the parameter and minimize other sources of registration error. To this end, first the camera parameters were evaluated with a rigid phantom, followed by the CBCT image parameters, and finally the registration parameters, and the investigation was then translated to cadaver studies for validation in a more realistic anatomical context.

E. Performance Evaluation in Phantom

The geometric accuracy of image-based registration was evaluated as a function of the following factors of system geometry, image quality, and computational speed.

1) Dependence of Registration Accuracy on Geometric Pose—The effect of geometric pose on RPD was investigated as a function of both the distance to target (d) and the baseline between image pairs (b). The first concerns the magnification of RPD as mentioned above, and the second holds implications for how video frames should be sampled relative to the speed of endoscope motion. We hypothesized that a larger baseline improves the SfM reconstruction (as shown in the vision literature [38]) and thereby improves registration. However, we further hypothesized upper and lower limits to this improvement due to lack of sufficient motion at the lower limit (small b) and lack of features at the upper limit (large b).

2) Dependence of Registration Accuracy on Image Quality—Two principal factors of CBCT image quality were investigated in terms of the effect on video-CBCT registration

accuracy. The first involved the level of quantum noise in CBCT reconstructions, which is directly related (inverse square root dependence) to the radiation dose used in forming the image: reducing dose by a factor of 4 increases noise by a factor of 2. Dose was varied by adjustment of tube current (mA) over the range (0.1–6.5 mA) allowed by the C-arm prototype [10]. Other technique factors were fixed—e.g., 100 kVp, 200 projections and reconstruction by a modified FDK algorithm [39] with a reconstructed voxel size of 0.3 mm $\times$ 0.3 mm $\times$ 0.3 mm. The corresponding range in dose, D , was 0.5–38.4 mGy.

Another method to reduce the dose to the patient is to limit the number of projections acquired. Fewer projections would increase streaking associated with view sampling effects. In this case the CBCT images were acquired with our standard scan protocol, which involves 200 projection images, thus the level of streak artifacts evident in this paper is typical of that achieved with the prototype C-arm. While the number of projections is potentially a parameter to be investigated (motivated primarily by reduced radiation dose and reduced reconstruction time), previous work [12] shows that for the current system and application, low dose acquisition is better achieved by mA reduction (rather than reduction of the number of projections).

The second factor concerned spatial resolution, which can be freely adjusted in trade-off with quantum noise. As a simple investigation of postreconstruction smoothing, we applied a 3D Gaussian smoothing filter to the CBCT images (kernel width, σ , ranging 0–5 voxels), thereby blurring the images in a manner that reduced image noise while sacrificing detail and fine image features. We hypothesized that video-CBCT registration accuracy would suffer at low dose levels (i.e., high noise levels) and would benefit from a certain level of image smoothing (perhaps around an optimal trade-off between noise and spatial resolution).

3) Dependence of Registration Accuracy on Input Data Size—We further investigated the dependence of registration accuracy on algorithmic parameters that have a direct bearing on computational speed. Two primary factors were considered: the number of input features (N_f) and the number of polygons (N_{poly}) used in registration, as these determine the amount of time needed to compute the registration. The number of polygons used is equal to the initial number of polygons times the decimation percentage d_c , i.e., $N_{\text{poly}} = N_{\text{poly}}^{\text{init}}(1 - d_c)$. We hypothesized that fewer features and/or polygons will improve runtime and decrease accuracy. Quantifying the steepness of this dependence in the measurements reported below provided a guide to selecting parameters that satisfied requirements in geometric accuracy within the runtime constraints of an implementation that would be practical for eventual clinical use.

F. Performance Evaluation in Cadaver

The system was deployed in pre-clinical studies in which a fellowship-trained neurosurgeon performed endoscopic skull base target resection in a cadaveric head specimen as illustrated in Fig. 5. The cadaver was fixed using a mixture of phenol and formalin. After the fixation process the cadaver remains slightly moist. This preparation allows for a cadaver that remains flexible to closely simulate actual tissue, and the reflectance of (moist) endonasal

tissues is fairly realistic. The setup included the same components labeled in the benchtop setup of Fig. 1, along with the mobile C-arm prototype for intraoperative CBCT and the surgical navigation interface [40] combining real-time tracking and video-CBCT registration. The specimen was rigidly secured in a Mayfield skull clamp (Integra Corp., Plainsboro, NJ, USA) mounted to an X-ray compatible carbon fiber operating table.

The geometric accuracy of image-based and tracker-based video-CBCT registration was quantified by measurement of RPD in the cadaver. Unambiguous target points were created by gently piercing a 27-gauge needle through the thin layer of bone at the tuberculum sellae and the floor of the sella turcica. The targets were large enough (~1 mm diameter) to be identified clearly in both CBCT and endoscopic video. Each target was manually identified in CBCT to define its location (repeated localization providing sub-voxel accuracy in mean target location). The targets were also manually identified for each video frame of endoscope video in which they appeared. We grouped a total of 31 independent views of the four targets (pinholes) created on the clivus in one cadaveric specimen. A passive articulated arm was used to support the endoscope, facilitate stable recording of video data and avoid potential synchronization errors between the tracker and the endoscope.

III. Results

The results below first summarize the performance of the image-based method measured as a function of factors of system geometry, image quality, and computational load, providing quantitation of hypothesized trends and a guide to parameter selection. The image-based method is then tested in a pre-clinical cadaver study in comparison to conventional tracker-based registration. The camera calibration used for both the tracker-based and image-based registration had 0.8 pixels mean reprojection error. The error of the tracking system (NDI Virca) used for both registration methods is reported to have localization error of 0.25 mm [41]. To evaluate the (intra-observer) reproducibility in the definition of $\mathbf{t}_{im}$ and $\mathbf{t}_{3D}$ point selection, both $\mathbf{t}_{im}$ and $\mathbf{t}_{3D}$ were segmented ten times in independent trials for each fiducial, showing an overall standard deviation of 1.48 pixels and 0.06 mm, respectively. Projecting both of these through the RPD equation at an distance to target of $d = 15$ mm yields an increase of 0.170 mm, which is well below the RPD reported below.

A. Performance Evaluation in Phantom

1) Dependence of Registration Accuracy on Geometric Pose—As detailed above, the rigid red skull phantom was used in analysis of registration accuracy versus factors of system geometry. Fig. 6 presents the RPD measured as a function of the baseline distance and camera-to-target distance. The effects of magnification are evident, as shown by the positive slope in RPD (d), consistent with the stated hypothesis. Another clear trend is that larger baseline distances between video image pairs improve the RPD over the range shown. The linear fit in each case follows the model $\sin(AE) * d = RPD$. Measurements at $b = 2.0$ mm and $d = 10$ mm were not possible, because the images were too disparate. In each case, the image-based video-CBCT registration shows improved accuracy in comparison to the conventional tracker-based registration.

Fig. 7 shows the measurements of RPD as a function of image pair baseline distance (at a fixed distance to target of $d = 15$ mm). Note that the baseline distance refers to in-plane motion along the x-axis of the image plane. The data suggest an operating range of 0.4–4 mm within which image-based registration outperforms tracker-based registration. Below this range, there was insufficient motion between frames to reconstruct a reliable point cloud, and above this range, there was an insufficient number of features to reliably compute correspondence and 3D registration. The lower limit of the endoscope movement is related to the epipolar constraint used to solve the structure from motion problem. However, the upper limit of the endoscope movement is a consequence of the chosen matching procedure.

As shown in Fig. 8, RPD measured as a function of baseline distance perpendicular to the imaging plane (i.e., along the z-axis) exhibits a larger operating range of 0.6–9 mm. The larger range is primarily attributed to the larger number of features remaining visible during endoscope motion. As in Fig. 7, below the lower limit of the operating range, there was insufficient motion to compute a reliable point cloud, and above the upper limit there was an insufficient number of features.

The results of Figs. 7 and 8 hold implications for the selection of video image pairs relative to the speed of endoscope motion. Using image pair SfM, the side-to-side motion should be no faster than 4 mm for every 33 ms (the typical frame rate of the video camera), corresponding to a velocity of 120 mm/s. In practice, the motion of the endoscope is much slower (~ 15 mm/s) when the surgeon is observing the surgical site, although it can be quite fast (~ 110 mm/s) when removing and reinserting the endoscope. At the other extreme ($b \sim 0.3$ mm, corresponding to ~ 9 mm/s), image pair SfM breaks down if the endoscope motion is too slow (i.e., nearly stationary). This may not be relevant in the context of freehand endoscopy but may become important for endoscopy with a tool-holder or robotic assistant, suggesting an additional “loop” in Fig. 3 by which the tracker informs the image-based registration system how much the endoscope has moved, and image pairs are selected not necessarily from successive frames, but at frames separated by a baseline distance within the operating range. However, this case would only occur if there were no motion during the initial registration, after which the camera location can be tracked with 2D-3D feature correspondences.

Fig. 9 shows a representative comparison of tracker-based and image-based video-CBCT registration accuracy (using a fixed 1 mm baseline distance for the latter). In this example, the mean PD for the tracker-based method was 19.3 pixels (with an interquartile range of 16.4 pixels), and the image-based method had a mean PD of 14.9 pixels (with an interquartile range of 10.5 pixels). The columns in Fig. 9 show: the air/tissue surface segmented from CBCT; an actual video image view overlaid by targets identified in the video and estimated target point locations of the CBCT; and the spatial distribution of the RPD interpolated from the measurement points over the entire field of view (interpolation by a multi-quadric radial basis function). The fairly small difference between the two methods is expected in this case because of the completely rigid phantom, but the results are consistent with the motivation for improved registration accuracy using a tracker-based initialization and an image-based refinement.

2) Dependence of Registration Accuracy on Image Quality—As detailed above, the anthropomorphic head phantom was used in analysis of registration accuracy versus factors of image quality, focusing on effects of image noise (dose) and spatial resolution. Fig. 10 shows the effect of dose, where we hypothesized that lower dose would degrade (i.e., increase) RPD due to increased noise in the CBCT air/tissue surface segmentation. While there is an appreciable increase in image noise at the lowest dose levels, we found that dose imparted little or no significant effect on RPD over the range considered. While the hypothesis may be correct, it does not apply over the range of dose deliverable by the CBCT C-arm. This is an advantageous finding in that image-based video-CBCT registration may be performed without loss of accuracy even at the lowest dose levels (e.g., 0.6 mGy) of intra-operative scanning. This corresponds to approximately 1/100th the dose of a typical diagnostic CT of the head (~ 52.0 mGy) [42].

Fig. 11 shows the effect of 3D image smoothing on video-CBCT registration accuracy. In these results, dose was fixed at a nominal value of 24.2 mGy. As hypothesized, a (fairly weak) trend is exhibited in which RPD increases (degrades) at high levels of image smoothing ($\sigma > \sim 2.5$ voxels) for which features of the air/tissue surface segmentation are lost. Following from the results in Fig. 11, where there was no observed dependency of RPD on dose over the range considered, there was similarly no optimum in smoothing—i.e., no value of σ below which RPD increased due to image noise. Under these circumstances, there appeared to be no benefit to image smoothing, although one might hypothesize conditions of increased image noise (e.g., still lower dose levels and/or large body sites) for which smoothing in the range ($0 < \sigma < 2.5$) may be beneficial to video-CBCT registration accuracy. In the results of both Figs. 10 and 11, image-based registration outperformed tracker-based registration at the nominal settings.

Fig. 12 presents an example endoscope view in the anthropomorphic phantom with tracker-based and image-based video-CBCT registration (nominal 2 mm baseline, 38.4 mGy, and no smoothing). In this case the PD for the tracker-based method exhibited a mean of 27.2 pixels (interquartile range of 11.9 pixels), and the image-based method gave a mean PD of 20.2 pixels (interquartile range of 6.3 pixels). Analogous to the results of Fig. 9, the results show a measurable improvement for the image-based registration, although the effect is fairly small ($\sim 0.25 - 0.59$ mm) due to the use of a highly rigid phantom.

3) Dependence of Registration Accuracy on Input Data Size—The effect of parameters associated with image-based registration runtime (viz., the number of features and the number of polygons) was investigated to determine the potential for increased computational speed without loss of registration accuracy. Fig. 13 summarizes the results. The number of features [Fig. 13(a)] demonstrated a fairly weak effect on runtime, governed by the search time of the kd-tree, $O(N_{\text{poly}} \log(N_{\text{poly}}))$ [43], where N_{poly} is the number of polygons in the tree. Since the search is called at every reconstructed feature point, the runtime performance is approximately $O(N_f N_{\text{poly}} \log(N_{\text{poly}}))$, where N_f is the number of features. In Fig. 13(a), N_{poly} was held constant ($\sim 3.5 \times 10^5$ polygons) while N_f was varied. The effect on runtime is fairly small above $N_f > \sim 20$. Fig. 13(b) shows that the number of features had a significant effect on RPD—specifically a sharp increase in RPD for fewer

than ~ 40 features, below which registration fails. The result is consistent with our hypothesis and suggests a nominal, stable operating condition of $N_f > \sim 100$.

Fig. 13(c) shows the strong effect of N_{poly} on runtime—approximately $O(N_f N_{\text{poly}} \log(N_{\text{poly}}))$ where N_{poly} is much larger than N_f and construction time of the kd-tree being $O(N_{\text{poly}} \log(N_{\text{poly}}))$. The runtime increases roughly in linear proportion to N_{poly} over the range examined. However, Fig. 13(d) shows that the RPD performance degrades rapidly with reduction in the number of polygons, with registration accuracy becoming worse than that of the tracker-based method for $N_{\text{poly}} > \sim 3.5 \times 10^5$.

B. Performance Evaluation in Cadaver

The results of Figs. 6–13 guided selection of nominal operating parameters—e.g., $b \sim 2$ mm, $d \sim 15$ mm, $D \sim 24.2$ mGy, $\sigma = 0$ voxels, N_f unchanged (i.e., no reduction in the number of features), and $d_c = 0$. Following such characterization of the underlying factors of registration and runtime performance, we evaluated the geometric accuracy of the system in a cadaver model. Fig. 14 summarizes the overall magnitude and range of RPD measurements in the cadaver study. The mean RPD for the tracker-based registration method was 1.82 mm (with 1.09 mm first quartile and 1.25 mm range). By comparison, the mean RPD for image-based registration was 1.28 mm (with 0.66 mm first quartile and 1.06 mm range). The improvement in RPD over the tracker-based method was statistically significant ($p < 0.001$). With regard to the statistical analysis, we previously showed that the distribution of RPD is not normal, so we applied the transformation to normality as described in [25]. We then applied a Student's t-test for statistical significance, and the data were trimmed of outliers based on the interquartile range on the transformed distribution. Two outliers were identified—one due to incorrect scale estimation and one due to an anomalous failure to converge.

Fig. 15 illustrates the results in images of the cadaver, showing example tracker-based and image-based registration from the 31 cases used in the experiment. In this instance, the PD for the tracker-based method had a mean of 25.3 pixels (with interquartile range of 1.3 pixels) and the image-based method had a mean of 12.9 pixels (interquartile range of 4.1 pixels). The overlay of video (red) and estimated CBCT (blue) target points in the video image shows a measurable improvement in localization accuracy, and the map of RPD(x, y) over the endoscope field of view suggests a significant improvement in video-CBCT registration accuracy.

IV. Discussion and Conclusion

The experiments detailed in this work quantify the measured dependencies of pose, image quality and input data size on the performance of image-based video-CBCT registration. We found, as supported by the computer vision literature, that an ample baseline distance between image pairs is required to have an accurate reconstruction and registration—specifically, $0.5 \text{ mm} < b < 4$. Furthermore, the system was found to be robust to various factors of CBCT image quality, including dose (down to the lowest levels deliverable by the CBCT C-arm prototype, $\sim 1/100$ th the dose of a diagnostic CT) and spatial resolution (additional image smoothing did not improve registration). Parameters affecting registration

runtime showed a weak dependence on the number of features (but a susceptibility to registration failure below $N_f \sim 40$) and a stronger dependence on the number of polygons ($N_{\text{poly}} > \sim 3.5 \times 10^5$). With an understanding of the factors governing video-CBCT registration accuracy, we conducted a cadaver study that demonstrated statistically significant improvement in registration accuracy for the image-based approach in comparison to the conventional tracker-based approach (Figs. 14–15).

The effect of system geometry on video-CBCT registration accuracy implies certain limitations on the speed at which the endoscope can be moved during acquisition of an image pair—specifically, less than 120 mm/s side-to-side (x-y) and 270 mm/s front-to-back (z), both of which are well above typical freehand motion of ~ 5 –50 mm/s. Additionally, motion in either the x-y or z directions is adequate for SfM point cloud generation and accurate registration provided that the motion provides an ample baseline distance between image pairs—approximately 0.4–0.6 mm or greater (up to ~ 4 –5 mm).

The effect of CBCT image quality on registration accuracy demonstrated that dose (i.e., noise) and spatial resolution in CBCT reconstructions exhibited little effect over the range considered. This showed the potential to operate at the lowest dose settings available on the C-arm (0.6 mGy) while still providing accurate registration. Other image quality factors not considered in the current work could diminish performance—e.g., streak artifacts from metallic components in the image, which would potentially confound a simple air/tissue surface segmentation in CBCT.

The effect of input data size on runtime and registration accuracy showed that it is possible to achieve some improvement in computational speed by reducing the input data; however, further work is needed to accomplish this in a way that does not adversely affect registration accuracy.

While a fair, direct comparison between closed-source methods is difficult to achieve and is beyond the scope of this current work, we offer a comparison of results reported in the literature [21], [22]. It is worth noting that such a comparison would best be conducted via multi-institutional collaboration using shared data/phantoms or via open-source/open-dataset so that each method may be applied fairly. In Luó *et al.* [21], an analysis of the pose error between an EM tracker and the video-registration was presented, with the EM tracker used as the ground-truth data. A mean translation error was reported between 0.679 and 0.875 mm with orientation error approximately 0.5° . For context, error at this level is within the 95% confidence interval of the Aurora EM tracker (Northern Digital Inc., Waterloo, ON, Canada) [44]; however, the model of the EM tracker was not reported. Similarly, the results presented above also surpass the tracker-base initialization. In Higgins *et al.* [22], the error was reported as the perpendicular distance between the biopsy needle and a metal bead target (conceptually similar to RPD), with an overall mean error of 1.97 mm, similar to the RPD reported above.

The results reported above help guide parameter selection in the implementation of image-based video-CBCT registration suitable for clinical use, and they provide the first demonstration of improved geometric accuracy in video-CBCT registration over the

conventional tracker-based approach. Of course, the work is not without limitations, areas of future improvement, and further investigation. One area of improvement in future work is the avoidance local minima in the optimization to allow for further refinement of the solution, as evident in the lack of additional improvement over the track-based method in the first phantom study. ASCK may be adapted to address the robustness of registration in future work. The current method performs SfM point cloud reconstruction from a single pair of video images and performs the 3D point-surface registration from only that one pair. This approach can be improved through incorporation of a collection of images with a bundle adjustment in the point cloud reconstruction. Clinical implementation would likely require tighter synchronization of the surgical tracker and the endoscopic video, mitigated in the experiments herein by using a stable tool holder for the endoscope and acquiring images using a static step and shoot. Furthermore, the phantoms and cadavers do not reflect the complexities and anatomical deformation in real surgical situations. Additionally, the image-based method presented in this work is best suited for applications in which rigid registration is sufficient to align the reconstructed anatomy to the intraoperative CBCT. For this reason, we focused on the registration of structures of the skull base, in which the motion, deformation and resection during surgery can be captured with an up-to-date CBCT scan. Note that this is a potentially significant improvement over the conventional context of tracking in preoperative CT, which fails to capture such intraoperative deformations and tissue excisions. Alternative advanced methods such as nonrigid SfM [45] could be employed to reconstruct and register more dynamic deformable structures (i.e., deformations occurring in the short time interval between the last CBCT and the current video acquisitions). Other areas of improvement include a streamlined camera calibration process, such as that reported in [46].

A final point of methodological note is the fact that image-based registration in the context of up-to-date intra-operative CBCT properly reflects and utilizes the current state of the anatomy, including surgical excisions of the anterior sphenoid sinus. Such is possible only to a limited extent in the context of registration to preoperative CT, where the real promise of image-based registration is only partly realized—i.e., the image features do not necessarily match once the intraoperative scene is perturbed due to tissue deformation and/or excision. Under conditions in which the anatomy is dramatically altered from the preoperative state, high-precision image-based registration would only be feasible in the context of intraoperative imaging, such as CBCT. A separate study involving comparison of image-based registration accuracy in the context of preoperative versus intraoperative imaging is planned for future work.

The limitations of video-CT registration using the traditional approach of external tracking are fairly well recognized [17], [47]. This work reports a two-fold advance: first via registration in the context of up-to-date CBCT (which properly reflects intraoperative change) and second via image-based registration that directly utilizes the video and CBCT data to perform a precise registration. Registration in preoperative CT alone is difficult or impossible following significant resection or deformation in surgery, motivating intraoperative imaging to utilize the wealth of information in the video data. The method reported above enables a higher level of accuracy than the conventional tracker-based approach, offering a robust real-time initialization by the tracker followed by a precise

image-based refinement to sub-mm accuracy. Furthermore, the work demonstrates that CBCT can provide the intraoperative imaging at a dose that is sufficiently low (~ 0.6 mGy per scan versus 52.0 mGy) to permit repeat intraoperative scanning and allow accurate registration in an up-to-date anatomical context.

Looking forward, we envision the approach implemented in a tightly integrated hybrid configuration whereby the tracker provides robust initialization to the registration, which is in turn refined as video data arrives and is processed in near real-time. The system would need to detect failures of the image-based method and fall back to the tracker-based method without disruption of the video stream and data overlay. As the registration is refined and evolves, however, the display would provide the most accurate visualization available at that time. This system should also eventually incorporate image-based tracking and bundle adjustment of collections of images in registration to the CBCT.

Acknowledgments

This work was supported principally by the National Institutes of Health under Grant R01-CA127444 with further support provided by the NSF ERC Grant EEC9731748, the Link Fellowship in Simulation and Training and the Medtronic Computer-Aided Surgery Research Award.

Academic-industry partnership in the development and application of the mobile C-arm prototype for CBCT is acknowledged, in particular collaborators at Siemens Healthcare (Erlangen Germany)—Dr. R. Graumann, Dr. G. Kleinszig, and Dr. C. Schmidgunst. Cadaver studies were performed at the Johns Hopkins Medical Institute, Minimally Invasive Surgical Training Center, with support and collaboration from Dr. M. Marohn and S. Eller (Department of Surgery, Johns Hopkins University) gratefully acknowledged.

References

1. Kassam AB, Prevedello DM, Carrau RL, Snyderman CH, Thomas A, Gardner P, Zanation A, Duz B, Stefkó ST, Byers K, Horowitz MB. Endoscopic endonasal skull base surgery: Analysis of complications in the authors' initial 800 patients. *J Neurosurg.* 2011; 114(6):1544–1568. [PubMed: 21166570]
2. Snyderman CH, Pant H, Carrau RL, Prevedello D, Gardner P, Kassam AB. What are the limits of endoscopic sinus surgery?: The expanded endonasal approach to the skull base. *Keio J Med.* 2009; 58(3):152–160. [PubMed: 19826209]
3. Heermann R, Schwab B, Issing P, Haupt C, Hempel C, Lenarz T. Image-guided surgery of the anterior skull base. *Acta Oto-Laryngologica.* 2001; 121(8):973–978. [PubMed: 11813906]
4. Wiltfang J, Rupperecht S, Ganslandt O, Nimsky C, Keßler P, Schultze-Mosgau S, Fahlbusch R, Neukam FW. Intraoperative image-guided surgery of the lateral and anterior skull base in patients with tumors or trauma. *Skull Base.* 2003; 13(01):21–29. [PubMed: 15912156]
5. Fried M, Parikh S, Sadoughi B. Image-guidance for endoscopic sinus surgery. *Laryngoscope.* 2008; 118(7):1287–1292. [PubMed: 18391767]
6. Ledderose GJ, Hagedorn H, Spiegl K, Leunig A, Stelter K. Image guided surgery of the lateral skull base: Testing a new dental splint registration device. *Comput Aid Surg.* 2012; 17(1):13–20.
7. Grauvogel TD, Soteriou E, Metzger MC, Berlis A, Maier W. Influence of different registration modalities on navigation accuracy in ear, nose, and throat surgery depending on the surgical field. *Laryngoscope.* 2010; 120(5):881–888. [PubMed: 20422680]
8. Strauß G, Koulechov K, Röttger S, Bahner J, Trantakis C, Hofer M, Korb W, Burgert O, Meixensberger J, Manzey D, Dietz A, Lüth T. Evaluation of a navigation system for ENT with surgical efficiency criteria. *Laryngoscope.* 2006; 116(4):564–572. [PubMed: 16585860]
9. Cohen N, Kennedy D. Revision endoscopic sinus surgery. *Otolaryngol Clin N Am.* 2006; 39(3): 417–435.

10. Siewerdsen JH, Daly MJ, Chan H, Nithiananthan S, Hamming N, Irish J. Computer assisted head and neck, and ENT surgery. *Int J Comput Assist Radiol Surg.* 2009; 4:71–80. [PubMed: 20033604]
11. Daly MJ, Siewerdsen JH, Moseley DJ, Jaffray DA, Irish JC. 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. [PubMed: 17089842]
12. Schafer S, Nithiananthan S, Mirota DJ, Uneri A, Stayman JW, Zbijewski W, Schmidgunst C, Kleinszig G, Khanna AJ, Siewerdsen JH. Mobile c-arm cone-beam CT for guidance of spine surgery: Image quality, radiation dose, and integration with interventional guidance. *Med Phys.* 2011; 38(8):4563–4574. [PubMed: 21928628]
13. Caire F, Gantois C, Torny F, Ranoux D, Maubon A, Moreau JJ. Intraoperative use of the medtronic o-arm for deep brain stimulation procedures. *Stereotact Funct Neurosurg.* 2010; 88(2):109–114.
14. Jackman AH, Palmer JN, Chiu AG, Kennedy DW. Use of intraoperative CT scanning in endoscopic sinus surgery: A preliminary report. *Am J Rhinol.* 2008; 22(2):170–174. [PubMed: 18416975]
15. Carelsen B, Grolman W, Tange R, Streekstra G, Van Kemenade P, Jansen R, Freling N, White M, Maat B, Fokkens W. Cochlear implant electrode array insertion monitoring with intra-operative 3d rotational X-ray. *Clin Otolaryngol.* 2007; 32(1):46–50. [PubMed: 17298312]
16. Siewerdsen JH, Moseley DJ, Burch S, Bisland SK, Bogaards A, Wilson BC, Jaffray DA. Volume CT with a flat-panel detector on a mobile, isocentric c-arm: Pre-clinical investigation in guidance of minimally invasive surgery. *Med Phys.* 2005; 32(1):241–254. [PubMed: 15719975]
17. Liu S, Gutierrez L, Stanton D. Quantitative evaluation for accumulative calibration error and video-CT registration errors in electromagnetic-tracked endoscopy. *Int J Comput Assist Radiol Surg.* 2011; 6:407–419. [PubMed: 20714934]
18. Daly MJ, Chan H, Prisman E, Vescan A, Nithiananthan S, Qiu J, Weersink R, Irish JC, Siewerdsen JH, Wong KH, Miga MI. Fusion of intraoperative cone-beam CT and endoscopic video for image-guided procedures. *Med Imag 2011: Visualizat, Image-Guided Procedures Model.* 2010; 7625(1): 762503.
19. Lapeer R, Chen MS, Gonzalez G, Linney A, Alusi G. Image-enhanced surgical navigation for endoscopic sinus surgery: Evaluating calibration, registration and tracking. *Int J Med Robot Comput Assist Surg.* 2008; 4(1):32–45.
20. Shahidi R, Bax M, Maurer C Jr, Johnson J, Wilkinson E, Wang B, West J, Citardi M, Manwaring K, Khadem R. Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. *IEEE Trans Med Imag.* Dec; 2002 21(12):1524–1535.
21. Luó X, Feuerstein M, Deguchi D, Kitasaka T, Takabatake H, Mori K. Development and comparison of new hybrid motion tracking for bronchoscopic navigation. *Med Image Anal.* 2012; 16(3):577–596. [PubMed: 21334250]
22. Higgins WE, Helferty JP, Lu K, Merritt SA, Rai L, Yu KC. 3d CT-video fusion for image-guided bronchoscopy. *Comput Med Imag Graph.* 2008; 32(3):159–173.
23. Mirota DJ, Uneri A, Schafer S, Nithiananthan S, Reh DD, Gallia GL, Taylor RH, Hager GD, Siewerdsen JH, Wong KH, DRH. High-accuracy 3d image-based registration of endoscopic video to c-arm cone-beam CT for image-guided skull base surgery. *Med Imag 2011: Visualizat, Image-Guided Procedures Model.* 2011; 7964(1):79 640J-1–79 640J-10.
24. Longuet-Higgins HC. Review lecture: The perception of music. *Proc R Soc London Ser B, Biol Sci.* 1979; 205(1160):307–22. [PubMed: 41250]
25. Mirota DJ, Wang H, Taylor RH, Ishii M, Gallia GL, Hager GD. A system for video-based navigation for endoscopic endonasal skull base surgery. *IEEE Trans Med Imag.* Apr; 2012 31(4): 963–976.
26. Prisman E, Daly MJ, Chan H, Siewerdsen JH, Vescan A, Irish JC. Real-time tracking and virtual endoscopy in cone-beam CT-guided surgery of the sinuses and skull base in a cadaver model. *Int Forum Allergy Rhinol.* 2011; 1(1):70–77. [PubMed: 22287311]
27. Vescan AD, Chan H, Daly MJ, Witterick I, Irish JC, Siewerdsen JH, Miga MI, Wong KH. C-arm cone beam CT guidance of sinus and skull base surgery: Quantitative surgical performance

- evaluation and development of a novel high-fidelity phantom. *Med Imag 2009: Visualizat, Image-Guided Procedures Model*. 2009; 7261(1):72610L.
28. Strobl, KH.; Sepp, W.; Fuchs, S.; Paredes, C.; Arbter, K. Institute of Robotics and Mechatronics, German Aerospace Center (DLR) DLR CalDe and DLR CalLab; Oberpfaffenhofen, Germany. Oct. 15, 2010;
 29. Lowe DG. Distinctive image features from scale-invariant key-points. *Int J Comput Vis*. 2004; 60:91–110.
 30. Delponte E, Isgrò F, Odone F, Verri A. Svd-matching using sift features. *Proc Int Conf Vis, Video Graph*. Jul.2005 :125–132.
 31. Wang H, Mirota D, Hager GD. A generalized kernel consensus based robust estimator. *IEEE Trans Pattern Anal Mach Intell*. Jan; 2010 32(1):178–184. [PubMed: 19926908]
 32. Newman TS, Yi H. A survey of the marching cubes algorithm. *Comput Graph*. 2006; 30(5):854–879.
 33. Fischler MA, Bolles RC. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM*. Jun.1981 24:381–395.
 34. Chou C, Barrett HH. Gamma-ray imaging in Fourier space. *Opt Lett*. Nov; 1978 3(5):187–189. [PubMed: 19684741]
 35. Fitzpatrick J, West J, Maurer J, CR. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imag*. Oct; 1998 17(5):694–702.
 36. Mirota, D.; Taylor, RH.; Ishii, M.; Hager, GD. Direct endoscopic video registration for sinus surgery. *Proc. SPIE Med. Imag. 2009: Visualizat., Image-Guided Procedures Model*; Feb. 2009; p. 72 612K-1-72 612K-8.
 37. van de Kraats EB, Penney GP, Tomaževi D, van Walsum T, Niessen WJ. Standardized evaluation methodology for 2-D-3-D registration. *IEEE Trans Med Imag*. Sep; 2005 24(9):1177–1189.
 38. Ma, Y.; Soatto, S.; Kosecka, J.; Sastry, S. *An Invitation to 3-D Vision*. New York: Springer Verlag; 2004. p. 107-227.
 39. Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. *J Opt Soc Am A*. 1984; 1(6): 612–619.
 40. Uneri A, Schafer S, Mirota DJ, Nithiananthan S, Otake Y, Taylor RH, Siewerdsen JH. Trek: An integrated system architecture for intraoperative cone-beam CT-guided surgery. *Int J Comput Assist Radiol Surg*. 2012; 7:159–173. [PubMed: 21744085]
 41. Polaris family of optical tracking systems—Specifications Oct. 30, 2012.
 42. Pantos I, Thalassinou S, Argentos S, Kelekis NL, Panayiotakis G, Efstathopoulos EP. Adult patient radiation doses from noncardiac CT examinations: A review of published results. *Br J Radiol*. 2011; 84(1000):293–303. [PubMed: 21266399]
 43. Ooi BC, Sacks-Davis R, McDonnell KJ. Spatial indexing in binary decomposition and spatial bounding. *Inf Syst*. 1991; 16(2):211–237.
 44. Aurora electromagnetic 3d tracking system accuracy and measurement volume Oct. 26, 2012, Northern Digital.
 45. Torresani L, Hertzmann A, Bregler C. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Trans Pattern Anal Mach Intell*. May; 2008 30(5):878–892. [PubMed: 18369256]
 46. Liu WP, Mirota DJ, Uneri A, Otake Y, Hager G, Reh DD, Ishii M, Gallia GL, Siewerdsen JH, DRH, Wong KH. A clinical pilot study of a modular video-CT augmentation system for image-guided skull base surgery. *SPIE Med Imag 2012: Image-Guided Procedures, Robotic Intervent Model*. 2012; 8316(1):831633.
 47. Schulze F, Bühler K, Neubauer A, Kanitsar A, Holton L, Wolfsberger S. Intra-operative virtual endoscopy for image guided endonasal transsphenoidal pituitary surgery. *Int J Comput Assist Radiol Surg*. 2010; 5:143–154. [PubMed: 20033497]

Fig. 1.

Benchtop setup for measurement of video-CBCT registration accuracy. The benchtop system is shown at (left), with details of various components shown at (right). Labels refer to: N (the tracker), CBCT (the CBCT data), R (the reference marker), X (the pointer), E (the endoscope), O (the optical center of the endoscope), and P (the phantom simulating a patient undergoing endoscopic skull base surgery).

Fig. 3.

System data flow diagram. Parameters of each process are highlighted in gray: (b) baseline distance between video images; (D, σ) dose and smoothing associated with the CBCT image; (N_f) the number of SIFT features; (d_c) the amount of decimation in the CBCT surface segmentation; (d_f, c) the distance and correlation between features; (s) the number of samples in motion estimation; (d) the distance from the endoscope to the target in the $imageT_{world}$ transformation; (N_i) the number of iterations in the 3D-3D video-CBCT registration.

Fig. 4.

The relationship between different measurements of TRE. The variable $\mathbf{o}$ is the optical center of the camera. The variable $\mathbf{t}_{im}$ is the target in the image and $\mathbf{t}_{3D}$ is the target in the CBCT volume. AE, PD, and RPD are the different measures of TRE in degrees, pixels, and millimeters, respectively.

Fig. 5. Cadaver study experimental setup. The labels refer to: N (the tracker), CBCT (the CBCT data), R (the reference marker), X (the pointer), E (the endoscope), O (the optical center of the endoscope), and P (the phantom—i.e., cadaveric head).

Fig. 6.

Registration error (RPD) increases as a function of distance to target (d) as a result of geometric magnification in projective geometry. The behavior is linear with the sine of the AE. The measurements demonstrate improved registration accuracy for the image-based approach compared to the tracker-based approach at all values of target distance (d) and image pair baseline distance (b). Throughout this manuscript, boxplots show the median (horizontal line), first and third quartile (lower and upper bounds of the rectangular box), and range (whiskers) of the data.

Fig. 7.

Registration error (RPD) measured as a function of the baseline between video image pairs. Camera-to-target distance (d) was fixed at 15 mm. The data suggest an operating range of 0.4–4 mm in-plane motion (indicated with blue vertical lines). The performance of conventional tracker-based registration is superimposed as green horizontal median and range lines. Within the operating range $0.4 \leq b \leq 4$ mm, the image-based approach outperforms the tracker-based approach.

Fig. 8.

Image-based registration method with z-axis motion at a distance of 15 mm shows the larger operation baseline of the out-of-plane motion from 0.6 to 9 mm (indicated with blue vertical lines).

Fig. 9.

Example comparison of video-CBCT registration using the conventional tracker-based method (top) and the proposed image-based method (bottom). Each row corresponds to the rigid red skull phantom. The first column shows the air/tissue isosurface segmented from CBCT. The second column is an endoscope video image overlaid by the video (red) and estimated CBCT (blue) location of target points. The third column shows the RPD interpolated over the endoscope field of view, illustrating the improvement in registration accuracy for the image-based method, as in Figs. 7 and 8.

Fig. 10.

Effect of dose (i.e., CBCT image noise) on video-CBCT registration accuracy. Over the range of dose considered, there is negligible effect on RPD. Although there is a visible increase in CBCT quantum noise at the lowest dose levels (fluctuations visible in both the axial slice images and the air/tissue surface segmentation), the surface registration appears to be robust down to the lowest dose deliverable by the C-arm. In each case, the image-based registration approach again outperformed tracker-based registration.

Fig. 11.

Effect of spatial resolution (i.e., a 3D image smoothing kernel) on video-CBCT registration accuracy. Smoothing is seen to reduce CBCT image noise at the cost of spatial resolution, and a small increase in RPD is observed over the range considered. Only at the highest levels of smoothing (major loss of image features in the segmented air/tissue boundary) does the RPD for the image-based degrade to a level at or greater than that of the tracker-based approach.

Fig. 12.

Example frame comparing tracker-based (top) and image-based (bottom) methods of video-CBCT registration in the anthropomorphic phantom. Dose and spatial resolution are fixed ($D = 38.4$ mGy and $\sigma = 0$ voxels). The first column shows the air/tissue isosurface segmented from CBCT. The second column shows an endoscope video image overlaid with video (red) and estimated CBCT (blue) target point locations. The third column shows the RPD interpolated over the endoscope field of view.

Fig. 13.

Computational load, runtime, and registration accuracy. (a) Reducing the number of features exhibits a minimal effect on runtime performance above a minimum of ~ 20 features, and (b) registration accuracy shows a fairly sharp threshold at ~ 40 features below which the image-based registration essentially fails. (c) Runtime exhibits a linear dependence on the number of polygons, and (d) arbitrary reduction in the number of polygons imparts a measureable increase in RPD (with performance inferior to the conventional tracker-based approach for fewer than $\sim 3.5 \times 10^5$ polygons).

Fig. 14.

Comparison of tracker-based and image-based registration accuracy in a cadaver study. A statistically significant improvement ($p < 0.001$) in RPD was measured for the image-based approach (mean RPD = 1.28 mm) in comparison to the tracker-based approach (mean RPD = 1.82 mm).

Fig. 15.

Example result of video-CBCT registration in cadaver. The tracker-based (top) and image-based (bottom) methods are overlaid by video (red) and estimated CBCT (blue) locations of four target points (pinholes pricked in the tuberculum sellae and floor of the sella turcica). The first column shows the air/tissue isosurface segmented from CBCT. The second column shows an example endoscope image overlaid with video and estimated CBCT target locations. The third column shows the RPD interpolated over the endoscope field of view, suggesting a modest but significant improvement in registration accuracy for the image-based approach.