Evaluation and Stability Analysis of Video-Based Navigation System for Functional Endoscopic Sinus Surgery on In-Vivo Clinical Data

Simon Leonard, Ayushi Sinha, Austin Reiter, Masaru Ishii, Gary L. Gallia, Russell H. Taylor, Gregory D. Hager

***Abstract*—Functional Endoscopic Sinus Surgery (FESS) is one of the most common outpatient surgical procedures performed in the head and neck region. It is used to treat chronic sinusitis, a disease characterized by inflammation in the nose and surrounding paranasal sinuses, affecting about 15% of the adult population [1]. During FESS, the nasal cavity is visualized using an endoscope, and instruments are used to remove tissues that are often within a millimeter of critical anatomical structures such as the optic nerve, carotid arteries, and nasolacrimal ducts. To maintain orientation and to minimize the risk of damage to these structures, surgeons use surgical navigation systems to visualize the 3D position of their tools on patients' preoperative CTs. This paper presents an image-based method for enhanced endoscopic navigation. The main contributions are: 1) a system that enables a surgeons to asynchronously register a sequence of endoscopic images to a CT scan with higher accuracy than other reported solutions using no additional hardware, 2) the ability to report the robustness of the registration, and 3) evaluation on in-vivo human data. The system also enables the overlay of anatomical structures, visible or occluded, on top of video images. The methods are validated on four different datasets using multiple evaluation metrics. First, for experiments on synthetic data, we observe a mean absolute position error of 0.21 mm and a mean absolute orientation error of 2.8° compared to ground truth. Second, for phantom data, we observe a mean absolute position error of 0.97 mm and a mean absolute orientation error of 3.6° compared to the same motion tracked by an electromagnetic tracker. Third, for cadaver data, we use fiducial landmarks and observe an average reprojection distance error of 0.82 mm. Finally, for in-vivo clinical data, we report an average ICP residual error of 0.88 mm in areas that are not composed of erectile tissue and an average ICP residual error of 1.09 mm in areas that are composed of erectile tissue.**

***Index Terms*—Navigation, evaluation, stability analysis, structure from motion, ICP, in-vivo data**

I. Introduction

CHRONIC rhinosinusitis (CRS) is a common condition in which the nose and the cavities around the nose (paranasal sinuses) become inflamed and swollen for more than 12 weeks. Between 1990 to 1992, 135.6 per 1,000 people were afflicted with CRS making it the second most common chronic disease/condition behind deformities or orthopedic impairments. Its incidence maybe rising. CRS is first treated medically. However, when medicines fail to improve patient quality of life, functional endoscopic sinus surgery (FESS) is indicated. Over 250,000 FESS procedures are performed annually in the United States making it the most common ambulatory surgery performed in the head and neck region in adults [2]. FESS is a minimal invasive surgery where the drainage pathways of the paranasal sinuses are viewed with a 4 mm rigid endoscope and enlarged or unplugged with slender cutting instruments. The paranasal sinuses and their drainage pathways are adjacent to vital structures like the brain, eyes, the vision nerve, nerves that enable eye movement, and major blood vessels that supply the brain; that is, the surgical site is separated from critical nerves and arteries by bones that are less than 1 mm thick. Fig. 1 is an intraoperative view looking into the left sphenoid sinus from the left ethmoid sinus. A small oval opening was made through the face of the sphenoid bone exposing the sphenoid sinus. The bulge of the carotid artery (blood vessel that supplies the brain) and the optic nerve (vision nerve) projecting into the sphenoid sinus can be seen just behind and adjacent to the opening (Fig. 1). If the carotid artery is struck during surgery, the likelihood of death or catastrophic stroke is extremely high. If the optic nerve is injured, the patient would most likely become blind. Since vital structures are intimately associated with the surgical site and are hidden from direct visualization by bone, and the paranasal sinuses often have to be entered blindly, sinus surgeons require a strong understanding of the paranasal sinuses and skull base anatomy. In addition they need to remain oriented and understand their exact location within the sinuses at all times. This is challenging because the ethmoid sinuses have a complex pneumatization pattern that partially develops in a stochastic fashion. That is, a patient's anatomy cannot be inferred from a prior. Instead, it has to be imaged and learned fresh for each patient. Additionally, sinonasal pathology obscures anatomical landmarks and, thus, the need to maintain orientation and control risks have led to the development of surgical navigation.

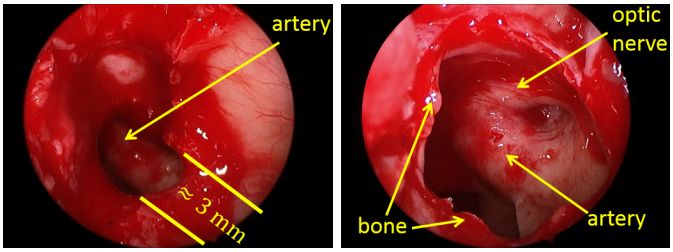

Fig. 1: Exposed carotid artery after breaching the sphenoid bone: The left image shows an initial hole through the bone with the artery resting a few millimeters behind. The right image shows a wider hole that also exposes the optic nerve.

In this paper, we present an image-based system for enhanced FESS navigation. The system introduces the combination of a novel algorithm to evaluate the stability of the computed registration between an endoscope and CT data, a

system integration with clinical devices and clinically relevant data and results. The navigation workflow starts by recording a short sequence of images by moving an endoscope within the nasal airway, and using structure from motion (SfM) to compute a sparse 3D structure from the camera motion. Our system uses hierarchical multi-affine (HMA) [3] feature matching to provide robust matches under difficult endoscopic imaging conditions. The resulting 3D point cloud is registered to a 3D mesh of the sinus cavities segmented from CT using a trimmed iterative closest point (TriICP) algorithm. The transformation produced by TriICP is applied to the coordinate frame of the endoscope to register it to the CT frame. On average, our implementation is able to compute the full pipeline (SfM+TriICP) within 10 seconds and register the camera with submillimeter accuracy. Once the registration has been computed, we measure the stability of the registration to detect and report weak solutions. This enables the assignment of confidence in the computed registrations, allowing a user to understand or gauge the reliability of any given registration. This measure is extremely important in trying to reduce surgical errors that are caused due to overconfidence in navigation systems. Additionally, our system enables augmented reality by overlaying occluded 3D structures on top of the video images to alert surgeons of proximity to critical structures.

We evaluated our system with synthetic, phantom, cadaver, and in-vivo clinical data. The synthetic data was generated by simulating a teleoperated endoscope inside a 3D model of the sinus cavities. These simulations provide the ground truth telemetry of the endoscope which is used to evaluate the absolute accuracy of our system. The accuracy of our method in the phantom experiment is evaluated against navigation provided by an electromagnetic (EM) tracker. The accuracy in the cadaver experiments is obtained by inserting 27-gauge needles through the skull base and relating the 3D coordinates of the needle tips to their image coordinates. Finally, the accuracy of our method in experiments with in-vivo clinical data collected from an outpatient clinic is measured by evaluating the ICP registration error. In all datasets, our results demonstrate that our system achieves submillimeter registration error. For video sequences viewing erectile tissue, however, the mean registration error increases slightly above 1 mm due to discrepancies between the CT and video data.

II. Previous Work

FESS became widely adopted during the 1980s after the pioneering work of Messerklinger and Kennedy [4] due to its minimally invasive nature. However, FESS is a challenging procedure to perform. In order to improve the safety and efficiency of FESS, surgeons use navigation systems to register patients to their respective preoperative CT scans. This enables surgeons to locate themselves in the CT coordinate frame and gain more context cues than endoscopy allows, and to track the position of their tools and pointers inside patients' sinus cavities. These systems are credited to decrease intraoperative time, improve the surgical outcome, and reduce workload [5]. Current navigation systems, however, suffer from tracking errors greater than 1 mm [6]. Errors for image-guided FESS reported in literature are commonly upper bounded by 2 mm [7], [8], [9]. Unlike analogous industrial applications, there is no international standard that defines the accuracy of a surgical navigation system or standard procedure to assess it. In the literature, the definition of "accuracy" is commonly reported as the root-mean-square of a 3D fiducial localisation error. Paraskevopoulos *et al.* [10] present several types of accuracies including: software, imaging, system and navigation. They define the *navigation accuracy* as the global accuracy that accounts for all possible sources of errors (calibration, registration, software, imaging and digtizing) and, as a result, is typically larger than the others due to the propagation of errors. For example, the Stryker Navigation system uses stereoscopic cameras and pointer instruments with a measured accuracy of 0.72 mm but the navigation accuracy degrades to 1.45 mm when all sources of errors are considered. Novel registration methods have been investigated for existing navigation systems but only achieve 1.28 mm ± 1.09 mm [11].

These errors are too large when compared to the size of the nasal airway and sinus cavities and, more importantly, the thickness of the boundaries of the sinuses. For instance, the width of the airways near the nasal septum can range from 2-5 mm, but the fovea ethmoidalis, or the roof of the sinuses that separates the cavity from the brain, is on average about 0.5 mm thick. The lateral lamella, which separates the sinuses from the olfactory system, is about 0.2 mm thick [12]. The boundary between the sinuses and the optic nerve can be as low as 0.45 mm, increasing to an average of 0.91 mm closer to the eyeball [13]. Soft mucosa often makes up a large part of these measurements. For instance, the uncinate process, which has a mean thickness of 1.4 mm, is made up of 0.67 mm of medial mucosa, 0.6 mm of lateral mucosa, and only 0.16 mm of bone [14]. In comparison, beyond the sinus boundaries, the carotid artery has a mean measured diameter of 3 mm. In practice, errors reported by current navigation systems are too large and prevent high fidelity overlays of anatomical structures on top of endoscopic images.

A meta-analysis by Labruzzo et al. [15] demonstrates that proficiency with enhanced imaging technologies have contributed to a significant decrease in the rate of complications in FESS. Occurrences of major complications have decreased from 8% to 0.31%, and current estimates of minor complications range between 1.37-5.6%. However, Krings et al. [16] report that image-guided FESS procedures can also have an increased rate of complications due to overconfidence in the technology and reliance on navigation technologies to treat the most complex cases. A similar conclusion is also found in [17]. Although improving navigation technologies to submillimeter accuracy can help improve the outcome of these complex cases, it is extremely important to ensure that the navigation system knows when it fails. Our method not only improves registration to submillimeter accuracy, but is also able to evaluate the stability of the registration and report to the surgeon when the computed registration is not reliable. This can help avoid complications arising from overconfidence in the navigation system.

Several previous methods have estimated the extrinsic parameters of the endoscope by rendering several views of the sinus mesh extracted from CT and searching for the rendered

image that best matches the real video image. To measure the similarity between real and rendered endoscopic images, Luo et al. [18] proposed a robust similarity measure that accounts for illumination and contrast. In [19], a stochastic optimization algorithm is used to search for the most similar rendered view which resulted in mean reprojection errors of 2 mm. Sets of initial poses are determined by testing for collisions between the endoscope and the sinus cavities and ensuring that the initial guesses are physically plausible.

Our work is closely related to [20], [21], where a sparse 3D point cloud is computed from a sequence of endoscopic images and then registered to a 3D geometry derived from a CT scan. Our team also reported reprojection distance errors of 0.7 mm during a previous cadaver study [22]. Another related approach adapts an algorithm from monocular simultaneous localization and mapping (SLAM) [23].A different approach to this problem involves planning an endoscopic view to achieve greater registration accuracy [24], but this method would be ineffective in a nasal cavity where the degrees of freedom of the endoscope are severely limited.

The scale and pose of the of the reconstructed 3D models are initially estimated using principal components analysis and then refined using ICP [25]. The system presented in this paper, however, estimates the 3D structure from a greater number of images and uses improved feature matching. The recent development of 3D endoscopes [26] using a backward-looking catadioptric camera provides an alternative for 3D reconstruction. A similar system using a conventional endoscopic equipment is also presented in [27]. Also, 4 mm stereo endoscopes are available for ENT procedures, but the short baselines of these devices make depth estimation prone to large errors [28].

III. Method

Our video-CT system uses a sequence of endoscopic images, typically between 15 and 30, to compute a 3D point cloud using structure from motion (SfM) algorithm with sparse bundle adjustment. The resulting 3D point cloud and, as a byproduct, the sequence of 3D camera extrinsic parameters are registered to the sinus cavity mesh extracted from patient CT using TriICP algorithm with scale adjustment. Once the registration is computed, its stability can be evaluated, and if the registration is found to be stable, then the sequence of camera poses can be used to overlay anatomical structures segmented from the CT onto the camera images. Contrary to optical image-guidance systems where only the 3D position of a tool is displayed in the three anatomical planes of the CT, our video-CT method provides full 3D registration (position and orientation) and enables the overlaying of anatomical structures, visible or not, on top of video images.

Our system is implemented by Robot Operating System (ROS) [29] services on a server with 20 cores (dual Xeon E5-2690 v2, Intel, Santa Clara CA) and 4 GPUs (GeForce GTX Titan Black, Nvidia, Santa Clara, CA).

A. Feature Matching

Endoscopic images, such as those used during FESS, pose a unique challenge for SfM algorithms and computer vision in general. The eyepiece of the endoscope occludes approximately 50% of the images leaving a relatively small circular foreground area. Additionally, the moving light source, lack of texture in the visible tissues, large specularities, and high dynamic range all contribute to difficult feature matching. Finally, the principal axis of motion inside the sinus cavity is along the optical axis of the endoscope, meaning that feature matching must be particularly robust across scales.

Scale invariant feature transforms (SIFT) [30] and adaptive scale kernel consensus (ASKC) [31] have been used in [20], but the difficulty in obtaining a reliable 3D structure stems from the difficulty in computing numerous robust matches in endoscopic images. We are able to overcome this difficulty by using Speeded Up Robust Features (SURF) [32] to compute features and initial matches in our images, and by using the Hierarchical Multi-Affine (HMA) algorithm [3], which has demonstrated superior robustness to filter initial matches in surgical images. Although SIFT keypoints were more repeatable in our preliminary results, our argument for using SURF is the availability of a GPU application programming interface (API) to extract keypoints and descriptors, and to generate initial matches [33]. Using the GPU API on a sequence of thirty images saves on average 20 seconds of computation time compared to using the CPU API for SIFT.

Since HMA only processes a set of initial matches, the algorithm is feature-neutral and can use any type of keypoints or descriptors. Therefore, given a set of initial matches computed using the SURF API, the HMA algorithm is able to find clusters of matches by enforcing a local affine transformation for each cluster. Using local affine constraints preserves a greater number of initial matches by lowering the match threshold while filtering mismatches. Our C++ implementation of HMA only implements the basic version of the HMA algorithm and does not compute the local recovery phase that ensures the clusters are locally consistent.

B. Structure from Motion

The HMA matches are used to estimate the 3D structure and the camera motion [34]. Our SfM implementation is based on the openMVG library [35]. The SfM solution is refined by computing a sparse bundle adjustment [36] with constant intrinsic camera parameters. The resulting 3D structure and camera motion are defined up to an unknown scale. An initial estimate for the scale is computed from a electromagnetic tracker that is attached to the endoscope during endoscopy. Each time an image is grabbed, the position and orientation of the magnetic reference is recorded.

Let I_i define the image at time $i = 1, \ldots, N$ and let h_i be the distance between the positions of the endoscope at times $t = i$ and $t = 1$ as computed by the SfM algorithm. Similarly, let h_i^* be the distance measured by the electromagnetic tracker. The initial scale estimate between the frame i and frame 1 is determined by the ratio $s_i = h_i^*/h_i$, and the scale estimate for the entire sequence of N frames is determined by the average $s = \frac{1}{N-1}\sum_{i=2}^{N} s_i$.

C. Trimmed Iterative Closest Point with Scale Adjustment

Once the scaled structure and camera motion are available, the structure is registered to the 3D mesh extracted from

patient CT. The 3D mesh is obtained from CT data by using the grayscale model maker in 3D Slicer [37] with a threshold of -450 Hounsfield units. The resulting mesh is then processed to remove the vertices and triangles that comprise the skin, ear cartilages, and other structures that are unnecessary for our application. An initial registration guess is found by manually aligning the scaled structure with the 3D mesh. Finally, TriICP algorithm with scale is used to align the point cloud to the 3D mesh of the patient. Our implementation extends the Levenberg Marquardt implementation of the Point Cloud Library (PCL) [38] to optimize over the initial scale in addition to the rotation and translation components. Finally, we use a trimmed rejector with 85% threshold to reject 15% of outliers from the ICP registration error.

D. Stability Analysis

Assessing the uncertainty of a registration is an important element in the development of navigation systems. As important as it is to provide a system that can accurately inform surgeons of their location in the reference CT frame, it is more important for the system to know when it is wrong so that surgeons can be warned about when not to rely on the system. For example, optical tracking devices used by navigation systems typically report failures when markers are occluded or outside the field of view.

Our proposed system has two possible points of failure. The first is when the SfM computation fails to generate a structure or when the ICP algorithm fails to converge within a fixed number of iterations. Typical causes for these failures include large endoscope motions that generate few matches, or small endoscope motions that generate negligible binocular disparities. Albeit critical, these failures are easily detected since they imply that one of the algorithms fails to converge.

The second and more difficult type of failure to detect is when a "converged" registration solution is numerically unstable and possibly inaccurate. These failures typically stem from ambiguous registrations between the 3D structure and the CT data. For example, when the endoscope is placed in a tunnel-like region, the computed 3D structure does not provide enough geometric constraints to establish a reliable registration. In these cases, although small ICP residual errors suggest successful registrations, the solutions are unstable because the structures are able to slide inside the cavity without significantly affecting the ICP residual error.

We address this problem by introducing a method to evaluate the stability of the registration between the sparse 3D point cloud obtained from SfM (typically between 500 and 1,200 vertices) and the CT mesh. The method is based on analyzing how each 3D SfM point constrains the pose of the endoscope. In the following development, each 3D point enforces a constraint that is formulated as a linear equation.

Let $\mathbf{p}_i$ define the coordinates of the ith 3D point ($1 \leq i \leq N$) in the coordinate frame of the camera as illustrated in Fig. 2. We also define the unit vector representing the direction of the ray passing through $\mathbf{p}_i$ and the origin of the camera by $\mathbf{l}_i$. Finally, by using ray tracing, we scale the ray $\mathbf{l}_i$ by d_i and define $\mathbf{q}_i = d_i \mathbf{l}_i$ as the coordinates of the intersection between the ray $\mathbf{l}_i$ and the 3D mesh obtained from CT. At $\mathbf{q}_i$, the surface normal of the sinus cavity is obtained from the CT mesh and is denoted by $\mathbf{n}_i$. Then, for a point $\mathbf{o}_i$ lying on the same plane as $\mathbf{q}_i$, we have

$$(\mathbf{o}_i - \mathbf{q}_i)^T \mathbf{n}_i = 0, \tag{1}$$

and given a small rotation represented by the skew symmetric matrix $[\delta\mathbf{r}]_\times$ and translation $\delta\mathbf{t}$, its coordinates can also be obtained by

$$\mathbf{o}_i = \mathbf{q}_i + [\delta\mathbf{r}]_\times \mathbf{q}_i + \delta\mathbf{t}. \tag{2}$$

Substituting Eq. 2 in Eq. 1 and simplifying gives us

$$([\delta\mathbf{r}]_\times \mathbf{q}_i + \delta\mathbf{t})^T \mathbf{n}_i = 0, \tag{3}$$

which can be converted to

$$(-d_i[\mathbf{l}_i]_\times \delta\mathbf{r} + \delta\mathbf{t})^T \mathbf{n}_i = 0.$$

Dividing by d_i gives us

$$\begin{bmatrix} -\mathbf{n}_i^T[\mathbf{l}_i]_\times & \mathbf{n}_i^T/d_i \end{bmatrix} \begin{bmatrix} \delta\mathbf{r}^T & \delta\mathbf{t}^T \end{bmatrix}^T = 0 \tag{4}$$

$$\mathbf{a}_i^T \mathbf{x} = 0 \tag{5}$$

For $N > 6$, we obtain an overdetermined homogeneous system of linear equations $A\mathbf{x} = 0$. Given the pseudoinverse $A^+ = (A^T A)^{-1} A^T$, the condition number

$$\text{cond}(A) = \|A\|_2 \|A^+\|_2 \tag{6}$$

measures the closeness of A to rank deficiency and, in our context, the sensitivity or stability of the registration. Although the least-squares formulation to Eq. 4 is sensitive to outliers, its solution is never computed since only the sensitivity of the system of equations is evaluated. Effects of outliers on the condition number is further reduced by only using the inliers returned by the trimmed ICP algorithm.

Intuitively, each point $\mathbf{q}_i$ and its associated normal $\mathbf{n}_i$ contribute one equation $\mathbf{a}_i$ to constrain the motion of the camera. If Eq. 4 only has the trivial solution, $\mathbf{x} = 0$, then $\det(A^T A) \neq 0$, and any non-zero camera motion, $\mathbf{x} \neq 0$, will cause a residual error. Conversely, if Eq. 4 has non-trivial solutions, $\mathbf{x} \neq 0$, then the camera can be moved without causing residual errors.

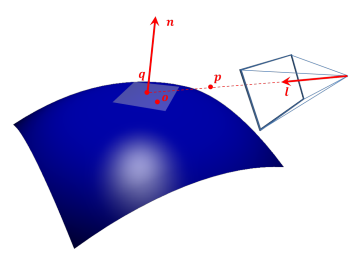

Fig. 2: Illustration of the geometry used for stability analysis.

IV. Results

We tested our system on four datasets. The first set consists of synthetic data that is generated by rendering 3D meshes of sinus cavities that have been textured using endoscopic images. The second dataset was obtained using a phantom head. The third dataset was obtained from a cadaver study, and was used in the experiments presented in [20]. The last dataset was obtained from examinations at an outpatient clinic under IRB NA_00074677. For each dataset we report registration errors that use the best available reference, and we analyze the stability of all the results.

A. Synthetic Images

The aforementioned protocol for clinical data does not provide the possibility to measure absolute accuracy of our

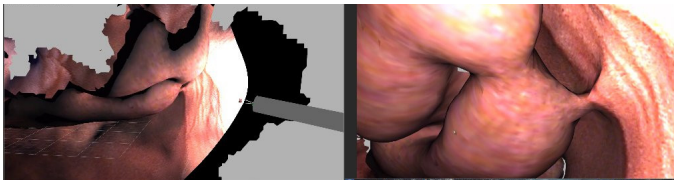

Fig. 3: Simulation of an endoscope inside the rendered sinus cavity: A view of the sagittal plane of the sinus cavity (left) with the endoscope looking at the middle turbinate. The endoscope is the gray cylinder with the light cone clearly visible in front of it. A synthetic image (right) generated by the simulated endoscope.

system. To overcome this limitation, we simulated an examination where a virtual endoscope is manually teleoperated in a textured model of the sinus cavity. We used video images from the clinical dataset to manually texture the inside of the 3D mesh extracted from a CT in the clinical dataset. Then, we simulated a 4 mm endoscope in Gazebo [39] with zero mean Gaussian noise ($\sigma = 1.5$ per RGB channel) and an attached light source to navigate within the virtual sinus cavities as illustrated in Fig. 3. To replicate the motion of a real endoscope, the virtual endoscope was constrained by the geometry of the cavity by enforcing collisions between the endoscope and the surrounding tissues.

This virtual endoscope was manually teleoperated inside the rendered sinus cavity using a SpaceNavigator (3Dconnexion, Munich, Germany). During the teleoperation, we recorded the synthetic images and the position and orientation of the endoscope. We used our video-CT registration pipeline with these synthetic images and compared the registered camera poses to the ground-truth provided by the simulator. We generated ten 5 second sequences each totaling 30 frames and found that the mean absolute error (MAE) of the camera positions was 0.21 mm (standard deviation 0.08 mm) and the MAE of the camera orientations was 2.8° (standard deviation 0.6°). The average TriICP residual error over the 10 sequences was 0.38 mm with a standard deviation of 0.15 mm.

Stability Analysis: We use the condition number of the matrix A (Eq. 6) as a measure of the stability of the registrations. Although a rank deficient matrix indicates a registration failure, this scenario rarely happens in practice and never did in our experiments. The condition number of the matrix A reflects the sensitivity of the registration to measurement errors and indicates the number of significant digits that are expected in a solution.

For each video sequence (composed of 30 frames), we computed the mean absolute position error between the registered and simulated camera poses. Then, for each sequence, we computed the stability of the registration by computing the average condition number using Eq. 6. These results are illustrated in Fig. 4 which plots the mean absolute position error of each sequence against its average condition number. Results of the linear regression indicated that there is a significant positive association between the mean absolute position error and the condition number ($p < 0.01$).

We employ a color scheme to encode the order of magnitude of the condition number, as illustrated in Fig. 5, to report

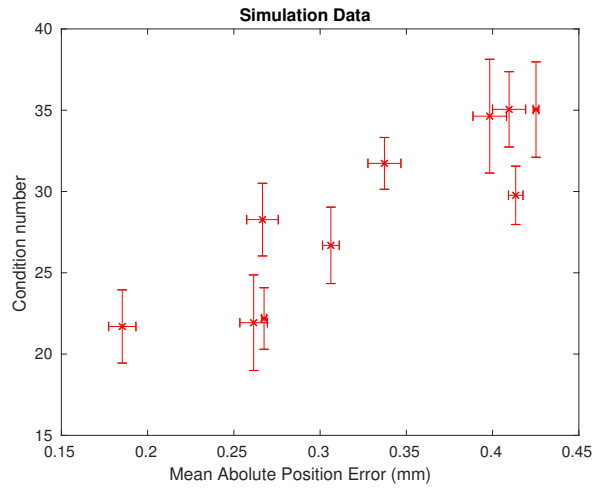

Fig. 4: Plot of the mean absolute position error versus the mean absolute condition number. For each frame in a sequence the registered position of the endoscope can be compared to the simulated position. The linear regression between them is significant ($p < 0.01$). The mean error and standard deviation for each video sequence is indicated on the horizontal axis.

the stability of a registration. In this set of sequences, the camera navigates under the middle turbinate in a tunnel-like environment as illustrated in Fig. 5. Initially, the position of the endoscope provides a wide and stable view of the sinus cavity, but as the endoscope is inserted toward the nasopharynx the airway becomes a narrow tunnel. The computed 3D structures in this area typically consist of two or three sidewalls that provide less stable constraints for the TriICP registration.

By combining the stability analysis with the residual errors of the registration, the system is able to validate the accuracy of its solutions. To determine the accuracy of a solution, we use the following inequality to bound the relative error [40]:

$$\frac{\|\Delta\mathbf{x}\|_2}{\|\mathbf{x}\|_2} < \left(\text{cond}(A)^2 \frac{\|\mathbf{r}\|_2}{\|A\|_2\|\mathbf{x}\|_2} + 2\text{cond}(A)\right)\left(\frac{\|E\|}{\|A\|}\right),$$

where $\Delta\mathbf{x}$ is the solution error, $\text{cond}(A)$ is the condition number of the matrix A (Eq. 6), $\mathbf{r} = -A\mathbf{x}$ is the residual, and E is a perturbation matrix of A. Two typical scenarios can be observed. If $\|\mathbf{r}\|$ is small, then the relative error is bounded by $\text{cond}(A)$. If $\|\mathbf{r}\|$ is large, then the relative error is bounded by $\text{cond}(A)^2$. We also note that if $\|\mathbf{r}\|$ is small, the solution of a linear system loses at most $\log_{10}(\text{cond}(A))$ accurate digits relative to the accuracy of the input.

Given that the 1920×1080 images are our only measurements, the coordinates of the image points are used to bound the accurate figures of our navigation system between one and four significant figures. Therefore, we expect the solutions of the navigation to have between one and four accurate figures from which $\log_{10}(\text{cond}(A))$ is subtracted. We used this simple algorithm to calculate the remaining significant figures of the registration and report the stability of each solution. For example, when $10 \leq \text{cond}(A) < 100$, the solution to Eq. 4 is typically limited to 1 or 2 accurate digits. When $\text{cond}(A) \geq 1000$, the solution loses more than three digits of accuracy and none of the digits of a solution are reliable.

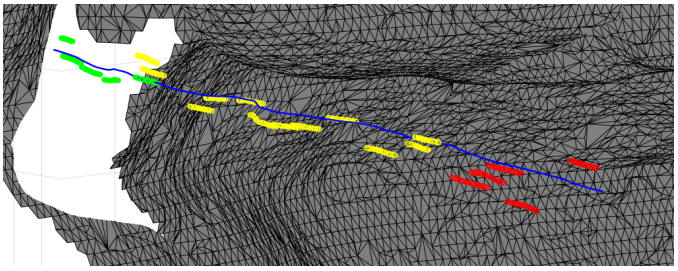

Fig. 5: Color coding representing the stability of each registration: A green circle represents an average condition number between 10 and 99 (solutions losing one significant figure). Yellow represents condition numbers between 100 and 999 (solutions losing two significant figures), and red indicates condition numbers above 1000 (solutions losing more than three significant figures). The real endoscope trajectory is displayed in blue.

This scenario does not imply that a registration solution has no accurate digit, but it means that the accuracy of the solution cannot be demonstrated. In fact, 80% of the solutions of the simulation data set with $\text{cond}(A) \geq 1000$ had at least one accurate digit.

B. Phantom Data

Our second dataset was generated using a sinus phantom (PHANCON Gmbh, Leipzig, Germany). We recorded videos inside the nasal cavity of the phantom and tracked the position and orientation of the endoscope using the same equipment used in clinical data collection (Section IV-D). This equipment consists of a small cart that holds a laptop, a DVI2USB 3.0 (Epiphan Video, Ottawa Canada) to collect 1920x1080 images at 30 frames per second, and an urora electromagnetic tracking system (NDI Waterloo, Canada) which is used to track the pose of the endoscope. The video input is connected to a 1288HD endoscopic camera (Stryker Kalamazoo, MI), and the magnetic reference of the EM tracker is clipped onto the endoscope. During data collection, raw video images and the position and orientation of the electromagnetic reference are time-stamped and saved to a ROS bag file. The phantom was also scanned at the Johns Hopkins Hospital.

Since the texture of the nasal cavities inside the phantom is uncharacteristically smooth compared to humans, we speckled a pinch of ground pepper within the cavities to aid the SfM computation. We recorded five examination sequences similar to those recorded in patients (Section IV-D) and compared the results to the motion tracked by the Aurora. Before comparing both trajectories, the rigid transformation between the tracking reference and the camera coordinate frame is computed using a hand-eye calibration algorithm [41]. The hand-eye calibration was followed by registration of the CT data to video images using fiducial markers on the phantom.

We compared the trajectory of our video-based navigation to the tracking system and obtained a mean absolute position error of 0.97 mm (standard deviation of 0.4 mm) and a mean absolute orientation error of 3.6° (standard deviation of 1.1°). Although registration errors computed for the phantom dataset are higher than those obtained with synthetic data, we are not able to ascertain the amount of errors in the phantom

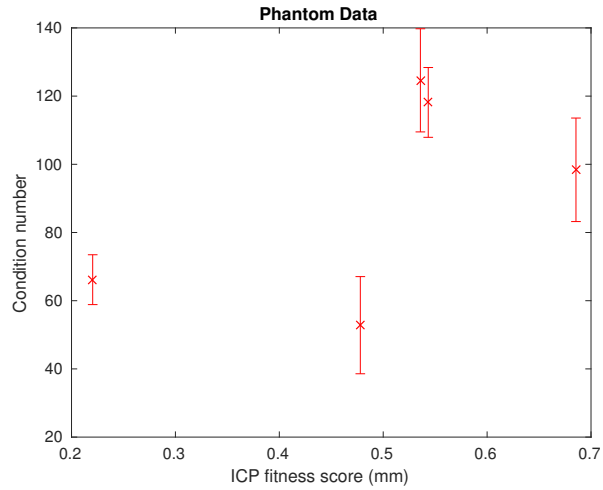

Fig. 6: Plot of the ICP residual error versus the mean absolute condition number. The linear regression between them is not significant ($p = 0.1$).

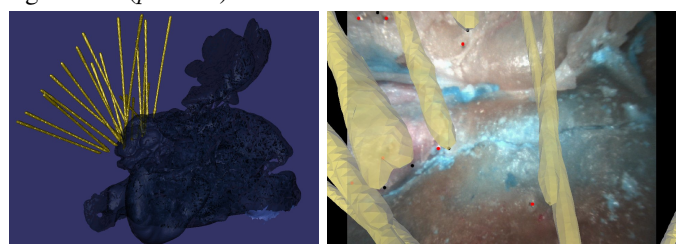

Fig. 7: Rendered view of the cadaver CT (left) with the needles inserted in the skull base. Overlay of the needles (right) on the cadaver images. Red dots correspond to the visible needle tips and the black dots correspond to the projection of the needle tips onto the image.

dataset that are imputable to the registration accuracy and the amount imputable to the accuracy of the Aurora. The average TriICP residual error for the five sequences was found to be 0.51 mm with a standard deviation of 0.17 mm.

Stability Analysis: We also report results for the stability analysis of the five sequences recorded in the phantom. Similarly to the synthetic sequences, Fig. 6 illustrates the relationship between the TriICP residual error and the condition number of Eq. 4, where a lower condition number indicates a more stable registration, and a lower fitness score indicates a more accurate registration. Contrary to the other data sets, however, the linear relationship between the $\text{cond}(A)$ and the registration error is not significant $p = 0.1$. We suspect that the main reason for this result is the relatively small size of phantom data set compared to the three other data sets.

C. Cadaver Data

In our next study, we measured registration errors using data collected from a cadaver. In the study, needles were inserted in the skull base of the cadaver to provide fiducial landmarks that can be detected in the CT as well as in the endoscopic video as illustrated in Fig. 7. These images were processed by our video-CT registration pipeline to estimate extrinsic parameters of the camera in the CT frame. The video sequences were limited to frames where a reasonable number of needle tips

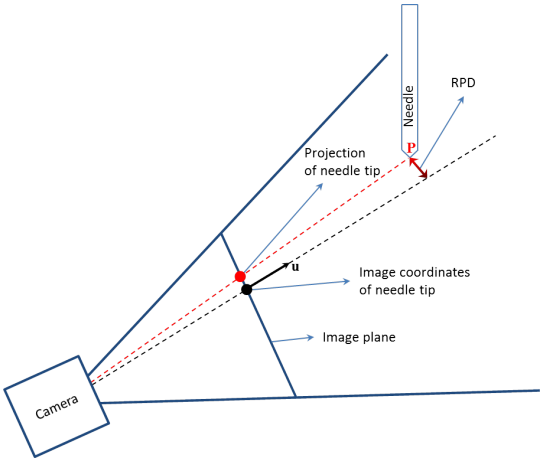

Fig. 8: Diagram representing the reprojection distance (RPD).

(5-6) could be visually identified and localized in the images. The needle tips were masked to prevent the extraction of SURF keypoints from the needle tips. The corresponding 3D coordinates of each needle tip was obtained by manually measuring the needle tips in the CT volume.

The 2D image coordinates of needle tips in the video sequences along with their corresponding 3D coordinates were used to measure the average reprojection distances (RPD) of our video-CT registration. As illustrated in Fig. 8, reprojection distances represent the Euclidean distance between the 3D coordinates of a needle tip $\mathbf{P}$ in the camera coordinate frame and the ray $\mathbf{u}$ that extends the reprojection of the same needle tip according to the following equation:

$$RPD = \|\mathbf{P} - \mathbf{u}(\mathbf{u} \cdot \mathbf{P})\|. \tag{7}$$

RPD was used in [20] to estimate the accuracy of the registration. In our study, we report similar results with minimum, maximum and average RPD per sequence reported in TABLE I. To compare the results across datasets, we report a TriICP average residual error computed over 6 sequences to be 0.43 mm with a standard deviation of 0.13 mm.

Also, we compute RPD based on the matched SURF keypoints. For a given video sequence, we extract SURF keypoints from two images and match the descriptors using the HMA algorithm. Then, using one of the two images as the reference, we use ray tracing to determine the 3D coordinates of the nearest CT data point to each keypoint. Finally, we use the nearest CT data point with the matched SURF keypoints in the second image to compute the RPD using Eq. 7.

Since these RPD errors represent relative errors they can be easily biased. For instance, two images taken from nearly identical camera poses can not only generate a large number of matches due to their similarity, but will also generate SURF matches with nearly identical keypoints that will result in very small relative RPD even for an arbitrarily bad video-CT registration. Given that these matches will have very low relative RPD, we argue that images taken from nearby cameras can favorably skew the results. To avoid this bias, we only use pairs of images composed of the first and last images in a video sequence. Although this does not guarantee that images from the two most distant cameras in a given sequence are used (in fact, this rarely happens), it generally involves matches that result from cameras that are a few millimeters apart. This results in a minimum, average and maximum RPD for each video sequence. Repeating this procedure for all sequences we obtained the results in the last row of TABLE I. Each video sequence shows 5 or 6 needle tips, and the reported error values are based on the RPD (millimeters) between the measured coordinates of each needle tip and the corresponding reprojected needle tip. Although the needles might be automatically detected as SURF keypoints, their main purpose is to serve as fiducial markers to evaluate RPD errors. For the entire cadaver data set, the average RPD is 0.81 mm with a standard deviation of 0.07 mm.

	Sequence					
	Seq01	Seq02	Seq03	Seq04	Seq05	Seq06
Min RPD	0.4185	0.2683	0.2086	0.1925	0.0824	0.0731
Avg RPD	0.9212	0.8277	0.8890	0.7529	0.8262	0.7669
Max RPD	1.2103	1.2023	1.1931	1.1386	1.2192	1.1819
Rel RPD	0.3208	0.3571	0.2260	0.2856	0.2164	0.1965

TABLE I: RPD values for the 6 cadaver sequences.

Finally, we also overlaid the needles segmented from CT on top of video frames after registration as illustrated in Fig. 7. This figure displays the projected needle tips with black dots and the measured tips with red dots. We note that metallic objects are prone to artifacts in CT images, and the diameter of the 27-gauge needles in Fig. 7 illustrates the difficulty in accurately measuring the 3D coordinates of the tips.

Stability Analysis: As before, we use properties of the matrix A in Eq. 4 to measure the stability of the registration. Since ground truth is not available for this dataset, we plot the average reprojected distance (RPD) versus the average condition number of each sequence (Fig. 9). As with results from the synthetic and phantom datasets, we note a general positive correlation between accurate registrations and their stability. Stability numbers are particularly good for the cadaver dataset. We suspect that the combination of the particularly feature-rich texture of the cadaver skull-base and predominant lateral movements of the endoscope contributed to wide and stable 3D reconstructions, resulting in more stable registration results. Results of the linear regression indicated that there is a significant positive relation between the RPD error and the condition number, cond(A) ($p = 0.02$).

D. Clinical Data

For the in-vivo experiments, we collected data from several patients during preoperative examinations under an IRB approved protocol. Data was collected using equipment described in Section IV-B. On average, the data collections lasted about 90 seconds per patient. These 90 seconds include the time used to clean the lens, and to insert the endoscope in through both nostrils. Therefore, each recording session provides roughly 15 to 25 seconds of useful data. After the examination, the camera was calibrated using CALTag [42] and a perspective camera model. The examined areas in the nasal cavity are patient specific, but the middle turbinates of all patients were examined, and in some cases the endoscope was inserted all the way to the nasopharynx.

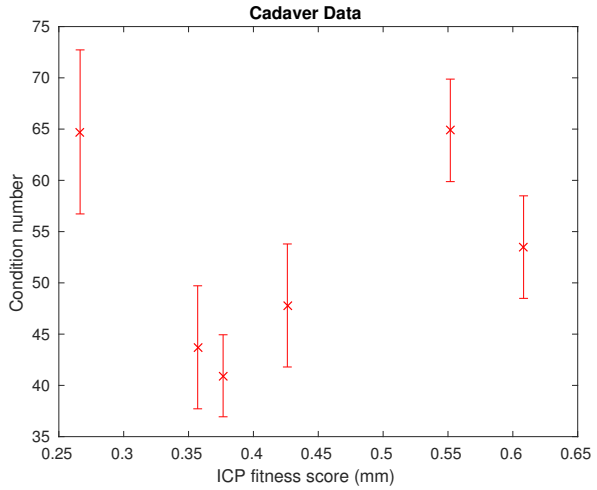

Fig. 9: Plot of the RPD mean error versus the mean absolute condition number for the cadaver data. The linear regression between them is significant ($p = 0.02$).

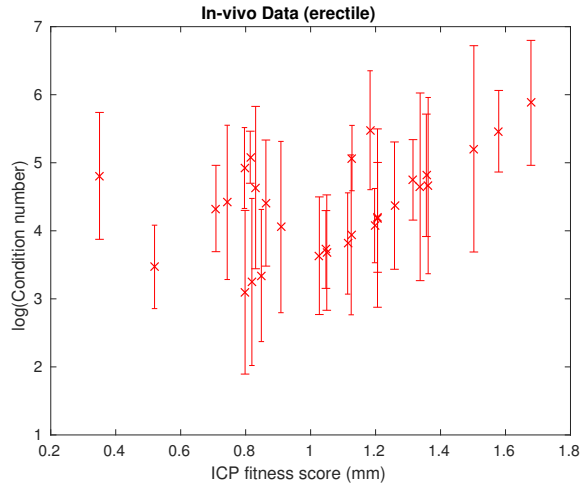

Fig. 10: Plot of the ICP residual error versus the log mean absolute condition number when viewing erectile tissue. The linear regression between them is significant ($p < 0.01$).

Several areas observed during the examination are composed of erectile tissue. These structures include the middle turbinate (MT) and the nasal septum (S). Large discrepancies between the geometry of these structures in CT and in the endoscopic video can be caused by the nasal cycle and because patients are generally decongested for endoscopy but not during CT acquisition [21]. Therefore, they present a significant challenge for rigid registration algorithms like ICP.

For each patient, the video segments during which the endoscope is inside the nasal cavity were edited to form several video sequences, each 1 second long and containing 30 to 35 images (nominal frame rate of 30 frames per seconds). We selected four patients enrolled in our study to generate a total of 52 video sequences. Each sequence was processed independently by the video-CT registration pipeline. For 88% of the sequences, the SfM algorithm was able to generate a structure that could be registered to the CT data. The 12% failures were caused because the SfM algorithm either failed to converge (9%) or generated a useless structure (3%) that could not be used to initialize a TriICP registration.

We measured the accuracy of our system on the clinical dataset set using two different methods. First, we report the residual error of the TriICP algorithm. Since we use a trimmed rejector with a ratio of 85% inliers, the reported errors correspond to the error of the 85th percentile best registered points. We further divide our measured errors according to whether the sequence involves erectile or non-erectile tissue. For non-erectile tissues, the 85th percentile mean absolute error was 0.88 mm with a standard deviation of 0.3 mm. For erectile tissues, the 85th percentile mean absolute registration error was 1.1 mm with a standard deviation of 0.32 mm.

Second, we compute RPD based on the matched SURF keypoints as presented in Section IV-C. Unlike the cadaver dataset, the clinical dataset does not provide fiducial landmarks that can be used to evaluate the absolute accuracy of the registration. Therefore, we adopt the approach of using the SURF keypoints as relative fiducial markers. This results in a minimum, average and maximum RPD for each video sequence. Repeating this procedure for all sequences we obtained the average minimum RPD, average RPD, and average maximum RPD presented in TABLE II. From the observation between RPD using needles and SURF keypoint in Section IV-C, we postulate that the average minimum relative RPD in TABLE II is indicative of the minimum RPD. One caveat in this direct comparison, however, is the nature of the motion in both datasets. Whereas the endoscope mainly moves laterally in the cadaver dataset, the endoscope mainly moves forward/backward in the in-vivo dataset which causes smaller disparities and, potentially, better results.

Average Min RPD	Average RPD	Average Max RPD
0.21mm ($\pm$0.08 mm)	0.32 ($\pm$0.28 mm)	0.71 ($\pm$0.21 mm)

TABLE II: Average ($\pm$ std. dev.) RPD for clinical data.

Stability Analysis: Stability analysis for the in-vivo dataset produces two plots: one for erectile tissue (Fig. 10) and the other for non-erectile tissue (Fig. 11). Due to the absence of ground truth, we plot the ICP residual error against the condition number for each sequence. For both erectile and non-erectile tissues, results of the linear regression indicate that there is a significant positive association between the residual TriICP error and the condition number ($p < 0.01$).

V. Discussion and Conclusion

FESS has become a very common and effective treatment for chronic rhinosinusitis. Yet, the proximity of critical anatomical structures combined with surgical tools offering little manipulability make these interventions very delicate. As navigation technologies evolve they are poised to play an increasingly important role during these procedures. To this day, however, state of the art navigation systems are struggling to break the 2 mm accuracy barrier. Given the scale of the nasal and sinus cavities and their boundaries, errors of this magnitude cannot be overlooked.

In this paper, we have presented a system capable of registering endoscopic video to the CT of a patient with submillimeter accuracy. Our system uses 30 to 35 frames, roughly

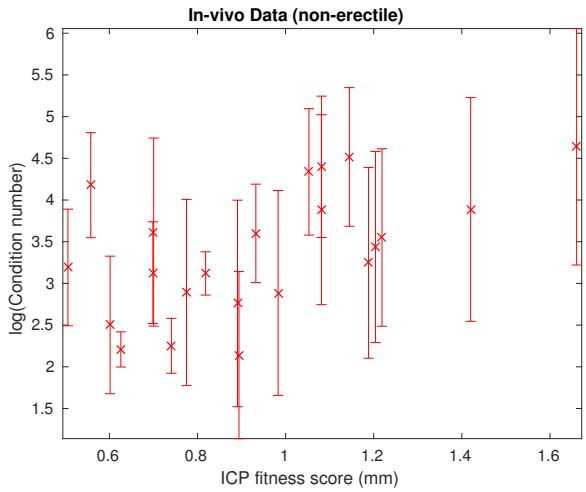

Fig. 11: Plot of the ICP residual error versus the log mean absolute condition number when viewing non-erectile tissue. The linear regression between them is significant ($p < 0.01$).

one second of video, to compute structure from motion with bundle adjustment. The resulting structure and motion are registered to patient CT using trimmed ICP with scale. We tested our video-CT registration pipeline on four different datasets. Not surprisingly, the synthetic dataset provides the best results with a mean absolute position error of 0.21 mm and a mean absolute orientation error of 2.8°. The importance of these results is that given the availability of a ground truth, they provide a gold standard for evaluating the other results.

Results from the cadaver study are similar to those already published in [20] with an RPD of 0.8 mm. We note, however, that the cadaver data is prone to produce better results than clinical data due to several reasons. First, the texture of the cadaver skull base is extremely feature-rich since the underlying vasculature is clearly visible due to injected red and blue latex. This allows for better SURF descriptor detection, and therefore better HMA matches. Second, the skull base is a fairly flat surface and more suitable for horizontal and vertical camera motion, which is harder to achieve in the nasal cavity. Finally, the endoscopy video collected in the cadaver only included the data from the center of the scope, and excluded the information near the edges where there is high distortion. Therefore, this dataset does not include areas of the video frames where relatively more erroneous feature matches could be computed due to lens distortion, leading to better results. The combination of these observations results in a more dense and accurate structure, and therefore, registration.

One of the challenges that our system faces during FESS is presented by possible discrepancies between pre-operative CTs and intra-operative video sequences. Such discrepancies can be significant in structures like the nasal turbinates due to the presence of erectile tissue which facilitates alternating partial swelling and contraction of the turbinates. This process is called the nasal cycle, and the mean duration of the cycle is about 2 hours [43]. Additionally, patients are administered decongestants before endoscopy to allow smooth insertion of tools into the nasal cavity, and reduce patient discomfort during the procedure. Further discrepancies can arise due to nasal congestion caused by exposure to allergens, viruses and other irritants at the time of CT acquisition. Although these discrepancies can be mitigated under a protocol that requires patients to be decongested before their CT scan, this protocol is often not followed. Another related challenge is the robustness to the significant appearance changes caused by a surgical procedure. Recent research on image descriptors that are more robust to anatomical changes might be possible alternatives to SIFT/SURF descriptors [44].

Our results demonstrate submillimeter registration errors in the absence of erectile tissue, whereas the presence of erectile tissue increases the error to 1.09 mm. We deliberately evaluated the registration of erectile tissues separately because, although they are our less accurate results, they represent cases that can be avoided if the proper decongestion protocol is followed during CT acquisition. In summary, our results have demonstrated that our system is capable of registering endoscopic video to CT data with submillimeter accuracy, and of computing the system's confidence in the registration based on the stability of the registration.

Our stability analysis of registration results also provides an improvement over state of the art navigation systems as it enables surgeons to assess the accuracy of the solution presented by the navigation system. The notion of uncertainty is often key to solving localization and navigation problems but these values are seldom reported to surgeons. As reported by [16], surgeons will typically rely on navigation systems for the most delicate aspects of procedures and, therefore, should be aware of the limitations of the navigations systems.

In this paper, we used *four* different metrics to evaluate our video-CT registration. Yet, other than the results computed on the synthetic dataset, it is difficult to obtain an absolute reference that is accurate. RPD with needles cannot be used in clinical studies, and it has limited accuracy due to metal artifacts in CT data. For example, the 27-gauge needles (0.4128 mm diameter) used in our experiment have a measured diameter of 2 mm in the CT data due to artifacts. These artifacts did not pose a serious problem for our cadaver study since the needles were inserted in the base of the skull, which is flat surface that is several centimeters wide. These artifacts, however, would pose a greater challenge if the needles were inserted in the nasal cavity since they are only 3-4 mm wide, and inserting needles in this area would corrupt the CT data.

Our future work will focus on three remaining areas. First, the ICP algorithm currently requires an initial guess that must be made manually. We hope to automate this step in the future. Research presented in [45] addresses a similar problem by using dynamic filtering of EM tracking data to initialize an image-based registration algorithm. Second, a better evaluation score than the ICP residual error for clinical data will help us assess the quality of registration better. Third, we hope to improve registration in the presence of erectile tissue by taking the deformation in these tissues into consideration.

References

[1] N. Bhattacharyya, "Contemporary assessment of the disease burden of sinusitis," *Am. J. Rhinol. Allergy*, vol. 23, pp. 392–395, Aug 2009.

[2] R. M. Rosenfeld, J. F. Piccirillo, S. S. Chandrasekhar, I. Brook, K. Ashok Kumar, M. Kramper, R. R. Orlandi, J. N. Palmer, Z. M. Patel, A. Peters, S. A. Walsh, and M. D. Corrigan, "Clinical practice guideline

(update): adult sinusitis," *Otolaryngol. Neck Surgery: Off. J. Am. Acad. Otolaryngol. Neck Surg*., vol. 152, pp. S1–S39, Apr. 2015.

[3] G. Puerto-Souza and G.-L. Mariottini, "A fast and accurate feature-matching algorithm for minimally-invasive endoscopic images," *IEEE Trans. Med. Electron*., vol. 32, pp. 1201–1214, July 2013.

[4] D. W. Kennedy, "Functional endoscopic sinus surgery. Technique," *Arch. Otolaryngol. (Chicago, Ill.: 1960)*, vol. 111, pp. 643–649, Oct. 1985.

[5] G. Strauss, E. Limpert, M. Strauss, M. Hofer, E. Dittrich, S. Nowatschin, and T. Lth, "Evaluation of a daily used navigation system for FESS," *Laryngo- rhino- otologie*, vol. 88, p. 776781, December 2009.

[6] K. Lorenz, S. Frhwald, and H. Maier, "The use of the BrainLAB Kolibri navigation system in endoscopic paranasal sinus surgery under local anaesthesia. An analysis of 35 cases," *HNO*, vol. 54, no. 11, pp. 851–860, 2006.

[7] J. N. Al-Swiahb and S. H. Al Dousary, "Computer-aided endoscopic sinus surgery: a retrospective comparative study," *Annals Saudi Medicine*, vol. 30, no. 2, pp. 149–152, 2010.

[8] M. P. Fried, J. Kleefield, H. Gopal, E. Reardon, B. T. Ho, and F. A. Kuhn, "Image-Guided Endoscopic Surgery: Results of Accuracy and Performance in a Multicenter Clinical Study Using an Electromagnetic Tracking System," *The Laryngoscope*, vol. 107, pp. 594–601, May 1997.

[9] R. Metson, R. E. Gliklich, and M. Cosenza, "A comparison of image guidance systems for sinus surgery," *The Laryngoscope*, vol. 108, pp. 1164–1170, Aug. 1998.

[10] D. Paraskevopoulos, A. Unterberg, R. Metzner, J. Dreyhaupt, G. Eggers, and C. R. Wirtz, "Comparative study of application accuracy of two frameless neuronavigation systems: experimental error assessment quantifying registration methods and clinically influencing factors," *Neurosurg. Rev*., vol. 34, pp. 217–228, Apr. 2010.

[11] J. Hong and M. Hashizume, "An effective point-based registration tool for surgical navigation," *Surg. Endosc*., vol. 24, pp. 944–948, Apr 2010.

[12] J. Kainz and H. Stammberger, "The roof of the anterior ethmoid: A place of least resistance in the skull base," *Am. J. Rhinol*., vol. 3, no. 4, pp. 191–199, 1989.

[13] H. Tao, Z. Ma, P. Dai, and L. Jiang, "Computer-aided three-dimensional reconstruction and measurement of the optic canal and intracanalicular structures," *The Laryngoscope*, vol. 109, no. 9, pp. 1499–1502, 1999.

[14] G. Berger, E. Eviatar, T. Kogan, and R. Landsberg, "The normal uncinate process: histology and clinical relevance," *Eur. Arch. Oto-Rhino-Laryngology*, vol. 270, no. 3, pp. 959–964, 2013.

[15] S. V. Labruzzo, N. Aygun, and S. J. Zinreich, "Imaging of the Paranasal Sinuses: Mitigation, Identification, and Workup of Functional Endoscopic Surgery Complications," *Otolaryngol. Clin. North Am*., vol. 48, pp. 805–815, Oct. 2015.

[16] J. Krings, D. Kallogjeri, A. Wineland, K. Nepple, J. Piccirillo, and A. Getz, "Complications of primary and revision functional endoscopic sinus surgery for chronic rhinosinusitis," *Laryngoscope*, vol. 124, no. 4, pp. 838–845, 2014.

[17] K. Stelter, B. Ertl-Wagner, M. Luz, S. Muller, G. Ledderose, V. Sie dek, A. Berghaus, S. Arpe, and A. Leunig, "Evaluation of an image-guided navigation system in the training of functional endoscopic sinus surgeons. a prospective, randomised clinical study," *Rhinol*., vol. 49, p. 429437, October 2011.

[18] X. Luo, H. Takabatake, H. Natori, and K. Mori, "Robust Real-Time Image-Guided Endoscopy: A New Discriminative Structural Similarity Measure for Video to Volume Registration," in *Information Processing in Computer-Assisted Interventions*, no. 7915, pp. 91–100, Springer Berlin Heidelberg, June 2013.

[19] Y. Otake, S. Leonard, A. Reiter, P. Rajan, J. H. Siewerdsen, G. L. Gallia, M. Ishii, R. H. Taylor, and G. D. Hager, "Rendering-Based Video-CT Registration with Physical Constraints for Image-Guided Endoscopic Sinus Surgery," *Proc. SPIE–the Int. Soc. for Opt. Eng*., vol. 9415, Feb. 2015.

[20] D. Mirota, A. Uneri, S. Schafer, S. Nithiananthan, D. Reh, M. Ishii, G. Gallia, R. Taylor, G. Hager, and J. Siewerdsen, "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," *IEEE Trans. Med. Electron*., vol. 32, pp. 1215–1226, July 2013.

[21] S. Leonard, A. Reiter, A. Sinha, M. Ishii, R. H. Taylor, and G. D. Hager, "Image-based navigation for functional endoscopic sinus surgery using structure from motion," in *SPIE Proceedings Vol. 9784: Medical Imaging 2016: Image Processing*, Mar. 2016.

[22] D. Mirota, H. Wang, R. Taylor, M. Ishii, G. Gallia, and G. Hager, "A System for Video-Based Navigation for Endoscopic Endonasal Skull Base Surgery," *IEEE Trans. Med. Electron*., vol. 31, pp. 963–976, Apr. 2012.

[23] D. Burschka and G. D. Hager, "V-GPS(SLAM): vision-based inertial system for mobile robots," in *2004 IEEE International Conference on Robotics and Automation*, vol. 1, pp. 409–415 Vol.1, Apr. 2004.

[24] M. R. Robu, P. Edwards, J. Ramalhinho, S. Thompson, B. Davidson, D. Hawkes, D. Stoyanov, and M. J. Clarkson, "Intelligent viewpoint selection for efficient ct to video registration in laparoscopic liver surgery," *Int. J. Comput. Assist. Radiol. Surg*., vol. 12, pp. 1079–1088, Jul 2017.

[25] D. Burschka, M. Li, M. Ishii, R. H. Taylor, and G. D. Hager, "Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery," *Med. Image Analysis*, vol. 9, pp. 413–426, Oct. 2005.

[26] C. Schmalz, F. Forster, A. Schick, and E. Angelopoulou, "An endoscopic 3d scanner based on structured light," *Med. Image Analysis*, vol. 16, no. 5, pp. 1063 – 1072, 2012.

[27] H. N. D. Le, H. Nguyen, Z. Wang, J. Opfermann, S. Leonard, A. Krieger, and J. U. Kang, "An endoscopic 3d structured illumination imaging system for robotic anastomosis surgery," in *SPIE Brain: Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XVI*, 2018.

[28] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov, "Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery," *Med. Image Analysis*, vol. 17, no. 8, pp. 974 – 996, 2013.

[29] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.

[30] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis*., vol. 60, pp. 91–110, Nov. 2004.

[31] H. Wang, D. Mirota, and G. D. Hager, "A Generalized Kernel Consensus-Based Robust Estimator," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 32, pp. 178–184, Jan. 2010.

[32] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst*., vol. 110, pp. 346–359, June 2008.

[33] G. Bradski *Dr. Dobb's J. Softw. Tools*, 2000.

[34] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2004.

[35] P. Moulon, P. Monasse, and R. Marlet, "Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion," in *IEEE International Conference on Computer Vision*, pp. 3248–3255, Dec. 2013.

[36] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment - a modern synthesis," in *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ICCV '99, (London, UK, UK), pp. 298–372, Springer-Verlag, 2000.

[37] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3d Slicer as an image computing platform for the Quantitative Imaging Network," *Magn. Reson. Imaging*, vol. 30, pp. 1323–1341, Nov. 2012.

[38] R. B. Rusu and S. Cousins, "3d is here: Point Cloud Library (PCL)," in *2011 IEEE International Conference on Robotics and Automation*, pp. 1–4, May 2011.

[39] N. Koenig and A. Howard, "Design and use paradigms for Gazebo, an open-source multi-robot simulator," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings*, vol. 3, pp. 2149–2154 vol.3, Sept. 2004.

[40] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[41] F. C. Park and B. J. Martin, "Robot sensor calibration: solving ax=xb on the euclidean group," *IEEE Trans. Robot. Autom*., vol. 10, pp. 717–721, Oct 1994.

[42] B. Atcheson, F. Heide, and W. Heidrich, "CALTag: High precision fiducial markers for camera calibration," in *15th International Workshop on Vision, Modeling and Visualization*, November 2010.

[43] C. Rohrmeier, S. Schittek, T. Ettl, M. Herzog, and T. S. Kuehnel, "The nasal cycle during wakefulness and sleep and its relation to body position," *Laryngoscope*, vol. 124, pp. 1492–7, June 2014.

[44] M. Ye, E. Johns, B. Walter, A. Meining, and G. Yang, "An image retrieval framework for real-time endoscopic image retargeting," *Int. J. Comput. Assist. Radiol. Surg*., vol. 12, no. 8, pp. 1281–1292, 2017.

[45] X. Luó, T. Reichl, M. Feuerstein, T. Kitasaka, and K. Mori, "Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation," in *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part III*, ACCV'10, (Berlin, Heidelberg), pp. 409–421, Springer-Verlag, 2011.