1 Specific Aims

It is estimated that there are more than 200,000 functional endoscopic sinus surgeries (FESS) procedures performed annually in the United States at a cost of several billion dollars annually. As the name implies, all of these procedures are performed under endoscopic guidance, and a large fraction employ surgical navigation systems to visualize critical structures that must not be disturbed during the surgery.

Although navigation is widely employed for FESS, its capabilities are far from optimal. In particular, the sinuses contain structures that are smaller than a millimeter in size, and yet delineate critical anatomy such as the optic nerve or the carotid artery. However, the accuracy of navigation is 2 mm under near ideal conditions [1–5]. As a result, navigation can provide a qualitative sense of location, but final confirmation of anatomic structures ultimately relies on the surgeon's ability to interpret and relate the CT image to the endoscopic view [6]. This process, which is further complicated when the anatomy is distorted or otherwise altered by surgery, requires time, skill and experience and can lead to errors in judgement that adversely affect outcome.

The goal of this project is to overcome these limitations using images from the viewing endoscope. We first propose to improve registration accuracy through in-situ registration of computed tomography (CT) imagery to intraoperative endoscopic images. Improved accuracy has immediate significance as it will allow the surgeon to better visualize and discriminate critical structures, thus reducing the potential for surgical error and improving workflow. The second enhancement is computation of the shape of viewed surfaces directly from endoscopic video sequences. The significance of this enhancement is an improved ability to assess the progress of surgery and a reduced need for intraoperative imaging. Together, these capabilities will provide improvements in visualization that will enhance safety, reduce complications and morbidity, and improve clinical workflow.

The key innovation in our approach is the use of the images from the endoscope itself to perform registration and surface reconstruction. The fundamentals of computing geometric information from video have been well known for decades, but recent advances in computational vision have now made these methods practical to apply at scale. In prior work [7–14], we have developed prototype algorithms to show that these advances can be translated into the medical domain. These results set the stage for the broader, systematic development and exploration of video-CT registration and video-based reconstruction described in this proposal.

We have assembled a team of experienced engineering and clinical investigators to carry out the four specific aims of this project:

Specific Aim 1: Develop video-CT registration algorithms that are accurate to CT resolution.

We propose to develop video-CT registration algorithms that operate within 10 seconds and have an accuracy of 0.5 mm, comparable to the resolution of intraoperative CT scans.

Specific Aim 2: Develop methods for surface shape estimation from endoscopic images.

We propose to develop algorithms that are able to compute a surface reconstruction from video to an accuracy of 0.5 mm so that anatomic changes and surgical progress can be measured at any point of a procedure.

Specific Aim 3: Perform comparative evaluation of video-CT-based navigation on patient data.

We will perform algorithm evaluation on data from cadaver models, and then retrospectively on patient data acquired during during FESS procedures using an intraoperative CT-based navigation system.

Specific Aim 4: Assess the accuracy and reliability of intraoperative surface estimation on patient data.

Using the patient data acquired in Aim 3, we will evaluate the effectiveness of surface reconstruction from endoscopic video in representative clinical settings.

Successful execution of this project will demonstrate a *new paradigm* for surgical navigation and visualization, namely the fusion of widely available endoscopic video with traditional navigation approaches. We expect many of these ideas to apply immediately to other areas of endoscopic endonasal surgery. More broadly, optical imaging devices are widely used in many areas of surgery, and surgical navigation systems have become an indispensable tool for procedures in neurosurgery, craniofacial surgery, spine surgery and orthopedic surgery. Thus, successful image-based registration and reconstruction algorithms have potential use in these areas of clinical practice, and therefore a high potential for translation into broader clinical practice.

2 Research Strategy

2.A Significance

Figure 1: The Brainlab Brainsuite iCT™ intraoperative CT and navigation system at Johns Hopkins Bayview Campus.

Clinical Background Chronic rhinosinusitis, defined as inflammation in the nasal passages and surrounding paranasal sinuses present for more than twelve weeks, is a source of significant health care expenditure. Its prevalence is estimated at 146 per 1,000 people. It results in 18 to 22 million physician visits per year and its direct treatment cost is conservatively estimated at $3.4 to 5 billion annually with indirect costs exceeding $14 billion. It is typically treated medically using anti-inflammatory drugs and antibiotics. When medical therapy fails, surgery is performed to improve sinus aeration and drainage. Roughly 200,000 sinus surgeries are performed in the United States per year making it one of the more common surgeries performed. A large proportion of these are Functional Endoscopic Sinus Surgery (FESS) which, as the name implies, makes use of an endoscope to visualize the procedure. Indications for FESS include blockages due to chronic sinusitis, tumors appearing in the sinuses, and, more recently, staging for access to the anterior skull base.

The paranasal sinuses are in close proximately to the brain, carotid artery, and eyes; these structures are separated from the nose by thin bone. This bone can be as thin as a few hundred microns and, in the case of the carotid artery and optic nerve, can be missing, making the surgery high risk. Injury to these structures is catastrophic and is considered a "Never event." Because the disease primarily affects quality of life as opposed to morbidity and mortality, the risk must be lowered significantly to place the risk to benefit ratio in the acceptable range.

Surgical navigation systems, which allow for the real time tracking and localization of surgical instruments with respect to surrounding anatomic structures, are thus essential for safety and are increasing becoming integral components of the sinus surgery workflow, particularly during situations where the surgical anatomy is distorted either through the disease process itself or through previous surgery or trauma [6]. Navigation also plays an integral role in extended endoscopic procedures where anterior skull bases pathologies, such as tumors and cerebrospinal fluid leaks are addressed. In short, the evolution of this field is now driven in part by technical advancements such as computer-aided surgical navigation.

A typical surgical navigation system consists of a computer workstation, navigational tracking device, and associated tools with marker devices whose pose (position and orientation) is continuously measured relative to a tracking device. Through a registration between the patient anatomy and a preoperative image, a navigation system allows a surgeon to visualize the location of a tracked tool relative to the the preoperative image. Navigation systems were first developed in the 1980's for neurosurgical applications (e.g., [15–17]), and have now been applied in many surgical fields, including neurosurgery [18–26], craniofacial surgery [27–31], ENT [32–35], spine surgery [36–39], and general orthopedic surgery [40–45]. Applications in other surgical disciplines, such as minimally-invasive hepatic surgery [46–48], kidney surgery [49–51], and other procedures (e.g., [52]), are a subject of current research. Commonly used tracker technologies include specialized optical tracking systems [20, 38, 53–56], conventional cameras [57, 58], electromagnetic sensors [59, 60], acoustic sensors [24, 53], and mechanical linkages [16, 21, 24]. Currently, systems based on specialized optical devices such as the Polaris^{TM} or OptoTrak^{TM} (Northern Digital, Waterloo, Ont.) are the most accurate [61], but recent advances have placed electromagnetic tracking on nearly equal footing in some applications.

Limitations of Current Techniques Modern navigation systems suffer from two limitations. First, empirical studies report that most navigation systems today achieve position errors on the order of 2 mm [1–5], **a number that has not changed in over a decade.** This number is large compared with many anatomic structures in the sinuses. Indeed, Lapeer et al. [3] specifically states that **current registration accuracy of 2 mm is *inadequate* to support effective fused video-CT display in sinus surgery**, suggesting that errors of this magnitude are a limitation to progress of the field.

The clinical impact of poor accuracy in navigation is expressed well in a paper by Cohen and Kennedy which states "The technology is still susceptible to displacement of the registration and computer malfunction, and *repeated visual confirmation of registration should be performed during surgery ..*" (emphasis added). An apt analogy would be to think of navigating with a GPS system where the accuracy was at the level of blocks, rather than streets. The GPS would help, but knowing where you were would still require constant reference to a map.

What is the source of this error? The accuracy of traditional navigation is a consequence of the indirect nature of the tool-to-anatomy calculation. Suppose $\mathbf{F}_t$ represents the coordinate transformation of the tool tip relative to the tool marker and $\mathbf{F}_a$ is the coordinate system of some anatomic feature in CT coordinates. Then the relation between tool and anatomy, $\mathbf{F}_{at}$, is given by

$$\mathbf{F}_{at} = \mathbf{F}_{ca}^{-1}\mathbf{F}_{ct}\mathbf{F}_{tr}^{-1}\mathbf{F}_t \tag{1}$$

Each of the elements on the right hand side is subject to error: $\mathbf{F}_{tr}$ is measured by the tracking system; $\mathbf{F}_{ct}$ and $\mathbf{F}_t$ are computed by calibration and are assumed to be constant; and $\mathbf{F}_{ca}$ relies on accurate location of markers in the CT image.

A second limitation of current systems is that they continue to display the same image information, even as the anatomy is altered during surgery. As a result, the relationship between the endoscopic view and the navigation view is slowly lost over time, making interpretation of navigation information progressing more difficult. This has led to a recent interest in intra-operative cone-beam [12] or CT imaging as a way to update the visualization during surgery. Indeed, the importance of intraoperative imaging has led manufacturers to introduce systems such as the BrainLab Brainsuite iCT^{TM}system (shown in Figure 1) and the Medtronic O-Arm^{TM}system. While these systems undoubtedly increase the safety of surgery, they also introduce additional radiation exposure to the patient and significant additional cost. The increased cost may in turn limit the access of this technology to patients due to the financial pressures in the current health care climate.

Significance The significance of our work is the introduction of a paradigm shift in surgical navigation by using a device present in *every endoscopic surgery*, namely the endoscope, to improve registration and visualization of anatomy. This will have numerous positive impacts. Most importantly, our work will provide an inexpensive, non-invasive, radiation-free method to enhance registration accuracy at any point of the procedure. Enhancements in registration will reduce ambiguity for the surgeon during surgery, enhancing confidence, and improve workflow by reducing the need to re-register or re-image the patient. The endoscope will also be used as a measurement device to update anatomic models during a procedure. This not only will improve the ability of the surgeon to visualize the progress of the surgery, but it will accrue additional benefits to the patient and hospital, as it may reduce the level of radiation exposure and cost by eliminating the need for intraoperative CT imaging.

Finally, as discussed above, surgical innovation is, to some degree, co-dependent with technical innovation. The introduction of an accurate, intraoperative registration and surface revision facility would almost surely in turn spur dependent innovations in visualization and surgical technique currently not possible as noted previously [3]. We believe such advances will broaden the applicability of minimally invasive surgery of the sinuses and the anterior skull.

2.B Innovation

Computational vision had made remarkable strides during the past decade, to the point that it is now possible to mine online image archives for hundreds or thousands of images [62], to compute the spatial locations from which they were taken, and to fuse that information into a 3D model of the scene. It is also now possible to process video sequence in real time to perform mapping and 3D reconstruction [63]. The more recent advent of high-definition digital imaging in endoscopic surgery offer similar opportunities in medicine. By translating advances in computational vision into endoscopic sinus surgery, we will transform the endoscope from a visualization device to an instrument for *quantitative 3D measurement*. We refer to this approach as *quantitative endoscopy* (QE). With QE, endoscopic measurements can be combined with traditional imaging modalities, such as CT, providing new capabilities such as enhanced navigation, tissue surface reconstruction, and fused image visualization.

The specific innovation of the proposed project lies in the development of QE to create a *constantly updated reference for surgical navigation and visualization*. Existing methods rely largely on a pre-surgical image and registration. Changes during surgery are not measurable, except via intra-operative imaging and re-registration. However, this updated imaging comes at significant time cost, interruption of the clinical workflow, radiation exposure, and thus real cost to patient and physician.

Prior Work: Several groups, including our own, have attempted to use information from the endoscope to improve registration. Roughly, these may be broken down into methods that determine the camera pose by minimizing a difference measure between an observed image and one predicted from the preoperative data [35, 64–68], and geometry-based methods that first reconstruct partial 3D models from multiple video images and then compute a registration using 3D-to-3D techniques [69–71]. Thus far, most reported methods for endoscopic surgical navigation have been of the first type. An important limitation of these methods is that they require robust methods for predicting image appearance to be accurate. Our work has explored methods of the second type [7–14, 73].

Stereo and multi-view reconstruction in computer vision is heavily studied [74–77]. Reported work on surface reconstruction from endoscopic video is less common. Co-PI Taylor performed early work [78, 79] using hierarchical correlation. Most recent work has focused on the challenges of dense soft tissue reconstruction from stereo [80–88], there there has been some investigation of monocular reconstruction [89]. To date, there are no stereo endoscopic devices that are widely used in endonasal surgery, a situation unlikely to change[1] in the near future. Thus, we have chosen to focus on applying multi-view reconstruction techniques to traditional endoscopic images to **update** a prior model. It is important to note that these methods would be enhanced by stereo endoscopy should it become available.

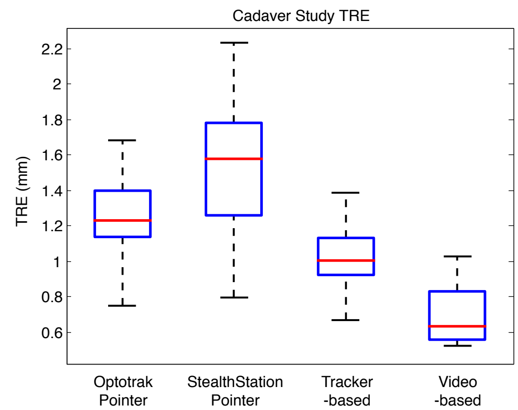


Figure 2: Target registration error (TRE) for cadaver study #2 for our laboratory Optotrak based system, a Medtronic StealthStation, and Video-based and Video-tracking based registration. The rank statistics from top to bottom are the maximum, third quartile, median, first quartile and the minimum.

Technical Innovations: Our prior results demonstrate the promise of QE for both navigation enhancement and reconstruction. Our first innovation is the use of direct, local measurement of $\mathbf{F}_{at}$ through high-accuracy registration of video imagery to surfaces in a pre-operative CT scan (*video-CT registration*), thus providing a new, high-accuracy solution to enhance the precision of navigation (Aim 1). In recent work, we have demonstrated that video-CT registration is able to improve the accuracy of traditional navigation systems by more than a factor of two to submillimetric levels (Figure 2). Further work is proposed to improve this to a target level of 0.5 mm, to enhance the speed of the process, and to develop online quality assessment metrics for registration accuracy.

Our second innovation is the development of high-accuracy multi-view reconstruction methods from sequences of biomedical images (*intraoperative surface reconstruction*), thus providing a non-invasive, radiation-free means of measuring changes to anatomy intraoperatively (Aim 2). In recent work, we have demonstrated the feasibility of producing a high accuracy surface reconstruction from video acquired by a tracked endoscope as shown in Figure 3.

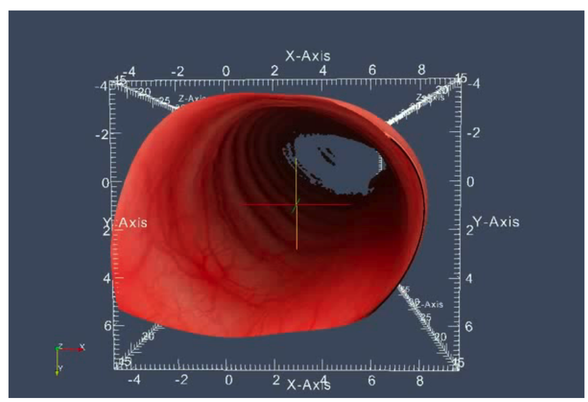


Figure 3: A full 3D reconstruction of a pediatric airway from video imagery acquired with a tracked endoscope.

Although the technical bases for video-CT registration and intraoperative surface reconstruction are relatively well-understood, the application of these ideas to medicine has been limited by the complexity of biomedical endoscopic images. Our initial results suggest that these hurdles can be overcome and, as further described in Section 2.C, practical video-enhanced navigation is within reach.

Clinical Innovations: Our project will provide three unique and significant innovations for endoscopic sinus surgery. First, video-CT registration provides a means for improving the usability of existing navigation technology in sinus surgery *with no additional cost or equipment*, and with *minimal disruption to the surgical workflow.* We believe this will lead to reductions in the time necessary to perform surgery (by reducing or eliminating the time taken to use traditional navigation methods), and reduce the likelihood of surgical errors. Second, surface reconstruction from endoscopic images to *compute anatomic changes intraoperatively* as surgery progresses, thus providing a new way of monitoring surgical progress, again improving, time-efficiency, patient safety, and offering the potential for cost reduction. Finally, improvements in

[1]We have in fact evaluated two generations of stereo endoscopy systems by Visionsense (Orangeburg, NY) both technically and clinically. Although they are improving rapidly and are exciting to use, such systems are still noticeably inferior to traditional endoscopes.

registration will provide accuracies that make it possible to definitively visualize thin bony structures such as the covering of the carotid artery. We hypothesize that such accuracies will enable new surgical approaches would be otherwise impossible to safely pursue.

Our evaluation aims (Aims 3 and 4) are specifically designed to translate video-CT registration and video-based surface revision into clinically relevant and practically usable tools that will support these clinical innovations.

2.C Approach

We have structured this project into four aims. The first two aims are focused on developing technical capabilities, specifically registration (Aim 1) and reconstruction (Aim 2). Aim 3 describes data collection and evaluation methodologies for video-CT registration. Aim 4 extends this evaluation to include the effectiveness of surface reconstruction.

2.C.1 Aim 1: Develop video-CT registration algorithms that are accurate to CT resolution

Aim Objective: This aim will develop video to CT registration algorithms and surface reconstruction algorithms, each of which operates in less then 10 seconds and provides registration accuracy of 0.5 mm or better. This time budget, which is an *upper* bound that we expect to improve on, was chosen because it is comparable to the time necessary to manipulate tracked tools in and out of the field when using current navigation systems, and far less than the time that would be necessary for reconstruction on intraoperative CT systems. We have chosen this level of accuracy as it is comparable to commonly available CT image resolution.

Preliminary Results: We have developed and evaluated a prototype method for video-CT registration with support from NIH R21 EB005201 [9–11, 13, 72, 73]. The system consists of the components shown in Figure 4. Although not shown, the reconstruction component is composed of two modules: robust motion estimation, and point reconstruction. System processing proceeds as follows. Image features are detected and matched in two temporally adjacent images. These matching pairs are then used to estimate the camera motion using a robust estimator we have developed [10]. Once the camera motion is estimated, the 3D locations of the matched features are reconstructed. The reconstructed 3D surface points are then passed to the 3D-3D registration component.

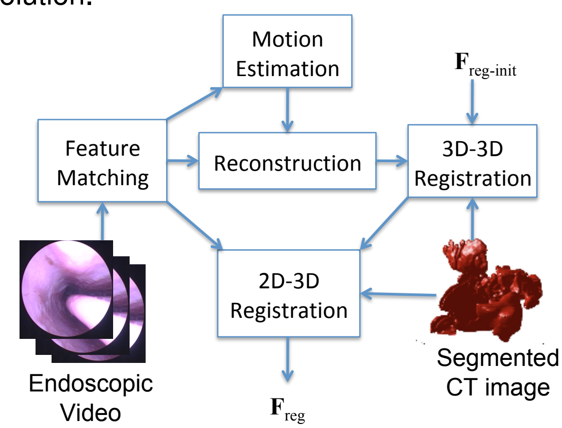


Figure 4: Motion estimation and registration chain implemented during prior R21 grant.

The 3D-3D registration component is composed of two modules, a renderer, and a robust registration algorithm. First, the tissue surface is segmented from the CT. An initial camera placement is chosen. The renderer is used to compute the visible section of the surface. The visible surface and reconstructed points are then registered with a modified Trimmed Iterative Closest Point algorithm [11, 90]. Once an initial registration is established, it is tracked using 2D-3D (viewpoint-based) registration methods [91]. In [13], we have shown results that indicate: 1) higher registration accuracy with video-CT methods than with traditional methods (see Figure 5), and 2) sub-millimetric error in while computing an updated registration through multiple video frames with a mean translation error of $0.84mm$.

Task 1.1: Implement registration at speeds consistent with use in surgical workflow. In this task, we will re-implement the existing software chain shown in Figure 4 to enhance with computational performance and further improve the accuracy and reliability of core algorithmic components. Our development will make extensive use of our CISST software development environment [92, 93]. Briefly, CISST provides support for efficient, rapid software development by providing libraries that support multi-threaded numerical processing, standard coordinate transformations, registration, calibration, and so forth. CISST also "wraps" existing software libraries such as OpenCV [94] which support image processing. A staff software engineer will provide the support necessary to implement and extend the existing processing chain. Our performance objective will be to perform the existing processing chain in no more than 10 seconds with the same reliability and accuracy of the existing Matlab implementation. A breakdown of timing[2] is shown in Table 1. The following provides specific details on the proposed implementation and how it differs from the current prototype.

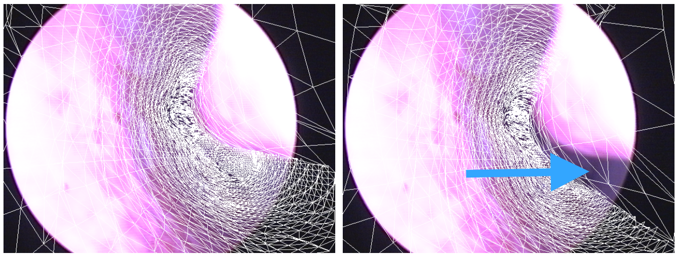


Figure 5: A comparison of our registration method (left) and traditional navigation using an Optotrak (right) at the end of a video sequence. The arrow indicates an obvious error in the latter.

Image Selection: The current implementation makes use of a fixed time offset between pairs of images for which feature matches are computed. However, the accuracy and reliability of the initial registration solution depends on the motion between the images, and the number of pairs used. The minimum number of images necessary for initial reconstruction and registration is 2 (i.e. one pair); thus 3-5 well-chosen images can provide a suitably redundant solution. These 3-5 images must be chosen from that roughly 30-200 that will be captured during 1-3 seconds of video identified for registration. We have recently developed a method for automated choice of video frames to produce well-constrained epipolar estimates [72]. This method will be further refined and implement to provide enhanced frame selection for registration.

Feature Matching: Feature detection and matching is currently the slowest element of the prototype system, largely due the poor scalability of the SVD-SIFT method that was used [100]. We will instead make use of the SURF [101] image matching. We have already performed initial studies of SURF-based matching on existing sinus images and have found it operates in 0.75 sec and provides matching performance comparable to the current SVD-SIFT matcher. We will augment SURF matching with an internal motion consistency constraint. Suppose that, for three images, we have the following location pairs for matches of the same feature in three images: (q_1, q_2), (q_2, q_3) and (q_1, q_3). Let $d_{i,j} = q_i - q_j$. Then it should be the case that $d_{1,2} + d_{2,3} = d_{1,3}$. If any match is in error, this will not be true. Further, if matching is ambiguous, we can use this constraint to choose the best (in terms of consistency) match.

Motion Estimation: Camera motion is currently be estimated using our already well-validated ASKC method [73] followed by a standard continuous optimization to compute camera pose. A non-optimized Java implementation of the latter has been tested and found to operate in approximately 5 sec. for 5 images; we expect a C++ implementation to perform at least a factor of two better.

Function	Time
Video Acquisition	3 sec
Feature Matching	1 sec
Motion Estimation	2 sec
Initial Registration	2 sec
Registration Refinement	2 sec
Total	10 sec

Table 1: Proposed module times for registration system operation.

The current implementation establishes the initial registration on frame pairs. It is well-known that multi-frame registration [102, 103] will produce a higher quality estimate. We will adapt a multi-frame estimation procedure that we have developed for airway reconstruction to this application to drive registration error below the target 0.5 mm threshold. The structure of this system is depicted in Figure 6.

Registration: We will estimate the initial 3D-3D registration using our current method described in [11]. An initial registration estimate will be provided by an tracking target attached to the endoscope (see Section 2.C.3). Registration refinement will be performed as description in [13]. Matlab implementations of these modules take roughly 30 seconds to execute, however performance improvements of 30-50x with C re-implementation and GPU acceleration is common.

Potential Pitfalls and Alternatives: The primary risk in this task is that the overall implementation is unable to perform in real time with the required reliability and accuracy. There are however several further computational optimizations possible, for example using the recent FAST feature detection framework [104] (reported times as low as 1.08 ms/frame), using GPU accelerated libraries [105], multi-scale methods, and careful exploitation of system sparsity in the optimizations. Robustness can be enhanced using our recent results on meta-matching methods [106] and robust registration [11]. Finally, we can also investigate the use of dense (i.e. pixel level) multi-view reconstruction to provide additional data for registration (see Figure 8c).

Task 1.2: Registration Quality Assessment The focus of this task the creation of methods to perform online analysis of the accuracy of the video-CT registration. The role of these methods is to ensure that information is not

[2] It is worth noting that there are now systems that report performance of several of the component elements of our architecture in real time [95–99] on non-biomedical images.

provided to the clinician until an appropriate level of reliability is reached. We propose to develop and evaluate two methods, one analytical and one statistical.

Analytical Quality Control: This method measures projection error in the image, computes the corresponding pose error numerically, and evaluates the solution for statistical consistency.

Consider a CT fiducial with location $p_i \in \Re(3)$ which appears in an endoscope image acquired at time k (Figure 7). Let $q_i^k \in \Re(2)$ denote the image location of p_i in image k. Given an estimated video-CT registration, $\hat{M}^k$, the an estimated projection $\hat{q}_i^k$ of p_i can be computed using

$$\hat{q}_i^j = \Pi(\hat{M}^j, p_i) \tag{2}$$

where Π represents the process of camera projection. The projection error is thus $e_i^k = \hat{q}_i^k - q_i^k$.

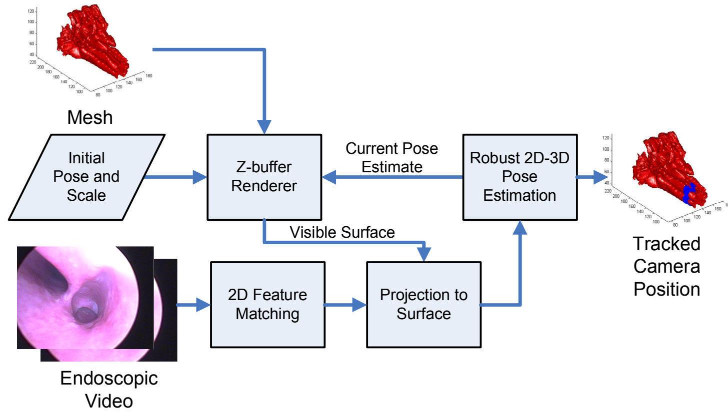


Figure 6: Data flow of the proposed dense bundle adjustment algorithm.

We assume the fixed parameters of the projection geometry (image distortion and camera internal parameters) are fixed and known through calibration. If we consider the second order statistics of e, it follows that $\Lambda_e = \Lambda_{\hat{q}} + \Lambda_q$ where $\Lambda_{\hat{q}}$ is the covariance of projected image locations and Λ_q models (measured) feature localization error. Note that Λ_e is observed and Λ_q can be estimated offline empirically. Carrying through a first order linearization of (2) and accounting for other noise sources (specifically CT and video location inaccuracies), we can relate $\Lambda_{\hat{q}}$ to the camera pose (contained in M) as

$$\Lambda_{\hat{q}} \approx J_M \Lambda_M (J_M)^t + J_p \Lambda_p (J_p)^t \tag{3}$$

where J_M is the Jacobian matrix of (2) with respect to camera pose M, J_p is Jacobian of (2) with respect CT image location p and Λ_p is the second order statistics for fiducial localization error in CT. This linearization can be extended to n observed points, in which case J_M is $2n$ by 6 and J_p is $2n$ by $3n$. If there is a unique registration solution, J_M must be full rank and therefore invertible in the least squares sense. Let $K = (J_M^t J_M)^{-1}(J_M)^t$ denote this inverse. Recalling that $\Lambda_e = \Lambda_{\hat{q}} + \Lambda_q$, we can now solve for the unknown registration error statistic, Λ_M, as

$$\Lambda_M \approx K \left[\Lambda_e - \Lambda_q - J_p \Lambda_p (J_p)^t\right] K^t \tag{4}$$

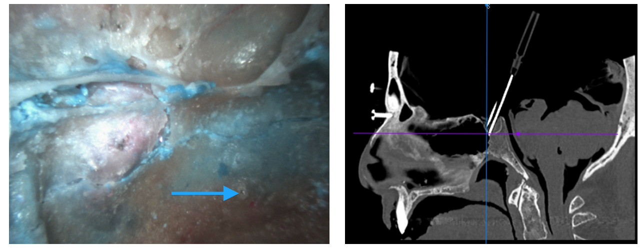
The result is an estimate of the camera localization error as a function of registration error in 2D-3D correspondences. Thus, given imputed matches in CT and video, we can calculate an estimate of registration error.

This analysis will be extended to account for the ambiguity of video-CT matches. For example, high curvature points in the surface typically produce visible image features that are constraining in all three directions. Conversely, smooth surfaces may produce visible features due to surface texture, but the local surface geometry only provides one degree of surface constraint – i.e. constraint along the surface normal.

Statistical Quality Control: Equation (3) can be exploited in another way. Given a desired level of registration quality, Λ_M, it is possible to compute $\Lambda\hat{q}$ and thus to *test* if the value of Λ_e computed from match data is consistent with the corresponding analytical value. A closely related approach is to note that there are many more images available than are practically usable for one registration step. Thus, it is possible to compute several independent registration solutions *in parallel.* We expect the variance of this "sampling" of $\hat{M}$ to be comparable to the corresponding analytically computed values; indeed it is quite likely that one or more of these solutions is measurably superior based on any of the previously developed measures. We will explore both analytical and sampling based approaches as a means of creating a method of internal quality control of registration solutions. The evaluation of these solutions will also provide a means of evaluating the performance and accuracy of the underlying registration chain.

Evaluation data will be recorded from both cadaver studies and from live surgeries (see Aim 3). For these purposes, a small number of data sets will be selected as engineering data. In both cases, we will perform a careful manual analysis to establish correspondence between video and CT data sets (Figure 7). Given the manually observed location q_i^k and the corresponding value p_i for at least three locations, it is possible to compute a gold standard estimate of the observing camera location M directly using well-established methods [91]. This estimate can be directly compared to the result of the video-CT registration algorithm and an error computed. Note that this error will still be subject to errors in manual location of the fiducial in both video and CT images, and thus must be corrected for these factors using methods similar to those described in our evaluation methodology (Aim 3, below).

Pitfalls and Alternatives: We already have experience with both the analytical [72] and statistical [12] approaches to accuracy evaluation. The primary risk in this work is thus not the development of the methods, but the ability to evaluate of their effectiveness. If the proposed manually developed goal-standard is inadequate, we will also acquire cadaver data using a micron precision robot available in our laboratory (JHU steady-hand robot (SHR) [107–109]). The endoscope will be tracked using a high-precision optical measurement system of our own design [13]. The SHR is known to be at least two orders of magnitude more precise than optical tracking. Thus the initial registration can be compared with the optotrak system, and the relative motion of the endoscope can be compared against the robotic baseline.



(a) **(b)**

Figure 7: A 23 Ga needle tip in optical (left) and CT (right) images. These provide a fiducial that can be localized with sub-millimeter precision in both video and CT for gold standard registration.

2.C.2 Aim 2: Develop methods for surface shape estimation from endoscopic images

Aim Objective: The goal of this aim is to develop tools that would allow the computation of an updated surface model for areas that have been modified during surgery. The accuracy goal 0.5 mm. Normally, changes to the anatomy could only be visualized using additional imaging (e.g. intraoperative CT) during the procedure. Typically, such imaging is performed at the end of the case, to ensure the surgical goals are met, or at a point where the accuracy of navigation is called into question. It is assumed that the area of interest (e.g. the face of the sphenoid sinus) has been indicated in the pre-operative CT prior to the procedure. Using the registration methods of Aim 1, it is thus possible to determine the corresponding area of any registered endoscope image using (2). This area will then be reconstructed from multiple endoscopic images.

Preliminary Results: Our group has extensive background in reconstruction of anatomic surfaces from video imagery in cardiac applications [80], urologic applications [110], and in retinal applications [111]. In unpublished work, we have developed a dense surface reconstruction algorithm based on the camera motion estimated from feature matches as described in Aim 1. Figure 8a and 8b, shows the results for a pair of images (resolution 640x480 pixels) taken from a porcine data set. Figure 3 shows a reconstruction of a pediatric airway using an extension of these methods to multiple views. Analytical estimates of reconstruction error using established methods [74] suggest that such reconstructions would have an accuracy of approximately 0.3 mm RMS at a distance of 2 cm with a 1 mm lateral motion of the endoscope. Empirical evaluation of our airway reconstruction data against a CT image shows an observed error of 0.22 mm - 0.3 mm, in good agreement with this estimate.

Methods: The problem of interest here is relatively unique in the computer vision literature because we can assume that 1) we always know a prior surface for the viewed image (either from the CT or from a prior reconstruction) and 2) the modification of the current surface is a process of tissue *removal* from the prior surface. Thus, we pose this problem as one of *locally deforming* the prior surface to the current surface and will adapt the methods used to reconstruct the airway in Figure 3 to this problem.

Given a set of 3D points $(p_1 \dots p_k)$ in a reference frame M^0, the surface viewed from M^0 can be represented using a thin plate spline function using their image coordinates $(q_1 \dots q_k)$ as control points. We adopt the thin-plate spline parameterization described in [87, 112], which maps image points to distance in the camera Z direction along their associated ray. Given control points $(q_1 \dots q_k)$, the mapping function $TPS(\cdot; W)$ is defined by parameter vector $[W = (w_1, \dots w_k, r_1, r_2, r_3)^T$, where the Z value of an image point q is given by

$$TPS(q; W) = r_1 + r_2 x + r_3 y + \sum_{i=0}^{k} w_i \mathbf{U}\left(\|q_i - q\|\right) \quad \text{where} \quad \mathbf{U}(x) = x^2 log(x). \tag{5}$$

We now take M^0 as a fixed reference frame for reconstruction. Given initial estimates of the camera matrices $\mathcal{M} = \{M^1, M^2, ...M^n\}$ and W from the methods of Aim 1, we simultaneously correct both using an image-based, dense brightness model. For a 3D point p, the brightness of its projection in the i^{th} image is given by $I_i(\Pi(M^i, p))$. If we assume brightness constancy at the surface (not true in practice; we revisit this below), the image of an arbitrary point p should have the same intensity when projected into any image. By selecting a uniform mesh of surface points $\mathcal{P} = \{p_1, p_2, \ldots, p_N\}$ with corresponding projections $\mathcal{Q} = \{q_1, q_2, \ldots q_N\}$ in M^0, we can define a photometric error term D_k, $k = 1, 2, \ldots n$ as

$$D_i^j = \Phi(I_0(q_i) - I_j(\Pi(M^j, q_i TPS(W, q_i)))), \quad q_i \in \mathcal{Q} \tag{6}$$

where Φ is an error measure. Summing D over all values of j and i yields an objective function to optimize.

We further assume that feature matches computed in Aim 1, having come from fiducial markers or robust matching, are accurate. The mapping function $TPS(\cdot; W)$ and camera matrices in $\mathcal{M}$ should preserve the location of feature matches across images as well. These terms can be written as equality constraints when minimizing (6). These two conditions can be combined leading to a Karush–Kuhn–Tucker matrix [113]. The final system can then be solved to improve both the camera positions $\mathcal{M}$ and the spline model W for the surface while preserving feature matches.

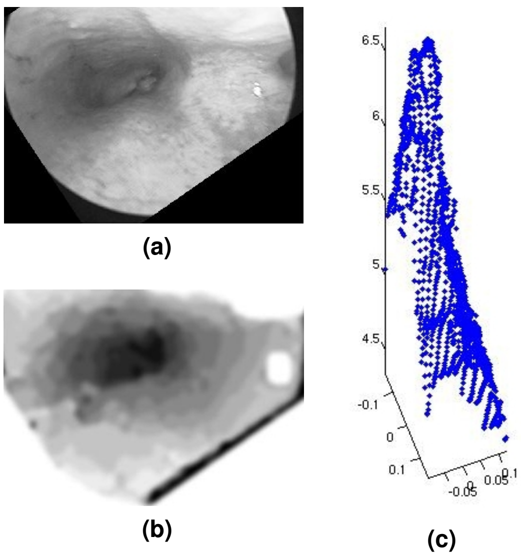


Figure 8: 8a is a sinus image from a porcine data sets, and 8b is corresponding dense stereo disparity map. 8c is the reconstructed sinus cavity for 8b subsampled by a factor of 100.

There are two refinements additional to be considered. First, (6) assumed image constancy at the surface, which in the context of endoscopy and other in-vivo procedures is not true. In general, reflectance modeling is a very difficult problem to solve, but, In the case of an endoscope, the light source is co-located with the camera. As a result, estimates of the camera positions contained in $\mathcal{M}$ provide both lighting and viewing direction. A viewing ray through p_i makes a corresponding angle θ_i with the surface normal of the spline model at p_i. We then replace (6) with an approximation of the Phong reflection model:

$$\alpha(cos(\theta_i) + \exp(cos(\theta_i)^{\beta}) - 1) \tag{7}$$

Here, α models the overall albedo of the point the β term accounts for the additional fall-off of specular reflection as $cos(\theta_i)$ decreases.

Second, the reconstruction is in the frame of an arbitrarily selected camera image. In order to be used for surface revision, this camera location much be known with high accuracy relative to the initial CT image – i.e. it must be accurately registered. This implies that either: 1) the image shares feature points with undisturbed anatomy in other images, in which case the video-CT registration can be extended to this frame; or, 2) navigation data is used to “bridge” one or more recent, nearby registration solutions to this point. Our focus will be to use method 1, but we will fall back to method 2 when shared feature points are not available.

To evaluate the performance of the system during the engineering phase, we will use sequestered cadaver and live surgery data acquired as described in Aim 3. We will make use of paired CT data sets taken before and after alteration of the anatomy together with corresponding video data. Areas of modification will be identified by first registering (using fiducials) the pairs of CT images, then performing change detection at the air/tissue boundary. We will perform video registration and surface modeling video data acquired before and after alteration. Using a gold-standard fiducial-based registration (see Aim 1), it will be possible to automatically compare the revision surface to the true CT surface and to thus tabulate errors. Our goal is to achieve agreement with CT to 0.5 mm RMS.

Potential Pitfalls and Alternatives: Potential pitfalls include: 1) difficulties in developing convergent and stable optimization methods; 2) difficulties in developing measures of image similarity that are insensitive to lighting variation; and 3) inadequate registration accuracy producing poor epipolar geometry. With regard to 1, in our prior work [80], we have shown that spline-based (parametric) representations can be optimized using continuous, direct methods and are able to tolerate significant levels of lighting change. If these methods prove unstable, we will instead turn to energy-based discrete global optimization methods as an alternative [114, 115]. We will also explore robust variants of the optimization procedure as described. With regard to 2, there are several well-known techniques that have been shown to be extremely tolerant to lighting variation, including standard image normalizations [74], mutual information [116], and explicit light-field modeling [117]. With regard to 3, we do not expect this to be an issue given the accuracy requirements we have already demonstrated in our prior results.

2.C.3 Aim 3: Perform comparative evaluation of video-CT-based navigation on patient data

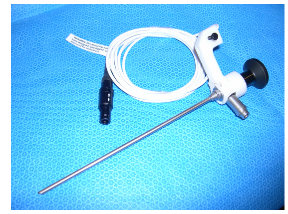


Figure 9: Left, a high-precision optically tracked endoscope developed for cadaver data collection and algorithm validation. Right, a more recent EM tracked scope for use in airway data collection.

Aim Rationale: The principle question addressed by this aim is whether video-CT registration can improve registration results to a clinically significant level, which we have chosen as 0.5 mm. This question will be answered through statistical analysis of measurements derived from patient data acquired during FESS procedures. This protocol will be prototyped in cadaveric subjects. Only when expected accuracy improvements and adequate levels of reliability are observed under laboratory conditions will testing with human subjects proceed.

Prior Results In prior work under R21EB008490 and R21EB005201, the PI has developed a data collection system that allows synchronized capture of endoscopic video and motion tracking data. These data collection systems have included the development of endoscopes tracked using both electromagnetic (EM) and optical tracking technologies as shown in Figure 9. In other projects, engineering and cadaver phantom data as well as intraoperative data has been recorded with the system.

A novel cadaveric model for navigation and visualization validation (further described below) has also been developed (Figure 10). The cadaver model is created by purchasing specimens with enhanced vasculature supplied by the Maryland Board. A craniotomy is performed and critical structures such as the optic nerve, carotid artery, and the sella turcia are outlined with 27 gauge needles as shown in Figure 10. After implanting the needles into the bone or tissue, Great Stuff (Dow Chemical Company, Midland, Michigan, USA) was used to secure and support the needles. The needles form a “gold standard” baseline against which the measurements that will be taken in-vivo can be assessed.

The results of a pilot system evaluation experiment on this model are reported in in a journal article currently under review [14]. The key result has been to demonstrate that target registration error (TRE, defined formally below) using video-CT methods is measurably improved over traditional methods (p-value less than 0.001) in a pilot study. An example of TRE improvement from one experiment is shown in Figure 2.

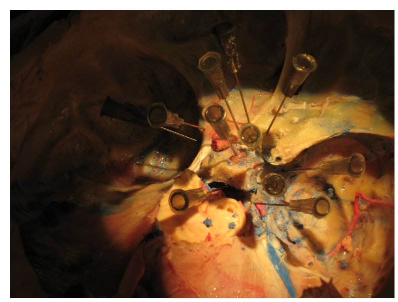


Figure 10: The cadaver model developed for cadaver data collection showing the placement of fine needles in and about the sinus cavities.

Methods Testing and validation of the algorithms described in Aim 1 will be performed on cadaver models using an expansion of the protocol established in [13, 14]. We will make use of an optically tracked endoscope similar to that shown in Figure 9, but modified to improve ergonomics. In particular, the modified endoscope will allow the camera to rotate freely about the endoscope optical axis. The rotation between the camera and the endoscope will be monitored automatically using the optical marker visible in the endoscope image. This design will be consistent with the tracked endoscope used in a typical clinical workflow with the Brainlab navigation system. It will be continually evaluated and modified as necessary during the initial phase of cadaver data acquisition. Prior to use, the endoscope will be calibrated using established methods [13]. This include an optical calibration of the endoscope itself as well as a geometric calibration relating the optical center of the endoscope to the tracker coordinate system. Data collection will only proceed when we have a calibration that is accurate to within $1/2$ pixel in image space and 0.5 mm in 3D space. Note that the latter is *not an element in the registration chain we are testing*, but provides a secondary means of accuracy evaluation as further described below.

Cadaveric data will be collected in a manner that simulates as closely as possible the live patient data collection. The heads will be prepared as described in [14] with needle tips just below the mucosal surface and imaged with a voxel size of at least 0.5x0.5x0.5mm. Registration will be performed using a commercial navigation system (Medtronic StealthStation) available in our laboratory. The endoscope will be run through the nasal cavity and a video-CT registration will be performed. An experienced surgeon will then perform a total spheno-ethoidectomy on the head exposing the ethmoid cavity and sphenoid sinus. The surgeon will then use the navigation system to navigate to each of the needle tips using the CT image based on each registration solution, and the navigation system and visual location of the needle tip will be recorded. Each target will be localized at least six times. This will comprise Cadaver Data Set 1 (CD1).

Following this the endoscope will be again be run through the sinuses, with particular attention paid to the sphenoid sinus. The needles will then be advanced to be visible to the endoscope. Also, in anticipation of studies based on operative data in Aim 4, CT-visible markers (1mm gold beads) will be placed as well using a temporary adhesive. The needle tips and the fiducials will be touched at least six times using a pointer under endoscopic feedback and the navigation system position will be recorded. This will comprise data set CD2. Finally, the head will undergo a second CT image, and the process of running the endoscope and touching the needle tips will be repeated six times. This will comprise data set CD3.

Using the acquired data, we will create test sequences from the ethmoid air cells and the sphenoid sinus. Particular attention will be paid to clearly defined anatomic structures such as the middle and lateral optocarotid recess, optic nerve, vidian canal sella turcia, anterior ethmoid artery, and the carotid artery. We will seek a minimum of 3 independent image sets from CD1, CD2 and CD3. For each of these data sets, we will perform video-CT registration on these sequences. These registrations and reconstructions will then be subjected to analysis.

Surgical Recordings will be acquired from endoscopic endonasal procedures to characterize algorithm performance on in-vivo data. Data acquisition will be performed using a BrainLab Brainsuite iCT^{TM}intraoperative CT system available at Johns Hopkins Bayview. Our CISST software is compatible with the Brainlab IGTLink interface which will allow us to record tracked tool positions during the surgery. Our video recording software will be adapted to ensure correct time-stamping of recorded data from the Brainlab system and tested on an engineering phantom prior to use. We propose to acquire data from a population of approximately 10 patients per year for up to four years (total 40) from a population undergoing endoscopic sinus surgery at Johns Hopkins Bayview (see Human Subjects section). CT images and navigation data will be captured from a Brainlab Brainsuite iCT ^{TM}system during the course of the procedure with minimal disruption to the normal workflow. Corresponding endoscopic video data will also be captured. All patient data will be anonymized so that no patient identification remains in the processed data set.

The data acquisition protocol will mirror that described above. Prior to the procedure, the patient will be CT imaged and registered. Using a navigated endoscope, the sinuses will be explored endoscopically, producing an initial video data set. The surgeon will also use a pointer tool to repeatedly touch a small set of well-identified anatomic landmarks, and temporarily implanted 1 mm gold beads while video and navigation data is being recorded. Together, this will comprise Live Data (LD) set #1. These landmarks form the surrogate for the needles used in the prior experiment. The FESS will proceed normally until the sphenoid sinus is reached. At this point, the same process of endoscopic exploration and navigation will be performed with care to include the modified anatomy. In particular, several points in areas of reconstructed anatomy will be touched with a navigated probe. This is data set LD2. The patient will undergo a second CT image, and a data set consisting of endoscopic video and navigation data matching LD2 will be acquired, forming data set LD3. This second CT scan is part of the standard patient workflow and is performed near the end of surgery to ensure surgical goals were obtained.

Post-procedure, the visual and CT coordinates of the landmark points will be manually annotated. Visual reconstruction will be performed on independent subsets of the video data.

Analysis: Will compare traditional navigation and video-CT-based navigation using three error metrics. The first is TRE_1 as described by Fitzpatrick and West [118]:

$$TRE_1 = \left\| \mathbf{p}_{CT} - \left({}^{CT}T_{Navigation} \right) \mathbf{p}_{pointer} \right\| \tag{8}$$

where $\mathbf{p}_{CT}$ is the target segmented from the CT, ${}^{CT}T_{Navigation}$ is the transformation from the tracking system to the CT as computed with the fiducial points and $\mathbf{p}_{pointer}$ is the current pointer location. The second is TRE_2 as described in Mirota et al. [14]

Metric	Description
TRE_1	Metric for evaluating accuracy of traditional pointer-based methods
TRE_2	Metric for evaluating video-based registration methods
NGE	Same as TRE_2, with the target is not visible in the endoscope image forcing reliance on navigation

Table 2: A summary of the metrics used in our experimental evaluation and their intended use.

$$TRE_2 = \left\| \mathbf{p}_{CT} - \left(\mathbf{t} + \mathbf{r} \left(\frac{\mathbf{r} \cdot (\mathbf{p}_{CT} - \mathbf{t})}{\mathbf{r} \cdot \mathbf{r}} \right) \right) \right\| \quad \text{where} \quad \mathbf{r} = RK^{-1}\mathbf{q}_{image} - \mathbf{t}, \tag{9}$$

where R and $\mathbf{t}$ represent the camera position and K is the camera internal parameters (together comprising M above). The vector $\mathbf{q}_{image}$ is the point segmented in the video image. Equation 9 effectively creates a virtual pointer along the ray from the endoscope with end point equal to the closest point to the target along the ray.

The third metric is NGE which is the relative distance between a target segmented in the CT data and the location the navigation system reported as the location of the target. Mathematically, NGE is the same as equation 9, but NGE is distinct from TRE in that it applies when the needle tips are not visible in the endoscope and so only the surgeon's hand-eye coordination and the navigation system define the location of the target. Table 2 summarizes the error metrics and their relevance to navigation accuracy. With these measures established the primary questions we will investigate are

1) Is TRE$_2$ for video-CT registration improved over traditional navigation in cadaveric data using fine needles as markers?

2) Is TRE$_2$ for video-CT registration improved over traditional navigation in cadaveric data using implanted beads as markers?

3) Is TRE$_2$ for video-CT registration improved over traditional navigation in patient data using temporarily implanted fine beads as markers?

Q1 represents a comparison of TRE$_2$ using our gold standard under laboratory conditions. Q2 is designed to simulate the human trial and determine the relationship of the bead data to the gold standard measurements. Q3 is then applies what we have learned to the human data and represents the central finding of this aim. The interpretation of its significance will be based on the findings in Q1 and Q2. As part of these tests, we will also calculate and report RMS registration error for each method.

To address Q1, we will repeat the analysis described in [14]. We assume that the three orthogonal components of both TRE_1 and TRE_2 are independent, normally disturbed variables with zero mean. $\phi = \|TRE_2\|^2$ is therefore a measure of total variance, so we test the null hypothesis that the expected mean ϕ values are equal to elucidate the relative performance of our navigation systems. Since ϕ represents the sum of three chi-squared variables, we use the zero-skewness log transform to normality to account for this. We will employ a mixed linear model for data analysis and will account for three major sources of variation: i) a subject term which accounts for experimental variability caused by differences between subjects; ii) a pin/marker term which accounts for experimental variability due to differences between pins; this accounts for differences between target registration error between pins due to the relationship between target registration error and fiducial registration error and variability between the surgeons' ability to touch pins in the nose due to anatomic constraints of the cadaver specimens; and iii) residual error. Residual error will be a function of the scatter in data caused by the collection method, i.e., test subjects touching pins with a pointer, in addition to the traditional sources of error attributed to regression analysis. The pin term is nested within the heads term. To this random effects model we will add fixed covariates to elucidate the effects of navigation methods on target registration error.

To compare the navigation systems, we calculate the marginal means treating the factor variables as balanced and then perform post hoc Wald tests corrected for multiple comparisons using Bonferroni's method. The post hoc tests perform hypothesis tests on the differences in transformed squared errors between methods. We will also compare ϕ to $\|NGE\|^2$ using a two-factor analysis of variance (ANOVA) for this comparison. The squared errors serve as the dependent variables. Two factors (navigation method and error type) will be used as independent variables. The two factors will be allowed to interact. We assume a global alpha of 0.05. Monte Carlo sample size calculations using the assumptions of video accuracy of 0.5 mm (SD 0.5mm); surgical navigation accuracy 2.0 mm (SD 2.0); scatter in subjects term of 6.76 mm2 and the ability to localize 5 markers per head shows a power of 0.79 with 15 subjects, 0.92 with 20 subjects and 0.97 with 25 subjects. These values were based on previous cadaver work. We have conservatively prepared to test 25 subjects, but are prepared to increase this number to 40 if scatter in live subjects is greater than anticipated.

In addition to our study on video based registration we will use our data to perform several related assessments. We will study the accuracy to which we can compute relative motion of the endoscope to determine the need for a tracked endoscope and further define the accuracy of commercial navigation systems based on intraoperative CT using the methods described by Strau et. al [5]. We will also investigate workflow by computing how much time is spent during surgery performing imaging and navigation, and how that compares with our expect time for video-CT registration and video-CT reconstruction.

Pitfalls and Alternatives: We have phrased the testing procedure using direct measurables (visual and CT targets). However, this does not provide a direct measure of registration or motion accuracy. An alternative is to compute a relative measure of motion accuracy. If manually acquired endoscope motion data is inadequate for evaluation purposes, we will also acquire cadaver data using a micron precision robot available in our laboratory (JHU steady-hand robot (SHR) [107–109]) or the endoscope will be tracked using a high-precision optical measurement system of our own design [13]. Another alternative is to perform several statistically independent registration solutions to the same video frame, and to analyze the variability of the solution, similar to the methods described in Aim 1, Task 2.

2.C.4 Aim 4: Assess the accuracy and reliability of intraoperative surface estimation on patient data

Rationale: The endpoint of this aim will be to establish the performance of QE-based reconstruction relative to the current clinical standard: intraoperative CT imaging. CT imaging requires exposure to radiation and has a voxel accuracy of no more than 0.5mm typically. The methods of this aim will determine if video reconstruction techniques produce surface models that agree with CT to 0.5mm across the three live patient data sets (L1, L2, and L3) collected in Aim 3.

Methods: Each of the intraoperative videos from set L1, L2, and L3 will be sampled into a set of short (1-4 second) segments of video data. Each segment will be considered an independent sample which should agree with a corresponding CT surface model. The CT surface, S_{ct} will be computed by applying a threshold at the air/tissue boundary (approx. 500 Hounsfield Units) followed by post-processing using the marching cubes algorithm [119] to create a polygon isosurface. Within S_{ct}, a uniform sampling grid of sample points will be established.

Reconstruction will be performed on an endoscopic video segment producing S_{ve}. Using the registration solutions provided by Aim 3, S_{ve}will be mapped to CT coordinates, at each sample point, p, the distance from p to S_{ve} be measured and the RMS value over all sample points, e, will be computed. For the purposes of comparison, an RMS CT error will be computed by acquiring multiple scans of the cadaver phantoms, and performing the same error calculations between CT surfaces for all $n(n-1)$ independent scan pairs.

Surface reconstruction accuracy will be evaluated relative to three registration solutions: that provided by the navigation system (R1) yielding error e_1, that provided by video-CT registration (R2) yielding error e_2, and a local registration solution (R3) computed directly between S_{ve} and S_{ct} using the 3D-3D registration algorithm of Aim1 yielding error e_3. We would expect that $e_1 > e_2 > e_3$. Our goal is to see $e_3 \leq 0.5$ mm and $e_2 \leq 1.0$ mm (i.e. the composition of 0.5 mm registration error and 0.5 mm RMS reconstruction error). Additional questions include: 1) is performance consistent on unaltered (L1) and altered (L2, L3) anatomy? 2) Is e_2 (resp. e_3) the same between $L2$ and $L3$? 3) How does e_2 vary with distance from a registered frame?

Pitfalls and Alternatives: The primary risk in this aim is that we are unable to compute a clear result due to the dependence of the surface reconstruction on the structure of the images. If we observe high variability, we will make use of the proposed quality control (Aim 1, Task 2), to first select frames that produce reliable registration results prior to testing surface reconstruction. We may also choose to vary the length of the sequences processed, and the distance of the chosen reconstruction from the last good registration solution.

2.C.5 Work Plan

Task	Y1	Y2	Y3	Y4	Y5
Aim 1	+	+	-		
Aim 2	-	+	+	-	
Aim 3	-	-	+	+	-
Aim 4		-	-	+	+

Table 3: Research and development time line for the five years of this project. A "+" indicates intensive work; a "-" indicates incremental development.

The work proposed herein follows a natural line of research and development as shown in Table 3. Our project staff will include the PI, Prof. Russell Taylor (CS), Prof. Jeff Siewerdsen (BME), Dr. Masaru Ishii, (Otolaryngology), Dr. Gary Gallia (Neurosurgery), two graduate students, a postdoctoral scholar, and staff engineers. PI Hager will manage the overall project. In addition, he will contribute specifically to the research of Aims 1 and 2 and collaborate with Siewerdsen to design the testing in Aims 3 and 4. The graduate students and postdoctoral scholar will be responsible for addressing the significant research questions of the proposal. Engineering staff will participate in the development of the software infrastructure for the system (Aims 1 and 2) and the data collection hardware and software for (Aims 3 and 4). Taylor will specifically contribute to the development of the registration and navigation system components (Aim 1). Drs. Ishii and Gallia will be responsible preparing the cadaver specimens and acquiring data. Dr. Ishii will manage the patient data collection and evaluation in conjunction with the postdoc (Aims 3 and 4).

3 References Cited

[1] M. Fried, J. Kleefield, H. Gopal, E. Reardon, B. Ho, and F. Kuhn, “Image-guided endoscopic surgery: Results of accuracy and performance in a multicenter clinical study using an electromagnetic tracking system,” *Laryngoscope*, vol. 107, no. 5, pp. 594–601, 1997.

[2] R. Metson, R. Gliklich, and M. Cosenza, “A comparison of image guidance systems for sinus surgery,” *Laryngoscope*, vol. 108, no. 8, pp. 1164–1170, 1998.

[3] R. Lapeer, M. S. Chen, G. Gonzalez, A. Linney, and G. Alusi, “Image-enhanced surgical navigation for endoscopic sinus surgery: evaluating calibration, registration and tracking,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 4, no. 1, pp. 32–45, 2008.

[4] R. Phillips, “(ii) the accuracy of surgical navigation for orthopaedic surgery,” *Current Orthopaedics*, vol. 21, no. 3, pp. 180 – 192, 2007.

[5] G. Strau, K. Koulechov, S. Röttger, J. Bahner, C. Trantakis, M. Hofer, W. Korb, O. Burgert, J. Meixensberger, D. Manzey, *et al.*, “Evaluation of a navigation system for ent with surgical efficiency criteria,” *The Laryngoscope*, vol. 116, no. 4, pp. 564–572, 2006.

[6] N. Cohen and D. Kennedy, “Revision endoscopic sinus surgery.,” *Otolaryngologic Clinics of North America*, vol. 39, no. 3, p. 417, 2006.

[7] Darius Burschka, Jason J. Corso, Maneesh Dewan, William W. Lau, Ming Li, Henry Lin, Panadda Marayong, Nicholas A. Ramey, Gregory D. Hager, Brian Hoffman, David Larkin, and Christopher J. Hasser, “ Navigating inner space: 3-D assistance for minimally invasive surgery ,” *Robotics and Autonomous System*, vol. 52, no. 1, pp. 5–26, 2005.

[8] Darius Burschka and Ming Li and Russell Taylor and Gregory D. Hager and Masaru Ishii, “Scale-Invariant Registration of Monocular Endoscopic Images to CT-Scans for Sinus Surgery,” *Medical Image Analysis*, vol. 9, pp. 413–439, October 2005.

[9] H. Wang, D. Mirota, G. Hager, and M. Ishii, “Anatomical reconstruction from endoscopic images: Toward quantitative endoscopy,” *American Journal of Rhinology*, vol. 22, pp. 47–51, January/February 2008.

[10] H. Wang, D. Mirota, M. Ishii, and G. Hager, “Robust Motion Estimation and Structure Recovery from Endoscopic Image Sequences With an Adaptive Scale Kernel Consensus Estimator,” in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–7, 2008.

[11] D. Mirota, R. H. Taylor, M. Ishii, and G. D. Hager, “Direct endoscopic video registration for sinus surgery,” in *Medical Imaging 2009: Visualization, Image-guided Procedures and Modeling. Proceedings of the SPIE*, vol. 7261, pp. 72612K–1 – 72612K–8, February 2009.

[12] A. Uneri, S. Schafer, D. Mirota, S. Nithiananthan, Y. Otake, S. Reaungamornrat, J. Yoo, W. Stayman, D. D. Reh, G. L. Gallia, J. Khanna, G. D. Hager, R. H. Taylor, G. Kleinszig, and J. H. Siewerdsen, “Architecture of a high-performance surgical guidance system based on c-arm cone-beam ct: software platform for technical integration and clinical translation,” in *Medical Imaging 2011: Visualization, Image-guided Procedures and Modeling. Proceedings of the SPIE*, vol. 7964, SPIE, February 2011. In press.

[13] D. Mirota, H. Wang, R. H. Taylor, M. Ishii, and G. D. Hager, “Toward video-based navigation for endoscopic endonasal skull base surgery,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009* (G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds.), vol. 5761 of *Lecture Notes in Computer Science*, pp. 91–99, Springer, 2009.

[14] D. Mirota, H. Wang, R. Taylor, M. ishii, G. Gallia, and G. D. Hager, “A system for video-based navigation for endoscopic endonasal skull base surgery,” 2011. Under review for publication in Trans. Medical Imaging.

[15] E. Watanabe, T. Watanabe, S. Manka, and et. al., “Three-dimensional digitizer (neuronavigator): new equipment for computed tomography-guided stereotaxic surgery.,” *Surg Neurol*, vol. 27, pp. 543–547, 1987.

[16] Y. Kosugi, E. Watanabe, J. Goto, T. Watanabe, S. Yoshimoto, K. Takakura, , and J. Ikebe, "An articulated neurosurgical navigation system using mri and ct images," *IEEE Transactions on Biomedical Engineering*, pp. 147–152, February 1988.

[17] H. Reinhardt, H. Meyer, and A. Amrein, "A computer-assisted device for intraoperative ct-correlated localization of brain tumors," *Eur. Surg. Res*, vol. 20, pp. 51–58, 1988.

[18] K. R. Smith, K. J. Frank, and R. D. Bucholz, "The neurostation - a highly accurate minimally invasive solution to frameless stereotactic neurosurgery," *Comput. Med. Imaging Graph.*, vol. 18, pp. 247–256, 1994.

[19] M. P. Heilbrun, S. Koehler, P. McDonald, V. S. W. Peters, , and C. Wiker, "Implementation of a machine vision method for stereotactic localization and guidance," in *Interactive Image-Guided Neurosurgery* (R. Maciunas, ed.), pp. 169–177, AANS, 1993.

[20] R. J. Maciunas, *Interactive Image-Guided Neurosurgery*. American Association of Neurological Surgeons, 1993.

[21] R. L. Galloway, C. A. Edwards, S. S. J. G. Thomas, , and R. J. Maciunas, "A new device for interactive, image guided surgery," in *Proc. SPIE Medical Imaging V*, 1991.

[22] G. H. Barnett, D. W. Kormos, D. P. C. P. Steiner, J. Weisenberger, F. Hajjar, C. Wood, , and J. McNally, "Frameless stereotaxy using a sonic digitizing wand: Development and adaptation to the picker vistar midical imaging system," in *Interactive Image-Guided Neurosurgery* (R. J. Maciunas, ed.), ch. 10, American Association of Neurological Surgeons, 1993.

[23] S. J. Zinreich, S. A. Tebo, D. M. Long, D. E. M. H. Brem, M. E. Loury, C. A. V. Kolk, D. W. K. W. M. Koch, , and R. N. Bryan, "Frameless stereotaxic intergration of ct imaging data: Accuracy and initial applications," *Radiology*, pp. 735–742, 1993.

[24] H. F. Reinhardt, "Neuronagivation: A ten years review," in *Computer-Integrated Surgery* (R. Taylor, S. LAvallee, G. Burdea, and R. Moegses, eds.), pp. 329–342, MIT Press, 1996.

[25] M. Scholtz, W. Konen, S. Tombrock, L. A. B. Fricke, M. v. During, A. Hentsch, L. Heuser, , and A. Harders, "Development of an endoscopic navigating system based on digital image processing," *Journal of Computer Aided Surgery*, vol. 3, no. 3, pp. 134–143, 1998.

[26] D. Dey, D. Gobbi, P. Slomka, K. Surry, and T. Peters, "Mixed reality merging of endoscopic images and 3d surfaces," *IEEE Transactions on Medical Imaging*, vol. 21, no. 1, pp. 23–30, 2002.

[27] C. Cutting, R. Taylor, R. Bookstein, D. Khorramabadi, B. Haddad, A. Kalvin, H. Kim, and M. Nox, "Computer aided planning and execution of cranofacial surgical procedures," in *Proc. IEEE Engineering in Medicine and Biology Conference*, 1992.

[28] C. B. Cutting, F. L. Bookstein, , and R. H. Taylor, "Applications of simulation and morphometrics, robotics in craniofacial surgery," in *Computer-Integrated Surgery* (R. H. Taylor, S. Lavallee, G. Burdea, , and R. Mosges, eds.), pp. 541–544, MIT Press, 1997.

[29] C. B. Cutting, B. Grayson, and H. C. Kim, "Precision multi-segment bone positioning using computer aided methods in craniofacial surgery applicationa," in *12'th IEEE Engineering in Medicine and Biology Conference*, IEEE, 1990.

[30] C. Burghart, R. Krempien, T. Redlich, A. Pernozzoli, H. Grabowsky, J. Muncherberg, S. H. J. Albers, C. Vahl, U. Rembold, and H. Worn, "Robot assisted craniofacial surgery: first clinical evaluation," *Computer Assisted Radiology and Surgery*, pp. 828–833, 1999.

[31] C. VanderKolk, S. Zinreich, B. Carson, N. Bryan, and P. Manson, "An interactive 3d-ct surgical localizer for craniofacial surgery," in *Craniofacial Surgery* (A. Montoya, ed.), 1992.

[32] L. Adams, J. M. Gilsbach, W. Krybus, D. Meyer-Ebrecht, R. Mosges, and G. Schlondorff, "Cas - a navigation support for surgery," in *3d Imaging in Medicine*, pp. 411–423, Springer-Verlag, 1990.

[33] L. Adams, A. Knepper, W. Krybus, D. Meyer-Ebrecht, G. Pfeiffer, R. Rueger, , and M. Witte, "Navigation support for surgery by means of optical position detection and real-time 3d display," in *Proceedings Computer Aided Radiology*, Springer Verlag, 1991.

[34] L. Adams, A. Knepper, W. Krybus, D. Meyer-Ebrecht, G. PFeifer, R. Ruger, , and M. Witte, "Orientation aid for head and neck surgeons," *Innovation et Technologie en Biologie et Medicine*, vol. 14, no. 4, pp. 409–424, 1992.

[35] D. Bartz, O. Gurvit, D. Freudenstein, H. Schiffbauer, and J. Hoffman, "Integration of navigation, optical and virtual endoscopy in neurosurgery and oral and maxillofacial surgery," in *3rd Caesarium – Computer Aided Medicine*, 2001.

[36] S. Lavallee, P. Sautot, J. Troccaz, P. Cinquin, , and P. Merloz, "Computer assisted spine surgery: a technique for accurate transpedicular screw fixation using ct data and a 3-d optical localizer," *Medical Robotics and Computer-Assisted Surgery*, vol. 2, pp. 315–32, 1994.

[37] P. Merloz, J. Tonetti, A. Eid, C. Faure, L. Pittet, M. Coulomb, P. Sautot, , and O. Raoult, "Computer-assisted versus manual spine surgery: clinical report," in *Proc. First Joint Conference of CVRMed and MRCAS* (E. Grimson and R. Mosges, eds.), vol. 1205, pp. 541–544, Springer, 1997.

[38] L. P. Nolte, J. Zamorano, Jiang, F. L. G. Want, E. Arm, , and H. Visurius, "A novel approach to computer assisted spine surgery," *in First Int. Symp. on Medical Robotics and Computer Assisted Surgery (MRCAS 94). Pittsburgh: Shadyside Hospital*, pp. 323–328, 1994.

[39] K. Cleary, "Workshop report: Technical requirements for image-guided spine procedures," 1999.

[40] A. M. DiGioia, D. A. Simon, B. Jaramaz, F. M. M. Blackwell, R. V. O'Toole, B. Colgan, , and E. Kischell, "Hipnav: Pre-operative planning and intra-operative navigational guidance for acetabular implant placement in total hip replacement surgery," *Computer Assisted Orthopedic Surgery*, 1996.

[41] F. Picard, A. Digioia, D. Sell, B. J. J. Moody, C. Nikou, R. LaBarca, , and T. Levison, "Computer-assosted navigation for knee arthroplasty: intra-operative measurements of alignment and soft tissue balancing," *in First Annual Meeting of CAOS International. Davos*, p. 114., 2001.

[42] D. A. Simon, B. Jaramaz, M. Blackwell, A. M. D. F. Morgan, M. D., E. Kischell, B. Colgan, , and T. Kanade, "Development and validation of a navigational guidance system for acetabular implant placement," in *Proc. First Joint Conference of CVRMed and MRCAS* (J. Troccaz, E. Grimson, , and R. Mosges, eds.), vol. 1205, pp. 583–592, Springer, 1997.

[43] M. Kunz, F. Langlotz, J. Strauss, W. Ruther, , and L.-P. Nolte, "Development and verification of an non-ct based total knee arthroplasty system for the lcs prosthesis," *in First Annual Meeting of CAOS International. Davos*, p. 131, 2001.

[44] G. van HellenMondt, M. deKleuver, and P. Pavlov, "Computer assisted pelvic osteotomies; clinical experience in 25 cases," in *First Annual Meeting of CAOS International*, p. 123, 2001.

[45] L. P. Nolte, H. Visarius, and et al, *Computer Assisted Orthopaedic Surgery*. Hofgrefe & Huber, 1996.

[46] A. J. Herline, J. D. Stefansic, S. L. H. J. P. Debelak, C. W. Pinson, R. L. Galloway, and W. C. Chapman, "Image-guided surgery: Preliminary feasibility studies of frameless stereotactic liver surgery," *Arch. Surg.*, vol. 134, pp. 644–650, 1999.

[47] J. Stefansic, A. Herline, Y. Shyr, W. Chapman, J. Fitzpatrick, B. Dawant, and R. J. Galloway, "Registration of physical space to laparoscopic image space for use in minimally invasive hepatic surgery," *IEEE Trans Med Imaging*, vol. 19, no. 10, pp. 1012–1023, 2000.

[48] S. Nicolau, L. Goffin, and L. Soler, "A low cost and accurate guidance system for laparoscopic surgery: validation on an abdominal phantom," in *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 124–133, ACM New York, NY, USA, 2005.

[49] O. Ukimura, I. Gill, M. Nakamoto, *et al.*, "Augmented reality visualization during laparoscopic urologic surgery: The initial clinical experience," *J Urol*, vol. 177, 2007.

[50] L. Su, B. Vagvolgyi, R. Agarwal, C. Reiley, R. Taylor, and G. Hager, "Augmented Reality During Robot-assisted Laparoscopic Partial Nephrectomy: Toward Real-Time 3D-CT to Stereoscopic Video Registration," *Urology*, 2009.

[51] B. Vagvolgyi, C. E. Reiley, G. D. Hager, A. W. Levinson, and L. Su, "Toward direct registration of video to computed tomography for intraoperative surgical planning during laparoscopic partial nephrectomy," in *World Congress of Endourology*, 2007.

[52] L. Soler, S. Nicolau, J. Fasquel, V. Agnus, A. Charnoz, A. Hostettler, J. Moreau, C. Forest, D. Mutter, and J. Marescaux, "Virtual reality and augmented reality applied to laparoscopic and notes procedures," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 1399–1402, 2008.

[53] R. Bucholz and K. Smith, "A comparison of sonic digitizers versus light emitting diode-based localization," *Interactive Image-Guided Neurosurgery*, pp. 179–200, 1993.

[54] R. H. Taylor, H. A. Paul, B. M. C. B. Cutting, W. Hanson, P. Kazanzides, B. Musits, A. K. Y.-Y. Kim, B. Haddad, D. Khoramabadi, , and D. Larose, "Augmentation of human precision in computer-integrated surgery," *Innovation et Technologie en Biologie et Medicine*, vol. 13, no. 4, pp. 450–459, 1992.

[55] S. Lavallee, J. Troccaz, P. Sautot, B. Mazier, P. Cinquin, P. Merloz, and J.-P. Chirossel, "Computer-assisted spinal surgery using anatomy-based registration," in *Computer-Integrated Surgery*, pp. 425–449, MIT Press, 1996.

[56] R. Hofstetter, M. Slomczycowski, M. Sati, , and L. P. Nolte, "Principles of precise fluoroscopy based surgical navigation," *in 4th International Symposium on CAOS*, p. 28, 1999.

[57] M. P. Heilbrun, P. McDonald, C. Wiker, S. Koehler, and W. Peters, "Stereotactic localization and guidance using a machine vision technique," *Stereotact Funct Neurosurg*, vol. 58, pp. 94–98, 1991.

[58] R. Evans, "Vislan computer-aided surgery," *IEE Review*, vol. 41, no. 2, pp. 51–54, 1995.

[59] X. Wu and R. Taylor, "A direction space interpolation technique for calibration of electromagnetic surgical navigation systems," in *Proceedings of the Sixth International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. II, pp. 215–22, Springer Verlag, 2003.

[60] F. Poulin and L. Amiot, "Electromagnetic tracking in the or: Accuracy and sources of intervention," in *Proc. CAOS USA*, pp. 233–235, 2001.

[61] F. Chassat and S. Lavalle, "Experimental protocol of accuracy evaluation of 6-d localizers for computer-integrated surgery: Application to four optical localizers," in *Medical Image Computing and Computer-Assisted Interventation MICCAI98*, pp. 277 – 284, 1998.

[62] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. Seitz, and R. Szeliski, "Building rome in a day," *Communications of the ACM*, vol. 54, no. 10, pp. 105–112, 2011.

[63] R. Newcombe and A. Davison, "Live dense reconstruction with a single moving camera," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 1498–1505, IEEE, 2010.

[64] K. Mori, D. Deguchi, J. Hasegawa, J. T. Y. Suenaga, H. Takabatake, and H. Natori, "A method for tracking the camera motion of real endoscope by epipolar geometry analysis and virtual endoscopy system," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 2208, pp. 1–8, Utrecht: Springer, 2001.

[65] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[66] N. Hata, M. Halle, S. Nakajima, R. K. P. Viola, and F. Jolesz, "Image guided microscopic surgery system using mutual information based registration," *in VBC. Hamburg*, 1996.

[67] W. M. Wells, P. Viola, and R. Kikinis, "Multi-model volume registration by maximization of mutual information," in *MRCAS 97*, pp. 55–62, 1997.

[68] D. Dey, P. Slomka, D. Gobbi, and T. Peters, "Mixed reality merging of endoscopic images and 3d surfaces," in *Medical Image Computing and Computer-Assisted Interventions*, vol. 1935, pp. 796–80, Springer Verlag, 2000.

[69] D. Burschka, M. Li, R. H. Taylor, and G. D. Hager, "Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery," in *MICCAI (2)*, pp. 413–421, 2004.

[70] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[71] B. Maintz and M. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[72] D. Abretske, D. Mirota, G. D. Hager, and M. Ishii, "Intelligent frame selection for anatomic reconstruction from endoscopic video," in *IEEE Workshop on Applications of Computer Vision 2009*, pp. 1–5, 2009.

[73] H. Wang, D. Mirota, and G. D. Hager, "A generalized kernel consensus based robust estimator," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 178–184, 2010.

[74] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[75] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Computer Vision and Patt. Recog.*, pp. I: 519–528, 2006.

[76] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," in *Proc. CVPR*, 2007.

[77] Y. Furukawa and J. Ponce, "Carved Visual Hulls for Image-Based Modeling," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 53–67, 2009.

[78] R. H. Taylor, J. Funda, B. Eldgridge, K. G. S. Gomory, D. LaRose, M. Talamini, L. Kavoussi, and J. anderson, "Telerobotic assistant for laparoscopic surgery.," *IEEE Eng Med Biol*, vol. 14, no. 3, pp. 279–288, 1995.

[79] R. H. Taylor, J. Funda, B. Eldridge, K. Gruben, D. LaRose, S. Gomory, M. T. MD, L. K. MD, and J. Anderson, "A telerobotic assistant for laparoscopic surgery," *in IEEE EMBS Magazine Special Issue on Robotics in Surgery*, pp. 279–291, 1995.

[80] W. W. Lau, N. A. Ramey, J. J. Corso, N. V. Thakor, and G. D. Hager, "Stereo-based endoscopic tracking of cardiac surface deformation," in *MICCAI (2)*, pp. 494–501, 2004.

[81] N. A. Ramey, J. J. Corso, W. W. Lau, D. Burschka, and G. D. Hager, "Real Time 3D Surface Tracking and Its Applications," in *Proceedings of Workshop on Real-time 3D Sensors and Their Use (at CVPR 2004)*, 2004.

[82] F. Mourgues and E. Coste-Maniere, "Flexible calibrations of acutated stereoscopic endoscope for overlay in robot assisted surgery," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002* (T. Dohi and R.Kikinis, eds.), vol. 1, pp. 25–34, Springer Verlag, 2002.

[83] P. Mountney, D. Stoyanov, A. Davison, and G.-Z. Yang, "Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery.," in *MICCAI (1)*, pp. 347–354, 2006.

[84] B. Lo, A. Chung, D. Stoyanov, and G. Mylonas, "Real-time intra-operative 3D tissue deformation recovery," in *ISBI 2008*, pp. 1387–1390, 2008.

[85] D. Stoyanov, A. Darzi, and G. Yang, "Dense 3D depth recovery for soft tissue deformation during robotically assisted laparoscopic surgery," *Lecture Notes in Computer Science*, pp. 41–48, 2004.

[86] D. Stoyanov, A. Darzi, and G. Yang, "A practical approach towards accurate dense 3D depth recovery for robotic laparoscopic surgery," *Computer Aided Surgery*, vol. 10, no. 4, pp. 199–208, 2005.

[87] R. Richa, P. Poignet, and L. Chao, "Three-dimensional motion tracking for beating heart surgery using a thin-plate spline deformable model," *Int. J. Rob. Res.*, vol. 29, pp. 218–230, February 2010.

[88] P. Mountney, D. Stoyanov, and G.-Z. Yang, "Three-dimensional tissue deformation recovery and tracking," *Signal Processing Magazine, IEEE*, vol. 27, pp. 14 –24, july 2010.

[89] M. Hu, G. Penney, M. Figl, P. Edwards, F. Bello, R. Casula, D. Rueckert, and D. Hawkes, "Reconstruction of a 3d surface from video that is robust to missing data and outliers: Application to minimally invasive surgery using stereo and mono endoscopes," *Medical Image Analysis*, 2010.

[90] D. Chetverikov, D. Svirko, D. Stepanov, and P. Krsek, "The trimmed iterative closest point algorithm," *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 3, pp. 545–548 vol.3, 2002.

[91] C. Lu, G. D. Hager, and E. J. Mjolsness, "Fast and globally convergent object pose from video images," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 22, no. 6, pp. 610–622, 2000.

[92] P. Kazanzides, A. Deguet, and A. Kapoor, "An architecture for safe and efficient multi-threaded robot software," in *IEEE Intl. Conf. on Tech. for Practical Robot Appl. (TePRA)*, (Boston, MA), pp. 89–93, Nov 2008.

[93] A. Kapoor, A. Deguet, and P. Kazanzides, "Software components and frameworks for medical robot control," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, (Orlando, FL), pp. 3813–3818, May 2006.

[94] G. Bradski, "THE OPENCV LIBRARY," *DOCTOR DOBBS JOURNAL*, vol. 25, no. 11, pp. 120–126, 2000.

[95] A. J. Davison, I. D. Reid, N. Molton, and O. Stasse, "Monoslam: Real-time single camera slam.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, 2007.

[96] K. Konolige and M. Agrawal, "FrameSLAM: From Bundle Adjustment to Real-Time Visual Mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1066–1077, 2008.

[97] G. Klein and D. Murray, "Improving the agility of keyframe-based SLAM," in *Proc. 10th European Conference on Computer Vision (ECCV'08)*, (Marseille), pp. 802–815, October 2008.

[98] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *ISMAR '07: Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, (Washington, DC, USA), pp. 1–10, IEEE Computer Society, 2007.

[99] B. Williams, G. Klein, and I. Reid, "Real-time slam relocalisation," in *ICCV 2007*, pp. 1 –8, oct. 2007.

[100] E. Delponte, F. Isgrň, F. Odone, and A. Verri, "SVD-matching using SIFT features," *Graphical Models*, vol. 68, no. 5-6, pp. 415–431, 2006.

[101] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," *LECTURE NOTES IN COMPUTER SCIENCE*, vol. 3951, p. 404, 2006.

[102] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment: A modern synthesis," in *VisionAlg99*, pp. 298–372, 1999.

[103] Z. Zhang and Y. Shan, "Incremental motion estimation through bundle adjustment," Tech. Rep. MSR-TR-01-54, Microsoft Research, May 2001.

[104] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in *European Conference on Computer Vision*, vol. 1, pp. 430–443, May 2006.

[105] J. Farrugia, P. Horain, E. Guehenneux, and Y. Alusse, “GPUCV: A FRAMEWORK FOR IMAGE PROCESSING ACCELERATION WITH GRAPHICS PROCESSORS,” in *Proceedings of the IEEE International Conference on Multimedia & Expo*, 2006.

[106] S. Seshamani, P. Rajan, R. Kumar, H. Girgis, T. Dassopoulos, G. Mullin, and G. D. Hager, “A meta registration framework for lesion matching,” in *MICCAI (1)*, pp. 582–589, 2009.

[107] R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stoianovici, P. Gupta, Z. Wang, E. Dejuan, and L. Kavoussi, “A Steady-Hand Robotic System for Microsurgical Augmentation,” *The International Journal of Robotics Research*, vol. 18, no. 12, pp. 1201–1210, 1999.

[108] I. Iordachita, A. Kapoor, B. Mitchell, P. Kazanzides, G. Hager, J. Handa, and R. Taylor, “Steady-hand manipulator for retinal surgery,” in *MICCAI Workshop on Medical Robotics*, (Copenhagen), pp. 66–73, 2006.

[109] I. Fleming, M. Balicki, J. Koo, I. Iordachita, B. Mitchell, J. Handa, G. Hager, and R. Taylor, “Cooperative robot assistant for retinal microsurgery,” *Medical Image Computing and Computer-Assisted Intervention –MICCAI 2008*, pp. 543–550, 2008.

[110] B. Vagvolgyi, C. Reiley, G. Hager, R. Taylor, and L. Su, “Augmented reality using registration of 3 computed tomography to stereoscopic video of laparoscopic renal surgery,” *Journal of Urology*, vol. 179, no. 4, pp. 241–241, 2008.

[111] M. Dewan, P. Marayong, A. M. Okamura, and G. D. Hager, “Vision-based assistance for ophthalmic microsurgery,” in *MICCAI (2)*, pp. 49–57, 2004.

[112] F. L. Bookstein, “Principal warps: Thin-plate splines and the decomposition of deformations,” *IEEE TPAMI.*, vol. 11, pp. 567–585, June 1989.

[113] H. W. Kuhn, “Nonlinear programming: a historical view,” *SIGMAP Bull.*, pp. 6–18, June 1982.

[114] V. Kolmogorov and R. Zabih, “What Energy Functions Can Be Minimized via Graph Cuts?,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, pp. 147–159, 2004.

[115] Y. Boykov, O. Veksler, and R. Zabih, “Fast Approximate Energy Minimization via Graph Cuts,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, pp. 1222–1239, 2001.

[116] I. Ernst and H. Hirschmüller, “Mutual information based semi-global stereo matching on the gpu,” in *ISVC (1)*, pp. 228–239, 2008.

[117] T. Zickler, S. Mallick, D. Kriegman, and P. Belhumeur, “Specularity removal in images and videos: A pde approach,” in *Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2000–2010, 2006.

[118] J. Fitzpatrick and J. West, “The distribution of target registration error in rigid-body point-based registration,” *Medical Imaging, IEEE Transactions on*, vol. 20, pp. 917–927, Sept. 2001.

[119] W. E. Lorensen and H. E. Cline, “Marching cubes: A high resolution 3d surface construction algorithm,” in *SIGGRAPH ’87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, (New York, NY, USA), pp. 163–169, ACM, 1987.

5 Human Subjects

The proposed project involves two classses of human subjects: patients who will undergo surgery and physicians who are performing the surgery. Here, we provide protocols for both patients (Protocol A) and surgeons (Protocol B).

5.A Protection of Human Subjects (Protocol A)

5.A.1 Risks to the Subjects

Proposed Involvement of Human Subjects:

Patients with chronic sinusitis and pituitary tumors receiving a total ethmoidectomy and sphenoidotomy with intraoperative imaging and navigation will be recruited for this study. All patients will have their standard workup and surgery. During the surgery we will record the video and navigation data streams. These data streams will be used for retrospective analysis. The video data stream is already being recorded for archival purposes. The modifications to the standard surgical protocol will be the use of a tracked endoscope, the placement of biocompatible radiolucent endonasal markers, and the formal identification of these markers with an image-guided probe. We anticipate that this will be a minor perturbation of the surgical workflow due to the heavy reliance of surgical navigation during these surgeries with the use of multiple tracked tools, and our previous experience with similar experiments in cadaver models. The preoperative and intraoperative images will be retained for further analysis. We expect that that this study will have minimal risk associated with it, since the workup and treatment of these patients will not be changed in any significant way.

Subject Population:

Patients with medical refractory chronic sinusitis requiring a total spheno-ethmoidectomy (endoscopy sinus surgery) to treat their sinus disease and pituitary patients requiring the same approach to remove their tumor will be recruited for this study. Inclusion criteria include that patients are sufficiently healthy so that the experimental overhead will not be a safety issue and that intraoperative imaging and navigation are appropriate from a clinic standpoint. We anticipate enrolling 15 to 20 patients per year for two to three years. Adults ranging in age from 21 to 99 years of age will be enrolled in this study. Patients who are deemed poor candidates for surgery for health or technical reasons will be excluded from the study. This exclusion will be at the discretion of the operating surgeons.

Exclusion Criteria

Subjects who are deemed unfit for safe surgery, for medical reasons, i.e., emergent surgery in an unstable patient, by the attending surgeons will be excluded for enrollment in this study. Patients whose surgery does not require intraoperative imaging and surgical navigation will be excluded from the study.

Rational for using Special Classes of Subjects

No special classes of subjects will be recruited for this study.

Collaborating Sites

None.

5.A.2 Sources of Materials

Research Material Obtained

Video recordings of the subject's surgery and computed tomographic imaging of the patient's maxillo-facial complex will be obtained and stored in a computer for further analysis. Endoscopic motion data taken during the surgery and will also be recorded and stored, as will the surgical navigation data. In addition standard navigation and surgical proficiency metrics will be recorded.

Recorded Human Subjects Data

Minimal demographic data, such as age, height, weight, and sex of the patient will be obtained. The medical diagnosis of the subject, and the findings of surgery will be recorded. We will also record any complications noted during surgery.

Data Linkage & Data Access

All data will be stored in a de-identified fashion on a secure server. Only members of the study group will have access to the collected data. Only the principle investigator and co-investigators will have access to the recruitment logs that will identify which patients were enrolled in the study.

Data Collection

Demographic data will be obtained by the examining physician after being enrolled into the study. Recordings from the endoscope will be stored in a digital format using video capture device and computer. Computed tomographic and magnetic resonance images will be pushed from the clinical system at Johns Hopkins Medicine to a secure server maintained within the Center for Computer-Integrated Surgical Systems and Technology of Johns Hopkins University. This server automatically de-identifies the image data and will maintain the data until recalled for data analysis.

5.A.3 Potential Risks

Physical and Psychological

We believe that this study will have minimal risk associated with it, since the workup up and treatment plan of the patient will not be changed in any fashion. Furthermore, the workflow during surgery will be perturbed only minimally- that is, the placement and identification of biocompatible markers, the identification of key surgical landmark and defining the surgical envelope will be done formally, so that the motion data capture during this events can be correlated. It is important to note that key portions of the endoscopic surgery are routinely recorded for medical documentation. We will access this video stream for our study. We will capture endoscopic coordinates using a tracked endoscope using the intra-operative navigation device, which is clinical designed for this purpose. We anticipate that minimal psychological risk will be associated with this study, since the patient work and patient flow during the procedures studied will not be altered in a perceptible fashion from a subject's standpoint. No additional intraoperative imaging will be performed for this study, i.e., it will rely on the planned presurgery/registration scan and the near completion confirmation of surgical goals scan. We have considerable experience using tracked endoscopes in clinical studies.

Legal

The subject's name and medical record number will be recorded and stored in the recruitment log maintained by the principle investigator. This list will be stored in a secure fashion. There is a potential that this list could be compromised.

Alternative Treatments

The proposed investigation is a research study not a diagnosis or treatment study; therefore there are no alternative treatments.

5.A.4 Adequacy of Protection Against Risks

Recruitment and Informed Consent

Patients requiring endoscopic sinus surgery performed using intraoperative imaging and computer-assisted navigation presenting to the Johns Hopkins Department of Otolaryngology- Head and Neck Surgery will be recruited for this study. Informed consent will be obtained by the attending physicians during a study consultation, during which the study and its risks and benefits will be discussed in detail. Literature regarding the study will also be made available to the patients and their families. After informed consent is obtained the patient's history will be reviewed with the patient and their referring physician to ensure they meet the eligibility requirements. Patients will be made aware of this study through advertisement within the Department. An IRB approved consent form will also be reviewed with the subject and the subject's family if indicated prior to obtaining consent. Consent will be obtained in a fashion compliant with the Johns Hopkins Institutional Review Board. Consent will be obtained for all subjects participating in the study. A signed consent form will be saved and serve as documentation that consent was obtained.

Protection Against Risk

All data will be stored in a de-identified fashion on a secure server to minimize the risk of breaches of confidentiality. We anticipate that storage of de-identified data coupled with a secure server will be highly effective at protecting the confidentiality of the study participants. A recruitment list and consent forms will be kept by the principle investigator in a double locked fashion to minimize access to these documents. Only the principle investigator and co-investigators will have access to this data. A number of precautions will be taken to ensure a subject's safety during enrollment in this study. All surgeries will be performed by a board certified Otolaryngologist- Head and Neck Surgeons skilled in endoscopic and skull base surgery. In the advent of an unexpected medical event the situation will be evaluated by the Attending surgeon and appropriate medical and/or surgical care instituted. Adverse events will be reported to the IRB as dictated by the IRB bylaws.

5.A.5 Potential Benefits of the Proposed Research to the Subjects and Others

This study will yield minimal benefit to the subject participating in the study, but may lead to a quantitative endoscopic method useful for endoscopic sinus surgery and endoscopic skull base surgery. If this technique is successful it may lead to a computer assisted surgery system with sufficient accuracy to permit a change in the approach methodology based on identification of surgical landmarks to a direct approach to the target of interest. This has the potential of reducing the number of structures exposed during surgery, which would reduce surgical morbidity, and more importantly significantly reducing the approach time. Since no changes will be made to the standard work-up and treatment of these patients and surgical workflow will be altered only minimally, we expect this study to have extremely low risk associated with it. Since this study has great potential to benefit future patients we anticipate the risk to benefit ratio to be acceptable.

5.A.6 Importance of the Knowledge to be Gained

Chronic sinusitis affects 146 patients per 1,000 Americans and results in 18 to 22 million office visits per year with direct treatment costs of $3.4 to 5 billion annually. Approximately 200,000 endoscopic sinus surgeries are performed per year with an increasing number of surgeries performed and relying on surgical navigation. This is due to increasing complexity of endonasal procedures and the proximity of the paranasal sinuses to critical structures such as the brain, eye, and carotid artery. Small errors in surgery can have catastrophic consequences. Because of the scale of this surgery and the consequences of mistakes considerable effort has been placed on improving current computer assisted surgical techniques, with the latest improvement being intraoperative imaging as an adjuvant to standard intraoperative navigation. Intraoperative navigation is expensive from a resource standpoint and therefore has limit chance of diffusing throughout the specialty especially in the climate of limited health care dollars. The goal of this research is to supplement standard navigation using quantitative endoscopy techniques to improve the navigation accuracy and add the ability of update the surgical model so that intraoperative imaging will not be required. The results of this research has the potential to greatly impact the delivery of health care by delivering the similar capabilities of intraoperative imaging at a fraction of the cost making these technologies available to a wide audience.

5.B Protection of Human Subjects (Protocol B)

5.B.1 Risks to the Subjects

Proposed Involvement of Human Subjects:

Otolaryngology- Head and Neck Surgery and Neurosurgery residents and attending physicians participating in endoscopic skull bases cadaver dissections will be recruited for this study. These physician's performance will be graded using classical computer assisted surgical navigation methods and those developed in this grant. Comparison of metrics of key tasks will allow us to assess subcomponents of our system.

Subject Population:

Attending physicians in the Department of Otolaryngology- Head and Neck Surgery and Neurosurgery at Johns Hopkins Hospital will be enrolled in this study. We anticipate being able to enrolling 4 faculty members in this study. Age range will be from the 20's to 40's.

Exclusion Criteria

None.

Rational for using Special Classes of Subjects

Attending physicians specializing in endoscopic surgical methods are used for this surgery to ensure that the subject population has a basic understanding of the surgical approaches and techniques.

Collaborating Sites

None.

5.B.2 Sources of Materials

Research Material Obtained

Video recordings of the subject's cadaver dissection as well performance metrics such as time to complete tasks and subjective skill set analysis by trained observers will be used to asses the subject's surgical performance.

Recorded Human Subjects Data

Minimal demographic data such as years of training, exposure to endoscopic skull base surgery, etc., will be recorded.

Data Linkage & Data Access

All data will be stored in a de-identified fashion on a secure server. Only members of the study group will have access to the collected data. Only the principle investigator and co-investigators will have access to the recruitment logs that will identify which subjects were enrolled in the study.

Data Collection

The investigators will obtain demographic data at the time informed consent is obtained. Data collection will occur during regularly scheduled cadaver dissections used for educational purposes. Metrics of surgical competence, i.e., time to perform a task and subjective measures of competence will be obtained during the cadaver dissection. This will be performed in a blocked fashion to permit comparison of standard navigation methods with those developed in this grant.

5.B.3 Potential Risks

Physical and Psychological

We believe that this study will have minimal risk associated with it. This data will be collected in a passive fashion during routine cadaver dissections currently being performed to teach and assess endoscopic surgery.

Legal

The subject's name will be recorded and stored in the recruitment log maintained by the principle investigator. This list will be stored in a secure fashion. There is a potential that this list could be compromised.

Alternative Treatments

This is not applicable.

5.B.4. Adequacy of Protection Against Risks

Recruitment and Informed Consent

Attending physicians at Johns Hopkins Medical Institute participating in endoscopic cadaver dissections will be recruited for this study. Informed consent will be obtained by on of the investigators of this grant. Consent will be obtained in a fashion compliant with the Johns Hopkins Institutional Review Board. Consent will be obtained for all subjects participating in the study. A signed consent form will be saved and serve as documentation that consent was obtained. An IRB approved consent form will be reviewed with the subjects prior to obtaining consent.

Protection Against Risk

All data will be stored in a de-identified fashion on a secure server to minimize the risk of breaches of confidentiality. We anticipate that storage of de-identified data coupled with a secure server will be highly effective at protecting the confidentiality of the study participants. A recruitment list and consent forms will be kept by the principle investigator in a double locked fashion to minimize access to these documents. Only the principle investigator and co-investigators will have access to this data. A number of precautions will be taken to ensure a subject's safety during enrollment in this study. All dissections will be performed using universal precautions under the supervision of surgeons specializing in endoscopic surgery. All tasks being evaluation will be taught to participating subjects prior to evaluation. Adverse events will be reported to the IRB as dictated by the IRB bylaws.

Potential Benefits of the Proposed Research to the Subjects and Others

This study will yield minimal benefit to the subject participating in the study, but may lead to a quantitative endoscopic method useful for endoscopic skull base surgery. If this technique is successful it may lead to a computer assisted surgery system with sufficient accuracy to permit a change in the approach methodology based on identification of surgical landmarks to a direct approach to the target of interest. This has the potential of reducing the number of structures exposed during surgery, which would reduce surgical morbidity, and more importantly significantly reducing the approach time. Since no changes will be made to the standard work-up and treatment of these patients and surgical workflow will be altered only minimally, we expect this study to have extremely low risk associated with it. Since this study has great potential to benefit future patients we anticipate the risk to benefit ratio to be acceptable.

5.B.5 Importance of the Knowledge to be Gained

See 5.A.5 above.

6 Inclusion of Women and Minorities

6.A Inclusion of Women and Minorities (Protocol A)

Targeted/planned subject group with respect to gender and ethnic groups

Subjects will be recruited from airway patients presenting to the Johns Hopkins Department of Otolaryngology-Head and Neck surgery. Since there is no racial or sexual predilection for this disease type we expect that our subject population will be reflective of community that we serve. Current census estimates that 64% of our patients will be White, 23.9% Black or African American, 0.3% American Indian or Alaska Native, 4.8% Asian, 0.1% Native Hawaiian or other Pacific Islander, and 5.7% Hispanic. We expect 51.6% of our patients to be female, and 48.4% to be male. These statistics were used to generate our Targeted/Planned enrollment Table in the following section.

Selection Criteria with respect to gender and ethnic group

There is no gender or ethnic preference for skull base pathology; therefore there will be no selection criteria for enrollment into this study based on gender or ethnicity. We expect racial and gender distribution of our subjects to reflect the distribution of patients presenting to the Johns Hopkins Hospital, which in general reflects the regional population distributions as a whole.

Criteria for excluding subjects based on gender or ethnic group

No subjects will be excluded from this study based on gender or ethnicity. Outreach programs for recruiting subjects based on gender or ethnic group will not be performed. The targeted and planned study cohort will be recruited by advertised to the patients presenting to the Departments of Otolaryngology- Head and Neck Surgery at Johns Hopkins Hospital. This population closely resembles the ethnic and gender diversity of the Baltimore-Washington area.

6.B Inclusion of Women and Minorities (Protocol B)

Targeted/planned subject group with respect to gender and ethnic groups

Subjects will be recruited from the attending pool in the Departments of Otolaryngology and Head and Neck Surgery and Neurosurgery at Johns Hopkins University. The gender and ethnic profile will depend on the attending pool at that time. However, we expect the population to be skewed towards males and White subjects. Due to the nature of the study and assuming that surgical skill is independent or race or gender we expect our skewed study population not to be a major factor in our data analysis.

Selection Criteria with respect to gender and ethnic group

We assume that there is no gender or ethnic effect with respect to surgical skill so there will be no selection criteria with respect to gender o ethnic group. We are however limited by our study population, since we require subjects trained in endoscopic skull base methodologies. We therefore expect a skewed study population.

Criteria for excluding subjects based on gender or ethnic group

No subjects will be excluded from this study based on gender or ethnicity.

7 Targeted/Planned Enrollment

7.A Targeted/Planned Enrollment (Protocol A)

Study Title: Effectiveness of Video-CT Registration for Sinus Surgery

Total Planned Enrollment: 40

TARGETED/PLANNED ENROLLMENT: Number of Subjects			
Ethnic Category	**Sex/Gender**		
	Females	**Males**	**Total**
Hispanic or Latino	3	2	5
Not Hispanic or Latino	16	19	35
Ethnic Category: Total of All Subjects *	19	21	40
Racial Categories			
American Indian/Alaska Native	0	0	0
Asian	1	0	1
Native Hawaiian or Other Pacific Islander	0	0	0
Black or African American	4	5	9
White	15	15	30
Racial Categories: Total of All Subjects *	20	20	40

* The “Ethnic Category: Total of All Subjects” must be equal to the “Racial Categories: Total of All Subjects.”

7.B Targeted/Planned Enrollment (Protocol B)

Study Title: Effectiveness of Video-CT Registration for Sinus Surgery

Total Planned Enrollment: 4

TARGETED/PLANNED ENROLLMENT: Number of Subjects			
Ethnic Category	**Sex/Gender**		
	Females	**Males**	**Total**
Hispanic or Latino	0	0	0
Not Hispanic or Latino	0	4	4
Ethnic Category: Total of All Subjects *	0	4	4
Racial Categories			
American Indian/Alaska Native	0	0	0
Asian	0	1	1
Native Hawaiian or Other Pacific Islander	0	0	0
Black or African American	0	0	0
White	0	3	3
Racial Categories: Total of All Subjects *	0	4	4

* The “Ethnic Category: Total of All Subjects” must be equal to the “Racial Categories: Total of All Subjects.”

8 Inclusion of Children

Children will not be included in this study for either protocol A or B. With respect to protocol A, the size of their nasal corridor usually is too small to permit expanded endonasal approach and resection at this time. With respect to protocol B, there are no children who are sinus surgeons.