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

Daniel J. Mirota, a Ali Uneri, a Sebastian Schafer, b Sajendra Nithianathan, b Douglas D. Reh, c Gary L. Gallia, d Russell H. Taylor, a Gregory D. Hager, a and Jeffrey H. Siewertsen a,b

a Department of Computer Science, Johns Hopkins University, Baltimore MD
b Department of Biomedical Engineering, Johns Hopkins University, Baltimore MD
c Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins Hospital, Baltimore MD
d Department of Neurosurgery and Oncology, Johns Hopkins Hospital, Baltimore MD

ABSTRACT

Registration of endoscopic video to preoperative CT facilitates high-precision surgery of the head, neck, and skull-base. Conventional video-CT registration is limited by the accuracy of the tracker and does not use the underlying video or CT image data. A new image-based video registration method has been developed to overcome the limitations of conventional tracker-based registration. This method adds to a navigation system based on intraoperative C-arm cone-beam CT (CBCT) that reflects anatomical change, in turn providing high-accuracy registration of video to the surgical scene. The resulting registration enables visualization of the CBCT and planning data within the endoscopic video. The system incorporates a mobile C-arm for high-performance CBCT, integrated with an optical tracking system, video endoscopy, deformable registration of preoperative CT with intraoperative CBCT, and 3D visualization. As in the tracker-based approach, in the image-based video-CBCT registration the endoscope is localized using an optical tracking system that provides a quick initialization followed by a direct 3D image-based registration of the video to the CBCT. In this way, the system achieves video-CBCT registration that is both fast and accurate. Application in skull-base surgery demonstrates overlay of critical structures (e.g., carotid arteries and optic nerves) and surgical target volumes with sub-mm accuracy. Phantom and cadaver experiments show consistent improvement in target registration error (TRE) in video overlay over conventional tracker-based registration – e.g., 0.92 mm versus 1.82 mm for image-based and tracker-based registration, respectively. The proposed method represents a two-fold advance–first, through registration of video to up-to-date intraoperative CBCT (overcoming limitations associated with navigation with respect to preoperative CT), and second, through direct 3D image-based video-CBCT registration, which together provide more confident visualization of target and normal tissues within up-to-date images and improved targeting precision.

Keywords: image registration, 3D visualization, image-guided surgery, cone-beam CT, intra-operative imaging, skull-base surgery, surgical navigation, video endoscopy, video fusion

1. INTRODUCTION

Complex head and neck procedures – tumor surgery in the skull-base in particular – require high accuracy navigation. Transnasal surgery to the skull-base uses the endonasal sinus passage to gain access to lesions of the pituitary and the surrounding ventral region.1 Accuracy is critical when operating in this region due to the proximity of critical structures, including the optic nerves and carotid arteries, violation of which could result in blindness or death. To assist in preservation of critical anatomy while facilitating complete target resection, surgical navigation systems involving tracking and visualization of a pointer tool have shown to support complex cases.2 Recent work3 extends such tracking capability to the endoscope, allowing registration of the endoscopic video scene to preoperative CT. However, such tracker-based approaches are limited by inaccuracies in the tracking system, with typical registration errors of 1–3 mm.4

In addition, conventional navigation based on preoperative CT is subject to gross inaccuracies arising from anatomical change (i.e., deformation and tissue excision) occurring during the procedure. As noted by Schlosser et al.,5


Proc. of SPIE Vol. 7964 79640J-1
intraoperative change limits applicability in the treatment of more complex disease. The work reported below attempts to answer the conventional limitations of accounting for anatomical change and tracker accuracy: first, by resolving the limitations of conventional tracker-based video registration through direct 3D image-based registration; and second, through registration of video to high-quality intraoperative C-arm cone-beam CT (CBCT) that properly reflects anatomical change. By incorporating CBCT and direct 3D image-based registration with video overlay, it is possible not only to improve surgical performance but also to lower the learning curve for skull-base surgery and provide a tool for surgical training via enhanced visualization.

Conventional tracker-based registration is performed by attaching a rigid body to an endoscope and using an optical or electromagnetic tracking system to measure the pose, rotation, and translation of the endoscope. With the pose of the endoscope known, the intrinsic and extrinsic camera parameters are applied along with the CT-to-tracker (or CBCT-to-tracker) registration. The camera is then registered to the CT (or CBCT) image data, which can be fused in alpha-blended overlay on the video scene. Such approaches have demonstrated the potential value of augmenting endoscopic video with CT, CBCT, and/or planning data, but are fundamentally limited by the accuracy of registration provided by the tracking system. The approach described below incorporates image-based video registration to directly register the video image (viz., a 3D video scene computed by structure-from-motion) to the CT or CBCT image data.

The image-based method is further integrated with a navigation system that combines intraoperative CBCT with real-time tool tracking, video endoscopy, preoperative data, video reconstruction, registration software, and 3D visualization software. The intraoperative CBCT is provided by a prototype mobile C-arm developed in collaboration with Siemens Healthcare (Erlangen, Germany) to provide high-quality intraoperative images with sub-millimeter spatial resolution and soft-tissue visibility. A newly developed open-source software platform enables image-based video registration to be integrated seamlessly with surgical navigation.

2. METHODS

2.1 System Setup

Surgical tools and an ENT video endoscope with infrared retro-reflective markers (Karl Storz, Tuttingen, Germany), [E Figure 1], were tracked using an optical tracking system (Polaris Vicra or Spectrum, Northern Digital Inc., Waterloo ON), [N Figure 1]. The endoscopic video was captured using a video capture card (Hauppauge, Hauppauge NY) and combined with other navigation subsystems using a software platform linking open-source libraries for surgical navigation (CISST, Johns Hopkins University) and 3D visualization (3D Slicer, Brigham and Women’s Hospital, Boston MA), [CBCT Figure 1]. Components of the system are illustrated in Figure 1.

![Figure 1: The bench-top system (left) and the relationship of the frame transformations (right). The labels refer to: N (the tracker), CBCT (the CBCT data), R (the reference marker frame), X (the pointer frame), E (the endoscope frame), O (the optical center of the endoscope), and P (the patient frame).](image)

2.2 Camera Calibration

Camera calibration is required for both the conventional tracker-based registration and the proposed image-based registration. A passive arm was used to hold the endoscope (to minimize tracker latency effects) in view of a checkerboard calibration pattern (2 × 2 mm² squares). The camera was calibrated using the German Aerospace Center (DLR) camera calibration toolbox. The DLR camera calibration toolbox computed the intrinsic camera parameters (focal length, principle point, and radial distortion), extrinsic camera parameters (the pose of camera) and the hand-eye calibration (the transformation from the optically tracked rigid body and the camera center).
2.3 Tracker-Based Registration

Following camera calibration, the tracker-based registration used the tracker to localize the endoscope, and the hand-eye calibration was applied to resolve the camera in tracker coordinates. The registration of patient to tracker was then applied to render the camera in CT (or CBCT) coordinates as shown in Eq. 1:

\[
CTT_{camera} = (CTT_{patient})(patientT_{reference})(trackerT_{reference})^{-1} (trackerT_{endoscope})(endoscopeT_{camera})
\]

where each transformation is illustrated with respect to the experimental setup in Figure 1. The limitations in accuracy associated with a tracker-based registration become evident in considering the number of transformations required in Eq. 1. Each non-tracker transformation is assumed to be fixed and rigid, but such is not always the case even with rigid endoscopes, and errors in each transformation accumulate in an overall registration error that is typically limited to \(\sim1-3\) mm under best-case (laboratory) conditions. In particular, \(CTT_{patient}\) is subject to errors arising from deformation during surgery, while \(patientT_{reference}\) and \((trackerT_{reference})^{-1}\) are subject to perturbations of the reference marker and other “aging” of the preoperative tracker registration over the course of the procedure.

2.4 Image-Based Registration

Instead of relying on the aforementioned tracker-based transformations, image-based registration builds a 3D reconstruction of the tracked 2D image features and registers the reconstruction directly to a 3D isosurface of the air-tissue boundary in CT (or CBCT). \(^{17}\) 2D image features selected by Scale Invariant Feature Transfrom (SIFT) \(^{18}\) are tracked in a pair of images. From the motion field created by the tracked 2D features over the pair of images the 3D location of each tracked feature pair is triangulated. This significantly simplifies the transformations of Eq. 1 to a single image-based transform:

\[
CTT_{camera} = (CTT_{patient})(cameraT_{patient})^{-1}
\]

or

\[
CBCTT_{camera} = (CBCTT_{patient})(cameraT_{patient})^{-1}
\]

In particular, \(CBCTT_{patient}\) is robust to inaccuracies associated with preoperative images alone (i.e., \(CTT_{patient}\)) since the CBCT image accurately accounts for intraoperative change. Similarly, \((cameraT_{patient})^{-1}\) uses up-to-date video imagery of the patient. While this approach has the potential to improve registration accuracy essentially to the spatial resolution of the video and CT (or CBCT) images (each better than 1 mm), the image-based approach is more computationally intense and requires a robust initialization (within \(\sim10\) mm of the camera location) to avoid local minima.

2.5 Hybrid Registration

We have developed a hybrid registration process that takes advantage of both the tracker-based registration (for fast, robust pose initialization) and image-based registration (for precise refinement based on 3D image features). As illustrated in Figure 2, the tracker-based registration streams continuously and updates the registration in real-time (but

Figure 2: System data flow diagram
with coarse accuracy), while the image-based registration runs in the background (at a slower rate due to computational complexity) and updates the registration to a finer level of precision as available (e.g., when the endoscope is moving slowly or is at rest with manually motion).

2.6 Phantom and Cadaver Studies

Phantom and cadaver studies were conducted to measure the accuracy of image-based video-CBCT registration and to gather expert evaluation of its integration and utility in CBCT-guided skull base surgery. The phantom study included the use of two anthropomorphic phantoms. The first was a simple, rigid head phantom for endoscopic sinus simulation shown in Figure 1 and detailed in Reference 19. The phantom was printed with a 3D rapid prototype printer based on segmentations of bone and sinus structures from a cadaver CT. It is composed of homogeneous plastic, presenting a simple, architecturally realistic simulation of anatomy of the sinuses, skull base, nasopharynx, etc. Because the phantom is simply plastic (in air), CBCT images exhibit very low noise and ideal segmentation of surfaces.

The second phantom (referred to as the anthropomorphic phantom, or simply, phantom #2) provided more realistic anatomy, imaging conditions, and CBCT image quality, because it more closely approximated a human head. Phantom #2 incorporated a natural skull within tissue-equivalent plastic (Randoc™ material) while maintaining open space within the nasal sinuses, nasopharynx, oral cavity, oropharynx, etc. The phantom therefore provided stronger attenuation and realistic artifacts in CBCT of the head. As a result, the isosurface segmented in CBCT was not as smooth or free of artifact as in the simple plastic phantom. The second phantom therefore demonstrated the effects of image quality degradation on image-based video-CBCT registration, and improvements in image quality (for example, improved artifact correction) is expected to yield better registration.

Each of the specimens was scanned preoperatively in CT, and structures of interest (fiducials, critical anatomy, and surgical targets) were segmented. Each was then registered to the navigation system using a fiducial-based registration with fiducial markers spread over the exterior of the skull. All experiments used a passive arm to hold the endoscope and avoid the latency effects of video and tracker synchronization. Video and CBCT images were registered according to the conventional tracker-based approach and the image-based approach. In each image the target registration error (TRE) was measured as the RMS difference in target points – viz., rigid “divots” in the phantoms and anatomical points identified by an expert surgeon for the cadavers.

Figure 3 shows the cadaver study experimental setup. The devices and reference frames are as previously labeled in Figure 1. The cadaver specimen was held in a Mayfield Clamp on a carbon fiber table. The C-arm provided intraoperative CBCT on request from the surgeon at various milestones in the course of approach to the skull base, identification of anatomical landmarks therein, and resection of a surgical target (typically the pituitary or clival volume).
To evaluate the registration, target points were manually segmented in both the video and CBCT images. Tracker measurements were recorded 10 times each with the endoscope held stationary by a passive arm to avoid tracker jitter and single-sample measurement bias. The ten measurements were then averaged and used to compute the TRE for the tracker-based method. The same video frame from the stationary endoscope was used as the first image of the pair used for the image-based method. A total of 90 target points were identified in the phantoms the targets were from within the right and left maxillary, ethmoid and sphenoid sinuses and the mouth, and a total of 4 target points were identified in the cadavers (viz., the left and right optic carotid recess in two specimens).

After segmenting the targets in both the video and CBCT, the camera pose estimates from either the tracker-based or image-based method were used as the origin of a ray, \(r\), through the point segmented in the video. The distance between this ray and the 3D point in the CBCT was computed as the TRE. Thus, for a given camera pose \((R, t)\) in the CBCT and camera intrinsic parameters \(k\), the ray, \(r\), was formed from the segmented point in the video, \(p_{\text{image}}\) is:

\[
r = Rk^{-1}p_{\text{image}} - t
\]

(3)

It then follows that the distance, \(d\), between the ray and the point in the CBCT \((p_{\text{CBCT}})\) is:

\[
d = \left| p_{\text{CBCT}} - \left( t + \left( \frac{r \cdot (p_{\text{CBCT}} - t)}{r \cdot r} \right) r \right) \right|
\]

(4)

Geometrical, \(d\), can be thought of as the distance between \(p_{\text{CBCT}}\) and the projection of \(p_{\text{CBCT}}\) on the ray, \(r\). We used, \(d\), as our TRE metric to be able to compare, tracker-based and image-based directly.

3. RESULTS

3.1 Phantom Studies

From a high quality CBCT, an isosurface was segmented by applying a threshold at the value of the air/tissue interface. The resulting polygon surface was smoothed and de-noised using the Visualization Toolkit (VTK) available within Slicer, shown in Figure 4 on the left. While there is some noise in the texture of the hard palate and base of the mouth, the majority of the surface is smooth and artifact-free, which allows better registration. The simple, rigid phantom experiment therefore begins to probe the fundamental performance limits of image-based in comparison to tracker-based

![Tracker-based](a)

![Image-based](d)

Figure 4: Comparison of tracker-based (top row) and image-based (bottom row) video-CBCT registration in the rigid phantom of Figure 1. The three views (left to right) are: the CBCT surface rendering, the endoscopic video scene, and the TRE colormap. Green spheres indicate the (true) endoscope image target locations. Blue spheres indicate the CBCT target locations.
registration. In Figure 4 are the manually segmented targets in the CBCT are indicated by blue spheres and the targets in the video are indicated by green spheres. The result in Figure 4 shows (from left to right) the surface rendering, endoscope image, and TRE colormap. Figure 4(a-c) present the tracker-based results, while Figure 4(d-f) show the image-based results. In the tracker-based results (4a-c) the misalignment of all of the target spheres is visible, whereas image-based registration shows lower error for all of the targets. The colormap shows more clearly the distribution of error over each of the targets. Note the consistent offset from the inaccuracy of the extrinsic camera parameters of the tracker-based approach versus the radial distribution of error in the image-based approach.

Figure 5 summarizes the overall distribution of all TRE measurements within the rigid plastic phantom. The mean TRE for the tracker-based registration method was 1.82 mm (with 1.58 mm first quartile and 2.24 mm range). By comparison, the mean TRE for image-based registration was 0.92 mm (with 0.61 mm first quartile and 1.76 mm range). The improvement in TRE over the tracker-based method was statistically significant (p < 0.001) with the majority of all targets visualized showing sub-millimeter precision.

Figure 6 shows an example registration in the anthropomorphic phantom. As in Figure 4, the top row corresponds to tracker-based registration, and the bottom row the image-based results, where the images illustrate target points within the oropharynx (composed of a rapid prototyped CAD model incorporated within the oral cavity and nasopharynx of the natural skull). Registration in this phantom presented a variety of realistic challenges – for example, a lack of texture on smooth surfaces and a prevalence of realistic CBCT reconstruction artifacts. Nonetheless, the image-based registration algorithm yielded an improvement over tracker-based registration as illustrated in Figure 6. The CBCT targets are shown by blue spheres, and the video targets as light red

![Figure 5: Summary of TRE measurements in the simple, rigid, plastic phantom (N=43, p < 0.001).](image)

![Figure 6: Comparison of tracker-based (top row) and image-based (bottom row) video-CBCT registration in the more realistic, anthropomorphic phantom. The three views (left to right) are: the CBCT surface rendering, the endoscopic video scene, and the TRE colormap. Light red spheres indicate the (true) endoscope image target locations. Blue spheres indicate the CBCT target locations.](image)
spheres. The correspondence of the spheres is shown with a light blue line connecting each pair. A distinct pattern of consistent offset is again observed for the tracker-based approach, compared to a radial distribution of offset in the image-based approach.

Shown in Figure 7 is the overall distribution of TRE measured in the anthropomorphic phantom. The TRE was measured within the nasal passage, the oropharynx and mouth. The CBCT artifacts and lack of surface texture is evident in an overall increase in TRE and the large variance of the distribution. The mean TRE for the tracker-based registration method was 2.54 mm (with 1.97 mm first quartile and 5.04 mm range). By comparison, the mean TRE for image-based registration was 1.62 mm (with 0.83 mm first quartile and 3.30 mm range). The results again demonstrate a significant improvement (p < 0.001) for the image-based registration approach.

### 3.2 Cadaver Study

The same experimental setup as for the phantom studies was repeated in a cadaver to evaluate the effectiveness of the algorithm in a more realistic, preclinical environment. The left and right optic carotid recess (OCR) were manually segmented in the CBCT image and the video for TRE analysis. Figure 8 shows the results of the tracker-based registration (Fig. 8a-b) and image-based registration (Fig. 8c-d). Again, blue spheres denote points segmented in CBCT and light red spheres, from the video. Note the closer alignment of points after video registration. Table 1 shows the mean and range of the tracker and image-based results, once again showing an improvement in the registration. Although the small number of points do not support a statistical comparison, the mean TRE (of the four OCR points) was 4.86 mm for the tracker-based approach, compared to 1.87 mm for the image-based approach.

![Figure 8: Comparison of tracker-based and image-based video-CBCT registration in cadaver. The top row shows the tracker-based results from our cadaver study with two different views (CBCT [left], live endoscope [right]). The light red spheres indicate the endoscope image target locations. The blue spheres indicate the CBCT target locations. The bottom row shows the image-based results.](image-url)
Table 1: Summary of TRE measurements in the cadaver study.

<table>
<thead>
<tr>
<th>TRE (mm) of the OCR (N=4)</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracker-based</td>
<td>4.86</td>
<td>4.11</td>
</tr>
<tr>
<td>Image-based</td>
<td>1.87</td>
<td>1.27</td>
</tr>
</tbody>
</table>

registration. A more complete statistical analysis in cadaver is the subject of ongoing work.

After registering the planning data from a preoperative CT to the intraoperative CBCT (as shown in Figure 9), endoscopic video and CBCT were registered via the tracker-based and/or image-based approaches to yield overlay of the planning data on video. Figure 9 shows the posterior aspect of the sphenoid sinus (the surgical target) with the optic nerves overlaid in blue and the carotid arteries in red. The orthogonal views show the standard views visualizing the location of the camera’s optical center within the CBCT. Registered to the CBCT is the preoperative CT with planning data overlay in color on the slices. The planning data enables surgical tasks to be defined ahead of time, such as resection volumes and dissection cut lines.

4. DISCUSSION

Schulze et al. noted that an “exact match of the virtual and the real endoscopic image is virtually impossible” since the source of the virtual image is a preoperative CT. However, the access to intra-operative CBCT makes possible an image-based registration in which the endoscopic video is matched to an up-to-date tomographic image that properly reflects anatomical changes (including deformations, tissue excisions, etc.) occurring during the procedure. Consider the video snapshot of the clivus, for example, as shown in Figure 8; it is possible that the location of the camera might be within the anterior aspect of the sphenoid, making the rendering of the current endoscopic location very difficult in the context of preoperative CT alone (i.e., with the anterior sphenoid wall still intact) However, with a current CBCT the source of the tomographic image is up to date, reflecting surgical resections, thus making it possible to render an isosurface that matches the endoscopic video.

A significant improvement (p < 0.001) in TRE was demonstrated in these studies, but this method is sensitive to local minima and there are cases in which there were insufficient matched features to produce a meaningful reconstruction.

Figure 9: The surgeon's view of the navigation system. Shown in the endoscope view (top right) are the carotid arteries (red) and the optic nerves (blue). The three orthogonal slices indicate the location of the endoscope in the CBCT with planning data overlaid in color.
Local minima can be produced from the isosurface containing artifacts, and poor feature matches can occur when there is insufficient texture in the video images. In addition, the current use of only two images for 3D video image reconstruction is limiting and not as robust as multi-view reconstruction. Areas of future work include investigation of using the entire image for registration instead of a sparse feature set and more tightly coupling the tracker data into the image-based registration. This would add more constraints to the registration and could simplify the formulation of the problem to a single image, thus making video frame-by-frame registration possible.

5. CONCLUSION

Conventional tracker-based video-CT registration provides an important addition to the image-guided endoscopic surgery arsenal, but it is limited by fundamental inaccuracies in tracker registration accuracy and anatomical change imparted during surgery. As a result, conventional video-CT registration is limited to accuracy ~1-3 mm in the best-case scenarios, and under clinical conditions with large anatomical change it can be rendered useless. The work reported here begins to resolve such limitations: first, by overcoming tracker-based inaccuracies through a high-precision 3D image-based registration of video (computed structure-from-motion) with CT or CBCT (air-tissue iso-surfaces); and secondly, by integration with intraoperative CBCT acquired using a high-performance prototype C-arm. Quantitative studies in phantoms show a significant improvement (p < 0.001) in TRE for image-based registration compared to tracker-based registration, with a hybrid arrangement proposed that initializes the image-based alignment with a tracker-based pose estimate to yield a system that is both robust and precise. Translation of the approach to clinical use within an integrated surgical navigation system is underway, guided by expert feedback from surgeons in cadaveric studies. These preclinical studies suggest that removing such conventional factors of geometric imprecision through intraoperative image-based registration could extend the applicability of high-precision guidance systems across a broader spectrum of complex head and neck surgeries.

ACKNOWLEDGMENTS

This work was supported by the Link Foundation Fellowship in Simulation and Training, Medtronic and the National Institutes of Health grant number R01-CA127144. Academic-industry partnership in the development and application of the mobile C-arm prototype for CBCT is acknowledged, in particular collaborators at Siemens Healthcare (Erlangen Germany) – Dr. Rainer Graumann, Dr. Gerhard Kleinszig, and Dr. Christian Schmidgundst. Cadaver studies were performed at the Johns Hopkins Medical Institute, Minimally Invasive Surgical Training Center, with support and collaboration from Dr. Michael Marohn and Ms. Sue Eller gratefully acknowledged.

REFERENCES


