John Lee
Computer Integrated Surgery II

Paper Critical Review: Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery

Paper Selection and Project Relevance

The ultimate goal of both our project and this paper is to use image-processing techniques to register monocular endoscopic video to CT data in sinus surgery. The overlap of goals was the primary reason I chose this paper; by using image processing to register and track the endoscope during sinus surgery, a surgeon can overcome the limitations of currently existing fiducial methods of tracking. This paper also provides an informative workflow for the image-based registration in the context of sinus surgery. While this paper seeks to solve the same goals as our project, we are using a different image-processing technique, namely optical flow to track contours, from the algorithm outlined in the paper. Nonetheless, the paper provides an image-processing algorithm that gives insight into reconstructing the 3D surface from endoscopy video that will be useful for our project.

Problem Summary

The sinuses are near the brain, eye, and major arteries, so high precision is necessary during sinus surgery. This paper seeks to find an accurate and rigorous registration and tracking method that would increase precision and reduce the risk of serious damage to any of these important structures. To do so, they outline two main goals. The first goal is to reconstruct 3D surface data from monocular intraoperative endoscopy video. The second goal is to register the surface data to preoperative CT image coordinates in order to track the tool with respect to anatomical landmarks in the CT images. The authors of the paper pursued this image-based registration approach as an alternative to currently existing fiducial-based registration and magnetic tracking methods. There are three main limitations with currently existing methods that the paper addresses. The first is the inability to register the surgical site to anatomical landmarks. This can potentially decrease navigational accuracy and lead to large registration error in regions of interest. The second limitation is the inability to account for anatomy changes during surgery. This limitation means that the navigation system and preoperative data become less useful as the surgery goes on. The final limitation is the inability to repetitively and autonomously register a patient. By using their image-processing approach, the paper seeks to overcome these limitations to create a high accuracy tracking system to be used in sinus surgeries.

Overview of Methods

The paper outlines the methods by which the authors reconstruct 3D data from endoscope video and register that data to the CT data. The overall workflow is as follows: scaled 3D reconstruction (system initialization; feature extraction; localization and mapping) and registration of endoscope data to CT scan (scale recovery for 3D reconstruction; ICP registration).

The 3D reconstruction is initialized by making a guess about the 3D structure based on a number of landmarks. This is accomplished with either an eight-point algorithm or manual feature selection. The eight-point algorithm finds an essential matrix, $\mathbf{E}$**,** from eight point correspondences. The essential matrix is defined as $p_i^* E p_i = 0$ for two corresponding camera projections p_i^* and p_i (in this case, two consecutive camera frames). From the essential matrix, the rotation matrix and translation vector between the two frames can be calculated as $\mathbf{E} = \mathbf{R} \cdot \mathrm{sk}\,(\mathbf{T})$. In manual feature selection, the surgeon selects three points with known correspondence to the CT-data to bootstrap the process.

The feature extraction and localization and mapping steps result in a fully reconstructed 3D surface. They use Hager and Belhumeur's sum of squared differences method to track point features, which they do not go through in the paper. They can combine these tracked features with camera motion data to reconstruct the 3D surface data. To find the motion data, they use Burschka and Hager's algorithm to iteratively find a rotation matrix and translation vector between two frames, which is accurate up to a scale. They can accomplish this using only three point correspondences between frames.

After finding the reconstruction data, they were able to recover the scale for 3D recovery. To do this, they started with a rough estimate of the camera position that was provided by the system. Using that information, they estimated the portion of the CT data surface that was in view of the camera. Then using the estimated CT surface and the reconstructed surface from the previous steps, they calculated a covariance matrix between the two. They were able to reconstruct the scale between the two point clouds by aligning them with respect to a support plane defined by the eigenvectors of the covariance matrix.

At this point, they have reconstructed and localized the 3D data from the endoscope, which is scaled and aligned with the CT data. The last step is to register this 3D data with the CT data; they accomplish this using an iterative closest point algorithm (ICP). They use a rigid registration method since the anatomy of the sinus and nasal cavities consists mostly bony tissues. They use a covariance variation of the k-D tree ICP algorithm. In the covariance tree, each subspace is defined in the orthogonal coordinate system of the eigenvectors centered at the center of mass of the point set.

Overview of Results

The authors were able to provide and verify their results by first looking at the accuracy of their reconstructed points. They tested this by selecting a single point with varied distances, locations, and motions and noting the accuracy of the reconstructed 3D surface. For targets within 2 cm, they found that the reconstruction was accurate to 1.2 mm. For targets within 1 cm, they found that the reconstruction was accurate to 0.3 mm.

In addition to the accuracy of reconstructed points, they also found that they could track real sinus structures in an ex vivo porcine test. In this test, they manually hand selected feature points to track. They were able to see by inspection that the tracking was accurate for the ex vivo case.

Finally, they looked at the accuracy of the CT registration by performing ICP. They found that their method was accurate to 0.65 mm, which is comparable to 0.40

mm for fiducial methods. Additionally, their method was able to provide more informative registration error of the target region for the surgery.

Assessment

The paper was able to successfully accomplish its two goals of reconstructing surface data from endoscopy video and registering that data to the CT data in order to track the probe. They showed that this method is comparable in accuracy with currently existing methods. They further showed that their image processing based tracking technique was able to provide error measurements that the currently existing methods do not provide. As for their methodology, it seems that their feature tracking method and surface reconstruction were able to handle most of the image environments that are found in sinus surgeries. For example, they outlined a multi plane approach for surface reconstruction of complex structures.

Some of the shortcomings of this paper were that while it addresses the limitations of the currently existing fiducial methods, it appears to be susceptible to some of these limitations too. More specifically, their method is still susceptible to changes in anatomy that might affect the quality of the endoscopy video, such as excessive bleeding. However, in this case, recalibration is quicker and more accurate than fiducial based tracking. Their method may also have difficulty with certain images where it is hard to find easily discernable point features. This is a possibility for sinus surgery since the contrast and lighting may sometimes make it hard to distinguish features. Finally, their paper would have benefited from a more thorough and quantitative analysis of the ex vivo porcine results.