SCALE-INVARIANT REGISTRATION OF MONOCULAR ENDOSCOPIC IMAGES TO CT-SCANS FOR SINUS SURGERY

Paper by D. Burschka et al.

John Lee

JOHNS HOPKINS UNIVERSITY

LABORATORY FOR Computational Sensing + Robotics

Project Description

- Image Processing for Video-CT Registration in Sinus Surgery
- Use contour detection and optical flow algorithms to reconstruct surface of monocular endoscopic video
- Register reconstructed surface to CT data for intraoperative probe tracking

Paper Selection

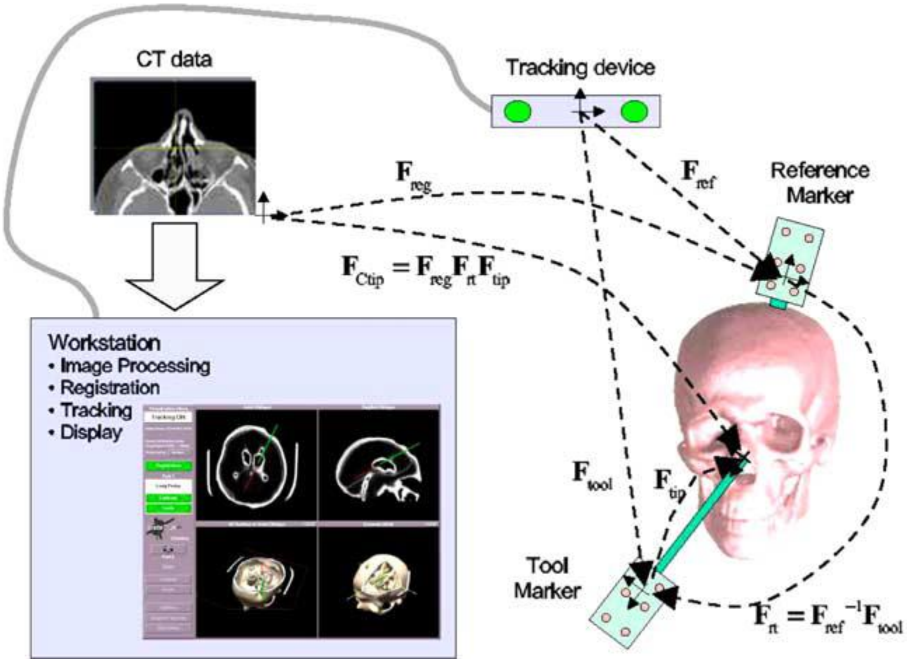
- “Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery” seeks to solve the same problem as our project
- They use monocular sinus endoscopic images as opposed to other imaging techniques such as fluoroscopy, x-ray, or stereoscopic endoscope
- Uses different image processing from our project

Problem Summary

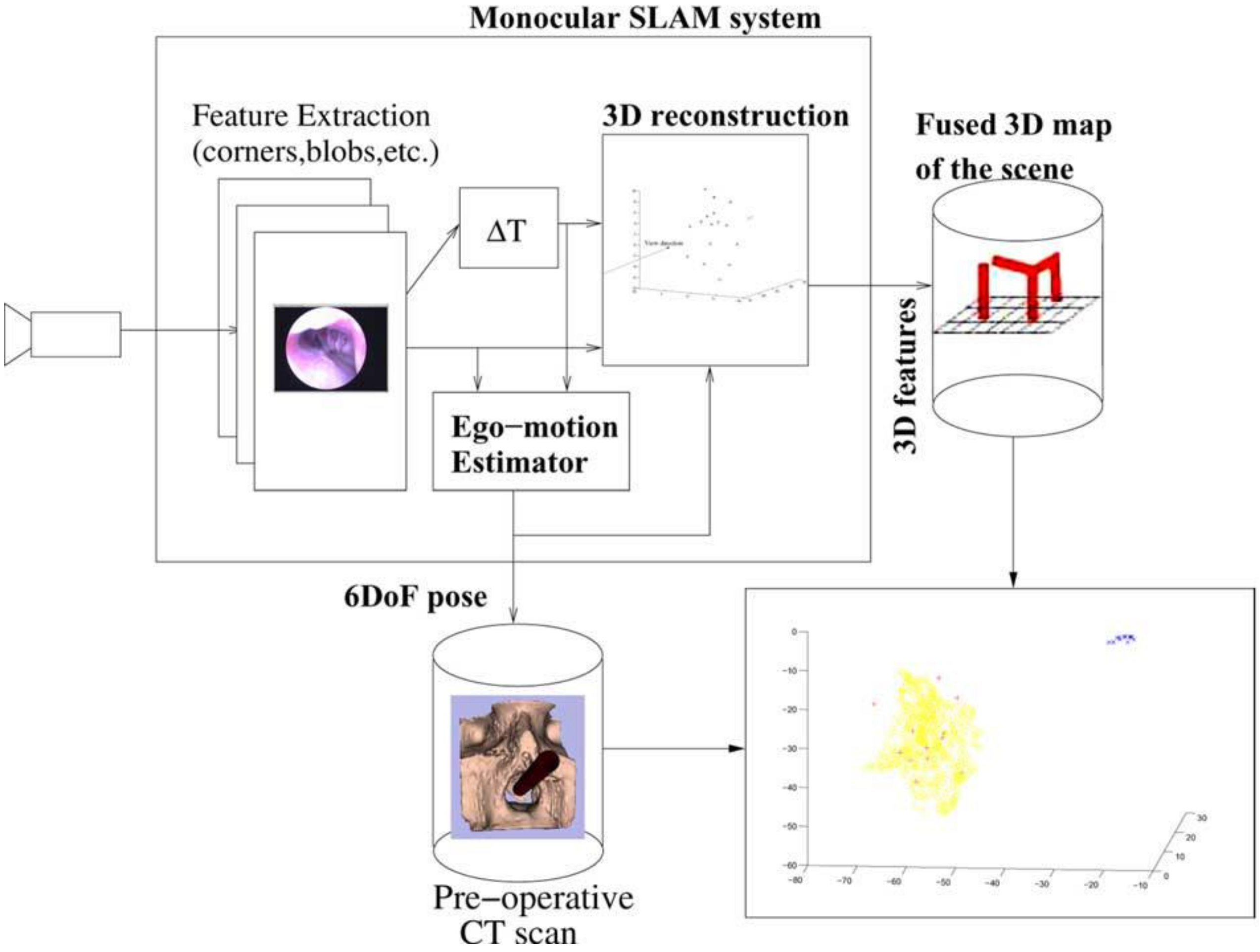
- The sinuses are near the brain, eye, and major arteries, so high precision is necessary during surgery
- This main goals of the paper are to:
  - Reconstruct the 3D surface geometry from monocular endoscopic images
  - Register the camera location to pre-operative CT image coordinates to track the endoscope

Problem Summary

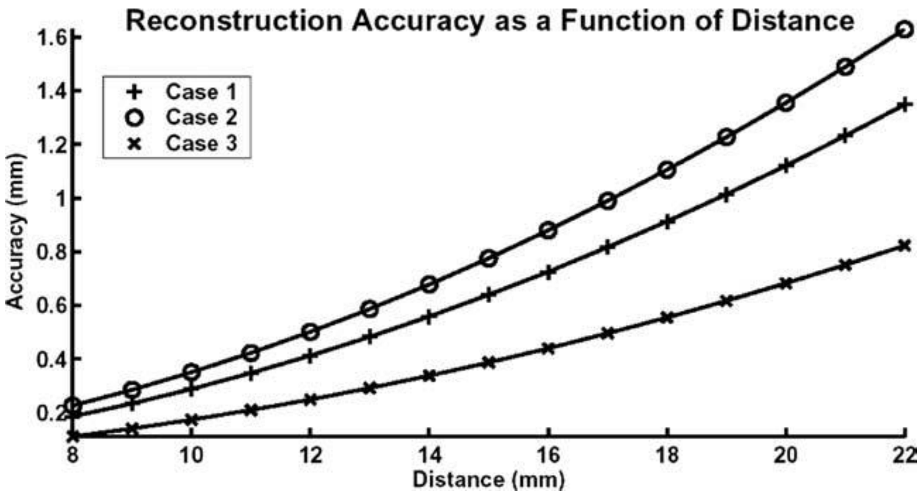
- Current tracking methods use a navigational tracking device and external fiducials
- Limitations in the context of sinus surgeries:
    - Can’t register to anatomical landmarks
    - Can’t account for anatomical changes during surgery
    - Can’t autonomously and repetitively register a patient



Architecture

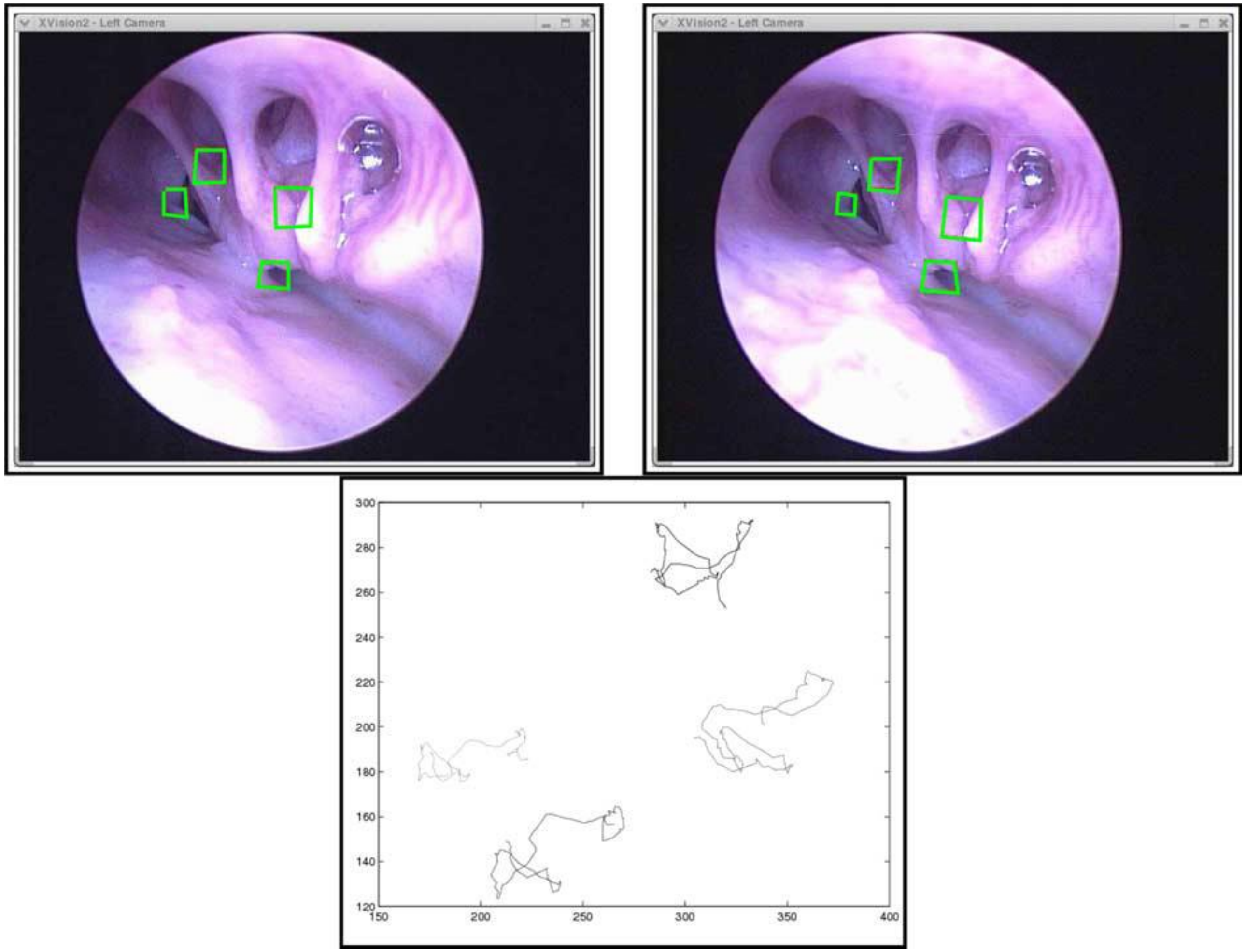


Key Result Summary



- After ICP registration, the average error between a set of selected points was 0.65 mm, compared to 0.40 mm in a fiducial based registration with four fiducials on the surface of the brain
- Able to robustly track ex vivo using a variety of anatomical structures such as significant vessel structures

Methods: Tracking



Methods: Initialization

- System initialized with eight-point algorithm or manual feature selection
  - In eight-point algorithm, you can find the essential matrix of a system from eight-point correspondences.
    - Essential matrix, $\tilde{\mathbf{E}}$, is defined as $p_i^* \tilde{\mathbf{E}} p_i = 0$ for two corresponding camera projections p_i^* and p_i (in this case, two consecutive camera frames)
    - Essential matrix provides rotation matrix $\tilde{\mathbf{R}}$ and translation vector $\mathbf{T}$ such that $\tilde{\mathbf{E}} = \tilde{\mathbf{R}} \cdot \mathrm{sk}(\mathrm{T})$
  - In manual feature selection, the surgeon selects three points with known correspondence to the CT-data to bootstrap the processing

Methods: Localization and Mapping

- Camera motion must be estimated simultaneously with reconstruction
- One way of recovering the motion between two camera frames is by using the eight-point algorithm from above
- In some cases, there may be fewer than eight points to match. They use Burschka and Hager's method for camera localization and mapping with only three point correspondences
- Brief overview of their algorithm follows

Methods: Localization and Mapping

- Each 3D point P_i is represented as a direction vector $n_i = \frac{p_i}{\|p_i\|}$ and distance to the point D_i such that $P_i = D_i \cdot n_i$. Since the scale m of reconstruction may be unknown, use $\lambda_i = \frac{D_i}{m}$
- For current frame $\{P_i\}$ and next frame $\{P_i^*\}$, estimate $\mathbf{R}$ and $\mathbf{T}$ as such.

$$\bar{P} = \frac{1}{n}\sum_{i=1}^{n} P_i \, , \bar{P}^* = \frac{1}{n}\sum_{i=1}^{n} P_i^*$$

$$P_i' = P_i - \bar{P}_i, P_i'^{\,*} = P_i^* - \bar{P}^*$$

$$\mathbf{M} = \sum_{i=1}^{n} P_i'^* P_i^{\mathbf{T}} \, , [UDV^{\mathrm{T}}] = \mathrm{svd}(\mathbf{M})$$

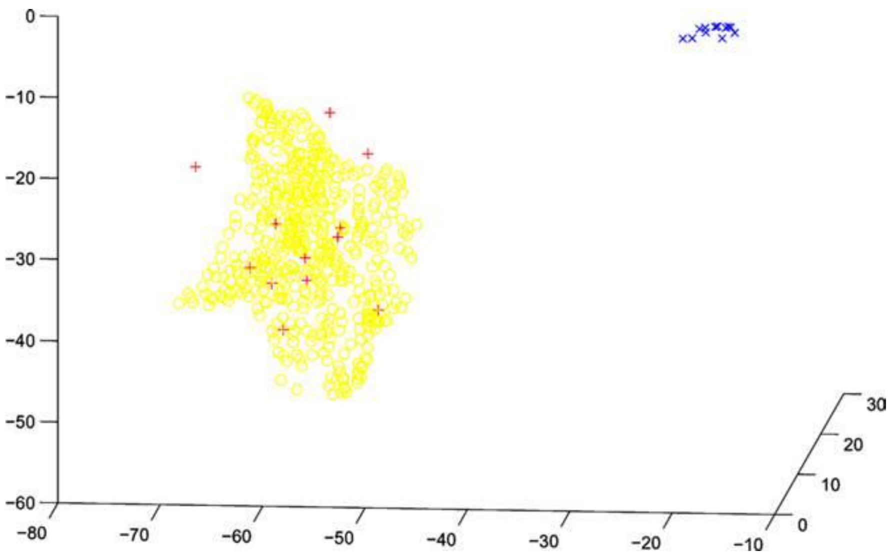
$$\boldsymbol{R} = V \cdot U^{\mathrm{T}}, \mathbf{T} = \bar{P}^* - \mathbf{R}\bar{P}$$

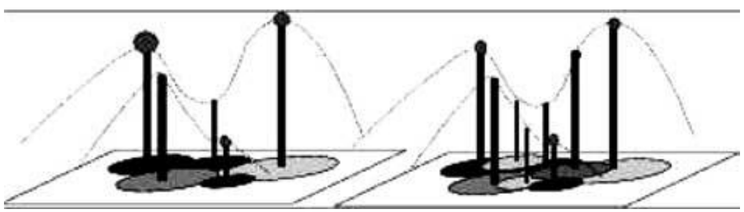
- For each new image, start with initial guess for λ_i set to the previous distance, then iterate to find true $\mathbf{R}, \mathbf{T}, \lambda_i$

Methods: Scale Recovery

- The system has an estimate for the current camera position. Using this estimate, they carve out a portion of the CT surface that they expect is currently visible to the camera
- Look at the covariance matrix between point cloud from selected CT region and the current camera reconstruction.
  - The two eigenvalues of this matrix from the larger eigenvalues define the supporting plane, and the third eigenvector describes the depth variation
  - Using the eigenvalues and eigenvectors, they recover the scale and rotation between the two point clouds with respect to the supporting plane

Methods: Scale Recovery





- Fig. 10. After the alignment along the normal vector to the supporting plane the scale is roughly recovered, but rotation around the normal vector is possible.

Fig. 11. Distance to the supporting plane is used as a pseudo-image representation to match the sparse reconstruction (left) to the dense point cloud (right).

Methods: ICP

- Perform ICP (iterative closest point) between the two point aligned point clouds for registration
- They use a rigid registration as opposed to a deformable registration since the anatomy of the nasal and sinus cavity is mostly bony tissue
- They use a covariance tree variation of a k-D tree as their data structure

Relevance

- As said earlier, this paper seeks to solve the same problem we are trying to solve using the same kind of imaging data
- They provide alternative image processing techniques that we hope to learn from and build upon

Assessment

Pros

- Overcomes the limitations that come with fiducial based tracking
- Ability to get registration error of a target region that isn’t possible in current methods
- Comparable results to fiducial based methods

Cons

- Still susceptible to large anatomical changes e.g. bleeding that covers the camera
- More rigorous and quantitative testing of ex vivo tracking results

Conclusion

- They have shown that it is feasible to register and track an endoscope using image processing techniques
- We plan to implement a comparable workflow with equal or better results using different types of image features

Questions?