Evaluation of CT Registration with Image-Based Sinus Reconstruction

CIS II Final Report
Project 12

Jan Mangulabnan
under the mentorship of
Dr. Roger Soberanis and Professor Mathias Unberath

May 11, 2023

1 Introduction

1.1 Clinical Motivation

Nasal obstructions are a common clinical problem that can significantly impact patients' quality of life. Septoplasty and turbinate reduction are two of the most common surgical interventions used to address this problem which are shown in Figure 1.

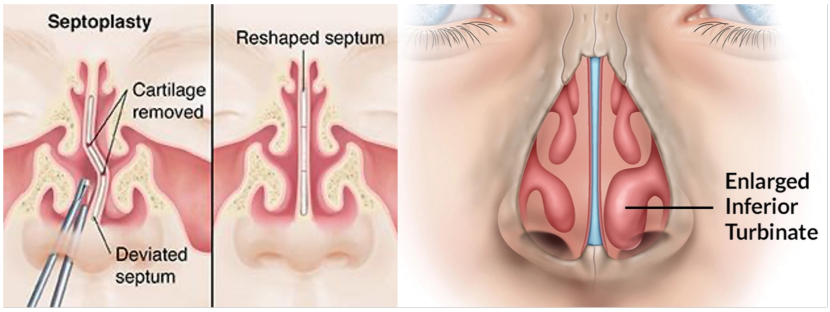

Figure 1: Example patient conditions that may cause nasal obstructions. Anatomical diagram from [2] of deviated septum which may suggest septoplasty surgery is needed to reshape the anatomy (left) and diagram from [1] of enlarged turbinate in the sinus cavity that may motivate turbinate reduction surgery (right).

Clinical studies have provided evidence towards the positive impact of these interventions in patients' quality of life [7, 5]. However, these assessments rely on subjective measures such as patient-reported symptoms and there is limited quantitative evidence to support the effectiveness of these surgeries [11]. This creates a need for objective evaluation methods to inform clinicians about which patients are most likely to benefit from surgical intervention and provide tools for longitudinal assessment of the surgical outcomes.

Quantitative evaluation of the patient sinus anatomy requires geometric information of the nasal cavity. Computed tomography (CT) scans provide this information, however, these scans are expensive and expose patients to harmful radiation. This motivates an image-centered approach that leverages routine endoscopy procedures that physicians use to examine patients. An accurate 3D model of the sinus anatomy is required to retrieve clinically relevant parameters, such as aperture and volume, which can be used for quantitative assessment of patient anatomy.

1.2 Prior Work

Liu et. al [9] developed a pipeline to generate a dense 3D reconstruction of the sinus anatomy from endoscopic video as shown in Figure 2.

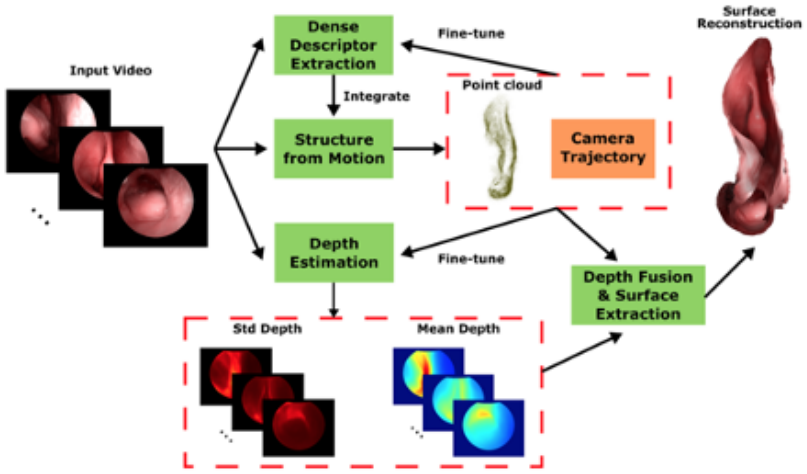

Figure 2: Dense reconstruction pipeline adapted from [9].

The pipeline utilizes a Structure from Motion (SfM) algorithm [12] which matches common visible points in input images to produce a 3D structure of the object in the image. The visible points are matched using dense feature descriptors which are extracted using a deep learning model. These dense descriptors are then integrated into SfM to produce a dense point cloud of the anatomy and the camera trajectory for each frame in the input sequence.

The input sequence is also used to estimate the depth of the images from the camera using a deep learning model [8]. The depth estimation, point cloud, and camera trajectories are then used in a depth fusion method [3] to produce a 3D reconstruction of the sinus cavity. Currently, there is no assessment framework towards the correctness of the image-based 3D reconstruction.

1.3 Goals

The main goal of this project was to implement a quantitative framework to evaluate the accuracy of the dense reconstruction based on the ground truth CT. The development of this assessment and framework would enable further research towards the usage of the dense reconstruction pipeline for sinus anatomy in clinical settings. The deliverables of this project are listed as follows:

1. (Minimum) Implement a registration framework to evaluate the image-based 3D reconstruction of the sinus anatomy with respect to the corresponding CT image.
   (a) Integrate multiple registration methods to align meshes.
   (b) Report error evaluation metrics between dense reconstruction and CT with visualizations.
2. (Expected) Analyze the influence of uncertainty in the reconstruction pipeline, and evaluate the resulting reconstruction with respect to the CT.
   (a) Adjust depth fusion step in pipeline using inverse weighting of uncertainties, removal of outliers, and removal of larger depth estimates.
   (b) Evaluation of resulting dense reconstructions.
3. (Updated Maximum) Transactions in Medical Imaging paper extension for MICCAI 2020 Dense Reconstruction paper.
   (a) Process additional cadaveric data in dense reconstruction pipeline and evaluate.
   (b) Ablation experiments of dense reconstruction pipeline.
4. (Original Maximum, now Future Work) Integrate robot kinematics in the registration process and evaluate results with respect to the CT.

2 Technical Summary

The implemented registration framework follows the workflow as shown in Figure 3 and is optimized for data collected from cadaveric experiments for the quantitative sinus endoscopy project as described in Appendix B.

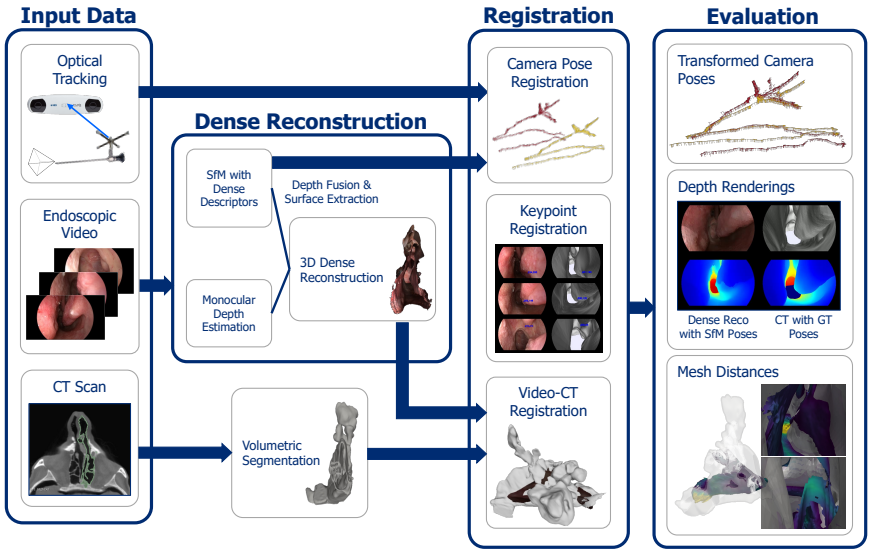

Figure 3: Registration framework

2.1 Data Preprocessing

Registration is dependent on input data of the dense reconstruction (DRECO) pipeline output (processed image sequence, estimated camera trajectories from SfM, and fused mesh of sinus anatomy), the ground-truth anatomical structure of the sinus from the corresponding CT scan and optical tracking data of the endoscope camera. The collected video sequence was preprocessed for use in the DRECO pipeline by curating the images to isolate the subsequence of frames that capture the sinus cavity. The input image frames were also downsampled and undistorted based on checkerboard camera calibration. The preprocessed images were then used as input to the DRECO pipeline. The CT scans were processed using 3D Slicer [6] to segment both the attached marker spheres and sinus anatomy and further instructions can be found in Appendix C.

2.2 Registration Framework

Direct rigid registration of sampled point clouds from the DRECO and CT meshes fails since the dense reconstruction only constructs a portion of the segmented anatomy present in the CT mesh. Therefore, this framework required advanced options including camera pose, keypoint, and coherent point drift registration to align the meshes. The GitHub repository for this codebase can be found at `https://git.lcsr.jhu.edu/sinusendoscopy/video_ct_registration`.

2.2.1 Camera Pose Registration

The segmented anatomy marker spheres are required to obtain the ground-truth positions of the endoscope camera in CT space. Based on the segmentation, the centers of these spheres are extracted using a sphere-fitting algorithm to register the CT to the tracked marker geometry of the anatomy recorded by the NDI Polaris tracker (${}^{T}T_{R}$) to obtain the transformation (VT_{R}). A checkerboard hand-eye calibration was also done to determine the transformation between the endoscope camera and endoscope marker sphere geometry (${}^{E}T_{C}$). These transformations along with the tracked endoscope (represented as ${}^{T}T_{E}$) may then be used to compute the ground-truth position of the camera in CT space (${}^{V}T_{C}$) using Equation 1.

$$ {}^{V}T_{C} = ({}^{V}T_{R})({}^{T}T_{R})^{-1}({}^{T}T_{E})({}^{E}T_{C}) \tag{1} $$

The tracked positions were also manually adjusted by comparing the recorded images and CT renderings when large errors were observed (camera position was outside of the anatomy). This alignment transformation was applied at the beginning of the chain in Equation 1.

The DRECO pipeline estimates the camera trajectories of each input image frame in the SfM algorithm which can be matched to the corresponding ground-truth position of the endoscope for rigid registration. The resulting transformation was then used to transform the DRECO mesh to CT space and initialize the iterative closest point algorithm for Video-CT registration.

The estimated camera trajectories were observed to have large variations compared to the ground-truth, reducing the accuracy of the overall registration. Due to these errors, local reconstructions of the sinus anatomy were also evaluated based on a section of the original input sequence as shown in Figure 4.

2.2.2 Keypoint Registration

Manual keypoint selection was also explored as an option to register the meshes. The user may manually select corresponding 2D keypoints from the input images and CT renderings which will compute the equivalent 3D point on the meshes to be used for registration.

2.2.3 Coherent Point Drift Registration

In addition to rigid registration methods, the Coherent Point Drift (CPD) registration algorithm [10] was also investigated. CPD is a probabilistic method integrated for rigid and affine point cloud registration between sampled points from the dense reconstruction and CT meshes. This method optimizes registration based on the most likely shape of the DRECO mesh within the CT structure, considering that the computed mesh only represents the section of the sinus anatomy visible in the input video.

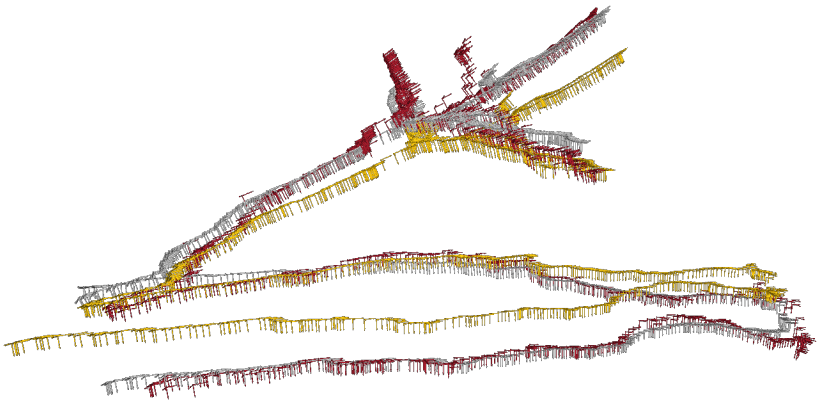

Figure 4: Registered camera poses where red represents the ground-truth tracked trajectory, orange is the registration result of the entire sequence, and gray shows the results from section-based registration.

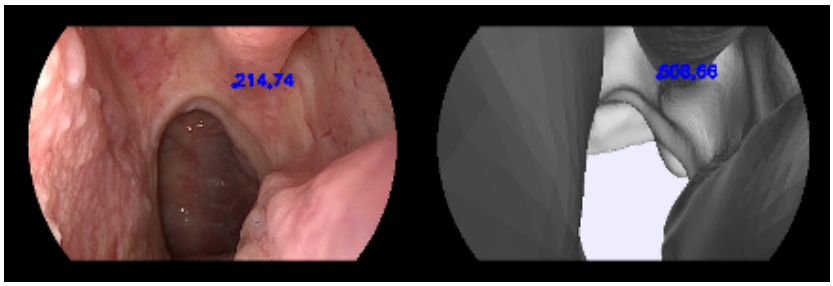

Figure 5: Example of manually selected keypoints from an input image to sample points from the reconstructed mesh (left) and corresponding selected point in CT rendering (right) used for registration.

2.3 Depth Fusion in Dense Reconstruction

The dense reconstruction pipeline utilizes depth estimates [8] in addition to the SfM point cloud and camera trajectories to generate the 3D structure. This information is integrated into a fusion method which resolves variation between the estimates of common points in multiple frames of the input sequence [3]. The fusion method currently weights every estimate equally; however, the points in the sinus anatomy that are further away from the camera when the image is captured is shown to have more uncertainty as seen in Figure 6.

We hypothesized that this uncertainty may be introducing errors which are propagated into the reconstruction at the fusion step. To further investigate, I generated reconstructions using inverse weighting based on the standard deviation to reduce the influence of uncertain depth estimates. I also removed outliers

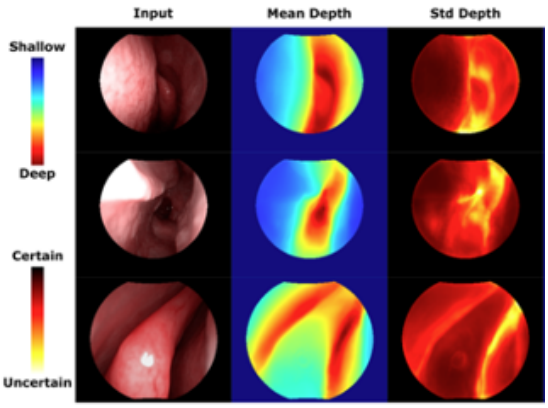

Figure 6: Adapted from [9]. Heat map of mean and standard deviation of depth estimates with corresponding input image. The deeper mean depth estimation, meaning further away from the camera at capture, exhibits higher uncertainty.

outside of the 68th percentile (one standard deviation) of uncertainty and mean depth estimations. These alternate reconstructions were also evaluated using the implemented registration framework.

3 Results

3.1 Evaluation Metrics

Video-CT registration was evaluated based on the estimated and ground-truth camera poses, depth renderings of the dense reconstruction and CT, and the distance between the registered DRECO and CT meshes. The camera poses were compared by scaling and transforming the estimated trajectories from the SfM algorithm using the registration transformation to report the mean translation and rotation error (Equation 2).

$$\theta = cos^{-1}\left(\frac{trace(R_{SfM}^{-1}R_{GT}) - 1}{2}\right) \tag{2}$$

The depth was rendered for the dense reconstruction at the estimated SfM camera poses and CT at the ground-truth poses. While the DRECO mesh was scaled to the CT based on the norm of the camera poses during registration, these renderings were compared using Scale-Invariant Mean Squared Error [4] to eliminate scaling as a potential source of error. The difference between each pixel i was computed (Equation 3) in order to compute the scale-invariant depth error for each frame (Equation 4) where y represents the DRECO rendering and y^* is the ground-truth CT rendering.

$$d_i = log(y_i) - log(y_i^*) \tag{3}$$

$$D(y, y^*) = \frac{1}{n}\sum_i (d_i)^2 - \frac{1}{n^2}\left(\sum_i d_i\right)^2 \tag{4}$$

The weighted-average of this error across the sequence was reported to account for pixels in the renderings where no surface was present. Example renderings are shown in Figure 7. The mean distance between meshes was also computed by finding the closest point from the DRECO to CT mesh after applying the registration transformation.

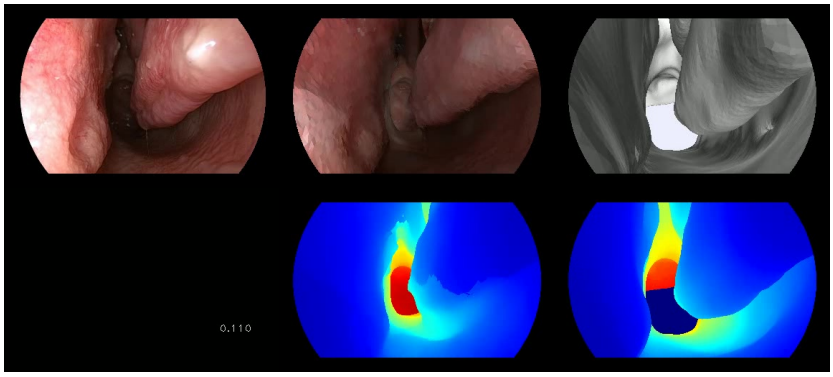

Figure 7: Example of color and depth renderings of dense reconstruction from estimated camera trajectory (middle column) and CT renderings from ground-truth trajectory (right column) with the corresponding endoscope image shown on the left.

3.2 Registration Options

In order to examine both global and local registrations, sections of the input image sequence were also reconstructed for mesh to mesh comparison. The registered local meshes are shown in Figure 8.

The results of registration using manually selected keypoints are shown in Figure 9. While the layered renderings seem to show adequate alignment of the internal sinus structures, the overall Video-CT registration of the meshes exhibit larger errors in the transformed camera poses (8.99 mm translation error). Considering this method also requires manual intervention, keypoint registration was not investigated further for Video-CT registration.

Table 1 displays the mean errors of the various registration algorithms. Rigid camera pose registration using ICP had the lowest translational error, ranging from 1 to 3mm differences, whereas both Coherent Point Drift algorithms had translational errors on the magnitude of centimeters for the transformed camera

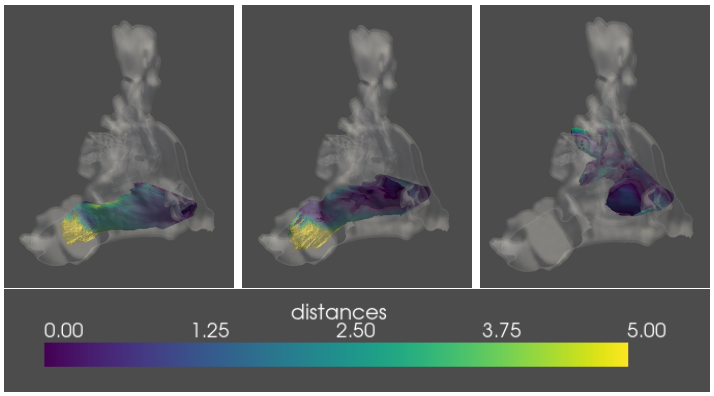

Figure 8: Local reconstructions from three sections of the input image sequence.

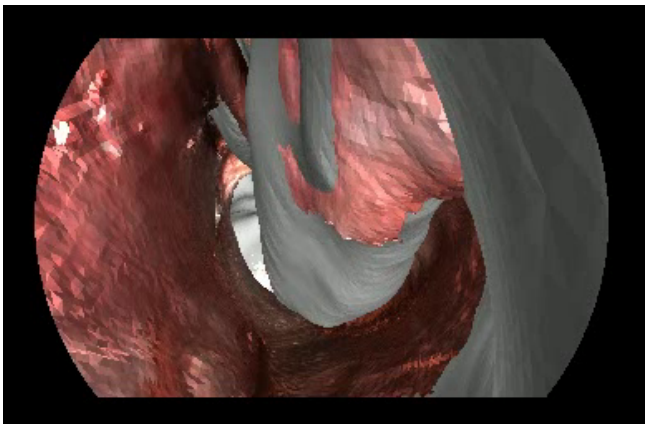

Figure 9: Layered mesh rendering from keypoint registration.

poses. All three registration results have significant rotational error, ranging from 12 to 21 degrees. These errors are the same for each type as only the camera position was transformed.

The CPD rigid and affine registration algorithms have lower errors in the mean distance between meshes but based on visual inspection of the layered meshes (Figure 10) and depth renderings, this does not seem to mean that the anatomy is more aligned. The smaller magnitudes may be a result of the intricate sinus anatomy as the closest points between the meshes do not necessarily correspond to the same points within the sinus cavity.

It is expected that the camera pose + ICP registrations have lower pose errors and the CPD registrations have lower mesh errors because these algorithms compute the transformation by minimizing those parameters. The scale invariant depth error serves as a metric independent of the registration. Since the camera

Table 1: Comparison of various registration types using the entire image sequence (indexes 0 - 1059) and multiple sections for local registration reported as the mean across poses, pixels, and sampled points for camera pose, scale invariant depth, and mesh distance errors, respectively.

Registration Type	**Sequence Indexes**	**Camera Pose**		**Scale Invariant Depth Error**	**Mesh Distance (mm)**
		Translation Error (mm)	Rotation Error (deg)		
Camera Pose + ICP	0 - 1059	3.30	12.56	0.49	1.69
	0 - 220	1.38	14.03	0.422	2.885
	221 - 450	1.08	14.95	0.384	3.098
	451 - 1059	1.81	21.06	0.472	0.873
CPD Rigid	0 - 1059	13.43	12.56	0.51	1.38
	0 - 220	41.45	14.03	1.097	1.028
	221 - 450	36.61	14.95	0.692	0.927
	451 - 1059	4.97	21.06	0.550	0.886
CPD Affine	0 - 1059	18.86	12.56	0.67	1.09
	0 - 220	21.52	14.03	1.047	0.890
	221 - 450	22.94	14.95	0.727	0.775
	451 - 1059	7.42	21.06	0.503	1.005

pose + ICP registration type had the smallest error in the depth renderings, this algorithm was used to further investigate adjustments to the depth fusion step in the dense reconstruction pipeline.

3.3 Depth Fusion Adjustment

The resulting reconstructions using different adjustment schemes are shown in Figure 11. There were no obvious visible differences in the shape of the sinus anatomy, except for missing surfaces when outliers were removed. These meshes were also reconstructed with sections of the sequence and evaluated using the registration framework. The reported errors are shown in Table 2.

There does not seem to be significant differences in the reported error metrics for the reconstructions with the adjusted depth fusion. This implies that the structure of the sinus anatomy remains mostly similar to that of the original reconstruction. This disproves the original hypothesis that the uncertainty in the depth estimation contributes to the structural differences with respect to the CT in the depth fusion step in the dense reconstruction pipeline. Further research is necessary to isolate the main source of this error and is planned for future work towards quantitative sinus endoscopy.

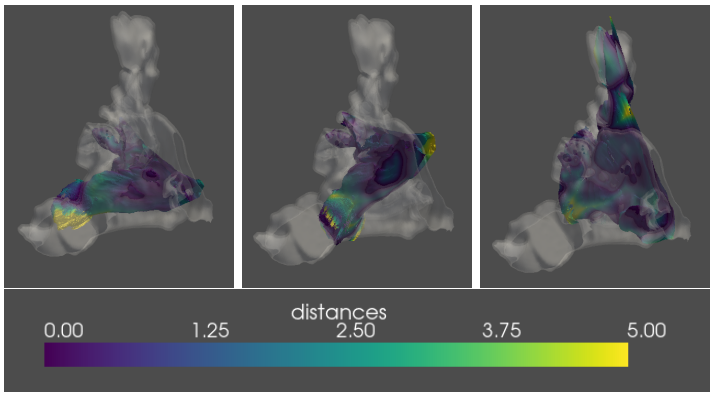

Figure 10: Layered mesh visualization using different types of registration methods: camera pose and ICP (left), CPD rigid (middle), and CPD affine (right).

Figure 11: The resulting dense reconstructions with different adjustment schemes (from left to right): original, weighting by uncertainty, large depths removed, and outliers removed.

4 Management Summary

During the course of this project, I was able to construct a registration framework for repeatable evaluations of the dense reconstruction pipeline output and ground truth CT data, which was the minimum deliverable. The finalized repository supports multiple registration options including initialization using registered camera poses and manual keypoint selection as well as iterative closest point and coherent point drift algorithms for point-cloud registration of the dense reconstruction and ground-truth CT structure. Based on the registration scheme used, this codebase also provides an evaluation based on frame-by-frame depth renderings and the registered meshes.

I was also able to complete the expected deliverable of evaluating different weighting schemes based on the depth fusion step in the dense reconstruction pipeline using the aforementioned registration framework. The mesh was reconstructed using inverse weighting of the uncertainties, removal of outliers (outside

Table 2: Fusion Adjustment Results

Sequence Indexes	**Fusion Adjustment**	**Scale Invariant Depth Error**	**Mesh Distance (mm)**
0 - 1059	Original	0.490	1.691
	Weighted	0.491	1.674
	Large Depths Removed	0.493	1.786
	Outliers Removed	0.514	1.710
0 - 220	Original	0.422	2.885
	Weighted	0.425	2.890
	Large Depths Removed	0.418	3.129
	Outliers Removed	0.408	2.675
221 - 450	Original	0.384	3.098
	Weighted	0.380	3.102
	Large Depths Removed	0.385	2.974
	Outliers Removed	0.413	2.661
451 - 1059	Original	0.472	0.873
	Weighted	0.482	0.874
	Large Depths Removed	0.533	0.950
	Outliers Removed	0.528	0.937

of a percentile of the uncertainty), and removal of larger depths. Considering that our original hypothesis did not seem to improve the registration evaluation, I plan to continue investigating potential sources of error in the pipeline that may occur during the depth fusion.

The original maximum deliverable was to integrate robot kinematics in the registration framework, however, this was changed to a Transactions in Medical Imaging (TMI) paper extension of the original dense reconstruction conference paper for MICCAI 2020 [9]. Due to the complex nature of the registration, this was not completed during this course but I plan to continue working on this project to finalize this submission. The TMI paper will include evaluations based on this registration framework of at least two additional cadaveric subjects and ablations to the pipeline.

Overall, completion of this CIS II project helped solidify my understanding of our current efforts towards quantitative sinus endoscopy. I was able to become acquainted with the details of our dense reconstruction pipeline and cadaveric data collection procedure by implementing this registration framework and I believe that this knowledge will help facilitate my future contributions towards vision-based navigation for endoscopic procedures.

Appendices

A GitHub Code Repository

The codebase for Video-CT registration is stored on the Sinus Endoscopy GitHub repository at: `https://git.lcsr.jhu.edu/sinusendoscopy/video_ct_registration`
The documentation for usage can be found in the README.md file.

B Data Specification

This registration framework requires data collection in the format specified at the following link: `https://git.lcsr.jhu.edu/sinusendoscopy/video_ct_registration/-/blob/main/DataSpec.md`

C Segmentation Instructions

Detailed instructions for how to use 3D slicer for segmentations is also included in the GitHub repository to ensure proper formatting for use in this registration framework. This can be found at the following link:
`https://git.lcsr.jhu.edu/sinusendoscopy/video_ct_registration/-/blob/main/HowToSegment.md`

References

[1] Sept. 2022. URL: `https://www.stlsinuscenter.com/common-sinus-problems/inferior-turbinate-hypertrophy/turbinate-reduction-surgery/`.

[2] URL: `https://www.saintlukeskc.org/health-library/nasal-surgery-septoplasty`.

[3] Brian Curless and Marc Levoy. "A volumetric method for building complex models from range images". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996, pp. 303–312.

[4] David Eigen, Christian Puhrsch, and Rob Fergus. "Depth map prediction from a single image using a multi-scale deep network". In: *Advances in neural information processing systems* 27 (2014).

[5] Maija L Hytönen et al. "Does septoplasty enhance the quality of life in patients?" In: *European archives of oto-rhino-laryngology* 269 (2012), pp. 2497–2503.

[6] Ron Kikinis, Steve D Pieper, and Kirby G Vosburgh. "3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support". In: *Intraoperative imaging and image-guided therapy*. Springer, 2013, pp. 277–289.

[7] Michelle Lavinsky-Wolff et al. "Effect of turbinate surgery in rhinoseptoplasty on quality-of-life and acoustic rhinometry outcomes: a randomized clinical trial". In: *The Laryngoscope* 123.1 (2013), pp. 82–89.

[8] Xingtong Liu et al. "Dense depth estimation in monocular endoscopy with self-supervised learning methods". In: *IEEE transactions on medical imaging* 39.5 (2019), pp. 1438–1447.

[9] Xingtong Liu et al. "Reconstructing Sinus Anatomy from Endoscopic Video – Towards a Radiation-Free Approach for Quantitative Longitudinal Assessment". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by Anne L. Martel et al. Cham: Springer International Publishing, 2020, pp. 3–13. ISBN: 978-3-030-59716-0.

[10] Andriy Myronenko and Xubo Song. "Point set registration: Coherent point drift". In: *IEEE transactions on pattern analysis and machine intelligence* 32.12 (2010), pp. 2262–2275.

[11] DG Roblin and Ronald Eccles. "What, if any, is the value of septal surgery?" In: *Clinical Otolaryngology & Allied Sciences* 27.2 (2002), pp. 77–80.

[12] Johannes L Schonberger and Jan-Michael Frahm. "Structure-from-motion revisited". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 4104–4113.