



# Endoscopic Navigation in the Absence of CT Imaging

Ayushi Sinha<sup>1</sup>(✉), Xingtong Liu<sup>1</sup>, Austin Reiter<sup>1</sup>, Masaru Ishii<sup>2</sup>,  
Gregory D. Hager<sup>1</sup>, and Russell H. Taylor<sup>1</sup>

<sup>1</sup> The Johns Hopkins University, Baltimore, USA  
sinha@jhu.edu

<sup>2</sup> Johns Hopkins Medical Institutions, Baltimore, USA

**Abstract.** Clinical examinations that involve endoscopic exploration of the nasal cavity and sinuses often do not have a reference image to provide structural context to the clinician. In this paper, we present a system for navigation during clinical endoscopic exploration in the absence of computed tomography (CT) scans by making use of shape statistics from past CT scans. Using a deformable registration algorithm along with dense reconstructions from video, we show that we are able to achieve submillimeter registrations in in-vivo clinical data and are able to assign confidence to these registrations using confidence criteria established using simulated data.

## 1 Introduction

Endoscopic explorations of the nasal cavity and sinuses are generally not accompanied by a reference computed tomography (CT) image since CT image acquisition exposes patients to high doses of ionizing radiation and is, therefore, avoided unless necessary. Clinicians performing the exploration must rely entirely on the endoscopic camera for visualization and, therefore, must cope with restricted field of view. In order to reduce reliance on experience or memory and to provide additional context information, we have developed a system that enables navigation without the need for accompanying patient CT or other similar imaging and associates a confidence measure to the navigation being provided. Further, our system does not introduce any additional devices than those already used in clinical endoscopic exploration. Therefore, the clinician is not responsible for anything in addition to the endoscope.

Most navigation systems that have been developed are intended for surgical use [1, 2]. For surgical navigation, there is almost always access to preoperative CT scans, which have high contrast between air, bone, and soft tissue. This allows surgeons to better understand their location, the proximity to surrounding bones and soft tissue, and the thickness of surrounding bones, enabling them to make more informed decisions during surgery and prevent harm to critical structures nearby, like the brain, eyes, optic nerves, carotid arteries, etc.

The main difference between these previous methods and the method presented here is the absence of patient specific CT scans. In order to make up

for this absence, we utilize past CT scans to build statistical shape models of relevant structures. Statistically derived shapes are then deformably registered to dense reconstructions of anatomy visible in endoscopic video, and statistical confidence measures are automatically assigned to the registrations. The registration accomplishes two tasks simultaneously. First, it aligns the endoscopic video to the statistically derived shape, giving the clinician more information about where surrounding structures may be. Second, it deforms the statistically derived shape to fit the structure obtained from video and, in effect, *estimates the patient CT*. The confidence measure further informs the clinician on when and how much the navigation system can be trusted, and also allows the navigation system to attempt to improve itself if its current registration estimate has low confidence. We perform two experiments to evaluate our framework. First, we establish that our framework can compute submillimeter registrations and reliably assign confidence to the registrations using simulated data. Second, we evaluate our framework on in-vivo clinical data, and use the confidence criteria to assign confidence to the registrations.

## 2 Method

To build statistical shape models (SSMs), we automatically segment 53 publicly available head CTs [3–6] by transferring 3D meshes extracted from manually created labels in a template CT image to the 53 CTs using deformation fields produced by an intensity-based CT-CT registration algorithm [7]. With some improvement to these initial segmentations using the method described in [8], we obtain reliably segmented structures in all CTs along with reliable correspondences. These correspondences allow us to build SSMs of the segmented structures using established methods like principal component analysis (PCA) [9]:

$$\Sigma_{\text{SSM}} = \frac{1}{\mathbf{n}_s} \sum_{j=1}^{\mathbf{n}_s} (\mathbf{V}_j - \bar{\mathbf{V}})^T (\mathbf{V}_j - \bar{\mathbf{V}}) = [\mathbf{m}_1 \dots \mathbf{m}_{\mathbf{n}_s}] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{\mathbf{n}_s} \end{bmatrix} [\mathbf{m}_1 \dots \mathbf{m}_{\mathbf{n}_s}]^T, \quad (1)$$

where  $\mathbf{V}_j$  is the stacked vector of vertices,  $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_{\mathbf{n}_v}]^T$ , for the  $j$ th mesh,  $\bar{\mathbf{V}}$  is the mean shape computed by averaging the  $\mathbf{n}_v$  corresponding vertices over  $\mathbf{n}_s$  shapes,  $\bar{\mathbf{V}} = \frac{1}{\mathbf{n}_s} \sum_{j=1}^{\mathbf{n}_s} \mathbf{V}_j$ , and  $\Sigma_{\text{SSM}}$  is the shape covariance matrix. An eigen decomposition of  $\Sigma_{\text{SSM}}$  produces the principal modes of variation,  $\mathbf{m}$ , and the mode weights,  $\lambda$ , which represent the amount of variation along the corresponding  $\mathbf{m}$  (Eq. 1). PCA enables any new shape,  $\mathbf{V}^*$ , that is in correspondence with the shapes used to build the SSM, to be estimated using  $\bar{\mathbf{V}}$ ,  $\mathbf{m}$  and  $\lambda$ :  $\tilde{\mathbf{V}}^* = \bar{\mathbf{V}} + \sum_{j=1}^{\mathbf{n}_m} s_j \mathbf{w}_j$ , where  $\tilde{\mathbf{V}}^*$  is the estimated  $\mathbf{V}^*$ ,  $1 \leq \mathbf{n}_m < \mathbf{n}_s$  is some specified number of modes,  $\mathbf{w}_j = \sqrt{\lambda_j} \mathbf{m}_j$  are the weighted modes of variation, and  $s_j$  are the shape parameters in units of standard deviation (SD) which can be obtained by projecting the mean subtracted  $\mathbf{V}^*$  onto the weighted modes.

These shape parameters,  $\mathbf{s} = \{s_j\}$ , can be incorporated into probabilistic models of registration to enable optimization over  $\mathbf{s}$  in addition to other registration parameters [10]. In particular, we evaluate the deformable extension of the

generalized iterative most likely oriented point (G-IMLOP) algorithm, an iterative rigid registration algorithm [11]. The generalized *deformable* iterative most likely oriented point (GD-IMLOP) algorithm extends G-IMLOP, which incorporates an anisotropic Gaussian noise model and an anisotropic Kent noise model to account for measurement errors in position and orientation, respectively [11]. Assuming both position and orientation errors are zero-mean, independent and identically distributed, the match likelihood function for each oriented point,  $\mathbf{x}$ , transformed by a current similarity transform,  $[a, \mathbf{R}, \mathbf{t}]$ , is defined as [11]:

$$f_{\text{match}}(\mathbf{x}; \mathbf{y}, \Sigma_{\mathbf{x}}, \Sigma_{\mathbf{y}}, \kappa, \beta, \hat{\gamma}_1, \hat{\gamma}_2, a, \mathbf{R}, \mathbf{t}) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|} \cdot c(\kappa, \beta)} \cdot e^{-\frac{1}{2}(\mathbf{y}_{\mathbf{p}} - a\mathbf{R}\mathbf{x}_{\mathbf{p}} - \mathbf{t})^T \Sigma^{-1}(\mathbf{y}_{\mathbf{p}} - a\mathbf{R}\mathbf{x}_{\mathbf{p}} - \mathbf{t}) - \kappa \hat{\mathbf{y}}_{\mathbf{n}}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}} + \beta \left( (\hat{\gamma}_1^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}})^2 - (\hat{\gamma}_2^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}})^2 \right)}. \quad (2)$$

This function finds the  $\mathbf{y} = (\mathbf{y}_{\mathbf{p}}, \hat{\mathbf{y}}_{\mathbf{n}})$  that maximizes the likelihood of a match with  $\mathbf{x} = (\mathbf{x}_{\mathbf{p}}, \hat{\mathbf{x}}_{\mathbf{n}})$ .  $\Sigma = \mathbf{R}\Sigma_{\mathbf{x}}\mathbf{R}^T + \Sigma_{\mathbf{y}}$ , where  $\Sigma_{\mathbf{x}}$  and  $\Sigma_{\mathbf{y}}$  are the covariance matrices representing the measurement noise associated with  $\mathbf{x}$  and  $\mathbf{y}$ ,  $\kappa = \frac{1}{\sigma^2}$  is the concentration parameter of the orientation noise model, where  $\sigma$  is the SD of orientation noise, and  $\beta = e^{\frac{\kappa}{2}}$  controls the anisotropy of the orientation noise model along with  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$ , which are the major and minor axes that define the directions of the elliptical level sets of the Kent distribution on the unit sphere [11, 12].  $\hat{\mathbf{y}}_{\mathbf{n}}$ ,  $\hat{\gamma}_1$ ,  $\hat{\gamma}_2$  are orthogonal and  $e \in [0, 1]$  is the eccentricity of the noise model.

Correspondences are computed by minimizing the negative log likelihood of  $f_{\text{match}}$  [10]. The main difference in the correspondence phases of G-IMLOP and GD-IMLOP is that GD-IMLOP computes matched points on the current *deformed* shape. Outlier rejection is performed after each correspondence phase. Under the assumption of generalized Gaussian noise, the square Mahalanobis distance is approximately distributed as a chi-square distribution with 3 degrees of freedom (DOF) [11]. Therefore, a match is labeled an outlier if this distance exceeds the value of a chi-square inverse cumulative density function (CDF) with 3 DOF at some probability  $p$ . That is, if for any corresponding  $\mathbf{x}$  and  $\mathbf{y}$ ,  $(\mathbf{y}_{\mathbf{p}} - a\mathbf{R}\mathbf{x}_{\mathbf{p}} - \mathbf{t})^T \Sigma^{-1}(\mathbf{y}_{\mathbf{p}} - a\mathbf{R}\mathbf{x}_{\mathbf{p}} - \mathbf{t}) > \text{chi2inv}(p, 3)$ , then that match is an outlier. Here, we set  $p = 0.95$ . Matches that are not rejected as outliers using this test, are evaluated for orientation consistency. Here, a match is an outlier if  $\hat{\mathbf{y}}_{\mathbf{n}}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}} < \cos(\theta_{\text{thresh}})$ , where  $\theta_{\text{thresh}} = 3\sigma_{\text{circ}}$  and  $\sigma_{\text{circ}}$  is the circular SD computed using the mean angular error between all correspondences.

Matches that pass these two tests are inliers and a registration between these points is computed by minimizing the following cost function with respect to the transformation and shape parameters [10]:

$$\mathbf{T} = \underset{[a, \mathbf{R}, \mathbf{t}], \mathbf{s}}{\text{argmin}} \left( \frac{1}{2} \sum_{i=1}^{n_{\text{data}}} \left( (\mathbf{T}_{\text{ssm}}(\mathbf{y}_{\mathbf{p}_i}) - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t})^T \Sigma^{-1}(\mathbf{T}_{\text{ssm}}(\mathbf{y}_{\mathbf{p}_i}) - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t}) \right) + \sum_{i=1}^{n_{\text{data}}} \kappa_i (1 - \hat{\mathbf{y}}_{\mathbf{n}_i}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}_i}) - \sum_{i=1}^{n_{\text{data}}} \beta_i \left( (\hat{\gamma}_{1i}^T \mathbf{R} \hat{\mathbf{y}}_{\mathbf{n}_i})^2 - (\hat{\gamma}_{2i}^T \mathbf{R} \hat{\mathbf{y}}_{\mathbf{n}_i})^2 \right) + \frac{1}{2} \sum_{j=1}^{n_{\text{m}}} \|s_j\|_2^2 \right), \quad (3)$$

where  $n_{\text{data}}$  is the number of inlying data points,  $\mathbf{x}_i$ . This first term in Eq. 3 minimizes the Mahalanobis distance between the positional components of the correspondences,  $\mathbf{x}_{\mathbf{p}_i}$  and  $\mathbf{y}_{\mathbf{p}_i}$ .  $\mathbf{T}_{\text{ssm}}(\cdot)$ , a term introduced in the registration phase, is a transformation,  $\mathbf{T}_{\text{ssm}}(\mathbf{y}_{\mathbf{p}_i}) = \sum_{j=1}^3 \mu_i^{(j)} \mathbf{T}_{\text{ssm}}(\mathbf{v}_i^{(j)})$ , that deforms the matched points,  $\mathbf{y}_i$ , based on the current  $\mathbf{s}$  deforming the model shape [10]. Here,  $\mathbf{T}_{\text{ssm}}(\mathbf{v}_i) = \bar{\mathbf{v}}_i + \sum_{j=1}^{n_m} s_j \mathbf{w}_j^{(i)}$ , and  $\mu_i^{(j)}$  are the 3 barycentric coordinates that describe the position of  $\mathbf{y}_i$  on a triangle on the model shape [10]. The second and third terms minimize the angular error between the orientation components of corresponding points,  $\hat{\mathbf{x}}_{\mathbf{n}_i}$  and  $\hat{\mathbf{y}}_{\mathbf{n}_i}$ , while respecting the anisotropy in the orientation noise. The final term minimizes the shape parameters to find the smallest deformation required to modify the model shape to fit the data points,  $\mathbf{x}_i$  [10].  $\mathbf{s}$  is initialized to 0, meaning the registration begins with the statistically mean shape. The objective function (Eq. 3) is optimized using a nonlinear constrained quasi-Newton based optimizer, where the constraint is used to ensure that  $\mathbf{s}$  are found within  $\pm 3$  SDs, since this interval explains 99.7% of the variation.

Once the algorithm has converged, a final set of tests is performed to assign confidence to the computed registration. For position components, this is similar to the outlier rejection test, except now the sum of the square Mahalanobis distance is compared against the value of a chi-square inverse CDF with  $3n_{\text{data}}$  DOF [11]; i.e., confidence in a registration begins to degrade if

$$E_p = \sum_{i=1}^{n_{\text{data}}} (\mathbf{y}_{\mathbf{p}_i} - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t})^T \Sigma^{-1} (\mathbf{y}_{\mathbf{p}_i} - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t}) > \text{chi2inv}(p, 3n_{\text{data}}). \quad (4)$$

If a registration is successful according to Eq. 4, it is further tested for orientation consistency using a similar chi-square test by approximating the Kent distribution as a 2D wrapped Gaussian [12]. Registration confidence degrades if

$$E_o = \sum_{i=1}^{n_{\text{data}}} \begin{bmatrix} \cos^{-1}(\hat{\mathbf{y}}_{\mathbf{n}_i}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\gamma}_{1_i}^T \mathbf{R} \hat{\mathbf{y}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\gamma}_{2_i}^T \mathbf{R} \hat{\mathbf{y}}_{\mathbf{n}_i}) \end{bmatrix}^T \begin{bmatrix} \kappa_i & 0 & 0 \\ 0 & \kappa_i - 2\beta_i & 0 \\ 0 & 0 & \kappa_i + 2\beta_i \end{bmatrix} \begin{bmatrix} \cos^{-1}(\hat{\mathbf{y}}_{\mathbf{n}_i}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\gamma}_{1_i}^T \mathbf{R} \hat{\mathbf{y}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\gamma}_{2_i}^T \mathbf{R} \hat{\mathbf{y}}_{\mathbf{n}_i}) \end{bmatrix} > \text{chi2inv}(p, 2n_{\text{data}}), \quad (5)$$

since  $\hat{\mathbf{y}}_{\mathbf{n}_i}$  must align with  $\hat{\mathbf{x}}_{\mathbf{n}_i}$ , but remain orthogonal to  $\hat{\gamma}_{1_i}$  and  $\hat{\gamma}_{2_i}$ .  $p$  is set to 0.95 for **very confident** success classification. As  $p$  increases, the confidence in success classification decreases while that in failure classification increases.

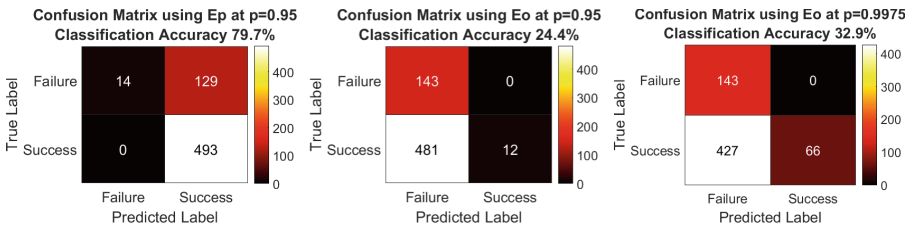
### 3 Experimental Results and Discussion

Two experiments are conducted to evaluate this system: one using simulated data where ground truth is known, and one using in-vivo clinical data where ground truth is not known. Registrations are computed using  $\mathbf{n}_m \in \{0, 10, 20, 30, 40, 50\}$  modes. At 0 modes, this algorithm is essentially G-IMLOP with an additional scale component in the optimization.

### 3.1 Experiment 1: Simulation

In this experiment, we performed a leave-one-out evaluation using shape models of the right nasal cavity extracted from 53 CTs. 3000 points were sampled from the section of the left out mesh that would be visible to an endoscope inserted into the cavity. Anisotropic noise with SD  $0.5 \times 0.5 \times 0.75 \text{ mm}^3$  and  $10^\circ$  with  $e = 0.5$  was added to the position and orientation components of the points, respectively, since this produced realistic point clouds compared to in-vivo data with higher uncertainty in the z-direction. A rotation, translation and scale are applied to these points in the intervals  $[0, 10] \text{ mm}$ ,  $[0, 10]^\circ$  and  $[0.95, 1.05]$ , respectively. 2 offsets are sampled for each left out shape. GD-IMLOP makes slightly more generous noise assumptions with SDs  $1 \times 1 \times 2 \text{ mm}^3$  and  $30^\circ$  ( $e = 0.5$ ) for position and orientation noise, respectively, and restricts scale optimization to within  $[0.9, 1.1]$ . A registration is considered successful if the total registration error (tRE), computed using the Hausdorff distance (HD) between the left-out shape and the estimated shape transformed to the frame of the registered points, is below 1 mm. The success or failure of the registrations is compared to the outcome predicted by the algorithm. Further, the HD between the left-out and estimated shapes in the same frame is used to evaluate errors in reconstruction.

Results over all modes, using  $p = 0.95$ , show that  $E_p$  is less strict than  $E_o$  (Fig. 1), meaning that although  $E_p$  identifies all successful registrations correctly, it also allows many unsuccessful registrations to be labeled successful.  $E_o$ , on the other hand, correctly classifies fewer successful registrations, but does not label any failed registrations as successful. Therefore, registrations with  $E_p < \text{chi2inv}(0.95, 3n_{\text{data}})$  and  $E_o < \text{chi2inv}(0.95, 2n_{\text{data}})$  can be **very confidently** classified as successful. The average tRE produced by registrations in this category over all modes was  $0.34 (\pm 0.03) \text{ mm}$ . At  $p = 0.9975$ , more successful registrations were correctly identified (Fig. 1, right). These registrations can be **confidently** classified as successful with mean tRE increasing to  $0.62 (\pm 0.03) \text{ mm}$ . Errors in correct classification creep in with  $p = 0.9999$ , where 3 out of 124 registrations are incorrectly labeled successful. These registrations can be **somewhat confidently** classified as successful with mean tRE increasing



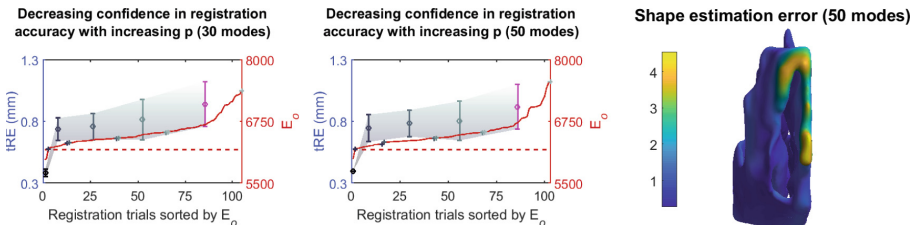
**Fig. 1.** Left: using only  $E_p$ , all successful registration pass the chi-square inverse test at  $p = 0.95$ . However, many failed registrations also pass this test. Using  $p = 0.9975$  produces the same result. Middle: on the other hand, using only  $E_o$ , no failed registrations pass the chi-square inverse test at  $p = 0.95$ , but very few successful registrations pass the test. Right: using  $p = 0.9975$ , more successful registrations pass the test.

slightly to  $0.78 (\pm 0.04)$  mm. Increasing  $p$  to 0.999999 further decreases classification accuracy. 10 out of 121 registrations in this category are incorrectly classified as successful with mean tRE increasing to  $0.8 (\pm 0.05)$  mm. These registrations can, therefore, be classified as successful with **low confidence**. The mean tRE for the remaining registrations increases to over 1 mm at  $1.31 (\pm 0.85)$  mm, with no registration passing the  $E_p$  threshold except for registrations using 0 modes. Of these, however, 0 are correctly classified as successful. Therefore, although about half of all registrations in this category are successful, there can be **no confidence** in their correct classification. Figure 2 (left and middle) shows the distribution of tREs in these categories for registrations using 30 and 50 modes.

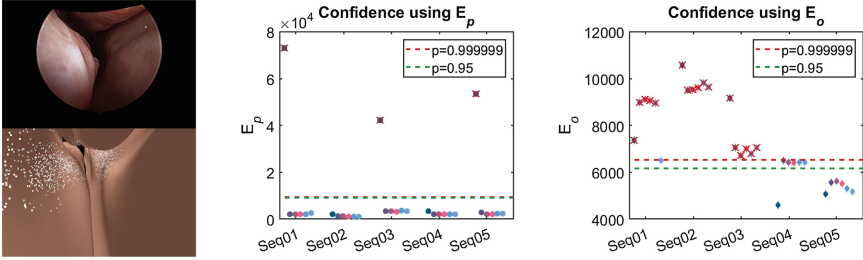
GD-IMLOP can, therefore, compute successful registrations between a statistically mean right nasal cavity mesh and points sampled only from part of the left-out meshes, and reliably assign confidence to these registrations. Further, GD-IMLOP can accurately estimate the region of the nasal cavity where points are sampled from, while errors gradually deteriorate away from this region, e.g., towards the front of the septum since points are not sampled from this region (Fig. 2, right). Overall, the mean shape estimation error was 0.77 mm.

### 3.2 Experiment 2: In-Vivo

For the in-vivo experiment, we collected anonymized endoscopic videos of the nasal cavity from consenting patients under an IRB approved study. Dense point clouds were produced from single frames of these videos using a modified version of the learning-based photometric reconstruction technique [13] that uses registered structure from motion (SfM) points to train a neural network to predict dense depth maps. Point clouds from different nearby frames in a sequence were aligned using the relative camera motion from SfM. Small misalignments due to errors in depth estimation were corrected using G-IMLOP with scale to produce a dense reconstruction spanning a large area of the nasal passage. GD-IMLOP is executed with 3000 points sampled from this dense reconstruction assuming noise with SDs  $1 \times 1 \times 2 \text{ mm}^3$  and  $30^\circ$  ( $e = 0.5$ ) for position and orientation



**Fig. 2.** Left and middle: mean tRE and standard deviation increase as  $E_o$  increases. The dotted red line corresponds to  $\chi^2_{inv}(0.95, 2n_{data})$ , below which registrations are classified **very confidently** as successful. Beyond this threshold, confidence gradually degrades. The pink bar indicates that none of these registrations passed the  $E_p$  test. Right: average error at each vertex computed over all left-out trials using 50 modes.



**Fig. 3.** Left: visualization of the final registration and reconstruction for Seq01 using 50 modes. Middle and right:  $E_p$  and  $E_o$  for all registrations, respectively, plotted for each sequence. Per sequence, from left to right, the plot points indicate scores achieved using 0-50 modes at increments of 10. Crossed out plot points indicate rejected registrations.

data, respectively, and with scale and shape parameter optimization restricted to within  $[0.7, 1.3]$  and  $\pm 1$  SD, respectively. We assign confidence to the registrations based on the tests explained in Sect. 2 and validated in Sect. 3.1.

All registrations run with 0 modes terminated at the maximum iteration threshold of 100, while those run using modes converged at an average 10.36 iterations in 26.03 s. Figure 3 shows registrations using increasing modes from left to right for each sequence plotted against  $E_p$  (middle) and  $E_o$  (right). All deformable registration results pass the  $E_p$  test as they fall below the  $p = 0.95$  threshold (Fig. 3, middle) using the chi-square inverse test. However, several of these fail the  $E_o$  test (Fig. 3, right). Deformable registrations on sequence 01 using 50 modes and on sequence 04 for all except 30 modes pass this test with **low confidence**. Using 30 modes, the registration on sequence 04 passes **somewhat confidently**. The rigid registration on sequence 04 (the only rigid registration to pass both  $E_p$  and  $E_o$ ) and all deformable registrations on sequence 05 pass this test **very confidently**. Although, the rigid registration on sequence 05 passes this test very confidently,  $E_p$  already labels it a failed registration. Successful registrations produced a mean residual error of 0.78 ( $\pm 0.07$ ) mm. Visualizations of successful registrations also show accurate alignment (Fig. 3, left).

## 4 Conclusion

We show that GD-IMLOP is able to produce submillimeter registrations in both simulation and in-vivo experiments, and assign confidence to these registrations. Further, it can accurately predict the anatomy where video data is available. In the future, we hope to learn statistics from thousands of CTs to better cover the range of anatomical variations. Additional features like contours can also be used to further improve registration and to add an additional test to evaluate the success of the registration based on contour alignment. Using improved statistics and reconstructions from video along with confidence assignment, this approach can be extended for use in place of CTs during endoscopic procedures.

**Acknowledgment.** This work was funded by NIH R01-EB015530, NSF Graduate Research Fellowship Program, an Intuitive Surgical, Inc. fellowship, and JHU internal funds.

## References

1. Mirota, D.J., Ishii, M., Hager, G.D.: Vision-based navigation in image-guided interventions. *Ann. Rev. Biomed. Eng.* **13**(1), 297–319 (2011)
2. Azagury, D.E., et al.: Real-time computed tomography-based augmented reality for natural orifice transluminal endoscopic surgery navigation. *Brit. J. Surg.* **99**(9), 1246–1253 (2012)
3. Beichel, R.R., et al.: Data from QIN-HEADNECK. The Cancer Imaging Archive (2015)
4. Bosch, W.R., Straube, W.L., Matthews, J.W., Purdy, J.A.: Data from head-neck\_cetuximab. The Cancer Imaging Archive (2015)
5. Clark, K., et al.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* **26**(6), 1045–1057 (2013)
6. Fedorov, A., et al.: DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. *PeerJ* **4**, e2057 (2016)
7. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54**(3), 2033–2044 (2011)
8. Sinha, A., Reiter, A., Leonard, S., Ishii, M., Hager, G.D., Taylor, R.H.: Simultaneous segmentation and correspondence improvement using statistical modes. In: *Proceedings of SPIE*, vol. 10133, pp. 101 331B–101 331B–8 (2017)
9. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models their training and application. *Comp. Vis. Image Underst.* **61**, 38–59 (1995)
10. Sinha, A., et al.: The deformable most-likely-point paradigm. *Med. Image Anal.* (Submitted)
11. Billings, S.D., Taylor, R.H.: Generalized iterative most likely oriented-point (G-IMLOP) registration. *Int. J. Comput. Assist. Radiol. Surg.* **10**(8), 1213–1226 (2015)
12. Mardia, K.V., Jupp, P.E.: *Directional statistics*. Wiley Series in Probability and Statistics, pp. 1–432. Wiley, Hoboken (2008)
13. Reiter, A., Leonard, S., Sinha, A., Ishii, M., Taylor, R.H., Hager, G.D.: Endoscopic-CT: learning-based photometric reconstruction for endoscopic sinus surgery. In: *Proceedings of SPIE*, vol. 9784, pp. 978 418–978 418–6 (2016)