



Endoscopic navigation in the clinic: registration in the absence of preoperative imaging

Ayushi Sinha¹ · Masaru Ishii² · Gregory D. Hager¹ · Russell H. Taylor¹

Received: 31 January 2019 / Accepted: 22 May 2019 / Published online: 31 May 2019
© CARS 2019

Abstract

Purpose Clinical examinations that involve endoscopic exploration of the nasal cavity and sinuses often do not have a reference preoperative image, like a computed tomography (CT) scan, to provide structural context to the clinician. The aim of this work is to provide structural context during clinical exploration without requiring additional CT acquisition.

Methods We present a method for registration during clinical endoscopy in the absence of CT scans by making use of shape statistics from past CT scans. Using a deformable registration algorithm that uses these shape statistics along with dense point clouds from video, we simultaneously achieve two goals: (1) register the statistically mean shape of the target anatomy with the video point cloud, and (2) estimate patient shape by deforming the mean shape to fit the video point cloud. Finally, we use statistical tests to assign confidence to the computed registration.

Results We are able to achieve submillimeter errors in registrations and patient shape reconstructions using simulated data. We establish and evaluate the confidence criteria for our registrations using simulated data. Finally, we evaluate our registration method on in vivo clinical data and assign confidence to these registrations using the criteria established in simulation. All registrations that are not rejected by our criteria produce submillimeter residual errors.

Conclusion Our deformable registration method can produce submillimeter registrations and reconstructions as well as statistical scores that can be used to assign confidence to the registrations.

Keywords Statistical shape models · Navigation for clinical endoscopy · Deformable registration · Shape estimation · Registration confidence

Introduction

Over 10% [14] of the population in the USA suffers from chronic rhinosinusitis (CRS). CRS is the prolonged inflammation and swelling in the nasal passages and sinuses which can cause interference with drainage, congestion, facial pain, etc [19]. Since CRS is a chronic disease, patients require regular clinic visits and care is provided over a lifetime. If medical treatment of CRS is unable to improve patient quality of life, surgery may be needed. Minimally invasive surgical treatment of CRS uses an endoscope to visualize the nasal

cavities and a navigation system accompanied by a reference computed tomography (CT) image of the patient to guide the surgeon. However, diagnosis and clinical examination of the nasal cavities are generally not accompanied by a CT scan since CT image acquisition exposes patients to high doses of ionizing radiation. Therefore, unless necessary, as in the case of surgery, CT image acquisition is avoided. This leaves clinicians performing examinations to depend entirely on the endoscopic camera, memory, and experience to identify both the anomaly and its optimal treatment. However, endoscopic cameras have limited field of view and certain regions of the sinuses exhibit pseudo-stochastic growth patterns that are too complicated to memorize [25] and can disorient even experienced clinicians.

In order to reduce clinicians' dependence on experience or memory and to provide additional context during clinical endoscopy, we present a method that enables navigation without the need for accompanying patient CT scan or other preoperative imaging and associates confidence estimates

✉ Ayushi Sinha
sinha@jhu.edu

¹ Laboratory for Computational and Sensing Robotics, The Johns Hopkins University, Baltimore, MD 21218, USA

² Department of Otolaryngology - Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, MD 21205, USA

with the navigation being provided. Further, our system does not introduce any additional devices than those already used in clinic. Therefore, the clinician is not responsible for anything in addition to the endoscope. Most navigation systems that have already been developed are intended for surgical use [10,13,27]. Surgical navigation through the nasal cavities is almost always accompanied by preoperative CT scans, which have high contrast between air, bone, and soft tissue. Navigation gives surgeons a better understanding of the location of the endoscope relative to anatomy not visible in endoscopic video. It also informs surgeons about their proximity to surrounding bones and critical structures like the brain, eyes, optic nerves, carotid arteries, etc., the thickness of surrounding bones, and other useful information. This enables surgeons to make more informed decisions during surgery and prevent harm to critical structures nearby.

The primary difference between prior methods for navigation and the method presented in this paper is the absence of patient specific CT scans. In order to compensate for this absence, we use past CT scans from population to build statistical shape models (SSMs) of relevant anatomical structures [17]. Statistically derived shapes, e.g., the mean shape, are then deformably registered to dense reconstructions of anatomy visible in endoscopic video, and statistical confidence measures are automatically assigned to the registrations [40]. The registration accomplishes two tasks simultaneously. First, it aligns the endoscopic video to the mean shape, giving the clinician more information about where surrounding structures may be. Second, it deforms the mean shape to fit the structure obtained from video and, in effect, *estimates the patient CT* [41]. The confidence measure further informs the clinician about when and how much the navigation system can be trusted and also allows the navigation system to attempt to improve itself if it has low confidence in its current registration estimate.

We perform experiments with simulated and in vivo data to evaluate our framework. We first establish that our framework can compute submillimeter registrations and reconstructions using simulated data. We also establish and evaluate our confidence assignment policy using simulated data. Finally, we evaluate our framework on in vivo clinical data and use the confidence criteria established in simulation to assign confidence to these registrations.

Prior work

There has been a lot of prior work in point-to-point and point-to-surface registration, and the Iterative Closest Point (ICP) algorithm [5,15] has emerged as one of the standard registration algorithms. ICP is a two-step algorithm that iterates between finding the closest point correspondences between point sets and finding the rigid transformation that best aligns

these correspondences. These two steps are repeated until convergence. Many different methods [35] have built upon this simple and elegant formulation to handle sparse and noisy data by adding outlier detection [12,33] and by reframing the problem in a probabilistic setting to incorporate noise estimates into the objective function [9,21,36]. Our method is formulated in a probabilistic framework similar to that of [9].

Several methods have also explored adding additional features to the registration process. For instance, normals have been used to disambiguate between point sets that are oriented differently [7,8,22,30,31], occluding contours have been used to match object edges [10], and color and texture information have been used to disambiguate between points extracted from different objects [20,24]. The method evaluated in this paper is a deformable extension of the registration method by [8] which, like ICP, is an iterative two-step algorithm and, unlike ICP, incorporates anisotropic noise in both position and orientation components in the match step as well as the align step.

Many deformable registration techniques have also been explored using, among others, stiffness parameters that are used to constrain deformations between correspondences [1] and displacement fields that optimize soft matches between points sets [32]. Our deformable algorithm extends the ideas of [23] by incorporating SSMs into the objective function. These SSMs describe the variance in a set of shapes with correspondences [17] and are used to drive the deformation in our method.

Methods

We build SSMs of anatomical structures by automatically segmenting 53 publicly available head CTs [4,11,16,18]. 3D meshes extracted from manually created labels in a template CT image [38] are transferred to the 53 CTs using deformation fields produced by an intensity-based CT-CT registration algorithm [2]. With some improvements to these initial segmentations [39], we obtain reliably segmented structures in all CTs along with reliable correspondences. These correspondences enable us to build SSMs of the segmented structures using established methods like principal component analysis (PCA) [17]:

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

where \mathbf{V}_j is the stacked vector of vertices, $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_{n_v}]^T$, for the j th mesh, $\bar{\mathbf{V}}$ is the mean shape computed by

averaging the n_v corresponding vertices over n_s shapes, $\bar{\mathbf{V}} = \frac{1}{n_s} \sum_{j=1}^{n_s} \mathbf{V}_j$, and Σ_{SSM} is the shape covariance matrix. An eigen decomposition of Σ_{SSM} produces the principal modes of variation, \mathbf{m} , and the mode weights, λ , which represent the amount of variation along the corresponding \mathbf{m} . 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 λ : $\hat{\mathbf{V}}^* = \bar{\mathbf{V}} + \sum_{j=1}^{n_m} s_j \mathbf{w}_j$, where $\hat{\mathbf{V}}^*$ is the estimated \mathbf{V}^* , $1 \leq n_m < n_s$ 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 modes, $s_j = \mathbf{m}_j^T (\mathbf{V}^* - \bar{\mathbf{V}}) / \sqrt{\lambda_j}$.

These shape parameters, $\mathbf{s} = \{s_j\}$, can be incorporated into probabilistic models of registration to allow for optimization over \mathbf{s} in addition to other registration parameters [41]. In particular, we evaluate the deformable extension of the generalized iterative most likely oriented-point (G-IMLOP) algorithm, an iterative rigid registration algorithm [8]. 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 [8]. The likelihood function of G-IMLOP finds the $\mathbf{y} = (\mathbf{y}_p, \hat{\mathbf{y}}_n)$ on a target model shape, ψ , that maximizes the likelihood of a match with $\mathbf{x} = (\mathbf{x}_p, \hat{\mathbf{x}}_n)$. 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 [8]:

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

where $\Sigma = \mathbf{R}\Sigma_x\mathbf{R}^T + \Sigma_y$, Σ_x and Σ_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 σ is the SD of orientation noise, and $\beta = e^{\frac{\kappa}{2}}$ controls the anisotropy of the orientation noise model along with $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$, which are the major and minor axes that define the directions of the elliptical level sets of the Kent distribution on the unit sphere [8,29]. $\hat{\mathbf{y}}_n$, $\hat{\mathbf{y}}_1$, $\hat{\mathbf{y}}_2$ are orthogonal, and $e \in [0, 1]$ is the eccentricity of the noise model.

At each iteration, correspondences are computed by minimizing the negative log likelihood of f_{match} [41] for each current inlier. 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 [41]. Using the current inliers, both the position and orientation noise parameters are updated using the residual match errors. The covariance of position noise is updated by the mean square distance between the inlying matches divided by the spatial dimensionality [37]. The concentration parameter, κ , of orientation noise is updated using an approximation to its maximum likelihood estimate [3,29]:

$$\kappa \approx \frac{\bar{R}(3 - \bar{R})}{1 - \bar{R}^2}. \quad (2)$$

\bar{R} evaluates the weighted sum of the position and orientation terms used to estimate κ , weighted by a user-defined parameter, w [6]:

$$\bar{R} = \frac{1-w}{n} \sum_{i=1}^{n_{\text{data}}} \hat{\mathbf{y}}_{n_i}^T \mathbf{R} \hat{\mathbf{x}}_{n_i} + \frac{w}{\alpha} \sum_{i=1}^{n_{\text{data}}} \tilde{\mathbf{y}}_{p_i}^T \mathbf{R} \tilde{\mathbf{x}}_{p_i},$$

where $\alpha = \sum_{i=1}^{n_{\text{data}}} \|\tilde{\mathbf{y}}_{p_i}\| \|\mathbf{R} \tilde{\mathbf{x}}_{p_i}\|$,

$$\tilde{\mathbf{x}}_{p_i} = \mathbf{x}_{p_i} - \frac{1}{n_{\text{data}}} \sum_{i=1}^{n_{\text{data}}} \mathbf{x}_{p_i} \quad \text{and} \quad \tilde{\mathbf{y}}_{p_i} = \mathbf{y}_{p_i} - \frac{1}{n_{\text{data}}} \sum_{i=1}^{n_{\text{data}}} \mathbf{y}_{p_i}. \quad (3)$$

n_{data} is the number of inlying data points, \mathbf{x}_i , and position and orientation terms are equally weighted using $w = 0.5$. Estimating κ based on orientation match errors alone can progressively overestimate κ since it is often possible to find a nearly perfect orientation match for closed shape models [6]. Including position data prevents κ from growing too large.

Finally, inliers are updated based on the current set of correspondences. Under the assumption of generalized Gaussian noise, the square Mahalanobis distance is approximately distributed as a Chi-square (χ^2) distribution with 3 degrees of freedom (DOF) [8]. Therefore, a match is labeled an outlier if this distance exceeds the value of a χ^2 inverse cumulative density function (CDF) with 3 DOF at some probability p . That is, for any corresponding \mathbf{x} and \mathbf{y} , if

$$(\mathbf{y}_p - a\mathbf{R}\mathbf{x}_p - \mathbf{t})^T \Sigma^{-1}(\mathbf{y}_p - a\mathbf{R}\mathbf{x}_p - \mathbf{t}) > \text{chi2inv}(p, 3), \quad (4)$$

then that match is an outlier. We set $p = 0.95$ [8]. Matches that are not rejected as outliers using this test are evaluated for orientation consistency. Here, a match is an outlier if $\hat{\mathbf{y}}_n^T \mathbf{R} \hat{\mathbf{x}}_n < \cos(\theta_{\text{thresh}})$, where $\theta_{\text{thresh}} = 3\sigma_{\text{circ}}$ and σ_{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 inliers is computed by minimizing the following cost function with respect to the transformation and shape parameters [41]:

$$\begin{aligned}
T = \operatorname{argmin}_{[a, \mathbf{R}, \mathbf{t}], \mathbf{s}} & \left(\frac{1}{2} \sum_{i=1}^{n_{\text{data}}} \left((T_{\text{ssm}}(\mathbf{y}_{\mathbf{p}_i}) - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t})^T \Sigma^{-1} (T_{\text{ssm}}(\mathbf{y}_{\mathbf{p}_i}) \right. \right. \\
& \left. \left. - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t}) \right) \right. \\
& + \sum_{i=1}^{n_{\text{data}}} \kappa_i (1 - \hat{\mathbf{y}}_{\mathbf{n}_i}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}_i}) - \sum_{i=1}^{n_{\text{data}}} \beta_i \left((\hat{\mathbf{y}}_{1i}^T \mathbf{R}^T \hat{\mathbf{y}}_{\mathbf{n}_i})^2 \right. \\
& \left. \left. - (\hat{\mathbf{y}}_{2i}^T \mathbf{R}^T \hat{\mathbf{y}}_{\mathbf{n}_i})^2 \right) + \frac{1}{2} \sum_{j=1}^{n_m} \|\mathbf{s}_j\|_2^2 \right). \quad (5)
\end{aligned}$$

The first term in Eq. 5 minimizes the Mahalanobis distance between the positional components of correspondences, $\mathbf{x}_{\mathbf{p}_i}$ and $\mathbf{y}_{\mathbf{p}_i}$. $T_{\text{ssm}}(\cdot)$ is a transformation, $T_{\text{ssm}}(\mathbf{y}_{\mathbf{p}_i}) = \sum_{j=1}^3 \mu_i^{(j)} T_{\text{ssm}}(\mathbf{v}_i^{(j)})$, that deforms the matched points, \mathbf{y}_i , based on the current \mathbf{s} deforming ψ , $T_{\text{ssm}}(\mathbf{v}_i) = \bar{\mathbf{v}}_i + \sum_{j=1}^{n_m} \mathbf{s}_j \mathbf{w}_j^{(i)}$ [41]. $\mu_i^{(j)}$ are the 3 barycentric coordinates that describe the position of \mathbf{y}_i on a triangle on ψ [41]. 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 maintaining the anisotropy in the orientation noise. The final term minimizes the shape parameters to find the smallest deformation required to modify ψ to fit the data points, \mathbf{x}_i [41]. \mathbf{s} is initialized to 0, meaning the registration begins with the statistically mean shape. This objective function (Eq. 5) is optimized using a constrained nonlinear quasi-Newton-based optimizer, where the constraint ensures that each \mathbf{s} is within ± 3 SDs, since this interval explains 99.7% of the population variance.

Once the algorithm has converged, a final set of tests is performed to assess the confidence in the computed registration. For position components, this is similar to the test for outlier rejection, except now the sum of the square Mahalanobis distance over all sample points is compared against the value of a χ^2 inverse CDF with $3n_{\text{data}}$ DOF [8]; i.e., confidence in a registration degrades if

$$\begin{aligned}
E_p = \sum_{i=1}^{n_{\text{data}}} (\mathbf{y}_{\mathbf{p}_i} - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t})^T \Sigma^{-1} (\mathbf{y}_{\mathbf{p}_i} - a\mathbf{R}\mathbf{x}_{\mathbf{p}_i} - \mathbf{t}) \\
> \chi^2_{\text{inv}}(p, 3n_{\text{data}}) \quad (6)
\end{aligned}$$

for increasing p . If a registration is successful according to Eq. 6, it is further tested for orientation consistency using a similar χ^2 test by approximating the Kent distribution as a 2D wrapped Gaussian [29]. Registration confidence degrades if

$$\begin{aligned}
E_o = \sum_{i=1}^{n_{\text{data}}} \begin{bmatrix} \cos^{-1}(\hat{\mathbf{y}}_{\mathbf{n}_i}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\mathbf{y}}_{1i}^T \mathbf{R}^T \hat{\mathbf{y}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\mathbf{y}}_{2i}^T \mathbf{R}^T \hat{\mathbf{y}}_{\mathbf{n}_i}) \end{bmatrix}^T \begin{bmatrix} \kappa_i & 0 & 0 \\ 0 & \kappa_i - 2\beta_i & 0 \\ 0 & 0 & \kappa_i + 2\beta_i \end{bmatrix} \\
\times \begin{bmatrix} \cos^{-1}(\hat{\mathbf{y}}_{\mathbf{n}_i}^T \mathbf{R} \hat{\mathbf{x}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\mathbf{y}}_{1i}^T \mathbf{R}^T \hat{\mathbf{y}}_{\mathbf{n}_i}) \\ \sin^{-1}(\hat{\mathbf{y}}_{2i}^T \mathbf{R}^T \hat{\mathbf{y}}_{\mathbf{n}_i}) \end{bmatrix} > \chi^2_{\text{inv}}(p, 2n_{\text{data}}) \quad (7)
\end{aligned}$$

for increasing p , since $\hat{\mathbf{y}}_{\mathbf{n}_i}$ must align with $\hat{\mathbf{x}}_{\mathbf{n}_i}$, but remain orthogonal to $\hat{\mathbf{y}}_{1i}$ and $\hat{\mathbf{y}}_{2i}$. In [40], p was empirically set to 0.95 in both E_p and E_o tests for *very confident* success assignment based on simulated experiments. Confidence in registration success decreased for higher values of p . However, these thresholds were not evaluated in other simulated experiments, but directly used to assign confidence to registrations computed using in vivo clinical data where ground truth was not available for validation. In this paper, p will be chosen empirically in a simulated experimental setup and then evaluated in increasingly realistic or difficult simulated experiments and finally on in vivo clinical data.

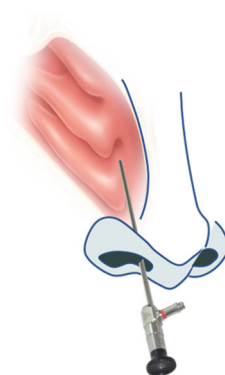
Experimental results and discussion

Experiments are conducted using two types of data in order to evaluate this method. The first is simulated data where ground truth is known, while the second is in vivo clinical data where ground truth is not known. Registrations using GD-IMLOP are computed using different number of modes. At 0 modes, this algorithm is essentially G-IMLOP with an additional scale component in the optimization. In all simulated experiments, we perform leave-one-out validation using shape models of the right nasal cavity (Fig. 1) extracted from 53 CTs.

For the purpose of this study, 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 coordinate frame of the registered points, is below 1 mm. The success or failure of the registrations is compared to the outcome predicted by our algorithm. This comparison will yield different results if the ground truth threshold of 1 mm is changed. We also compute the total shape error (tSE) as the HD between the left-out and estimated shapes in the same coordinate frame to evaluate errors in reconstruction.

Before describing each experimental setup, we would like to establish some expectations. Figure 2 shows the probabil-

Fig. 1 Endoscope inserted into the right nasal cavity



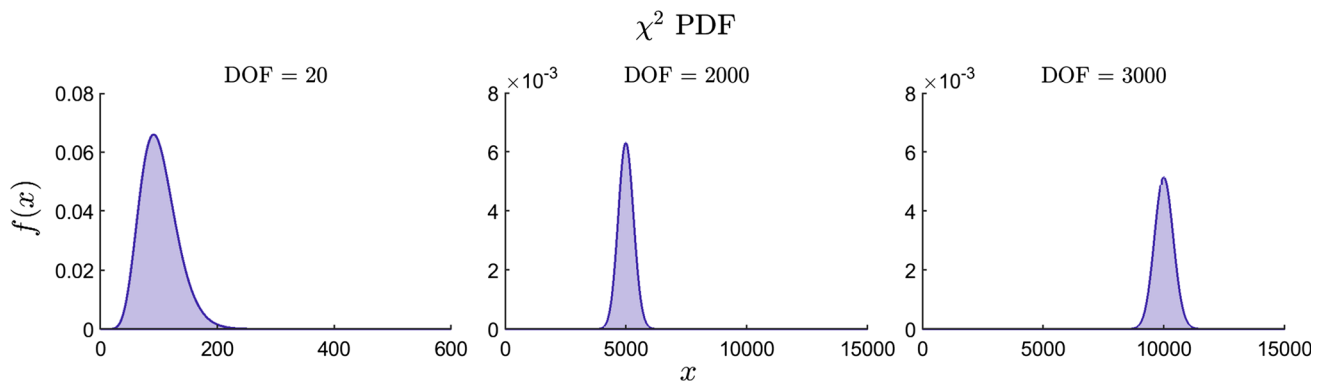


Fig. 2 Although the PDFs of χ^2 -distributions with smaller DOF are not very heavy-tailed (left), they become increasingly heavy-tailed as the DOF increases (note the y-axis values for the middle and right figures).

For 1000 sample points, E_o would be evaluated against a χ^2 -distribution with 2000 DOF (middle) and E_p against that with 3000 DOF (right), both of which have heavy-tailed PDFs

ity density functions (PDFs) of χ^2 -distributions with large DOF, such as we expect to see in our experiments. These PDFs are very heavy-tailed. Therefore, we expect the values of p to be large, i.e., closer to 1, since p is the value of the χ^2 CDF at E_p or E_o with the appropriate DOF. That is, p tells us the probability that the χ^2 -distribution with $3n_{\text{data}}$ or $2n_{\text{data}}$ DOF will take a value less than or equal to E_p or E_o , respectively. This also implies that we want the lower values of p to allow few registrations that our algorithm is extremely confident about into the class of successful registrations. As p increases, the number of registrations labeled successful will increase and so will the likelihood of introducing errors. Therefore, confidence in the computed registration will decrease as p increases with registrations rejected as failed at $p = 1$.

Establishing p

The first experiment using simulated data is used to establish the values of p that will be associated with varying levels of confidence. In this experiment, we want to introduce as few ambiguities as possible so that we can confidently establish the values of p . We sampled 1000 points uniformly from the left-out nasal cavity mesh. Since the 53 CT dataset had an isotropic resolution of $1 \times 1 \times 1 \text{ mm}^3$, isotropic noise with $1 \times 1 \times 1 \text{ mm}^3$ SD was added to the position component of the sampled points. Additionally, anisotropic noise with 20° SD and $e = 0.5$ was added to the orientation component. A rotation and translation sampled from the intervals $[0, 9]^\circ$ and $[0, 15] \text{ mm}$, respectively, were applied to these points. Two registrations per left-out shape were computed using GD-IMLOP assuming the same noise as in the generated sample points. $n_{\text{m}} \in \{0, 5, 10, \dots, 50\}$ modes are used in this experiment, but in all other experiments, we use $n_{\text{m}} \in \{0, 10, 20, \dots, 50\}$ modes.

GD-IMLOP produced submillimeter registrations and shape estimations as long as $n_{\text{m}} > 0$ modes were used (Figs. 3a, 6). The mean TRE is highest at 0 modes since the mean shape cannot deform to fit the sample points, leading to suboptimal matches and, therefore, registration. We do not compute a tSE at 0 modes since patient shape is not estimated at 0 modes. As the number of modes increases, the mean tRE and tSE show steady decline. Now, we want to evaluate whether our confidence criteria can accurately identify these successful registrations. Success or failure labels are assigned to each registration if they pass both the E_p and E_o tests at different values of p . These labels are compared against ground truth, which is established as described in the “Experimental results and discussion” section. The ROC (Fig. 3b) and precision-recall (Fig. 3c) curves show, as we expected, that few registrations pass our tests for lower values of p . As p increases, more registrations pass these tests. However, no failed registrations are labeled successful until p increases to 0.9999999. Finally, we observe that the tRE and tSE for registrations classified as successful at these values of p increase as p increases (Fig. 3d). Therefore, we can establish that our confidence in a successful registration decreases as p increases, culminating in a rejection at $p = 1$.

Simulated data without outliers

In the following experiments, we will try to simulate more realistic scenarios and use the values of p established in the previous section to assign confidence to the registrations computed. These assignments will be evaluated against ground truth. In this experiment, 3000 points were sampled from the section of the left-out mesh that would be visible to an endoscope inserted into the nasal cavity (Fig. 6a). Anisotropic noise with SDs $0.5 \times 0.5 \times 0.75 \text{ mm}^3$ and 10° with $e = 0.5$ was added to the position and orientation components of the points, respectively, since this produced

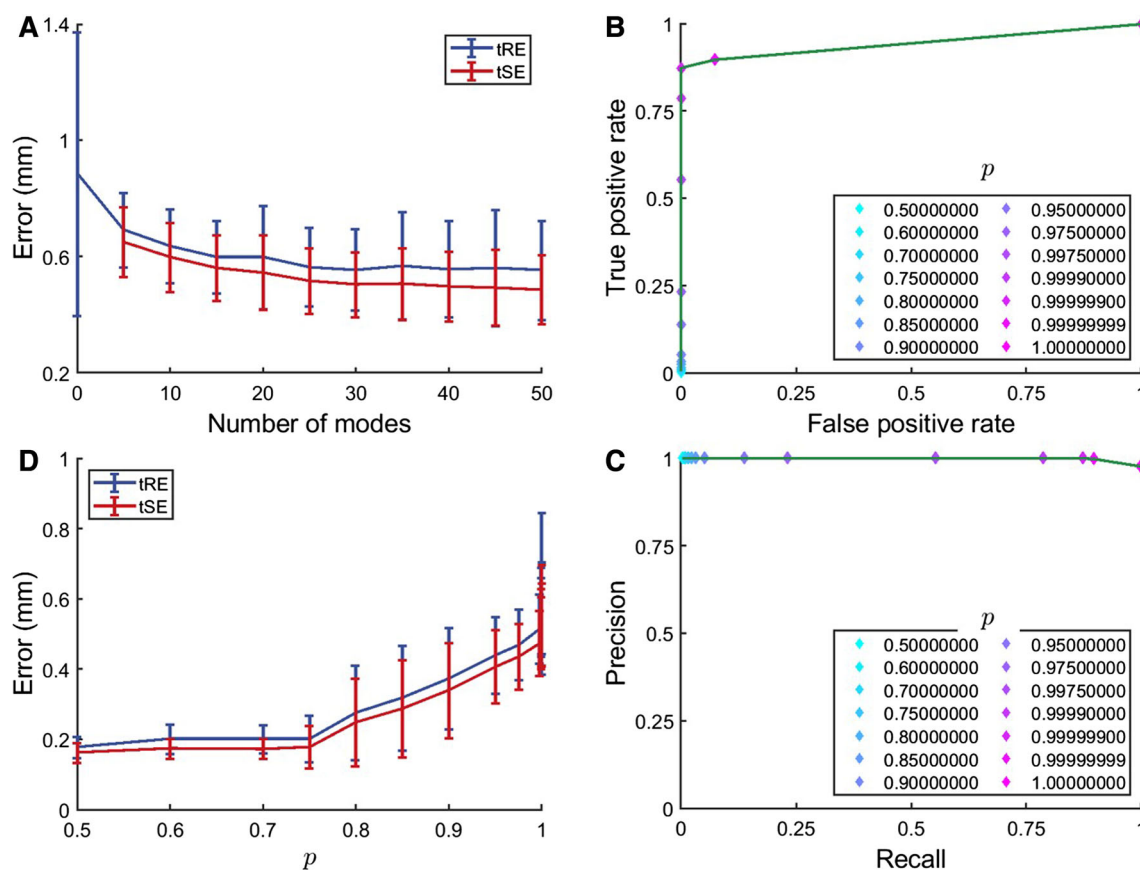


Fig. 3 Establishing p : submillimeter mean tRE and tSE for $n_m > 0$ (a), ROC (b) and precision-recall (c) curves for increasing p , and a steady increase in errors as confidence in registration degrades with increasing p (d)

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 $[0, 10]$ mm, $[0, 10]^\circ$, and $[0.95, 1.05]$, respectively. No outliers are added in this experiment. However, since in a real scenario, the exact noise in the data is not known, GD-IMLOP makes slightly more generous noise assumptions with SDs $1 \times 1 \times 2 \text{ mm}^3$ and 30° ($e = 0.5$) for position and orientation noise, respectively. Scale optimization is restricted to within $[0.9, 1.1]$.

The mean tRE and tSE produced by GD-IMLOP are again below 1 mm for $n_m > 0$, albeit higher than those produced in the previous experiment, as is expected due to the added unknowns (Fig. 4a). tSE is lower in regions closer to where points are sampled from (Fig. 6c) and grows as we move away from these regions. These errors also affect our confidence assignment with failed registrations classified as successful by our criteria at a slightly lower value of p than in the previous experiment at 0.9999 (Fig. 4b). This is also reflected by the lower precision values (Fig. 4c) at corresponding p values compared to the previous experiment (Fig. 3c). However, the average tRE of the three false positives at $p = 0.9999$ is $1.03 (\pm 0.05)$ mm which agrees with the somewhat low confidence implied by $p = 0.9999$ [40]. Finally, there is a

gradual increase in tRE and tSE for registrations classified as successful with increasing values of p , culminating in a drastic increase at $p = 1$ indicated by the large standard deviation (Fig. 4d), meaning that most failed registrations are rejected.

Simulated data with 10% outliers

This experiment is set up similarly to the experiment in the previous section, except that 10% of the simulated samples are randomly chosen to be outliers. Outliers are added by offsetting some of the generated samples' positions and orientations randomly in the interval $[5, 10]$ mm and $[5, 10]^\circ$, respectively.

The main similarity that results from this experiment have with results from the previous two is in the submillimeter mean tRE and tSE for $n_m > 0$ (Fig. 5a). Again, the tSE is lower within the nasal cavity since points were sampled from this region and grows larger away from this region (Fig. 6d). However, the success classification in this experiment is different from our expectations. Even for small values of p , the number of successful registrations classified as successful is high compared to the previous two experiments, as indicated by the high true positive or recall rate in Fig. 5b, c. Further,

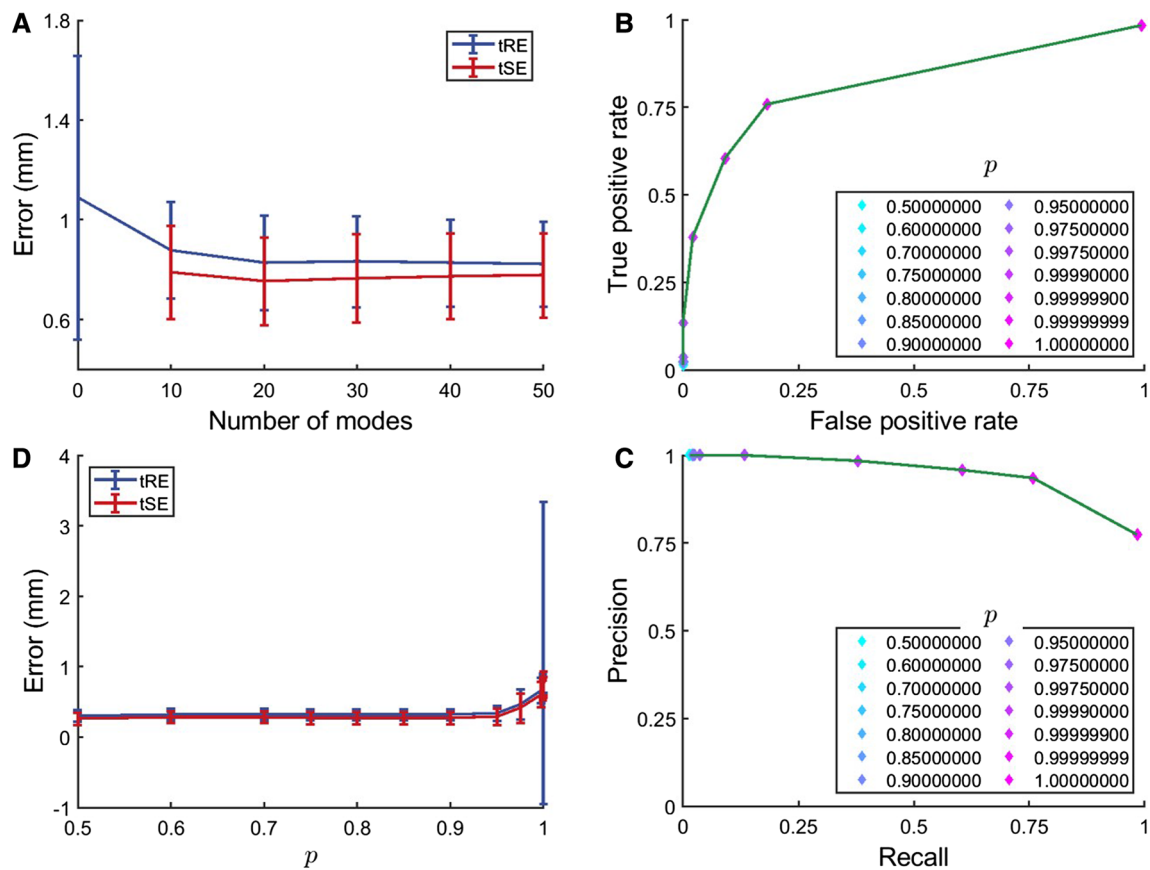


Fig. 4 Simulated data without outliers submillimeter mean tRE and tSE for $n_m > 0$ (a), ROC (b) and precision-recall (c) curves for increasing p , and a slow increase in errors as confidence in registration degrades with increasing p (d)

the false-positive rate is also higher at corresponding values of p compared to the previous two experiments, meaning that more failed registrations are classified as successful at all values of p (Fig. 5b). This also leads to a drop in precision at each p (Fig. 5c), although the drop is small due to the large number of successful registrations compared to failed registrations. Finally, the average tRE and tSE for registrations classified as successful with increasing p show gradual increase with large standard deviations at all values of p (Fig. 5d) due to the presence of misclassified failed registrations at each p . This is also clear from the smaller standard deviation at $p = 1$ (Fig. 5d) compared to the previous experiment (Fig. 4d) since not all failed registrations fell into this category.

Since the addition of outliers in our sampled points was the only difference between the setup for Exps. 4.2 and 4.3, we evaluate how well our algorithm was able to detect and exclude these outliers. We observed that our χ^2 confidence tests were performed on an average 2978 inlying sample points out of 3000. That is, on average, our algorithm was only able to detect about $0.73 (\pm 3.21)\%$ of the 10% outliers that were introduced in the sample points. Therefore, the final noise parameter estimation performed by our algorithm

includes on average 278 outliers. These outliers can cause the noise covariance estimates to be higher than the actual noise in the inliers, leading to lower κ estimates and, therefore, also lower E_o estimates. This allows a large percentage of successful registrations to be correctly classified even at lower values of p and also allows errors to be introduced.

In order to improve the reliability of our confidence assignment in realistic scenarios, our registration algorithm must perform more robust outlier detection. The $p = 0.95$ threshold used for outlier detection, as has been done in prior work [8], is likely too high. If χ^2 tests are used for outlier detection, then the choice of p needs to be thoroughly evaluated.

In vivo clinical data

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 by [34]. We use a self-supervised method that uses structure from motion (SfM) points [26] and relative camera motion to train a Siamese neural network to predict depth esti-

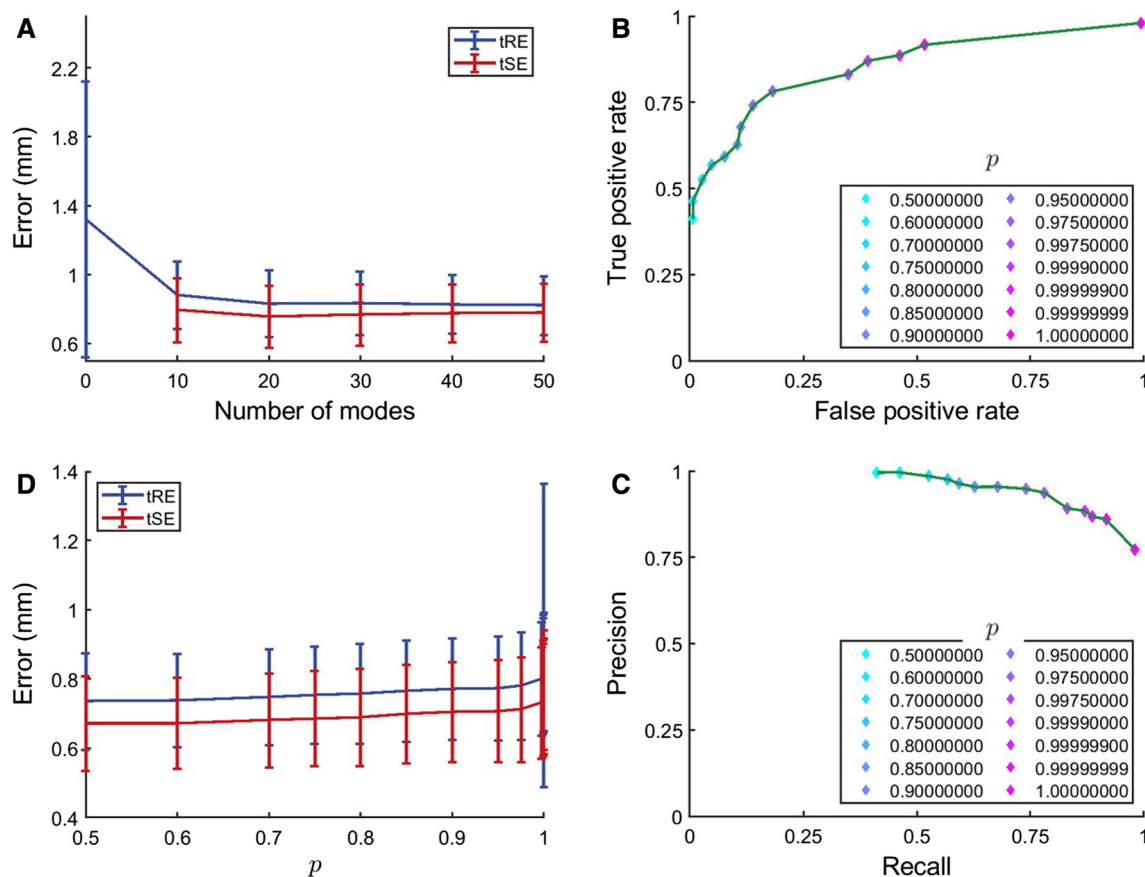


Fig. 5 Simulated data with 10% outliers submillimeter mean tRE and tSE for $n_m > 0$ (a), ROC (b) and precision-recall (c) curves for increasing p , and a very slow increase in errors as confidence in registration degrades with increasing p (d)

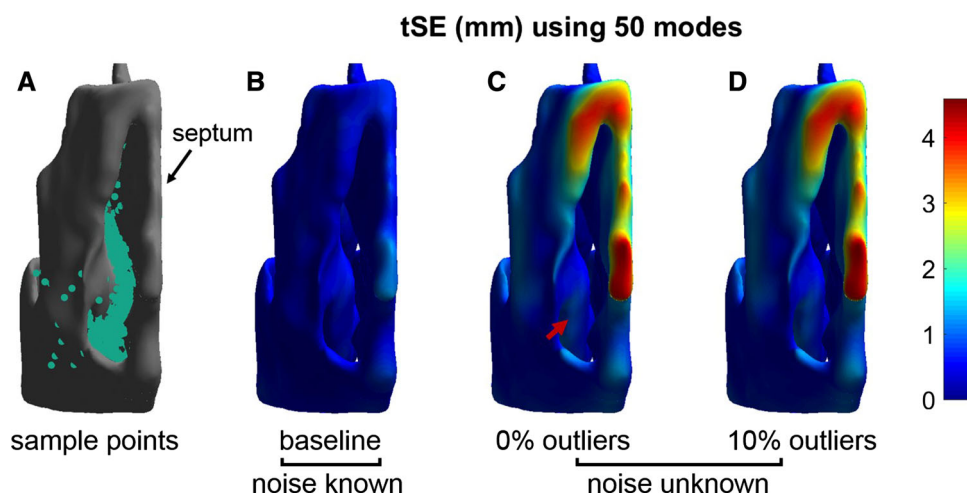


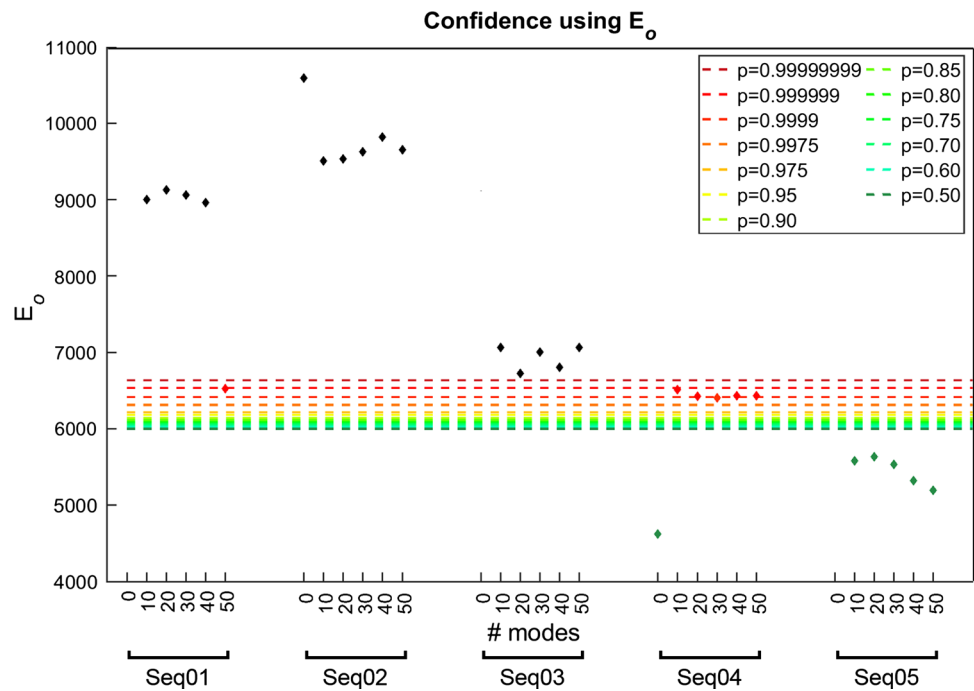
Fig. 6 a An example of points (green) sampled from within the nasal cavity. This rendering shows a view from the outside of the right nostril looking into the right nasal cavity at the inferior turbinate. Per vertex tSE for shapes reconstructed using $n_m = 50$ plotted on the mean right nasal cavity mesh: b tSE is low throughout the reconstructed shape

since points were sampled uniformly from the entire left-out shape in the baseline experiment (Establishing p); tSE is low inside the nasal cavity (red arrow) where points were sampled from for both the experiment with c 0% and d 10% outliers, but grows larger away from this region

mates at every pixel of a given frame [28]. 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 or relative camera motion estimation were corrected using G-IMLOP with scale to produce a dense reconstruction spanning a large area of the nasal

Fig. 7 E_o for all in vivo registrations plotted for each video sequence. Per sequence, from left to right, the data points indicate confidence scores achieved using $n_m \in \{0, 10, 20, 30, 40, 50\}$ modes. Data points are colored according to the value of p at which each registration passed the χ^2 test. Black data points indicate rejected registrations that failed the χ^2 test at $p = 1$



passage. Deformable registrations are then computed using GD-IMLOP with 3000 points sampled uniformly from this dense reconstruction, assuming noise with SDs $1 \times 1 \times 2 \text{ mm}^3$ and 30° ($e = 0.5$) for position and orientation data, respectively, and with scale and shape parameter optimization restricted to within $[0.7, 1.3]$ and ± 1 SD, respectively. Registrations are computed using $n_m \in \{0, 10, 20, 30, 40, 50\}$ modes.

All registrations run with 0 modes terminated at the maximum iteration threshold of 100, while those run using $n_m > 0$ modes converged at an average 10.36 iterations in an average 26.03 seconds. Figure 7 shows registrations that were computed with increasing number of modes from left to right for each sequence plotted against E_o . The rigid registrations (using 0 modes) for sequences 1, 3, and 5 did not pass any of the tests using E_p and are not plotted. The average residual error of these registrations was $3.1 (\pm 2.17) \text{ mm}$. These high residual errors are unsurprising since the average residual error produced by our algorithm is simply the average E_p over all inliers. Many of the remaining registrations fail the E_o tests (Fig. 7). The mean residual error of these registrations was $0.74 (\pm 0.18) \text{ mm}$. These low errors are not expected to lead to low E_o since E_o is entirely dependent on the alignment of the orientations. The rigid registration computed for sequence 4 and all deformable registrations in sequence 5 fell below the χ^2 threshold at $p = 0.50$ with a mean residual error of $0.83 (\pm 0.07) \text{ mm}$. Although these errors are higher than those from the registrations rejected by E_o , this can be explained by the fact that point sets can often achieve good positional alignment without good orientation

alignment. However, if such registrations are excluded, it is possible to end up with registrations where positional alignment is slightly worse but orientation alignment is better. Sequence 4 registration using 30 modes falls below the χ^2 threshold at $p = 0.9999$ with mean residual error of 0.72 mm while the remaining sequence 4 registrations along with the sequence 1 registration using 50 modes pass the χ^2 test at $p = 0.999999$ with mean residual error of $0.74 (\pm 0.04) \text{ mm}$. Figure 8 shows a qualitative evaluation of the registration and reconstruction produced for sequence 1 using 50 modes. The data points reconstructed from video (white) appear to align well with the reconstructed patient shape, and the reconstructed shape (right) resembles the anatomy in the video frame (left).

Conclusion

We show that GD-IMLOP is able to produce submillimeter registrations and reconstructions with simulated data and assign confidence to these registrations. Although the confidence assignment deteriorated in the presence of outliers, we show that this was because GD-IMLOP was unable to find and reject a majority of the outliers. GD-IMLOP was also able to compute submillimeter registrations on in vivo clinical data and assign confidence to these registrations. Further, it can accurately predict the anatomy where video data are available (Fig. 6).

In the future, we hope to learn statistics from thousands of CTs to better cover the range of anatomical variations. We will also work on improving outlier rejection so that the

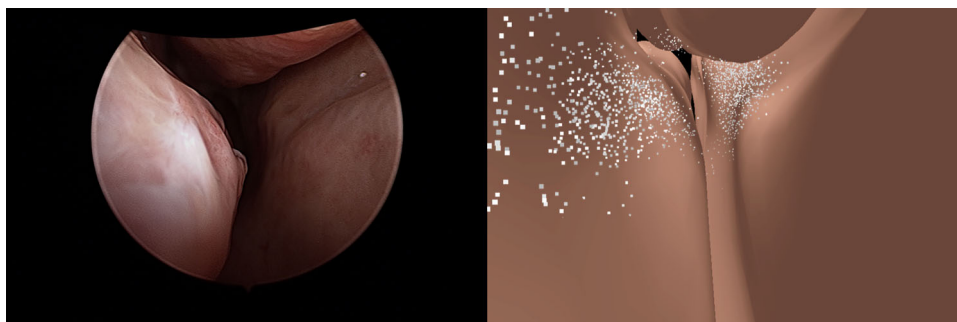


Fig. 8 Left: one of the frames from sequence 1 of nasal endoscopy video. Right: visualization of the final registration and reconstruction computed using 50 modes. The patient anatomy in video (left) and

the anatomy estimated using GD-IMLOP (right) appear similar, and the points from video (white) appear to align well with the estimated anatomy (right)

effect of outliers on both the registration as well as confidence assignment can be minimized. Another step forward would be to learn to correlate the bimodal distribution of confidence scores E_p and E_o with the tRE in order to predict a distribution of tREs that are most likely to occur given our confidence scores. This removes our current reliance on 1 mm as threshold for ground truth success label. This is important since registration accuracy requirements change based on location. For instance, this threshold can be higher near the opening of the nasal cavities since critical structures are not nearby. However, closer to the ethmoid bone which separates the nasal cavities and sinuses from the brain, this threshold should be much lower than 1 mm. This threshold also broadly affects our results. For instance, if we changed this threshold to 1.5 mm, our algorithm does not make any mistakes in any of our simulated experiments. Further, additional features like contours can also be used in addition to position and orientation data, as in [10], to further improve registration and to add an additional test to evaluate the success of the registration based on the quality of contour alignment. With these improvements along with more accurate reconstructions from video, it may be possible to extend this approach for use in place of CTs during endoscopic surgeries and further reduce patient expose to radiation.

Acknowledgements This work was funded by NIH R01-EB015530, JHU Provost's Postdoctoral Fellowship, and JHU internal funds. We would also like to acknowledge the comments and input from Seth D. Billings and Mathias Unberath.

Funding The National Institutes of Health provided a research Grant (NIH R01-EB015530) to conduct the study that yielded the clinical dataset used in our in vivo experiment. A. Sinha was supported partly by the Johns Hopkins University (JHU) Provost's Postdoctoral Fellowship and partly by other JHU internal funds.

Compliance with ethical standards

Disclaimer The concepts and information presented in this paper are based on research and are not commercially available.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standard All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the clinical study under JHU IRB NA_00074677.

References

1. Amberg B, Romdhani S, Vetter T (2007) Optimal step nonrigid ICP algorithms for surface registration. In: 2007 IEEE conference on CVPR, pp 1–8. <https://doi.org/10.1109/CVPR.2007.383165>
2. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* 54(3):2033–2044. <https://doi.org/10.1016/j.neuroimage.2010.09.025>
3. Banerjee A, Dhillon I, Ghosh J, Sra S (2005) Clustering on the unit hypersphere using von Mises–Fisher distributions. *J Mach Learn Res* 6:1345–1382
4. Beichel RR, Ulrich EJ, Bauer C, Wahle A, Brown B, Chang T, Plichta KA, Smith BJ, Sunderland JJ, Braun T, Fedorov A, Clunie D, Onken M, Riesmeier J, Pieper S, Kikinis R, Graham MM, Casavant TL, Sonka M, Buatti JM (2015) Data from QIN-HEADNECK. *Cancer Imaging Arch* <https://doi.org/10.7937/K9/TCIA.2015.K0F5CGLI>
5. Besl PJ, McKay ND (1992) A method for registration of 3-d shapes. *IEEE Trans Pattern Anal Mach Intell* 14(2):239–256. <https://doi.org/10.1109/34.121791>
6. Billings SD (2015) Probabilistic feature-based registration for interventional medicine. Ph.D. thesis, The Johns Hopkins University
7. Billings SD, Taylor RH (2014) Iterative most likely oriented point registration. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R (eds) *Medical image computing and computer-assisted intervention—MICCAI 2014*. MICCAI 2014. Lecture notes in computer science, vol 8673. Springer, Cham. https://doi.org/10.1007/978-3-319-10404-1_23
8. Billings SD, Taylor RH (2015) Generalized iterative most likely oriented-point (G-IMLOP) registration. *Int J CARS* 10(8):1213–1226. <https://doi.org/10.1007/s11548-015-1221-2>

9. Billings SD, Bector EM, Taylor RH (2015) Iterative most-likely point registration (IMLP): a robust algorithm for computing optimal shape alignment. *PLoS ONE* 10(3):1–45. <https://doi.org/10.1371/journal.pone.0117688>
10. Billings SD, Sinha A, Reiter A, Leonard S, Ishii M, Hager GD, Taylor RH (2016) Anatomically Constrained Video-CT Registration via the V-IMLOP Algorithm. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W (eds) *Medical image computing and computer-assisted intervention—MICCAI 2016*. MICCAI 2016. Lecture notes in computer science, vol 9902. Springer, Cham. https://doi.org/10.1007/978-3-319-46726-9_16
11. Bosch WR, Straube WL, Matthews JW, Purdy JA (2015) Data from Head-Neck_Cetuximab. *Cancer Imaging Arch*. <https://doi.org/10.7937/K9/TCIA.2015.7AKGJUPZ>
12. Bouaziz S, Tagliasacchi A, Pauly M (2013) Sparse iterative closest point. In: *Computer graphics forum*, vol 32. Wiley, pp 113–123. <https://doi.org/10.1111/cgf.12178>
13. Burschka D, Li M, Ishii M, Taylor RH, Hager GD (2005) Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. *Med Image Anal* 9(5):413–426. <https://doi.org/10.1016/j.media.2005.05.005>
14. Caulley L, Thavorn K, Rudmik L, Cameron C, Kilty SJ (2015) Direct costs of adult chronic rhinosinusitis by using 4 methods of estimation: results of the us medical expenditure panel survey. *J Allergy Clin Immunol* 136(6):1517–1522. <https://doi.org/10.1016/j.jaci.2015.08.037>
15. Chen Y, Medioni G (1992) Object modelling by registration of multiple range images. *Image Vis Comput* 10(3):145–155. [https://doi.org/10.1016/0262-8856\(92\)90066-C](https://doi.org/10.1016/0262-8856(92)90066-C)
16. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F (2013) The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 26(6):1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
17. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models: their training and application. *Comp Vis Im Underst* 61:38–59. <https://doi.org/10.1006/cviu.1995.1004>
18. Fedorov A, Clunie D, Ulrich E, Bauer C, Wahle A, Brown B, Onken M, Riesmeier J, Pieper S, Kikinis R, Buatti J, Beichel RR (2016) 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. <https://doi.org/10.7717/peerj.2057>
19. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, Cohen N, Cervin A, Douglas R, Gevaert P, Georgalas C, Goossens H, Harvey R, Hellings P, Hopkins C, Jones N, Joos G, Kalogjera L, Kern B, Kowalski M, Price D, Riechelmann H, Schlosser R, Senior B, Thomas M, Toskala E, Voegels R, de Wang Y, Wormald PJ (2012) EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. *Rhinology* 50(1):1–12. <https://doi.org/10.4193/Rhino50E2>
20. Francois R, Fablet R, Barillot C (2003) Robust statistical registration of 3d ultrasound images using texture information. In: *Proceedings 2003 international conference on image processing (Cat. No.03CH37429)*, vol 1, pp I–581. <https://doi.org/10.1109/ICIP.2003.1247028>
21. Granger S, Pennec X (2002) Multi-scale EM-ICP: a fast and robust approach for surface registration. In: Heyden A, Sparr G, Nielsen M, Johansen P (eds) *Computer Vision—ECCV 2002*. ECCV 2002. Lecture notes in computer science, vol 2353. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-47979-1_28
22. Granger S, Pennec X, Roche A (2001) Rigid point-surface registration using oriented points and an EM variant of ICP for computer guided oral implantology. *Tech. Rep. RR-4169*, INRIA
23. Hufnagel H, Pennec X, Ehrhardt J, Ayache N, Handels H (2009) Computation of a probabilistic statistical shape model in a maximum-a-posteriori framework. *Methods Inf Med* 48(04):314–319. <https://doi.org/10.3414/ME9228>
24. Huhle B, Magnusson M, Strasser W, Lilienthal AJ (2008) Registration of colored 3d point clouds with a kernel-based extension to the normal distributions transform. In: *2008 IEEE ICRA*, pp 4025–4030. <https://doi.org/10.1109/ROBOT.2008.4543829>
25. Kainz J, Stammberger H (1989) The roof of the anterior ethmoid: a place of least resistance in the skull base. *Am J Rhinol* 3(4):191–199. <https://doi.org/10.2500/105065889782009552>
26. Leonard S, Reiter A, Sinha A, Ishii M, Taylor RH, Hager GD (2016) Image-based navigation for functional endoscopic sinus surgery using structure from motion. In: *Proceedings of SPIE, medical imaging: image processing*, vol 9784, pp 97840V–7. <https://doi.org/10.1117/12.2217279>
27. Leonard S, Sinha A, Reiter A, Ishii M, Gallia GL, Taylor RH, Hager GD (2018) Evaluation and stability analysis of video-based navigation system for functional endoscopic sinus surgery on in vivo clinical data. *IEEE Trans Med Imaging* 37(10):2185–2195. <https://doi.org/10.1109/TMI.2018.2833868>
28. Liu X, Sinha A, Unberath M, Ishii M, Hager GD, Taylor RH, Reiter A (2018) Self-supervised learning for dense depth estimation in monocular endoscopy. In: Stoyanov D et al (eds) *Computer assisted robotic endoscopy—CARE 2018*. Lecture notes in computer science, vol 11041. Springer, Cham. https://doi.org/10.1007/978-3-030-01201-4_15
29. Mardia KV, Jupp PE (2008) *Directional statistics*. Wiley series in probability and statistics. Wiley, New York, pp 1–432. <https://doi.org/10.1002/9780470316979>
30. Min Z, Wang J, Song S, Meng MQ (2018) Robust generalized point cloud registration with expectation maximization considering anisotropic positional uncertainties. In: *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp 1290–1297. <https://doi.org/10.1109/IROS.2018.8593558>
31. Min Z, Wang J, Meng MQ (2019) Joint rigid registration of multiple generalized point sets with hybrid mixture models. *IEEE Trans Autom Sci Eng*. <https://doi.org/10.1109/TASE.2019.2906391>
32. Myronenko A, Song X (2010) Point set registration: Coherent point drift. *IEEE Trans Pattern Anal Mach Intell* 32(12):2262–2275. <https://doi.org/10.1109/TPAMI.2010.46>
33. Phillips JM, Liu R, Tomasi C (2007) Outlier robust icp for minimizing fractional rmsd. In: *6th international conference on 3D digital imaging and modeling*, pp 427–434. <https://doi.org/10.1109/3DIM.2007.39>
34. Reiter A, Leonard S, Sinha A, Ishii M, Taylor RH, Hager GD (2016) Endoscopic-CT: learning-based photometric reconstruction for endoscopic sinus surgery. In: *Proceedings of SPIE, medical imaging: image processing*, vol 9784, pp 978418–6. <https://doi.org/10.1117/12.2216296>
35. Rusinkiewicz S, Levoy M (2001) Efficient variants of the icp algorithm. In: *Proceedings of the 3rd international conference on 3D digital imaging and modeling*, pp 145–152. <https://doi.org/10.1109/IM.2001.924423>
36. Segal A, Haehnel D, Thrun S (2009) Generalized-ICP. In: *Proceedings of robotics: science and systems*, Seattle, USA. <https://doi.org/10.15607/RSS.2009.V.021>
37. Sinha A (2018) *Deformable registration using shape statistics with applications in sinus surgery*. Ph.D. thesis. The Johns Hopkins University
38. Sinha A, Leonard S, Reiter A, Ishii M, Taylor RH, Hager GD (2016) Automatic segmentation and statistical shape modeling of the paranasal sinuses to estimate natural variations. In: *Proceedings of SPIE, medical imaging: image processing*, vol 9784, pp 97840D–8. <https://doi.org/10.1117/12.2217337>
39. Sinha A, Reiter A, Leonard S, Ishii M, Hager GD, Taylor RH (2017) Simultaneous segmentation and correspondence improvement using statistical modes. In: *Proceedings of SPIE, medical*

- imaging: image processing, vol 10133, pp 101331B–8. <https://doi.org/10.1117/12.2253533>
40. Sinha A, Liu X, Reiter A, Ishii M, Hager GD, Taylor RH (2018) Endoscopic navigation in the absence of CT imaging. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G (eds) Medical image computing and computer assisted intervention–MICCAI 2018. MICCAI 2018. Lecture notes in computer science, vol 11073. Springer, Cham. https://doi.org/10.1007/978-3-030-00937-3_8
41. Sinha A, Billings SD, Reiter A, Liu X, Ishii M, Hager GD, Taylor RH (2019) The deformable most-likely-point paradigm. *Med Image Anal* 55:148–164. <https://doi.org/10.1016/j.media.2019.04.013>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.