

# Automatic segmentation and statistical shape modeling of the paranasal sinuses to estimate natural variations

Ayushi Sinha<sup>a</sup>, Simon Leonard<sup>a</sup>, Austin Reiter<sup>a</sup>, Masaru Ishii<sup>b</sup>, Russell H. Taylor<sup>a</sup> and Gregory D. Hager<sup>a</sup>

<sup>a</sup>Dept. of Computer Science, The Johns Hopkins University, Baltimore, MD 21218, USA;

<sup>b</sup>Dept. of Otolaryngology-Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA

## ABSTRACT

We present an automatic segmentation and statistical shape modeling system for the paranasal sinuses which allows us to locate structures in and around the sinuses, as well as to observe the variability in these structures. This system involves deformably registering a given patient image to a manually segmented template image, and using the resulting deformation field to transfer labels from the template to the patient image. We use 3D snake splines to correct errors in this initial segmentation. Once we have several accurately segmented images, we build statistical shape models to observe the population mean and variance for each structure. These shape models are useful to us in several ways. Regular registration methods are insufficient to accurately register pre-operative computed tomography (CT) images with intra-operative endoscopy video of the sinuses. This is because of deformations that occur in structures containing erectile tissue. Our aim is to estimate these deformations using our shape models in order to improve video-CT registration, as well as to distinguish normal variations in anatomy from abnormal variations, and automatically detect and stage pathology. We can also compare the mean shapes and variances in different populations, such as different genders or ethnicities, in order to observe differences and similarities, as well as in different age groups in order to observe the developmental changes that occur in the sinuses.

**Keywords:** paranasal sinuses, segmentation, statistical shape model, registration, erectile tissue, turbinates, variance, natural variation

## 1. DESCRIPTION OF PURPOSE

Automatic segmentation of medical images is important for several reasons. Manual segmentation is tedious and not scalable to large datasets. Automatic segmentation allows us to extract shapes from a large number of images, and therefore to study large populations. These segmented structures can be used as landmarks in multimodal registration. However, for our application, video-CT registration during minimally invasive surgery through the paranasal sinuses, simply having segmented structures is insufficient. Several structures in the sinuses exhibit high variability. The superior, middle and inferior turbinates, which reside in the nasal passage, contain erectile tissue. These turbinates undergo periodical deformation as they complete their nasal cycle, which is the alternating partial congestion and decongestion of the nasal cavities due to the expansion and contraction of the turbinates.<sup>1</sup> These contractions are further exaggerated by decongestants administered to patients before surgery to facilitate easy movement of tools in the sinuses. These changes in structure make registration error prone. In order to tackle this problem, we need models explaining structure variability in addition to high quality segmentations.

---

Further author information: (Send correspondence to Ayushi Sinha)

Ayushi Sinha: E-mail: asinha8@jhu.edu

Simon Leonard: E-mail: sleonard@jhu.edu

Austin Reiter: E-mail: areiter@cs.jhu.edu

Masaru Ishii: E-mail: mishii3@jhu.edu

Russel H. Taylor: E-mail: rht@jhu.edu

Gregory D. Hager: E-mail: hager@cs.jhu.edu

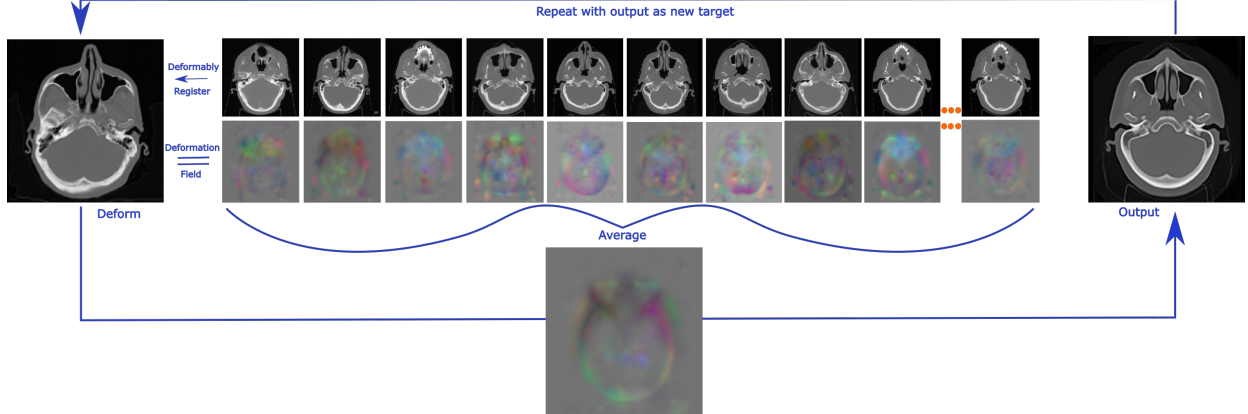


Figure 1. Template creation pipeline: All input images are deformably registered to one target image, which is then deformed by the mean of the deformation fields resulting from the registration. The colors in the deformation fields represent the direction of the vectors, whereas the intensity of the colors indicates the magnitude of the vectors. Deforming the target image by the mean deformation field takes the target image towards the mean of our input images. This process is iterated with the output image as the new target image to register to. Individual variation from the initial target image decreases with every iteration, and the resulting output gets closer to the true average of the input set of images.

Our aim, therefore, is to obtain accurate segmentations with correspondences for a large number of patient images, build shape models, and learn useful information from these models. In order to achieve these goals, we built a template image<sup>2</sup> (Fig. 1) which we manually segmented under the supervision of an otolaryngologist, and deformably registered all given patient images to this template. We then automatically segmented patient images by deforming the template segmentations to each patient space using the deformation fields resulting from registration. This initial segmentation was further improved using gradient vector flow (GVF) to drive mesh vertices closer to edges in the patient image.<sup>3</sup> Since we deform one shape to a collection of images, we are guaranteed correspondence between the shapes. Finally, once we had accurate segmentations for several images, we built statistical shape models (SSMs)<sup>4</sup> for each segmented structure, and acquired statistics for our patient population. Each of these steps is detailed in the next section.

## 2. METHODS

### 2.1 Template creation

In order to minimize the influence of individual variation in our registration, we built a template image which represents the mean of a collection of images. We use a standard template building technique (Fig. 1), which requires one fixed image, called the target image, and  $n$  moving images of the same dimensions and resolution as the target image. The  $n$  moving images are deformably registered to the target image using ANTs registration software,<sup>5</sup> resulting in deformation fields that can transform the target image to each of the  $n$  moving images. If we compute the mean of these deformation fields and transform the target image by this mean deformation field, it takes the target image towards the space of the mean of the  $n$  moving images.<sup>2</sup> This process can be repeated several times using the output from the previous iteration as the new target image. Each iteration further reduces the individual variation of the target image, and moves it closer to the population mean. We iterated this process five times to obtain our template image. The result is a highly symmetrical image of the sinus (Fig. 1). We hand segmented our template image under the supervision of an expert. We use this segmented template to automatically segment any given patient image.

### 2.2 Automatic segmentation

Given a new patient image, we deformably register it to our segmented template image, using ANTs registration software.<sup>5</sup> We extract surface meshes<sup>6</sup> from our template segmentations, and use the deformation fields from deformable registration to move our template meshes to patient space. In order to correct errors in this initial

---

**Algorithm 1** Order neighboring vertices

---

```
1: procedure ORDER_NEIGHBORING_VERTICES(Mesh  $m$ ,  $v[]$ ) ▷  $v[]$  is a vector of vectors
2:   for each vertex  $i$  in  $m$  do
3:      $f \leftarrow \text{find\_faces}(i)$  ▷ Find faces incident on  $i$ , ordered clockwise
4:     for  $j = 0 \rightarrow f.size()$  do ▷ For each face in  $f$ 
5:        $v[i][j] \leftarrow$  vertex at the end of half-edge9 facing vertex  $i$ 
6:   return  $v[]$ 
```

---

---

**Algorithm 2** Sample vertices

---

```
1: procedure SAMPLE_VERTICES(Mesh  $m$ ,  $i$ ,  $n$ ,  $s[]$ ,  $v[]$ ) ▷ Sample  $n$  vertices in a spiral starting at  $i$ 
2:    $visited.resize() \leftarrow m.vertices.size()$ 
3:    $s.resize() \leftarrow n$ 
4:    $visited[] \leftarrow 0$  ▷  $visited$  initialized to 0
5:    $s[0] \leftarrow i$ 
6:    $k \leftarrow 0$ 
7:    $x \leftarrow 0$ 
8:   for  $j = 1 \rightarrow n$  do
9:     if  $k < v[i].size()$  then
10:      if  $!visited[i][k]$  then ▷ Visit all unvisited vertices in one-ring neighborhood of  $i$ 
11:         $s[j] \leftarrow v[i][k]$ 
12:         $visited[v[i][k]] = 1$ 
13:         $k \leftarrow k + 1$ 
14:     else ▷ Visit all unvisited neighbors of vertices in one-ring neighborhood of  $i$ 
15:        $k \leftarrow 0$  ▷ Once the two-ring neighborhood is exhausted, proceed to the next
16:        $x \leftarrow x + 1$  ▷ Repeat until  $n$  vertices have been sampled
17:        $i \leftarrow s[x]$ 
18:   return  $s$ 
```

---

segmentation, we use gradient vector flow (GVF), which uses gradient based energies in the image to pull vertices in the mesh towards edges in the image.<sup>3</sup> The vertices in the mesh are treated as control points on a 3D snake spline, which is evolved using GVF.<sup>7</sup> Since it is not possible to sequentially order vertices in 3D the way it can be done in 2D, we need to find a way to sample and order the vertices that make up our snake splines. Simple blob-like structures have been approximated with high accuracy using parametrized ellipsoids, which can then be deformed.<sup>8</sup> However, structures in the sinuses can often be too complex to be approximated using parametrized ellipsoids. We take a very simple approach to solving this problem, without having to parametrize a complex shape, or having to build and store large adjacency matrices containing per vertex neighborhood information.

We use a simple but efficient sampling structure for the control points on our 3D snake splines. We sample points in a spiral around each vertex (Algorithms 1, 2), which allows us to store the points on the curve in a sequential order in a vector. These vertices can therefore be moved using simple vector-matrix operations. The user can specify how many points must be sampled at each iteration, which defines the length of the curve. These curves have consistent internal and external directions which make it easy to define internal and external energies<sup>7</sup> on the spline (Fig. 2). The only points of concern with this sampling structure are the source and end points on the spiral. These two points

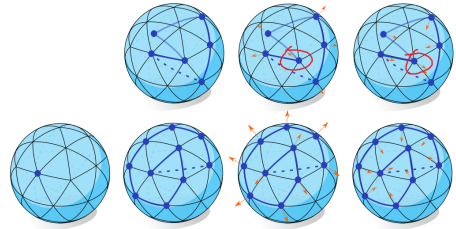


Figure 2. Sampling structure (blue dots represent sampled vertices, blue lines indicate adjacency; dotted lines imply that connecting points should not be attracted to each other; translucent lines cross through the shape): Random sampling (top) causes inconsistent internal and external directions. The middle sphere shows a vertex with external energy pointing inwards, and the right sphere shows one with internal energy pointing outwards. Our spiral sampling maintains internal and external direction consistency.

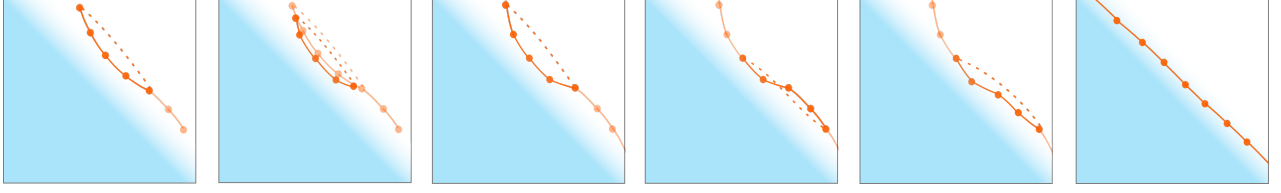


Figure 3. From left to right, we see how the vertices in our snake spline move towards the edge. Since we do not want corner points to get drawn inward toward each other, we do not allow corner points to move. However, they are updated during following iterations when they are not corner points. After several iterations all vertices are on the closest edge.

are attracted to each other despite being far away from each other spatially due to our closed-loop snake spline formulation. We resolve this by updating all points except the source and end points at each iteration. These points are updated in other sampling iterations when they are not source or end points (Fig. 3). GVF then guides these control points toward edges in the corresponding image.<sup>7</sup> This method gives us highly accurate segmentations, evaluated in the next section (Table 1, Fig. 4, 5). We are working on further improving segmentation by utilizing more information in our images and meshes. For instance, we can use pixel intensity in addition to image gradients. We can also use vertex orientation and mean shape information to guide segmentation.

### 2.3 Statistical Shape Models (SSMs)

Once we have high quality segmentations in images of several healthy individuals, we study the statistics of each segmented structure in this population. The standard way of doing so is by using principal component analysis (PCA).<sup>10</sup> In order to compute shape statistics, we first align and center all the meshes, and describe each mesh by a column vector called the shape vector,

$$V = \begin{bmatrix} v_1 \\ \vdots \\ v_{n_v} \end{bmatrix},$$

where  $n_v$  is the number of vertices in the mesh. Since the shapes are homologous, that is, in correspondence with each other, we compute the mean shape,  $\bar{V}$ , simply by computing the average of the shape vectors,

$$\bar{V} = \frac{1}{n_s} \sum_{i=1}^{n_s} V_i,$$

where  $V_i$  is the shape vector for the  $i$ th mesh, and  $n_s$  is the number of meshes. Next, we compute the shape covariance matrix as

$$\Sigma = \frac{1}{n_s} \sum_{i=1}^{n_s} (V_i - \bar{V})^T (V_i - \bar{V}).$$

Eigen decomposition of this matrix,  $\Sigma$ , gives us the modes and mode weights of shape variation.

$$\Sigma = [m_1 \dots m_{n_s}] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{n_s} \end{bmatrix} [m_1 \dots m_{n_s}]^T,$$

where  $m_i$  are the orthonormal set of eigenvectors that represent the modes of variation, and  $\lambda_i$  are the eigenvalues, or mode weights, that represent the amount of variation along the direction of each mode. Using the mean shape and the modes of variation, we can estimate any homologous shape using the formula

$$V = \bar{V} + \sum_{i=1}^{n_s} b_i m_i, \quad (1)$$

		Deformable Registration			GVF		
		Mean	Min	Max	Mean	Min	Max
Patient 1	RMS	0.4808	<b>0.0000</b>	4.0460	<b>0.3579</b>	0.0003	<b>2.2020</b>
	LMS	0.3726	<b>0.0001</b>	3.2085	<b>0.2913</b>	<b>0.0001</b>	<b>1.3471</b>
Patient 2	RMS	0.3419	<b>0.0001</b>	3.8089	<b>0.3401</b>	0.0002	<b>2.3474</b>
	LMS	0.4161	0.0002	4.5978	<b>0.3115</b>	<b>0.0001</b>	<b>3.7671</b>
Patient 3	RMS	0.3055	0.0002	<b>1.4191</b>	<b>0.2583</b>	<b>0.0000</b>	1.5183
	LMS	0.3286	0.0005	2.9758	<b>0.3089</b>	<b>0.0002</b>	<b>2.0215</b>
Mean	RMS	0.3761	<b>0.0001</b>	3.00914	<b>0.3188</b>	0.0002	<b>2.0226</b>
	LMS	0.3724	0.0003	3.5940	<b>0.3039</b>	<b>0.0001</b>	<b>2.3819</b>

Table 1. Mean, minimum and maximum vertex errors (in mm) for deformable registration and after GVF. RMS and LMS indicate right and left maxillary sinuses, and boldface indicates smaller errors.

where  $b_i$  are shape parameters computed as

$$b_i = m_i^T (V - \bar{V}). \quad (2)$$

Our SSM not only shows us how different structures vary across our sample population, but also reflects natural variations that occur periodically in some structures in the sinuses, as discusses earlier. This result is discussed in more detail in the next section.

### 3. RESULTS

We compare our segmentation results to manual segmentations. Table 1 shows that segmentations after GVF have smaller average error, and almost always have smaller maximum error. Fig. 4 shows edge maps of deformably registered maxillary sinuses (red) and those after snakes splines (green) compared against manual segmentations (blue). The green and blue edge-maps overlap almost perfectly (right) indicating that snakes splines minimize errors in segmentation. Fig. 5 also shows results from deformable registration and GVF on the right maxillary sinus. Errors noticeable after deformable registration (top) are clearly reduced after GVF (bottom).

We build our SSMs using these highly accurate segmentations. However, our point correspondences may contains errors, especially after GVF since we move the vertices in each shape independently to improve segmentations. We use this initial point correspondence to estimate a SSM using  $n_s - 1$  meshes. We then estimate

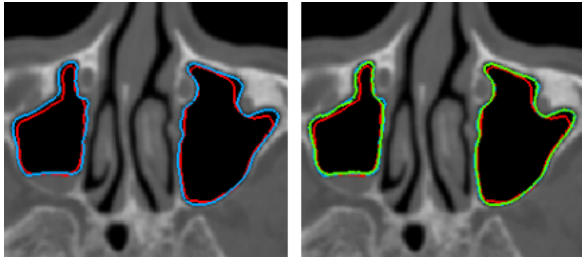


Figure 4. Left: Edge maps of hand-segmented left and right maxillary sinuses (blue) and deformably registered maxillary sinuses (red); Right: Edge map of maxillary sinuses after snakes splines (green) overlaps almost perfectly with the hand-segmented edge map.

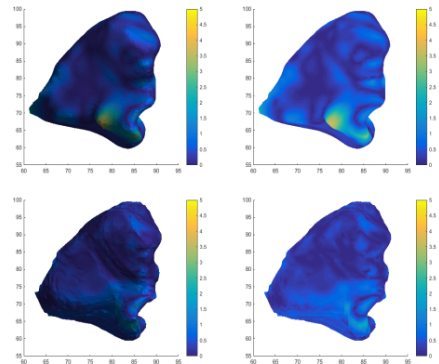


Figure 5. Top: Errors from deformable registration visualized on the mesh with lighting (left), to show structure, and without lighting (right), to show errors without distractions from specularities or shadows; Bottom: Errors after GVF, visualized similarly as above.

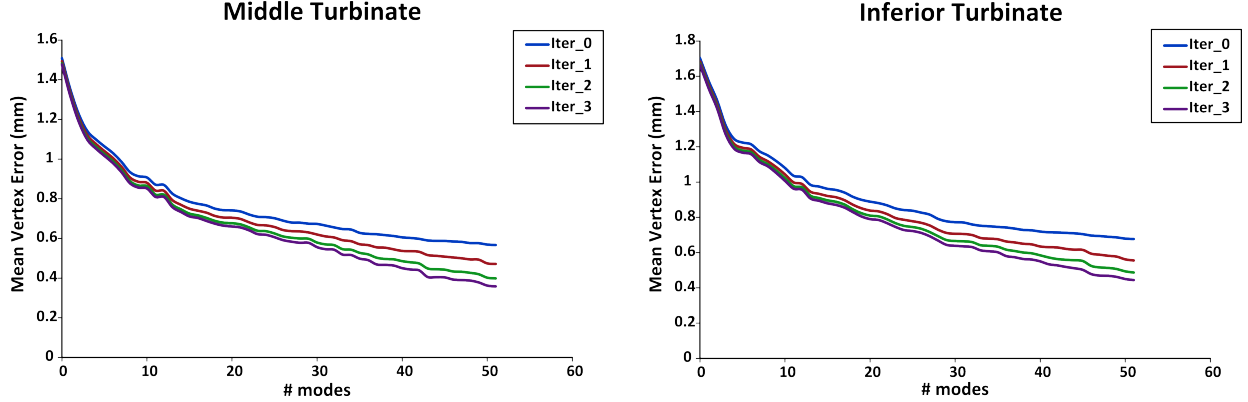


Figure 6. Improving vertex error with each iteration of the point correspondence refinement process<sup>11</sup> shown here for the middle turbinate (left) and the inferior turbinate (right).

the left-out mesh using Eq. 1 and 2, and move the vertices on the left-out mesh along the surface towards the corresponding vertices on the estimated mesh.<sup>11</sup> We repeat this resampling process for each of our meshes, and then recompute our SSM. We repeat this process several times, and at each iteration, we compute the distance between the left-out mesh vertices and the estimated mesh vertices. Averaging over all left-out meshes, we compute the mean vertex error of our SSM. We show in Fig. 6 that this error goes down with every iteration of the point correspondence improvement process.<sup>11</sup>

Finally, we evaluate whether or not our shape models can estimate natural variations in the turbinates. We know that the turbinates, superior, middle, and inferior, contain erectile tissue, and facilitate the nasal cycle by expanding and contracting periodically. Since each image in our dataset likely contains turbinates at different points in the nasal cycle, we can hypothesize that the variation we see in our population reflects the natural variation that turbinates undergo in each individual. In order to evaluate this hypothesis, we segment two CT images from the same patient, one pre- and the other post-operative, using the automatic segmentation method described in Section 2.2. Then, we project the skull from both the pre- and post-operative images onto the skull model, and obtain mode weights for each shape (Eq. 2). We can then use the mode weights and the shape model (Eq. 1) to estimate both the pre- and post-operative skulls. The mode weights for both the pre- and post-operative skulls are similar, and therefore, the two estimated shapes are also very similar. This is what we would expect since skulls do not exhibit change over a short period of time.

We repeat the same process with the inferior turbinates, and observe are large variation in the mode weights, and hence also in the estimated shapes. This, again, is as we expected, since the turbinates are likely at different points in their nasal cycle in the two CT images. Fig. 7 shows the lack in variation in the pre- and post-operative skulls, whereas Fig. 8 shows the similarity between population variation and natural variation in the inferior turbinates.

## 4. DISCUSSIONS

### Video-CT registration using natural shape variations in the turbinates

Since we can show that population variation is able to reflect natural variation in the turbinates, we hope to attempt to use this information to co-register pre-op patient CT and intra-op endoscopy video. The problems we face during this registration are due to topology changes that occur in the turbinates. During a CT scan, the patient's turbinates are at some stage of the nasal cycle. However, during surgery, the patient is administered decongestants to facilitate smooth insertion and movement of the endoscope and other surgical tools. This causes a very high amount of contraction in the turbinates. Since the endoscope sees the middle turbinate soon after entering the sinuses, we want to be able to use this structure to start our video-CT registration. We aim to use the modes of variation observed in the middle turbinate to optimize video-CT registration (Fig. 9).



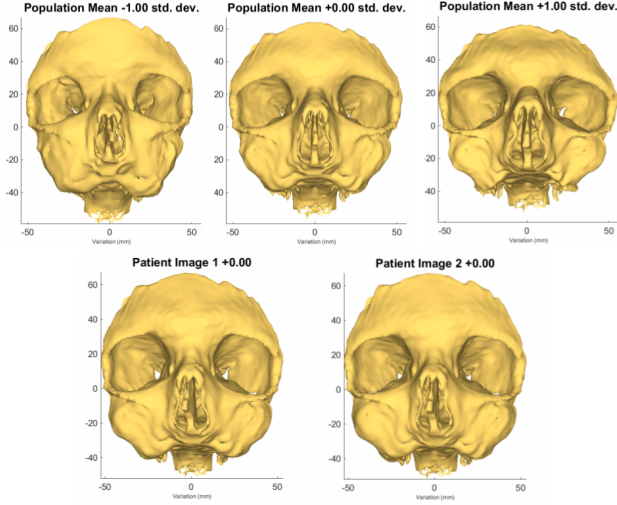


Figure 7. Top: Population variation in the skull model. The middle shape is the mean shape, the left shape shows mean shape with  $-1\sigma$ , where  $\sigma$  is the standard deviation, and the right shape is the mean shape with  $+1\sigma$ . Bottom: The left image shows the pre-op patient skull, and the right image shows the post-op patient skull. The two images show no, or negligible, difference, where minute difference can sometimes be observed due to errors in registration. However, we can see that the population variation is not reflected in the two patient images.

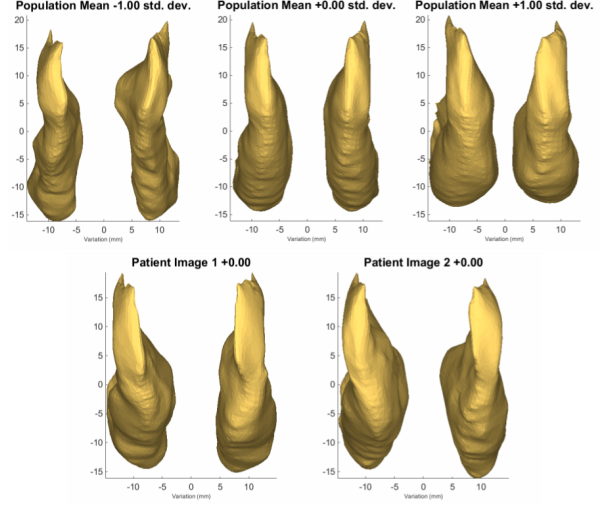


Figure 8. Top: Population variation in the inferior turbinate (IT) model. The middle shape is the mean shape, the left shape shows mean shape with  $-1\sigma$ , and the right shape is the mean shape with  $+1\sigma$ . Bottom: The left image shows the pre-op patient IT, and the right image shows the post-op patient IT. The two images show significant differences, allowing us to conclude that the population variation is reflected in the patient images.

## Normal vs abnormal variations

Since our SSMs are built from a set of patient images with ‘normal’ or disease free sinuses, they tell us what types and amounts of variations we should expect in each structure. This allows us to define the range of variations that can be described as normal, or as observed in a normal population. If, however, we observe an individual who demonstrates variations exceeding this normal range, we can attempt to quantify the amount by which this individual exceeds the normal range (Fig. 10), and relate this quantification to a grading scale for pathology. Further, we can evaluate this automatic grading scale against the manual grading scale currently in use, as well as against post-surgery patient quality of life.

## Population based statistics

Our current SSM is built from 53 images. Other than the scans, we did not have any additional patient information, such as age, gender, or ethnicity. Therefore, our SSM describes the statistics of a general population,



Figure 9. The image from video (left) looks less like the middle turbinate segmented from CT (top right), and more like the middle turbinate estimated from PCA (bottom right).

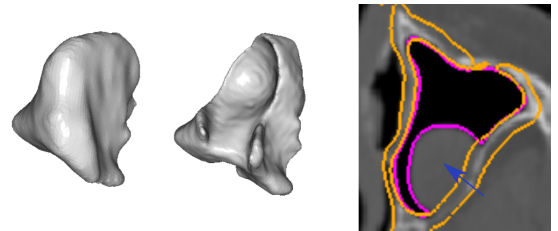


Figure 10. This image shows larger than expected difference between the mean (left) and patient (middle), implying disease. The volume between the orange and pink segmentations (right) can be related to the amount of disease.

not specific to age, gender, or ethnicity. We hope to collect more data in the future with additional patient information, and build SSMs for different age groups, genders, and ethnicities. Such specialized SSMs have several advantages. SSMs of different age groups will allow us to observe the developmental changes that structures in the sinus undergo as humans develop. It can also allow surgeons to adapt their tools to the size of structures present in different age groups. Similarly, SSMs of different genders and ethnicities also allow us to observe the differences and similarities in the different groups, again allowing surgeons to customize their care.

## 5. CONCLUSIONS

In conclusion, we present a system which is able to automatically segment images to sub-millimeter accuracy. We build statistical shape models that allow us to observe variability in the different structures in the sinus, allowing us to estimate the kinds of deformations that happen in the turbinates during the nasal cycle. We are currently working on improving video-CT registration accuracy using these shape models. We also hope to use these models to allow us to identify and stage pathology. We also hope to further improve our segmentations, and build higher quality population specific SSMs.

**This work has not been submitted elsewhere.**

## ACKNOWLEDGMENTS

This work was funded by NIH R01-EB015530: Enhanced Navigation for Endoscopic Sinus Surgery through Video Analysis.

## REFERENCES

1. M. Hasegawa and E. B. Kern, “The human nasal cycle,” in *Mayo Clinic Proceedings*, **51**, pp. 28–34, May 1977.
2. B. B. Avants, P. Yushkevich, J. Pluta, D. Minkoff, M. Korczykowski, J. Detre, and J. C. Gee, “The optimal template effect in hippocampus studies of diseased populations,” *NeuroImage* **49**(3), p. 2457, 2010.
3. C. Xu and J. L. Prince, “Gradient vector flow: A new external force for snakes,” in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, pp. 66–71, 1997.
4. G. Chintalapani, L. M. Ellingsen, O. Sadowsky, J. L. Prince, and R. H. Taylor, “Statistical atlases of bone anatomy: construction, iterative improvement and validation,” in *Medical Image Computing and Computer-Assisted Intervention*, pp. 499–506, 2007.
5. B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, “A reproducible evaluation of {ANTs} similarity metric performance in brain image registration,” *NeuroImage* **54**(3), pp. 2033–2044, 2011.
6. W. E. Lorensen and H. E. Cline, “Marching cubes: A high resolution 3d surface construction algorithm,” in *Computer Graphics, Transactions on, SIGGRAPH* **21**(4), pp. 163–169, 1987.
7. C. Xu and J. Prince, “Snakes, shapes, and gradient vector flow,” *Image Processing, IEEE Transactions on* **7**, pp. 359–369, Mar 1998.
8. R. Delgado-Gonzalo, N. Chenouard, and M. Unser, “Spline-based deforming ellipsoids for interactive 3d bioimage segmentation,” *Image Processing, IEEE Transactions on* **22**, pp. 3926–3940, Oct 2013.
9. K. Weiler, “Edge-based data structures for solid modeling in curved-surface environments,” *Computer Graphics and Applications, IEEE Transactions on* **5**, pp. 21–40, Jan 1985.
10. T. Cootes, C. Taylor, D. Cooper, and J. Graham, “Active shape models-their training and application,” *Computer Vision and Image Understanding* **61**(1), pp. 38–59, 1995.
11. S. Seshamani, G. Chintalapani, and R. Taylor, “Iterative refinement of point correspondences for 3d statistical shape models,” in *Medical Image Computing and Computer-Assisted Intervention*, pp. 417–425, 2011.