2D-3D Registration using Gradient-based MI for Image Guided Surgery Systems

Yeny Yim[1*], Xuanyi Chen[1], Mike Wakid[1], Steve Bielamowicz[2], James Hahn[1]
[1]Department of Computer Science, The George Washington University, Washington DC
[2]Division of Otolaryngology, The George Washington University, Washington DC
yenyyim@gwu.edu, xuanyi@gwmail.gwu.edu, mwakid@comcast.net,
gr8gosh@gmail.com, hahn@gwu.edu

ABSTRACT

Registration of preoperative CT data to intra-operative video images is necessary not only to compare the outcome of the vocal fold after surgery with the preplanned shape but also to provide the image guidance for fusion of all imaging modalities. We propose a 2D-3D registration method using gradient-based mutual information. The 3D CT scan is aligned to 2D endoscopic images by finding the corresponding viewpoint between the real camera for endoscopic images and the virtual camera for CT scans. Even though mutual information has been successfully used to register different imaging modalities, it is difficult to robustly register the CT rendered image to the endoscopic image due to varying light patterns and shape of the vocal fold. The proposed method calculates the mutual information in the gradient images as well as original images, assigning more weight to the high gradient regions. The proposed method can emphasize the effect of vocal fold and allow a robust matching regardless of the surface illumination. To find the viewpoint with maximum mutual information, a downhill simplex method is applied in a conditional multi-resolution scheme which leads to a less-sensitive result to local maxima. To validate the registration accuracy, we evaluated the sensitivity to initial viewpoint of preoperative CT. Experimental results showed that gradient-based mutual information provided robust matching not only for two identical images with different viewpoints but also for different images acquired before and after surgery. The results also showed that conditional multi-resolution scheme led to a more accurate registration than single-resolution.

Keywords: Registration, Image-Guided Surgery, Mutual Information, Multi-resolution, Preoperative CT, Intra-operative Endoscopic Image, Vocal fold

1. INTRODUCTION

Medialization laryngoplasty is a surgical procedure designed to restore voice in patients with vocal fold paralysis. The objective of the procedure is to place the vocal fold into a more medial position by implanting a patient-specific structural support lateral to the paretic vocal fold through a window cut in the thyroid cartilage. However, the failure rate of the procedure is as high as 24% even for experienced surgeons and the voice outcomes are dependent on the exact placement of the implant relative to the position of the underlying vocal fold[1]. During the surgery, the most challenging issues are to determine the optimal configuration and placement of implant.

In order to deal with these issues and thus reduce the revision rate and improve the outcome of the procedure, it is very helpful to provide an image-guided surgery (IGS) that allows the surgeon to place the implant in the desired

* yenyyim@gwu.edu;

location and assess the state of the vocal fold intra-operatively. For the image guided surgery of the vocal fold, it is necessary to register the 3D preoperative images to 2D endoscopic images so that the resultant shape of the vocal fold can be compared to the shape determined by a preoperative planning.

A few studies have recently suggested registering the 3D computed tomography (CT) data to the 2D endoscopic images[2-4]. Herferty et al.[4] proposed a CT-endoscopy registration method without camera tracking for image-guided bronchoscopy. The registration was performed by finding the viewpoint of the virtual camera which leads to the maximum mutual information (MI) between the bronchoscopic image and CT rendered image. The MI measure, which is calculated based on the joint probabilities of the intensities in the two images, has been successfully used for registering images from different modalities[5-7]. However, the use of naïve MI measure can lead to unstable and inaccurate result for registration of the vocal fold between 3D CT scans and 2D endoscopic image due to the difference of shape of the vocal fold and the lighting pattern in two images.

In this paper, we propose a novel registration method of the 3D preoperative CT scans to 2D endoscopic image during the medialization laryngoplasty, which provides image guidance for surgeons by employing a viewpoint matching with the gradient-based MI. We calculate the normalized MI (NMI) in the gradient images as well as the original images and optimize the viewpoint using the weighted sum of the two MI values. To emphasize the effect of the vocal fold and allow a robust rigid matching regardless of the different lighting patterns, we find the overlapping region of high gradient pixels in two images and assign more weight to the region during MI calculation. To find the viewpoint with maximum NMI, we apply a downhill simplex method in a conditional multi-resolution scheme which leads to a less-sensitive result to local maxima.

2. METHODS

We propose a novel registration method between the 3D preoperative CT and 2D postoperative endoscopic images using gradient-based MI. The 3D CT scan is aligned to 2D video image by finding the optimal viewpoint V of the virtual camera which corresponds to that of real camera. The goal of the viewpoint matching is to align the real endoscopic image to the rendered image of CT scans. The proposed method of this paper consists of the following two main steps; 1) surface extraction and generation of rendered image from 3D CT data, and 2) registration of the vocal fold with the gradient NMI-based viewpoint matching.

3.1 Surface extraction and generation of rendered image from 3D CT scans

To eliminate irrelevant structures outside the region of interests (ROI), the airway and vocal fold are segmented from the CT data using the 3D branch-based region growing[8] and the surfaces of the segmented regions are extracted using marching cube algorithm[9] during the preoperative planning stage. The extracted surface from the 3D CT data is then rendered onto the 2D images using iso-surface rendering.

For the surface extraction, three orthogonal axis-aligned views are first visualized from the CT data. From these 2D views, the surgeon can select a bounding region around the vocal fold to define the ROI. For this region, an isovalue which corresponds to the CT density values of the airway can be manually selected. Interactive manipulation of the isovalue helps to determine which value is most suitable to segment the vocal fold from the surrounding structures. A preview of the selected iso-surface within the ROI is given by a direct volume rendering[10]. Once the ideal isovalue for the vocal fold has been determined, a polygonal mesh is extracted using the marching cubes algorithm.

The 2D virtual image is rendered from the 3D CT data at a specific viewpoint of the virtual camera in order to register it to the 2D endoscopic image. To generate the 2D rendered image, the virtual camera is arbitrarily adjusted through the vocal fold in the 3D CT data. The surface of the vocal fold is rendered onto a 2D virtual image at each viewpoint $V = (X, Y, Z, \alpha, \beta, \gamma)$ where (X, Y, Z) are the positions of the viewpoint and (α, β, γ) are the Euler angles of the viewpoint along the x-, y-, and z- axes. During the rendering, the world coordinates of the 3D CT data are transformed into the 2D projected positions. The world coordinates of the surface points are calculated by multiplying the position and the sampling intervals of the x-, y- and z-directions. The camera coordinates are estimated by transforming the world

coordinates with the matrices for rotation and translation. The surfaces of the vocal fold are rendered based on a Lambertian shading model. This model can generate the CT rendered image which has similar intensity characteristic to the endoscopic image except small specular reflections and mucosal detail of the vocal fold.

3.2 Registration with the gradient NMI-based viewpoint matching

The registration of the 2D CT rendered image to the 2D endoscopic image is necessary in order to find the corresponding viewpoint of the 3D CT data given the endoscopic image. The registration is performed by matching the viewpoints of the endoscope and the virtual camera of the CT rendered image. Optimal viewpoint is determined by finding the viewpoint with the maximum similarity measure between the endoscopic image and the CT rendered image. The choice of an image similarity measure depends on the modality of the images to be registered.

The MI measure has been widely used for registration of the multimodality images. The MI measure between two images indicates the amount of information that one image contains about the other image, as shown in Eq. 1,

$$MI(A;B) = H(A) + H(B) - H(A,B) \tag{1}$$

where *H(A)* and *H(B)* are the marginal entropies of the two images, *H(A, B)* is the joint entropy of two images. The joint entropy which means the amount of uncertain information is calculated based on the joint probability *P(i, j)* of the intensities in the two images. Here, i and j is the pixel which belongs to the image A and image B, respectively. The probability is estimated by generating the 2D joint histogram that represents the combinations of gray values in each of the two images for all corresponding points. As the MI is maximized, the two images can be considered to be matched.

As the naïve MI tends to be sensitive to the amount of image overlap, it is difficult to register two images only with this measure robustly. A decrease in overlap reduces the statistical power of the probability distribution estimation and thus reduces the MI's ability to recover from larger initial misalignment. Therefore, we use an NMI that is independent of changes in the region of overlap in the two images. The NMI is calculated by dividing the joint entropy from the sum of two marginal entropies, as shown in Eq. 2. The maximization of NMI finds a transformation where joint entropy is minimized with respect to the marginal entropies.

$$NMI(A;B) = \frac{H(A) + H(B)}{H(A,B)} \tag{2}$$

Even though NMI is a robust measure for multi-modality registration, it can lead to an unstable registration result when it is applied to preoperative CT data and endoscopic images that have quite different illumination patterns. To find a corresponding viewpoint of the 3D CT data given the endoscopic images and allow a robust registration regardless of the differences in surface illumination, we propose a gradient NMI-based viewpoint matching. This method calculates the NMIs in the gradient images as well as in the original intensity images, and optimizes the viewpoint using the weighted sum of the two NMI values. The weighting factor for gradient image is set to 0.2 experimentally. The calculation of the gradient NMI helps to find the corresponding features in the two images with different illumination pattern.

To emphasize the effect of the vocal fold, we apply a locally adaptive weighting of NMI values. First, we calculate the gradient image using the Sobel operator[12] from the original image, as shown in Fig. 1(a) and (b), and find bounding box of the high gradient regions which correspond to the vocal fold from the CT rendered and endoscopic image, as shown in Fig. 1(c). Then, we find the overlap of the bounding boxes in two images, as shown in Fig. 1(d), and assign larger weight W_1 to the region than the weight W_2 for the remaining regions of the image. By applying these weighting factors to update 2D joint histogram, the vocal fold with higher gradients is emphasized during the calculation of NMI.

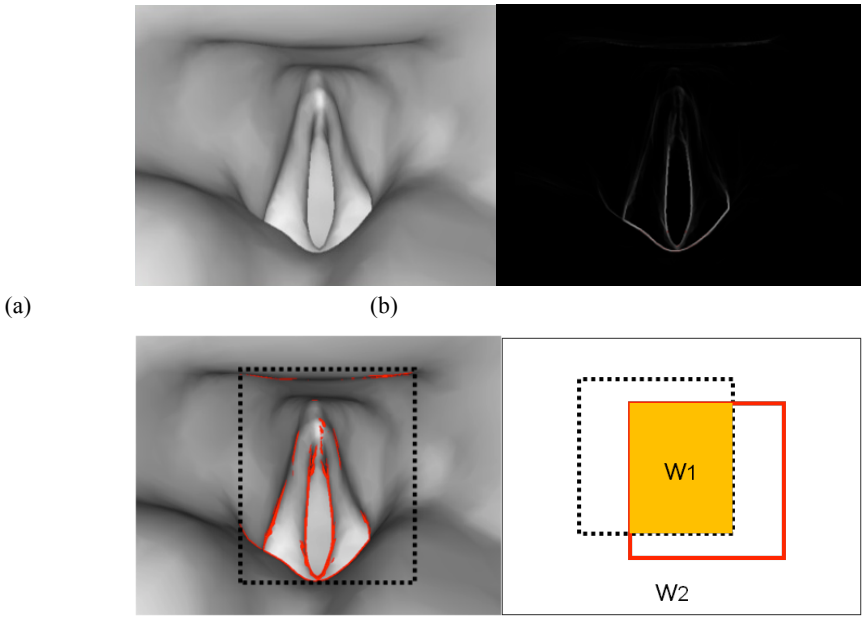

Fig 1. The calculation of the mutual information between the endoscopic image and (a) original rendered image of CT scans; (b) NMI is calculated in the gradient image as well as the original image; (c) The region of interest (indicated by dotted black line) with high gradient pixels (indicated by solid red contours) is detected; (d) The overlap of two high gradient regions (indicated by orange) is calculated and higher weight is assigned to the overlap region during calculation of NMI.

To optimize the registration process and thus find an optimal viewpoint with the largest NMI, we adapt downhill simplex method which is a well-defined optimization method and does not require derivative information. The method starts the optimization process with $N+1$ points, defining an initial simplex of N-dimensional parameter space. This simplex is deformed iteratively by reflection, expansion or contraction steps in order to find the optimum. The optimization converges when the fractional difference between the minimum and the maximum value evaluated at the vertices of the simplex is smaller than some threshold. We initialize the simplex with the origin which has the zero value for each parameter and with offsets of +15mm for translation and +20 degree for rotation in each of the parameter directions separately. The implementation of naïve downhill simplex method is based on the algorithm described in Press et al.[11].

To find the viewpoint with the optimal NMI while avoiding convergence to local maxima, the optimization is applied in a conditional multi-resolution scheme. First, the joint histograms are constructed at lower resolution by using only a fraction of the voxels in the image. After convergence at lower resolution, the optimization proceeds at higher resolution if the NMI measure is less than a pre-defined threshold. The optimization at lower resolution provides a good starting position for the optimization at higher resolution so that the NMI measure can converge to the global maxima using a multi-resolution scheme. The sub-sampling factors were set to 4 pixels at lower resolution and 2 pixels at higher resolution.

The gradient NMI-based viewpoint matching helps to find the corresponding features in the two images with different surface illumination by calculating the NMI in the gradient image as well as the original image. It also allows a robust registration result by applying a locally adaptive weighting of NMI values and emphasizing the effect of the vocal fold. The downhill simplex optimization in a conditional multi-resolution scheme enables to find the viewpoint with the optimal NMI while avoiding convergence to local maxima.

3. RESULTS

The proposed method was applied to register the preoperative CT to virtual endoscopic image which was generated from the postoperative CT. The CT images were obtained on a LightSpeed VCT Scanner (GE Medical Systems, US) from one normal subject and one patient with vocal fold paralysis. The image resolution was 512×512 and the slice thickness was 0.625 mm.

To validate the registration accuracy, we evaluated the sensitivity to initial viewpoint of preoperative CT. We tested the sensitivity on 1) the two rendered images from identical preoperative CT with different viewpoints and 2) the two rendered images from pre and postoperative CT. We assumed that the registration result was acceptable when the positional and angular error between target and source images were less than the threshold. This threshold value was set to 5 for the first test and 8 for the second test. The positional error was calculated by using the difference of the view positions (X, Y, Z) of two images, and the angular error was calculated by using the difference of the view angles (α, β, γ). In the first test, the initial offset which gives us good registration result was $-7.0 \text{ mm} \le X \le 7.0 \text{ mm}$, $-9.0 \text{ mm} \le Y \le 6.0 \text{ mm}$, and $-13.0 \text{ mm} \le Z \le 15.0 \text{ mm}$ for translation in x, y, z direction, and $-21.0° \le \alpha \le 16.5°$, $-22.5° \le \beta \le 25.5°$, and $-28.5° \le \gamma \le 25.5°$ for x, y, z rotation, respectively. In the second test, the offset was $-13.0 \text{ mm} \le X \le 11.0 \text{ mm}$, $-11.0 \text{ mm} \le Y \le 13.0 \text{ mm}$, and $-12.0 \text{ mm} \le Z \le 10.0 \text{ mm}$ for x, y, z translation, and $-6.0° \le \alpha \le 11.0°$, $-12.0° \le \beta \le 15.0°$, and $-13.0° \le \gamma \le 17.0°$ for x, y, z rotation.

Fig. 2 showed the effect of multi-resolution scheme to registration results. The difference image of Fig. 2(d) showed that conditional multi-resolution scheme provided better registration result than single-resolution scheme in Fig. 2(c). The maximum normalized MI between target and source images was 1.603 in Fig. 2(d) while the measure was 1.167 in Fig. 2(c).

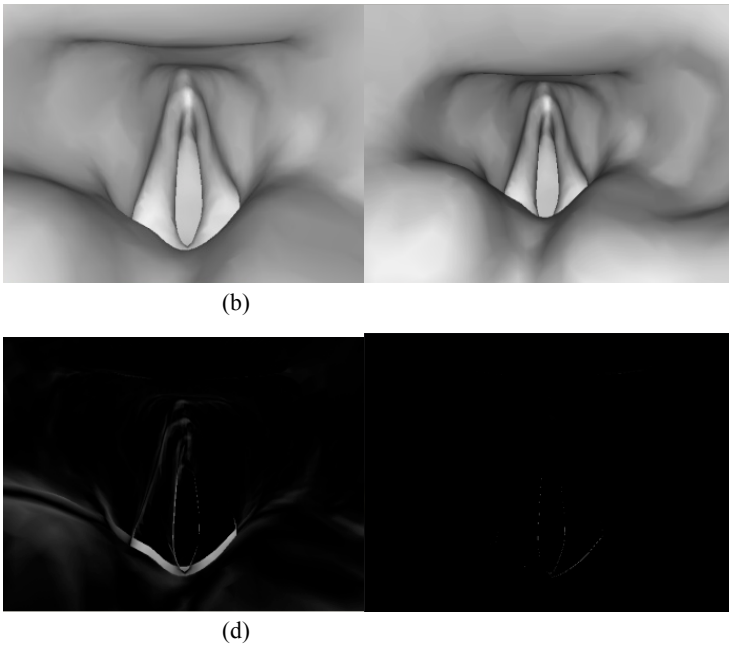

Fig 2. The comparison of the registration results between single-resolution and conditional multi-resolution scheme; (a) The rendered image of CT scans is the target image; (b) The image translated by 11mm along z axis from (a) is the source image; After registering (b) to (a), the difference between target and registered source image was calculated in (c) a single-resolution and (d) a conditional multi-resolution scheme.

4. CONCLUSIONS

We developed a robust and accurate registration method of 3D preoperative CT to 2D endoscopic images. The proposed viewpoint matching based on a gradient-based NMI can emphasize the effect of the vocal fold by assigning more weight to the high gradient region and allow a robust matching regardless of the differences in surface illumination. The optimization method in a multi-resolution scheme helps to optimize the viewpoint accurately by avoiding the convergence to the local maxima. Experimental results showed that the proposed gradient-based MI provided robust matching not only for the two identical images with different viewpoints but also for the two different images acquired before and after surgery. The results also showed that conditional multi-resolution scheme led to more accurate registration results than single-resolution scheme.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health - grant R01-DC007125 - to develop computer-based tools for medialization laryngoplasty.

REFERENCES

[1] Anderson, T. D., Spiegel, J. R., Sataloff, R. T., Thyroplasty revisions: frequency and predictive factors. J Voice. 17(3): 442-8 (2003).

[2] Deligianni, F., Chung, A. J., Yang, G. Z., Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: phantom validation. *IEEE Trans Med Imaging*. 25(11): 1462-71 (2006).

[3] Wang, X., Zhang, Q., Han, Q., Yang, R., Carswell, M. et al., Endoscopic video texture mapping on prebuilt 3-D anatomical objects without camera tracking. *IEEE Trans Med Imaging*. 29(6): 1213-23 (2009).

[4] Helferty, J. P., Sherbondy, A. J., Kiraly, A. P., Higgins, W. E., Computer-based system for the virtual-endoscopic guidance of bronchoscopy, Computer Vision and Image Understanding 108, 171–187 (2007).

[5] Viola, P., Wells III, W. M., Alignment by maximization of mutual information, International Journal of Computer Vision 24 (2), 137–154 (1997).

[6] Studholme, C., Hill, D. L. G., Hawkes, D. J., An overlap invariant entropy measure of 3D medical image alignment, Pattern Recognition 32 (1), 71–86 (1999).

[7] Maes, F., Vandermeulen, D., Suetens, P., Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information, Medical Image Analysis 3(4), 373–386 (1999).

[8] Yim, Y. and Hong, H., Correction of segmented lung boundary for inclusion of pleural nodules and pulmonary vessels in chest CT images. *Comput Biol Med*. 38(8): 845-57 (2008).

[9] Lorensen, W. E. Cline, H. E., Marching cubes: a high resolution 3D surface construction algorithm. *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. 21: 163-169 (1987).

[10] Levoy, M., Direct visualization of surfaces from computed tomography data *Proceedings of the SPIE*. 914: 828-841 (1988).

[11] Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., Numerical recipes in C, second edition. Cambridge University Press, Cambridge (1992).

[12] Gonzalez, R. C. and Woods, R.E., Digital image processing (1992).