Registration of 3D CT Data to 2D Endoscopic Image using a Gradient Mutual Information based Viewpoint Matching for Image-Guided Medialization Laryngoplasty

Yeny Yim, Mike Wakid and Can Kirmizibayrak

Department of Computer Science, The George Washington University, Washington, USA yenyyim@gwu.edu mwakid@comcast.net kirmizi@gwmail.gwu.edu

Steven Bielamowicz

The George Washington University Medical Center, Washington, USA gr8gosh@gmail.com

James Hahn

Department of Computer Science, The George Washington University, Washington, USA hahn@gwu.edu

Received 25 October 2010; Revised 30 November 2010; Accepted 10 December 2010

We propose a novel method for the registration of 3D CT scans to 2D endoscopic images during the image-guided medialization laryngoplasty. This study aims to allow the surgeon to find the precise configuration of the implant and place it into the desired location by employing accurate registration methods of the 3D CT data to intra-operative patient and interactive visualization tools for the registered images. In this study, the proposed registration methods enable the surgeon to compare the outcome of the procedure to the pre-planned shape by matching the vocal folds in the CT rendered images to the endoscopic images. The 3D image fusion provides an interactive and intuitive guidance for surgeon by visualizing a combined and correlated relationship of the multiple imaging modalities. The 3D Magic Lens helps to effectively visualize laryngeal anatomical structures by applying different transparencies and transfer functions to the region of interest. The preliminary results of the study demonstrated that the proposed method can be readily extended for image-guided surgery of real patients.

Copyright(c)2010 by The Korean Institute of Information Scientists and Engineers (KIISE). Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Permission to post author-prepared versions of the work on author's personal web pages or on the noncommercial servers of their employer is granted without fee provided that the KIISE citation and notice of the copyright are included. Copyrights for components of this work owned by authors other than KIISE must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires an explicit prior permission and/or a fee. Request permission to republish from: JCSE Editorial Office, KIISE. FAX +82 2 521 1352 or email office@kiise.org. The Office must receive a signed hard copy of the Copyright form.

Journal of Computing Science and Engineering, Vol. 4, No. 4, December 2010, Pages 368-387.

Categories and Subject Descriptors: Human Computing [Image Processing & Virtual Reality & Visualization]

General Terms: Image Registration, Visualization

Additional Key Words and Phrases: Image-guided surgery, Medialization laryngoplasty, Image Registration, Visualization, Vocal folds

1. INTRODUCTION

An estimated 7.5 million people in the United States have voice disorders, and onethird of the patients suffer from vocal fold paresis or paralysis each year [Statistics on voice, speech, and language]. Medialization laryngoplasty is a surgical procedure designed to restore voice in such patients. The objective of the procedure is to place the vocal folds into a more medial position by implanting a patient-specific structural support lateral to the paretic vocal folds 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 folds [Anderson et al. 2003]. The two most challenging issues are the determination of the optimal implant configuration and accurate placement of the implant. First, finding the precise configuration of the implant is important but difficult. It determines not only the degree of medialization but concurrently affects the vibratory characteristics of the vocal folds. Alterations in implant configuration are made by hand-carving the implant during surgery in response to phonation by the patient [Bielamowicz 2004]. The configuration of the implant is thus based on trial-and-error and voice outcomes are variable after a period of recovery. Another issue present is the difficulty in predicting the exact location of the underlying vocal folds since the vocal folds are located behind the thyroid cartilage. The anatomy of the thyroid cartilage varies in size and shape among individuals as does the precise position of the vocal folds. Furthermore, the vocal folds cannot be seen through the thyroplasty surgical window. The position of the implant window and the implant is selected based upon the perceived location of the underlying vocal folds. The location is therefore subject to a significant level of uncertainty.

In order to deal with the above two challenging issues and thus reduce the revision rate and improve the outcome of the procedure, it is very helpful to provide an imageguided surgery (IGS) system that allows the surgeon to place the implant in the desired location and assess the state of the vocal folds intra-operatively. The two main parts of these kinds of systems are the registration of the three-dimensional (3D) preoperative data to the two-dimensional (2D) intra-operative images, and the visualization of the registered images [Peters 2006] (see Fig. 1). For image-guided laryngoplasty, an accurate registration of the 3D magnetic resonance (MR) or computed tomography (CT) images to endoscopic images is necessary in order to compare the resultant shape of the vocal folds to the shape determined by a pre-operative planning system. An interactive visualization of the intra-operative images along with the pre-operative images is needed to provide guidance to the surgeon by presenting all available imaging modalities during the surgery. This can help the surgeon form a coherent understanding of the global relationship of the patient's anatomy, thereby allowing for



Figure 1. Image-guided surgery system whose key modules are registration and visualization. Preoperative data is registered to the intra-operative endoscope, and all these imaging modalities are fused into one and visualized. (I need to make this image into high-resolution image.)

a more accurate placement of the implant.

In this paper, we propose a novel registration method of the 3D pre-operative CT scans to 2D intra-operative patient image during the medialization laryngoplasty, which provides image guidance for surgeons by employing a viewpoint matching with the gradient-based mutual information (MI). This method can be used to track the position and orientation of the laryngoscope tip instead of using a tracker as well as to register the vocal folds that are located behind the thyroid cartilage. Since vision-based tracking is used for this study, we do not need additional tracking for the endoscope. The intra-operative endoscopic images are fused into the pre-operative CT data in order to help the surgeon form a coherent understanding of the global relationship of the patient's anatomy. The 3D Magic Lens provides an effective visualization tool so that the surgeon can place the implant at a more accurate position.

2. RELATED WORK

A number of IGS systems which include registration and visualization modules have been proposed [Jin et al. 2006; Image-guided surgery toolkit (IGSTK); Kelly 1986; Maurer et al. 1997; Darabi et al. 1997; Stefansic et al. 2000; Sato et al. 1998; Solomon et al. 2000; Klein et al. 2009; Wang et al. 2009; Deguchi et al. 2006; Hautmann et al. 2005; Helferty et al. 2007; Jin et al. 2009; Kirmizibayrak et al. 2010]. The Image-Guided Surgery Toolkit (IGSTK) was designed and developed to interface tracking devices with software applications incorporating medical images Image-guided surgery toolkit (IGSTK). This toolkit provides a common functionality for IGS applications, containing the tracking as well as the registration and the visualization. Most of the

IGS systems have employed the fiducial markers and camera tracking to relate the pre-operative data with the patient on the operating room (OR) table. Fiducial markers have been used to register these two different data by establishing their correspondence. They can be categorized into extrinsic invasive, extrinsic non-invasive, and intrinsic markers. Extrinsic invasive markers are usually fixed to the patient's large bones [Kelly 1986; Maurer et al. 1997], while extrinsic non-invasive markers are attached to the patient's skin [Darabi et al. 1997; Stefansic et al. 2000] or the surgical devices [Sato et al. 1998; Solomon et al. 2000]. The bone-fixed fiducial markers would cause potential damage to the thin and delicate laryngeal cartilage, and the skin-affixed or device-affixed markers may move significantly relative to the laryngeal cartilage.

As an alternative of the extrinsic markers, intrinsic markers such as anatomical landmarks, image intensity, and gradient can be used for tracking and registration of two imaging modalities. Numerous researchers have been proposed 2D-3D registration methods [Wang et al. 2009; Helferty et al. 2007; Dey and Napel 2006; Munbodh et al. 2008; Munbodh et al. 2009; Fu and Kuduvalli 2008; Tomaževič et al. 2006; Deligianni et al. 2006]. However, most of them were dedicated to register 3D pre-operative data to 2D X-ray images or ultrasound images. A few studies have recently suggested registering the 3D CT data to the 2D endoscopic images [Wang et al. 2009, Helferty et al. 2007, Deligianni et al. 2006]. Herferty et al. proposed a method for CT-endoscopy registration without camera tracking [Helferty et al. 2007]. The registration was performed by finding the viewpoint of the virtual camera which leads to the maximum similarity measure between the bronchoscopic image and CT rendered image.

For 3D image fusion of intra-operative images with 3D pre-operative data, augmented reality has been widely used [Azuma 1997]. Augmented reality has a constraint that its visualization is fixed to the pose of the camera system. Moreover, in image fusion of endoscopic images to the 3D data, the endoscopic device is limited by a small field of view and thus it is more difficult for the surgeon to recognize the 3D position of the endoscope relative to the surrounding anatomical structures. To deal with these limitations, a few alternative approaches were proposed which fuse intra-operative images onto 3D surface model [Wang et al. 2009; Dey et al. 2002; Paul et al. 2005]. These approaches enabled surgeons to see the fused patient information from arbitrary viewpoints without being limited by the location of the camera.

Volumetric CT and MR scans include unnecessary information that might occlude the important anatomical structures and thus obstruct the surgeon to understand the fused image exactly. To effectively visualize the region of interest (ROI) in the preoperative data, 2D [Eric et al. 1993; Perlin and Fox 1993] and 3D [Viega et al. 1996; Wang et al. 2005] Magic Lens filters were proposed. These filters were used to apply different transparencies and transfer functions for the ROIs and the rest of the volume. Joshi et al. [Joshi et al. 2008] proposed a visualization method which allows irregular cropping and focusing. Rieder et al. [Rieder et al. 2008] used a cylindrical cut geometry to avoid the occlusion of the surgical target from the surgeon. Even though these methods can deal with the occlusion problem, the information in the discarded regions is lost.

In the medical domain, IGS systems find their root in stereotactic neurosurgery, in

which the stereotactic frame is used as the standard reference system [Spiegel et al. 1947; Bergstrom and Greitz 1976; Brown 1979; Horsley and Clark 1908; Leksell 1951]. To date, image-guided technology has been applied to various medical scenarios including frame-based [Kelly 1986] and frameless [Roberts et al. 1986] stereotactic neurosurgery, spinal surgery [Lavallee et al. 1995], orthopedic surgery [Bettega et al. 2002], sinus surgery [Anon et al. 1994], brain biopsy [Maurer et al. 1993], breast lesion excision [Burns 1997], liver metastases, oral implantology [Birkfellner et al. 2001], maxillofacial surgery and prostate therapy [Onik et al. 1996]. Although a variety of IGS systems have been proposed thus far, few systems have been dedicated to image-guided medialization laryngoplasty to our knowledge. Our previous work proposed methods for surface reconstruction of the thyroid cartilage from the stereo camera images and registration of the reconstructed surface with one extracted from the CT data [Jin et al. 2006; Jin et al. 2009]. We also proposed an interactive visualization method which improves the surgeon's mental registration between the virtual and real worlds, thus guiding the surgeon for a more accurate surgical outcome [Kirmizibayrak et al. 2010].

3. METHODS

The proposed method of this paper consists of the following four main steps as outlined in Fig. 2; 1) surface extraction of the vocal folds from the 3D CT data using a marching cube algorithm [Lorensen and Cline 1987], 2) rendering of the CT scans using isosurface rendering, 3) registration of the vocal folds with the gradient-based MI, and 4) visualization of the registered images using 3D image fusion and 3D Magic Lens [Kirmizibayrak et al. 2010]. First, the airway and vocal folds are segmented from the CT data using the 3D branch-based region growing [Yim and Hong 2008] and the surfaces of the segmented regions are extracted using marching cube algorithm during the pre-operative planning stage. The extracted surface from the 3D CT data is then rendered onto the 2D images using isosurface rendering. The position and orientation of the viewpoint are pre-defined so that the viewpoint is very close to that of the endoscopic image. Third, the relationship between the CT rendered image and the intra-operative endoscopic image is found by using image based registration. In order to register the two input images, MI is chosen as a similarity measure [Wells et al. 1996; Viola 1997] and repeatedly calculated according to the different viewpoints. The viewpoint with the maximum MI is found among the viewpoints with various positions and angles during the optimization process. As a result of the registration, the optimal viewpoint of the 3D CT data is located and the 2D CT rendered image, which is generated with the viewpoint, is matched to the target endoscopic image. This step can be used to track the position and orientation of the laryngoscope tip by finding the corresponding viewpoint of the CT rendered image given the endoscopic image. Finally, the intra-operative images are projected onto the pre-operative CT data, fusing the acquired information from multiple imaging modalities into one combined and correlated view. A ROI including the airway and vocal folds is localized in the fused images and visualized using a 3D Magic Lens which is an interactive data exploration tool.



Figure 2. A pipeline of the proposed method of 3D pre-operative CT scans to a 2D intra-operative endoscopic image for image-guided medialization laryngoplasty.

3.1 Pre-Processing for the Registration

At the pre-operative stage, the airway and vocal folds are segmented from the CT data and the surface of the segmented regions are extracted in order to eliminate irrelevant structures outside the ROIs and render the isosurface of the vocal folds.

For the surface extraction, three orthogonal axis-aligned views are first visualized from the CT data, as shown by Fig. 3. From these 2D views, the surgeon can select a bounding region around the vocal folds to define the ROI. For this region, an isovalue which corresponds to the data values of the airway can be found by moving a slider bar in the user interface within the range of the CT density values. Interactive manipulation of the isovalue helps to determine which value is most suitable to segment the vocal folds from the surrounding structures. A preview of the selected isosurface within the ROI is given by a direct volume rendering [Levoy 1988]. Once the ideal isovalue for the vocal folds 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 folds in the 3D CT data. The surface of the vocal folds 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 folds are rendered based on a Lambertian shading model.



Figure 3. The user interface of the proposed system provides the 2D axis-aligned views including (a) axial view, (b) coronal view, (c) sagittal view and (d) 3D perspective view of the vocal folds that are extracted from the pre-operative CT scans. The yellow rectangle in each 2D view indicates the bounding region that is selected by the surgeon to obtain the surface mesh.

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 folds.

3.2 Registration of the Vocal Folds with the Gradient MI-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.

3.2.1 *NMI-based Viewpoint Matching.* To quantify the degree of similarity between intensity patterns in two images, various similarity measures have been used for image registration including MI, normalized MI (NMI), cross correlation (CC), and sum of squared intensity differences (SSD). The choice of an image similarity measure depends on the modality of the images to be registered. The MI and NMI measure have been successfully used for registration of the multimodality images [Wells et al. 1996; Studholme et al. 1999; Maes et al. 1999; Viola and Wells III 1997]. The MI measure between two images indicates the amount of information that one image contains about the other image, as shown in Eq. (1),

$$MI(A; B) = H(A) + H(A) - H(A, B)$$
(1)

$$H(A, B) = -\sum_{i \in A, j \in B} P(i, j) \log P(i, j)$$
(2)

where H(A) and H(B) are the marginal entropies of the two images, H(A, B) is the joint entropy of two images, and i and j is the pixel which belongs to the image A and image B, respectively. 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, as shown in Eq. (2). The probability p 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 [Studholme et al. 1999]. A decrease in overlap reduces the statistical power of the probability distribution estimation. Therefore, this reduces the MI's ability to recover from larger initial misalignment. For this reason, 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. (3).

$$NMI(A; B) = \frac{H(A) + H(B)}{H(A, B)}$$
(3)

The maximization of NMI finds a transformation where joint entropy is minimized with respect to the marginal entropies.

To optimize the registration process and thus find an optimal viewpoint with the largest NMI, several methods have been suggested including Powell's method, downhill simplex method, steepest descent, conjugate gradient, and simulated annealing [Press 1992]. In this study, we adapt Powell's method and downhill simplex method for the optimization of the NMI. Powell's method finds the minimum of the measure along a set of N, linearly independent directions, repeatedly. For the optimization of each direction, a one-dimensional line minimization method, the Brent algorithm, is used. The implementation of naïve Powell's method is based on the algorithm described in Press et al. [Press 1992]. In Powell's method, it is important to determine the order of each direction for the optimization due to differences in image resolution of different directions. The order of the directions which correspond to the transformation parameters is optimized over accuracy and robustness of the registration. Optimization of the x- and y-axes translation, X and Y, and of the z-axis rotation, γ , is better conditioned than optimization of the z-axis translation, Z, or of the x- and y-axis rotation, α and β [Maes et al. 1999]. We optimize the transformation parameters in each iteration in the order (X, Y, γ , Z, α , β). The downhill simplex method is a commonly used and well-defined optimization method which 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. The implementation of the naïve downhill simplex method is based on the algorithm given in Press et al [Press 1992]. We initialize the simplex with the origin which has the zero value for each parameter and with offsets of +15 mm for translation and +20 degree for rotation in each of the parameter directions separately.

3.2.2 Gradient NMI-Based Viewpoint Matching in Multi-resolution Scheme. Even though NMI is a robust measure for multi-modality registration, it can lead to an unstable registration result when it is applied to pre-operative CT data and endoscopic images that have quite different lighting patterns, as shown in Fig. 4. 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.

The gradient NMI-based viewpoint matching 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. 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 larger weight to the region during the calculation of the NMI. To calculate the gradient image, the Sobel operator, which is one of the most common edge detectors, is used [Gonzalez and Woods 1992].

To find the viewpoint with the optimal NMI while avoiding convergence to local maxima, the optimization methods are applied in a 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 by taking more voxels of the image into account. We apply a two-level multi-resolution hierarchy by sub-sampling the image with sampling factors f_x , f_y , and f_z along the x-, y- and z-axis dimensions using nearest neighbor (NN) interpolation. If the optimum at lower resolution is close to the optimum at full resolution, it is to be expected that most of the calculation will be performed at lower resolution and that the number of evaluations at full resolution. The optimization is done entirely at full resolution. The optimization at



Figure 4. The comparison of (a) the 2D rendered image from the pre-operative CT data and (b) the 2D virtual endoscopic image generated from the postoperative CT data.

Journal of Computing Science and Engineering, Vol. 4, No. 4, December 2010

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 multiresolution scheme.

The proposed viewpoint matching based on a gradient-based NMI in multi-resolution scheme 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.

3.3 Visualization of the Registered Images

The registration results provide a correlation between the surgeon's surgical view, endoscopic video of the airway and vocal folds, and the patient's pre-operative 3D CT. In a typical medialization laryngoplasty procedure, the surgeon must perform the difficult task of mentally registering this information together. For obtaining an intuitively combined and correlated view of the three imaging modalities, we apply the proposed viewpoint matching to fuse multiple 2D camera images with the volumetric data. In addition, we provide an interactive tool, called the 3D Magic Lens, to assist in guiding the surgeon to the correct implant location.

The 3D image fusion process is performed after the rendered image of the surface has been registered to the endoscopic video. Given two matched viewpoints, the images acquired from the camera are placed onto the surface extracted from the preoperative CT through projective texture mapping. The two images are correlated by overlaying the 2D endoscopic image onto the corresponding region of the extracted surface. The texture mapped surface is then recombined with the original volumetric information using ray casting based on the graphical processing unit (GPU) [Kruger and Westermann 2003; Lee et al. 2009]. We integrate the color rendered on the extracted surface model with the sampled color and opacity in the 3D CT data. Since the intra-operative view is merged onto the 3D surface model, the visualization is given through a virtual camera, which allows the user to view the fused information at the arbitrary position and orientation.

Coherent display of the data presents additional challenges since the fusion of multiple data sources can create visual clutter. Moreover, the surgeon may only be interested in specific information. To address these problems, we use the 3D Magic Lens paradigm, which gives the surgeon the ability to interactively control visual properties within a local region of the data [Kirmizibayrak et al. 2010]. The lens can be positioned virtually within the computer display to visualize a ROI within the data. The 3D Magic Lens is positioned over the anterior of the neck allowing the surgeon to visualize the structures of interest by applying partial transparency to the surface within the lens region. The anatomical structures such as thyroid cartilage, airway, and vocal folds can be further emphasized by applying a transfer function, which maps data values to a user-specified visual appearance, to the lens within the global context of the patient data, as shown in Figure 5.

The two main components of the proposed visualization method is the 3D image fusion of intra-operative camera images to the pre-operative CT and the 3D Magic Lens. The proposed method enables the surgeon to intuitively visualize correlated



Figure 5. A visualization of the larynx using the 3D Magic Lens. Two different lenses are used for the airway (indicated by green) and the thyroid cartilage (indicated by red). The anatomical structures of the larynx are visualized by applying different transparencies and transfer functions to the lens region and the rest of the dataset.

patient information as an alternative to mental reconstruction so they can be more confident in their decision making during the surgery.

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup and Evaluation Method

The proposed method was applied to register the 3D pre-operative CT to 2D intraoperative endoscopic images. The pre-operative CT scans were obtained on a LightSpeed VCT Scanner (GE Medical Systems, US) from one normal subject and one patient with vocal fold paralysis. The intra-operative endoscopic images were virtually generated according to a defined viewpoint in the post-operative CT scans which are correspondent to the pre-operative CT scans. The image resolution of the CT scans was 512×512 and the slice thickness was 0.625 mm.

The vocal folds were segmented in the pre-operative CT scans and the 3D surfaces of the segmented regions were extracted. The surface of the vocal folds was rendered from the pre-operative CT data by OpenGL rendering. The CT rendered image was registered to the 2D endoscopic image with the gradient NMI-based viewpoint matching. Visualization algorithms were implemented on the GPU to take advantage of the modern graphics hardware and achieve real-time performance. The EM tracker was used to track the stylus intra-operatively for visual feedback of 3D Magic Lens. The proposed method was developed at Microsoft Visual Studio 2008 with the support of standard libraries including MFC, OpenGL, and VTK [Schroeder et al. 2000].

For the viewpoint matching, we chose NMI as a similarity measure and calculated it in the intersection of the high gradient regions of the two 2D input images. The high gradient regions were determined by finding the bounding box for the pixels with the gradients larger than T. This threshold value was experimentally set to 0.1 to include the edge of the vocal folds while the maximum gradient is 1.0. The NMI was calculated by assigning the adaptive weights for the high gradient regions and the remaining portion of the image. In this study, the weights for the two different regions were set to 1.5 and 1.0 to emphasize the high intensity regions relative to the rest of the image.

To integrate the gradient and intensity information, the two NMI values which were calculated from the gradient and original images were weighted averaged, as shown in Eq.. The weighting factor for the gradient image C_G was experimentally set to 0.2 and that for the original image C_I was set to 0.8.

$$NMI_{Combined}(A; B) = c_G \times NMI_G(A; B) + c_I \times NMI_I(A; B)$$
(4)

To optimize the NMI measures, we applied two different methods: Powell's method and the downhill simplex method. The initial search space of the viewpoint was defined to 15 mm for translation and 20 degree for rotation in both optimization methods. The maximum number of the iteration for optimization was defined to 20 and 300. As Powell's method requires a set of one dimensional optimization, the iteration number for one set was defined to 20. On the other hand, the number of iterations in the downhill simplex method was selected to a relatively high value experimentally. The optimization was performed in a two-level resolution scheme. The NMI measures between the CT rendered image and the virtual endoscopic image were calculated at the sub-sampled positions and optimized for various viewpoints of the virtual camera until the measure converged to the optimum at lower resolution. Starting from the optimized viewpoint at the lower resolution level, the optimization was performed at higher resolution. In this study, the sub-sampling factors, f_x , f_y , and f_z , were set to 4 pixels at lower resolution and 2 pixels at higher resolution.

To validate the accuracy of the proposed registration method, we evaluated the sensitivity to the initial viewpoint of the pre-operative CT. We tested the sensitivity on 1) the two rendered images from identical pre-operative CT data with different viewpoints and 2) the two rendered images from pre- and post-operative CT data. For these two experiments, downhill simplex optimization method was used. We assumed that the registration result was acceptable when the positional and angular error between target and source images were less than a predefined threshold ε and λ . The threshold for positional error, ε , was set to 5 mm for the first test and 8 mm for the second test. The threshold for angular errors were calculated by using the difference of the view positions of the two images, and the difference of the view angles of the two images.

4.2 Results

4.2.1 Registration of the Vocal Folds with the Gradient MI-Based Viewpoint Matching. Figure 6 shows that the proposed gradient NMI-based viewpoint matching in multiresolution scheme led to a better registration result than the result using the singleresolution scheme.

Figure 7. shows that gradient NMI-based viewpoint matching led to robust registration results of the pre-operative CT data to the virtual endoscopic image in spite of the differences in surface illumination. Figure 7(b) shows the 2D CT rendered image that was translated by -3 mm along the x-dimension. The difference in images of Figure 7(e) and Figure 7(f), which indicate how well the two images are matched,



Figure 6. A comparison of the registration results between single-resolution and multi-resolution schemes; (a) the rendered image of the CT scans is the target image; (b) the image translated by 11 mm along z axis from (a) is the source image; after registering (b) to (a), the difference between the target and registered source image was calculated in (c) a single-resolution and (d) a conditional multi-resolution scheme.

show that the proposed gradient NMI-based viewpoint matching allowed more accurate results than the naïve NMI-based viewpoint matching.

The sensitivity to the initial viewpoint was evaluated using two different datasets. In the first test, the initial offset which results in good registration was -7.0 mm \leq X \leq 7.0 mm, -9.0 mm \leq Y \leq 6.0 mm, and -13.0 mm \leq Z \leq 15.0 mm for translation in x, y, z directions, and -21.0° $\leq \alpha \leq$ 16.5°, -22.5° $\leq \alpha \leq$ 25.5°, and -28.5° $\leq \gamma \leq$ 25.5° for x, y, z rotations, respectively. In the second test, the offset was -13.0 mm \leq X \leq 11.0 mm, -11.0 mm \leq Y \leq 13.0 mm, and -12.0 mm \leq Z \leq 10.0 mm for x, y, z translations, and -6.0° $\leq \alpha \leq$ 11.0°, -12.0° $\leq \beta \leq$ 15.0°, and -13.0° $\leq \gamma \leq$ 17.0° for x, y, z rotations.

4.2.2 Visualization of the Registered Images. Figure 8 shows an example of 3D image fusion. A camera image was fused onto the corresponding region of the phantom surface extracted from the pre-operative CT data. The stickers indicated by circles with three different colors were placed on the phantom to emphasize the placement of the image on the volume-rendered surface. Partial transparency was used within a lens region to reveal the thyroid cartilage and airway underneath the physically visible surface.

Figure 9 shows an example of the tracking based interaction for the 3D Magic Lens

Journal of Computing Science and Engineering, Vol. 4, No. 4, December 2010







Figure 7. The registration results of the vocal folds after applying the proposed gradient NMIbased viewpoint matching to the pre-operative CT scans and virtual endoscopic images: (a) the virtual endoscopic image generated from the post-operative CT data, (b) the 2D CT rendered image with a viewpoint translated from the ground truth by -3 mm along the x axis, (c) the registration result of (b) with the naïve NMI-based viewpoint matching, (d) the registration result of (b) with the proposed gradient NMI-based viewpoint matching, (e) the difference in images between (a) and (c), and (f) the difference in images between (a) and (d).

using two different tracking technologies. The left and right columns show the optical tracking and the EM tracking, respectively, which are used to point to the locations of the ROIs. The corresponding regions of the virtual and real images shown in the top and bottom row of Figure 9 were concurrently displayed in order to illustrate how a surgeon can select ROI locations by using trackers.

5. DISCUSSION AND CONCLUSION

We presented accurate and robust methods for registration of the pre-operative CT



Figure 8. A camera image (indicated by the white plane) was fused onto the corresponding region of the phantom surface from the pre-operative CT data.



Figure 9. Interactive selection of the 3D Magic Lens location on phantom models by two different tracking technologies: optical (a, c) and EM tracking (b, d). The corresponding regions in virtual (a, b) and real (c, d) images were indicated by red solid circles. The 3D Magic Lens displayed the thyroid cartilage selectively in the top left image and visualized the airway inside the lens in transparent green color in the top right image.

data to the intra-operative camera images and visualization of the registered images. The gradient NMI-based viewpoint matching allowed a robust matching of the vocal folds regardless of the differences in surface illumination. This method also can provide the vision-based tracking of the endoscope tip by finding the corresponding viewpoint between the two images. The proposed 3D image fusion led to combined and coherent viewing of information from the multiple imaging modalities by visualizing the fusion of the CT data and endoscopic images of the vocal folds. The 3D

Magic Lens helped the surgeon to place the implant in an accurate position by visualizing the ROI with a different transfer function and transparency to the surrounding structures. The results showed that the gradient NMI-based viewpoint matching allowed robust registration results of the pre-operative CT data to the virtual endoscopic image contrary to the naïve NMI-based viewpoint matching. In addition, the viewpoint matching in multi-resolution schemes led to better registration results than the result in single-resolution scheme.

In this preliminary study using the phantom models, we aimed to validate the results of the proposed method and determine whether this would be applicable to the image-guided surgery for real patients. The visual results in this study showed that the proposed registration and visualization methods can be used effectively for the phantom models. The results also showed the possibility that the proposed methods can be readily applicable to patients in the OR. For the image-guided surgery of patients, further study is needed to ascertain the complicated circumstances such as the movement and phonation of the patient, and deformation of the vocal folds.

For the registration of vocal folds between the pre-operative and intra-operative data, we applied gradient NMI-based viewpoint matching. Even though this method led to accurate and robust registration results for the two datasets which have different dimensions and were acquired before and during surgery, the deformation of the vocal folds was not taken into account. For the registration of the vocal folds, the deformable motion relative to the thyroid cartilage should be considered since the collisions of the vocal folds generate an elastic wave that propagates to the mucous, involving deformable motion [Montagnoli et al. 2006]. In order to intra-operatively assess the shape of the vocal folds after implant and compare it to the pre-planned shape, we need to deform the pre-operative CT data to the intra-operative surface of the patient after the implant. We are currently developing the feature-based deformable registration method of the 3D CT data to the 2D endoscopic images.

In the study, the virtual endoscopic images that were generated from the postoperative CT data were used for registration with the pre-operative data instead of the intra-operative endoscopic images. The virtual endoscopic images presented the medialized shape of the vocal folds after the implant and provided similar views to the intra-operative endoscopic images with some exceptions. These virtual images were generated from the CT data that were acquired at the resting stage of the vocal folds, for the patients not phonating, while the real endoscopic images are obtained from the patients during the operation. Furthermore, illumination patterns of the virtual endoscopic images tend to be different from those of the real endoscopic images. We plan to apply the proposed registration method to the real intra-operative endoscopic images and the pre-operative CT data, and validate the performance of the proposed registration method for the vocal folds.

We are also planning a user study to evaluate the visualization of the larynx using 3D Magic Lens, targeting inexperienced and experienced surgeons. This user study will be used to quantify the performance of the 3D Magic Lens for identifying anatomical landmarks or locating target structures.

ACKNOWLEDGEMENTS

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

REFERENCES

- ANDERSON, T. D., J. R. SPIEGEL, and R. T. SATALOFF. 2003. Thyroplasty revisions: frequency and predictive factors. J Voice. 17, 3, 442–8.
- ANON, J. B., S. P. LIPMAN, D. OPPENHEIM, AND R. A. HALT. 1994. Computer-assisted endoscopic sinus surgery. *Laryngoscope*. 104, 7, 901–5.
- AZUMA, R. 1997. Survey of Augmented Reality. Presence: teleoperators and virtual environments. 6, 4, 355–385.
- BERGSTROM, M. AND T. GREITZ. 1976. Stereotaxic computed tomography. AJR Am J Roentgenol. 127, 1, 167–70.
- BETTEGA, G., P. CINQUIN, J. LEBEAU, AND B. RAPHAEL. 2002. Computer-assisted orthognathic surgery: clinical evaluation of a mandibular condyle repositioning system. J Oral Maxillofac Surg. 60, 1, 27–34; discussion 34–5.
- BIELAMOWICZ, S. 2004. Perspectives on medialization laryngoplasty. Otolaryngol Clin North Am. 37, 1, 139–60, vii.
- BIRKFELLNER, W., P. SOLAR, A. GAHLEITNER, K. HUBER, F. KAINBERGER, ET AL. 2001. In-vitro assessment of a registration protocol for image guided implant dentistry. *Clin Oral Implants Res.* 12, 1, 69–78.
- BROWN, R. A. 1979. A computerized tomography-computer graphics approach to stereotaxic localization. J Neurosurg. 50, 6, 715–20.
- BURNS, R. P. 1997. Image-guided breast biopsy. Am J Surg. 173, 1, 9-11; discussion 12-3.
- DARABI, K., P. GRUNERT, AND A. PERNECZKY. 1997. Accuracy of intraoperative navigation using skin markers. Computer Assisted Radiology and Surgery: 920–924.
- DEGUCHI, D., K. AKIYAMA, K. MORI, T. KITASAKA, Y. SUENAGA, ET AL. 2006. A method for bronchoscope tracking by combining a position sensor and image registration. *Comput Aided Surg.* 11, 3, 109–17.
- DELIGIANNI, F., A. J. CHUNG, AND G. Z. YANG. 2006. 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.
- DEY, D., D. G. GOBBI, P. J. SLOMKA, K. J. M. SURRY, AND T. M. PETERS. 2002. Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas. *IEEE Transactions on Medical Imaging*. 21, 1, 23–30.
- DEY, J. AND S. NAPEL. 2006. Targeted 2D/3D registration using ray normalization and a hybrid optimizer. *Med Phys.* 33, 12, 4730–8.
- ERIC, A., E. A. BIER, M. C. STONE, K. PIER, W. BUXTON, ET AL. 1993. Toolglass and Magic Lenses: The See-Through Interface. Proc. of the 20th Ann.Conf. on Computer Graphics and Interactive Techniques 1, 73–80.
- FU, D. AND G. KUDUVALLI. 2008. A fast, accurate, and automatic 2D-3D image registration for image-guided cranial radiosurgery. *Med Phys.* 35, 5, 2180–94.
- GONZALEZ, R. C. AND R. E. WOODS. 1992. Digital image processing.
- HAUTMANN, H., A. SCHNEIDER, T. PINKAU, F. PELTZ, AND H. FEUSSNER. 2005. Electromagnetic catheter navigation during bronchoscopy: validation of a novel method by conventional fluoroscopy. *Chest.* 128, 1, 382–7.
- HELFERTY, J. P., A. J. SHERBONDY, A. P. KIRALY, AND W. E. HIGGINS. 2007. Computer-based system for the virtual-endoscopic guidance of bronchoscopy. *Comput Vis Image Underst.* 108, 1-2, 171–187.
- HORSLEY, V. AND R. CLARK. 1908. The structure and functions of the cerebellum examined by a new method. *Brain.* 31, 1, 45–124.

Journal of Computing Science and Engineering, Vol. 4, No. 4, December 2010

Image-guided surgery toolkit (IGSTK). Available from: http://www.igstk.org/.

- JIN, G., N. BAEK, J. K. HAHN, S. BIELAMOWICZ, R. MITTAL, ET AL. 2009. Image guided medialization laryngoplasty. Comput Animat Virtual Worlds. 20, 1, 67–77.
- JIN, G., S. J. LEE, J. K. HAHN, S. BIELAMOWICZ, R. MITTAL, ET AL. 2006. 3D surface reconstruction and registration for image guided medialization laryngoplasty. *Lect Notes Comput Sci.* 4291, 1, 761–770.
- JOSHI, A., D. SCHEINOST, K. VIVES, D. SPENCER, L. STAIB, ET AL. 2008. Novel interaction techniques for neurosurgical planning and stereotactic *IEEE Trans on Visualization and Computer Graphics.* 14, 6.
- KELLY, P. J. 1986. Computer-assisted stereotaxis: new approaches for the management of intracranial intra-axial tumors. *Neurology*. 36, 4, 535–41.
- KIRMIZIBAYRAK, C., M. WAKID, J. K. HAHN, AND S. BIELAMOWICZ. 2010. Interactive visualization for image guided medialization laryngoplasty. Proc. of Computer Graphics International.
- KLEIN, J., O. FRIMAN, M. HADWIGER, B. PREIM, F. RITTER, ET AL. 2009. Visual computing for medical diagnosis and treatment. Computers & Graphics 33, 554–565.
- KRUGER, J. AND R. WESTERMANN. 2003. Acceleration techniques for gpu-based volume rendering. 14th IEEE Visualization: 287–292.
- LAVALLEE, S., P. SAUTOT, J. TROCCAZ, P. CINQUIN, AND P. MERLOZ. 1995. Computer-assisted spine surgery: a technique for accurate transpedicular screw fixation using CT data and a 3-D optical localizer. J Image Guid Surg. 1, 1, 65–73.
- LEE, T. H., J. LEE, H. LEE, H. KYE, Y. G. SHIN, ET AL. 2009. Fast perspective volume ray casting method using GPU-based acceleration techniques for translucency rendering in 3D endoluminal CT colonography. *Comput Biol Med.* 39, 8, 657–66.
- LEKSELL, L. 1951. The stereotaxic method and radiosurgery of the brain. Acta Chir Scand. 102, 4, 316–9.
- LEVOY, M. 1988. Direct visualization of surfaces from computed tomography data Proceedings of the SPIE. 914, 828–841.
- LORENSEN, W. E. AND H. E. CLINE. 1987. Marching cubes: a high resolution 3D surface construction algorithm. Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques. 21, 163–169.
- MAES, F., D. VANDERMEULEN, AND P. SUETENS. 1999. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Med Image Anal.* 3, 4, 373–86.
- MAURER, C. R., J. J. MCCRORY, AND J. M. FITZPATRICK. 1993. Estimation of accuracy in localizing externally attached markers in multimodal volume head images. SPIE Medical imaging: image processing. 1898, 43–54.
- MAURER, C. R., JR., J. M. FITZPATRICK, M. Y. WANG, R. L. GALLOWAY, JR., R. J. MACIUNAS, ET AL. 1997. Registration of head volume images using implantable fiducial markers. *IEEE Trans Med Imaging*. 16, 4, 447–62.
- MONTAGNOLI, A. N., J. B. RUBERT, R. CAPOBIANCO, AND J. C. PEREIRA. 2006. Vocal folds vibrations with a three-dimensional deformable model. *The eighth IEEE International Symposium on Multimedia (ISM'06)*: 674–678.
- MUNBODH, R., H. D. TAGARE, Z. CHEN, D. A. JAFFRAY, D. J. MOSELEY, ET AL. 2009. 2D-3D registration for prostate radiation therapy based on a statistical model of transmission images. *Med Phys.* 36, 10, 4555–68.
- MUNBODH, R., Z. CHEN, D. A. JAFFRAY, D. J. MOSELEY, J. P. KNISELY, ET AL. 2008. Automated 2D-3D registration of portal images and CT data using line-segment enhancement. *Med Phys.* 35, 10, 4352–61.
- ONIK, G. M., D. B. DOWNEY, AND A. FENSTER. 1996. Three-dimensional sonographically monitored cryosurgery in a prostate phantom. J Ultrasound Med. 15, 3, 267–70.
- PAUL, P., O. FLEIG, AND P. JANNIN. 2005. Augmented virtuality based on stereoscopic reconstruction in multimodal image-guided neurosurgery: Methods and performance evaluation. *IEEE*

Transactions on Medical Imaging. 24, 11, 1500–1511.

- PERLIN, K. AND D. FOX. 1993. Pad An Alternative Approach to the Computer Interface. The 20th Ann. Conf. on Computer graphics and Interactive Techniques: 57–64.
- PETERS, T. M. 2006. Image-guidance for surgical procedures. Phys Med Biol. 51, 14, R505-40.

PRESS, W. H. 1992. Numerical recipes in C: the art of scientific computing. xxvi, 994 p.

- RIEDER, C., F. RITTER, M. RASPE, AND H. O. PEITGEN. 2008. Interactive Visualization of Multimodal Volume Data for Neurosurgical Tumor Treatment. *Computer Graphics Forum*. 27, 3, 1055–1062.
- ROBERTS, D. W., J. W. STROHBEHN, J. F. HATCH, W. MURRAY, AND H. KETTENBERGER. 1986. A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. J Neurosurg. 65, 4, 545–9.
- SATO, Y., M. NAKAMOTO, Y. TAMAKI, T. SASAMA, I. SAKITA, ET AL. 1998. Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization. *IEEE Trans Med Imaging*. 17, 5, 681–93.
- SCHROEDER, W., K. MARTIN, AND B. LORENSEN. 2000. Visualization toolkit: an object-oriented approach to 3D graphics.
- SOLOMON, S. B., P. WHITE, JR., C. M. WIENER, J. B. ORENS, AND K. P. WANG. 2000. Threedimensional CT-guided bronchoscopy with a real-time electromagnetic position sensor: a comparison of two image registration methods. *Chest.* 118, 6, 1783–7.
- SPIEGEL, E. A., H. T. WYCIS, M. MARKS, AND A. J. LEE. 1947. Stereotaxic apparatus for operations on the human brain. *Science*. 106, 2754, 349–50.
- Statistics on voice, speech, and language. Available from: http://www.nidcd.nih.gov/health/statistics/vsl.asp.
- STEFANSIC, J. D., A. J. HERLINE, Y. SHYR, W. C. CHAPMAN, J. M. FITZPATRICK, ET AL. 2000. Registration of physical space to laparoscopic image space for use in minimally invasive hepatic surgery. *IEEE Trans Med Imaging*. 19, 10, 1012–23.
- STUDHOLME, C., D. L. G. HILL, AND D. J. HAWKES. 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*. 32, 1, 71–86.
- TOMAŽEVIČ, D., B. LIKAR, AND F. PERNUŠ. 2006. 3-D/2-D Registration by Integrating 2-D Information in 3-D. *IEEE Transactions on Medical Imaging.* 25, 1.
- VIEGA, J., M. J. CONWAY, G. WILLIAMS, AND R. PAUSCH. 1996. 3D Magic Lenses. The 9th Annual ACM Symposium on UI Software and Technology. 1, 1, 51–58.
- VIOLA, P. AND P. E. WELLS III. 1997. Alignment by maximization of mutual information. International Journal of Computer Vision. 24, 2, 137–154.
- VIOLA, P., WELLS III, P. E. 1997. Alignment by maximization of mutual information. International Journal of Computer Vision. 24, 2, 137–154.
- WANG, L., Y. ZHAO, K. MUELLER, AND A. KAUFMAN. 2005. The Magic Volume Lens: An Interactive Focus+Context Technique for Volume Rendering. *IEEE Visualization 05*: 367–374.
- WANG, X., Q. ZHANG, Q. HAN, R. YANG, M. CARSWELL, ET AL. 2009. Endoscopic video texture mapping on pre-built 3-D anatomical objects without camera tracking. *IEEE Trans Med Imaging.* 29, 6, 1213–23.
- WELLS, W. M., 3RD, P. VIOLA, H. ATSUMI, S. NAKAJIMA, AND R. KIKINIS. 1996. Multi-modal volume registration by maximization of mutual information. *Med Image Anal.* 1, 1, 35–51.
- YIM, Y. AND H. HONG. 2008. Correction of segmented lung boundary for inclusion of pleural nodules and pulmonary vessels in chest CT images. *Comput Biol Med.* 38, 8, 845–57.



Yeny Yim is currently a post-doctoral researcher in the Department of Computer Science at the George Washington University, Washington DC. Her research interests include image registration, image-guided surgery, medical image processing, and computer graphics. She received her Ph.D. degree in Computer Science and Engineering from the Seoul National University, Seoul, Korea in 2010. She received her B.S. degree in Computer Engineering from the Kyungpook National University, Daegu, Korea in 2001.



Mike Wakid is currently a doctoral student in the Department of Computer Science at the George Washington University. He received his bachelor's degree in Computer Science from the University of Maryland, College Park in 2003 and his M.S. in Computer Science from the George Washington University in 2005. His research interests include medical visualization, image-guided surgery, real-time rendering, and animation.



Can Kirmizibayrak is currently a doctoral student the Department of Computer Science at the George Washington University. He received his bachelor's degree in Electrical-Electronics Engineering from Bogazici University, Turkey in 2003, and his master's degree in Telecommunications and Computers from The George Washington University in 2005. His research interests are medical visualization, image-guided surgery, interaction in biomedicine and computer graphics.





Steven Bielamowicz MD, FACS is Professor and Chief of the Division of Otolaryngology and Director of the Voice Treatment Center at The George Washington University. He received his medical degree from the Baylor College of Medicine and completed a residency in Otolaryngology-Head and Neck Surgery at UCLA. He completed a Neurolaryngology fellowship at the NIH and has an interest in the neurologic basis of voice disorders. Dr. Bielamowicz evaluates and treats patients with voice and swallowing disorders and has an active clinical research program within the field of neurolaryngology.

James Hahn is currently a full Professor in the Department of Computer Science at the George Washington University where he has been a faculty since 1989. He is the founding director of the Institute for Biomedical Engineering and the Institute for Computer Graphics. His areas of interests are: medical simulation, image-guided surgery, medical informatics, visualization, and motion control. He received his Ph.D. in Computer and Information Science from the Ohio State University in 1989, an M.S. in Physics from the University of California, Los Angeles in 1981, and a B.S. in Physics and Mathematics from the University of South Carolina in 1979.