Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2015 February 21; 9415: . doi:10.1117/12.2081732.

Rendering-Based Video-CT Registration with Physical Constraints for Image-Guided Endoscopic Sinus Surgery

Y. Otake^{a,d}, S. Leonard^a, A. Reiter^a, P. Rajan^a, J. H. Siewerdsen^b, G. L. Gallia^c, M. Ishii^c, R. H. Taylor^a, and G. D. Hager^a

^aDepartment of Computer Science, Johns Hopkins University, Baltimore MD, USA

^bDepartment of Biomedical Engineering, Johns Hopkins University, Baltimore MD, USA

^cDepartment of Otolaryngology – Head and Neck Surgery, Johns Hopkins University, Baltimore MD, USA

^dGraduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan

Abstract

We present a system for registering the coordinate frame of an endoscope to pre- or intra-operatively acquired CT data based on optimizing the similarity metric between an endoscopic image and an image predicted via rendering of CT. Our method is robust and semi-automatic because it takes account of physical constraints, specifically, collisions between the endoscope and the anatomy, to initialize and constrain the search. The proposed optimization method is based on a stochastic optimization algorithm that evaluates a large number of similarity metric functions in parallel on a graphics processing unit. Images from a cadaver and a patient were used for evaluation. The registration error was 0.83 mm and 1.97 mm for cadaver and patient images respectively. The average registration time for 60 trials was 4.4 seconds. The patient study demonstrated robustness of the proposed algorithm against a moderate anatomical deformation.

Keywords

Image-guided endoscopic surgery; rendering-based Video-CT registration; endoscopic sinus surgery

1. INTRODUCTION

Functional endoscopic sinus surgery (FESS) has evolved to represent the standard surgical treatment for sinonasal pathology [1, 2]. Navigation systems are used to maintain orientation during surgery because the anatomy is complex, highly variable due to stochastic embryologic development, and sinuses are adjacent to vital structures that have significant morbidities/mortality associated with them if injured. Commercial navigation systems, however, only provide qualitative sense of location because their 2 mm positioning errors under near ideal conditions [3] is insufficient for an environment featuring complex anatomical structures smaller than a millimeter, for example, for removal of structures within 1 mm thick bone right under the boundary to the brain, eye, major nerves or blood vessels. A recently proposed system using a passive arm for tracking the position and

orientation of the endoscope [4] demonstrated promising results (~1.0 mm error) in non-clinical experiments. The physical presence of the arm and its physical connection with the endoscope, however, is cumbersome and adversely affects the ergonomics and workflow.

Video-CT registration has been explored as a means of transcending the 2 mm accuracy limit by exploiting information contained in endoscope images [5]. The registration approaches are classified as reconstruction-based or rendering-based. The former [6] attempts to reconstruct a 3D surface from a series of images and then performs 3D-3D registration to a CT-derived surface. The latter [7] generates simulated views by rendering surface models derived from patient CT images. It searches for the camera pose that produces the simulated view that best matches to the real video image. Although both approaches can significantly improve the accuracy of navigation systems, there are several challenges that must be addressed before they reach clinical acceptance. First, the methods must be robust to variations of appearance of tissue surface and distortion of the anatomy. Second the initial registration needs to be automated and be fast enough so that it does not significantly impair clinical workflow.

This paper introduces a novel rendering-based video-CT registration system combining physical constraints, image gradient correlation, and a stochastic optimization strategy to improve robustness and provide an accurate and computationally efficient initial localization of the endoscope.

2. METHODS

2.1 Overview of the proposed method

The proposed registration method is summarized in Fig. 1. The framework follows the image-based 2D-3D registration [8] where an optimization algorithm searches for the camera pose that yields a rendered image (virtual endoscope) that has maximum similarity with the real endoscope image. As described below, the photometric rendering and similarity metric computation were implemented on GPU to accelerate the system.

2.2. Rendering of virtual endoscope

A traditional rendering algorithm [9] with a bidirectional reflectance distribution function (BRDF), shown in eq (1), was implemented on the GPU using CUDA (NVidia, Santa Clara CA)

$$BRDF_{Lamb}(p) = \sigma \frac{L \cos \theta_s(p)}{\pi R(p)^2} \quad (1)$$

where L is intensity of the light source, θ_s is the angle between the ray and surface normal, R is distance from light source to the surface and σ is a proportionality factor (albedo, camera characteristics, etc.). The surface normal was computed at the point where the ray intersects with an iso-surface of CT image defined by a specified threshold. A threshold of -300 HU was used in our experiments. Since the pixel size of the endoscope image is small compared to the CT voxel, a recursive search for the intersection point was used to avoid aliasing effects (Fig. 2).

2.3. Initialization using physical constraints

One unique characteristic of FESS is that the workspace is long, narrow and surrounded by complex structures of the nasal cavity. Thus the set of endoscope poses that are feasible during the surgery can be estimated a priori based on the patient-specific anatomical structure in CT data. This is accomplished by sampling possible poses and computing collisions between a 3D model of the anatomy and a 3D model of the endoscope (a cylinder with 4 mm diameter in this study). More specifically, an iso-surface of sinus structure was extracted by thresholding the CT data, and the workspace of the endoscope tip was roughly outlined. Then, n endoscope poses were randomly generated within workspace ($n=50,000$ in this study). The orientation of each pose was determined randomly around the insertion direction. Collision between the sinus surface and the endoscope model was detected by the RAPID package [10]. Due to the complex anatomical structure of the sinuses, the computed collision-free poses form multiple islands in the 6-dimensional pose space (i.e., 3-DOF translation and 3-DOF rotation). These were grouped into m clusters using K-means clustering ($m=20$ in this study) and barycenter of the poses in each cluster was used as initial seed points for optimization. Note that this initialization was performed pre-operatively based solely on the CT data and a 3D model of the endoscope.

2.4. Similarity metric

We used gradient correlation [11] as the similarity metric in this study.

$$GC(I_0, I_1) = \frac{1}{2} \left\{ NCC\left(\frac{d}{dx}I_0, \frac{d}{dx}I_1\right) + NCC\left(\frac{d}{dy}I_0, \frac{d}{dy}I_1\right) \right\} \quad (2)$$

$$NCC(I_0, I_1) = \frac{\sum_{(i,j)} (I_0 - \bar{I}_0)(I_1 - \bar{I}_1)}{\sqrt{\sum_{(i,j)} (I_0 - \bar{I}_0)^2} \sqrt{\sum_{(i,j)} (I_1 - \bar{I}_1)^2}} \quad (3)$$

The metric averages normalized cross correlation (NCC) between X-gradients and Y-gradients of the two images. In the endoscope image, the metric captures edges of occluding contours which are the main features in the virtual endoscope image derived from the CT volume. The gradient computation was implemented as a highly parallelizable kernel convolution on GPU.

2.5. Optimization

The optimization problem was formulated as follows.

$$\hat{T} = \underset{T \in S_T}{\operatorname{argmax}} GC(I_1, I_0(T(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z))) \quad (4)$$

where S_T represents the space of feasible camera poses (i.e., collision-free poses). The objective function has multiple local optima due to structures that create similar contours in the image. To improve robustness against these local optima, a derivative-free stochastic optimization strategy, covariance matrix adaptation - evolutionary strategy (CMA-ES) [12], was employed. After generating sample solutions in each generation, any samples that caused collisions were rejected and replaced with a new sample solution. Table 1

summarizes parameters used in the optimization. The CMA-ES is especially beneficial on a parallel hardware such as a GPU since it allows multiple samples to be evaluated in parallel as opposed to intrinsically sequential algorithms such as simulated annealing or Powell's method. Computational efficiency is further improved by simultaneously executing multiple optimization trials from m initial poses computed in Section 2.3. Computation time (which is proportional to the number of function evaluations) increases proportionally with m ; robustness against local optima also improves with increased m , since the density of the searched sample solution increased. The trade-off between the number of function evaluations and robustness generally holds in derivative-free optimization algorithms, thus we paid careful attention to its computational efficiency to improve robustness.

2.6. Experiments

Cadaver study—A cadaver specimen was used for quantitative evaluation of the proposed algorithm (Fig. 3) [13]. CT data ($0.46 \times 0.46 \times 0.5$ mm/voxel) and endoscope video images (1024×768 , 30 Hz) were acquired. 21 gauge syringe needles were placed in the sphenoid sinus wall prior to the CT scan to create landmarks for quantitative error analysis. An offline camera calibration was performed using the German Aerospace Center (DLR) CalLab [14].

Patent study—A retrospective study using CT data and endoscopic images recorded during surgery was performed to further investigate robustness and accuracy of the proposed method in the presence of texture of the mucosal layer and soft tissue deformation. The study protocol was approved by our institutional review board. Quantitative evaluation of the accuracy was challenging due to absence of landmarks that were identifiable both in CT and the endoscope image (such as needle tip in cadaver study). In this study, we used the uncinate process as a landmark, because, although identification of the exact point correspondence was not possible, the edges of the sheet-like structure always appears as a line feature in the endoscope image. We manually identified 5 points along the line of uncinate process in the 2D image and established correspondence between each point and closest point along the edge segmented in 3D to compute error.

Error metric—The target registration error (TRE) was defined as the distance between a landmark point in 3D and a ray emanating from the camera center in the direction of the target in the image [13].

3. RESULTS

3.1. Cadaver study

The result of the initialization process is summarized in Fig. 4. The collision-free poses were grouped into 20 clusters and barycenter of each cluster was used as an initial seed point in the optimization. The physical constraints based on collision detection restrict the search space which improved robustness of the optimization. The result is shown in Fig. 5. The average TRE over 7 landmark points was 0.83 mm.

3.2. Patient study

Proposed registration algorithm was tested on 60 consecutive endoscopic image frames. Fig. 6 shows the real endoscope images, predicted virtual endoscope images and the gradient images computed at the solution in three frames from the sequence. TRE over all frames was 1.97 ± 1.05 mm (mean $\pm$ std). The gradient-based similarity metric successfully captured the occluding contours in both images, which led the registration to converge to the pose that provides approximately the same view as the real endoscope, even in the presence of clear deformation of the structure at the middle (middle turbinate) between preoperative CT and intraoperative endoscope. Our algorithm based on a stochastic optimization strategy was robust against these local optima in this case. However, the landscape depends on various factors including lighting, degree of deformation, quality of CT data, etc. Our preliminary investigation on sensitivity of the registration accuracy to computation time (Table 2) demonstrated the trade-off between them. As we increase the number of multi-starts, the search takes longer time since larger number of sample solutions are evaluated, whereas the search converges to a better solution since the denser search finds a better local optimum in the multi-modal objective function space. Since the algorithm is highly parallelizable, runtime is expected to decrease proportionally to the number of parallel processor. Further investigation with different patients and different anatomy is underway.

4. CONCLUSIONS

We have presented a method that is able to register endoscopic video image to CT image. The method maximizes a gradient-based similarity metric between an observed image and a predicted image. Our primary contribution has been two-fold: 1) incorporation of physical constraint in the image-based registration algorithm for initialization and constraint of the search space, 2) utilization of a robust stochastic strategy in optimization by leveraging large number of function evaluations with GPU-acceleration. We have shown that tracking of the endoscope with the proposed algorithm can produce a registration in 4.4 seconds. Although this is too slow for real-time tracking, it is fast enough to produce an initial registration from which real-time tracking can begin using methods which may not be so robust for initial registration. Since the system relies on a rigid prior model, anatomical deformations, such as those produced by decongestants, create mismatches between predicted and observed images and generate local optima in objective function which could cause a failure. A possible solution to address this problem is intraoperative CT scanning [15] and/or simultaneous prediction of deformation and camera position.

ACKNOWLEDGEMENT

This work is partially supported by NIH-1R01EB015530-01 grant.

REFERENCES

1. Prulière-Escabasse V, Coste A. Image-guided sinus surgery. *European Annals of Otorhinolaryngology, Head and Neck Diseases*. 2010; 127(1):33–39.
2. Kassam AB, et al. Endoscopic endonasal skull base surgery: analysis of complications in the authors' initial 800 patients. *J Neurosurg*. 2011; 114(6):1544–68. [PubMed: 21166570]

3. Thoranaghatte R, et al. Landmark-based augmented reality system for paranasal and transnasal endoscopic surgeries. *Int J Med Robot.* 2009; 5(4):415–22. [PubMed: 19623600]
4. Lapeer RJ, et al. Using a passive coordinate measurement arm for motion tracking of a rigid endoscope for augmented-reality image-guided surgery. *Int J Med Robot.* 2013
5. Burschka D, et al. Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. *Medical Image Analysis.* 2005; 9(5):413–426. [PubMed: 16009593]
6. Mirota DJ, et al. Evaluation of a System for High-Accuracy 3D Image-Based Registration of Endoscopic Video to C-Arm Cone-Beam CT for Image-Guided Skull Base Surgery. *IEEE Trans Med Imaging.* 2013; 32(7):1215–26. [PubMed: 23372078]
7. Luo X, et al. Robust Real-Time Image-Guided Endoscopy: A New Discriminative Structural Similarity Measure for Video to Volume Registration. *Information Processing in Computer-Assisted Interventions.* 2013:91–100.
8. Otake Y, et al. Intraoperative image-based multiview 2D/3D registration for image-guided orthopaedic surgery: incorporation of fiducial-based C-arm tracking and GPU-acceleration. *IEEE Trans Med Imaging.* 2012; 31(4):948–962. [PubMed: 22113773]
9. Bosma MK, Smit J, Lobregt S. Iso-surface volume rendering. 1998
10. Gottschalk, S.; Lin, MC.; Manocha, D. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM; 1996. OBBTree: a hierarchical structure for rapid interference detection; p. 171-180.
11. Penney GP, et al. A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE Trans Med Imaging.* 1998; 17(4):586–95. [PubMed: 9845314]
12. Hansen N. The CMA Evolution Strategy: A Comparing Review. *Towards a New Evolutionary Computation.* 2006:75–102.
13. Mirota DJ, et al. A system for video-based navigation for endoscopic endonasal skull base surgery. *IEEE Trans Med Imaging.* 2012; 31(4):963–76. [PubMed: 22113772]
14. Strobl KH, et al. DLR CalDE and CalLab. Institute of Robotics and Mechatronics. German Aerospace Center (DLR). Available from: <http://www.robotic.dlr.de/callab>.
15. Chennupati SK, et al. Intraoperative IGS/CT updates for complex endoscopic frontal sinus surgery. *ORL J Otorhinolaryngol Relat Spec.* 2008; 70(4):268–70. [PubMed: 18487904]

Fig. 1.
Flowchart for the proposed registration framework

Figure 2. Iso-surface detection method, (a) concept of the level-of-detail (LOD) detection. Renderings (b) without and (c) with the LOD detection.

Fig. 3.
(a) Experimental setup and rendering of bones and needles inserted in the wall of sinus as landmark points in evaluation, (b) endoscope image with landmarks (tip of the needles)

Figure 4.

Initialization using collision detection. (a) Point cloud model of the sinus (cyan) and computed collision free poses clustered with K-means clustering. The line indicates center of one endoscope pose and each cluster colored in the same color, (b) virtual endoscope images at the barycenter pose of each cluster. Note that poses 8, 12, 14 shows nothing because the camera is very close to the wall, but not collided.

Fig. 5.
Result of cadaver study. (a) Real endoscope, (b) virtual endoscope image at the predicted pose, (c) gradient image computed in the optimization

Fig. 6.
Result of the proposed registration algorithm on image sequence from a patient study

Table 1

Summary of optimization parameter

Downsampled image size	160×128
Number of multi-starts	20
Population size	100
Search range ($t_x, t_y, t_z, \theta_x, \theta_y, \theta_z$)	(5 mm, 5 mm, 5 mm, 15° , 15° , 15°)

Table 2

Comparison of runtime and accuracy (50 registration trials were performed on one image from patient study.)

# of multi-starts (m)	1	10	20	40
# of function evaluations	1938	20303	36132	71155
Runtime (sec)	0.88	2.78	4.60	8.37
Median TRE (mm)	2.37	1.16	1.14	1.13