

Motion-compensated OCT imaging of laryngeal tissue

Sarah Latus^a, Marica Kulas^a, Johanna Sprenger^a, Debayan Bhattacharya^{a,b}, Philippe C. Breda^b, Lukas Wittig^b, Tim Eixmann^c, Gereon Hüttmann^c, Lennart Maack^a, Dennis Eggert^b, Christian Betz^b, and Alexander Schlaefer^a

^aHamburg University of Technology (TUHH), Institute of Medical Technology and Intelligent Systems, Hamburg, GER

^bUniversity Medical Center Hamburg-Eppendorf (UKE), Clinic for Ears, Nose and Throat, Hamburg, GER

^cMedical Laser Center Lübeck GmbH, Lübeck, GER

ABSTRACT

The increasing incidence of laryngeal carcinomas requires approaches for early diagnosis and treatment. In clinical practice, white light endoscopy of the laryngeal region is typically followed by biopsy under general anesthesia. Thus, image based diagnosis using optical coherence tomography (OCT) has been proposed to study sub-surface tissue layers at high resolution. However, accessing the region of interest requires robust miniature OCT probes that can be forwarded through the working channel of a laryngoscope. Typically, such probes generate A-scans, i.e., single column depth images, which are rather difficult to interpret. We propose a novel approach using the endoscopic camera images to spatially align these A-scans. Given the natural tissue motion and movements of the laryngoscope, the resulting OCT images show a three-dimensional representation of the sub-surface structures, which is simpler to interpret. We present the overall imaging setup and the motion tracking method. Moreover, we describe an experimental setup to assess the precision of the spatial alignment. We study different tracking templates and report root-mean-squared errors of 0.08 mm and 0.18 mm for sinusoidal and freehand motion, respectively. Furthermore, we also demonstrate the in-vivo application of the approach, illustrating the benefit of spatially meaningful alignment of the A-scans to study laryngeal tissue.

Keywords: Optical coherence tomography, Image based tracking, Motion compensation, Laryngeal carcinoma

1. INTRODUCTION

An increasing incidence and prevalence of laryngeal carcinoma have been reported in recent cancer statistics.¹ Squamous cell carcinomas of the head and neck originate in the epithelial layer, showing a thickening in case of pathology. Assessing these tissue variations is essential for optimal treatment planning. While white light endoscopy is already established for soft tissue analysis, interpreting the image data requires considerable experience, particularly when distinguishing between benign and malignant pathologies.² Thus, conspicuous lesions must still be confirmed by biopsy under general anesthesia and histopathologic examination, which is costly and time-consuming. Optical coherence tomography (OCT) has been proposed to improve the image based diagnosis of laryngeal lesions.³ OCT has already been applied in clinical practice, e.g., in ophthalmology or cardiology, and suggested for optical biopsy.⁴ In contrast to endoscopic RGB images, OCT enables the visualization of sub-surface tissue structures with a resolution of a few micrometers. Thus, the combination of endoscopic and OCT imaging promises a more comprehensive and non-invasive analysis of the epithelial layers.

Embedding fiber-based optics in medical instruments has been proposed for several applications.⁵ For example, OCT probes have been used to study soft tissue characteristics by means of the morphology⁶ or instrument interaction.⁷ To enable robustness and miniaturization, OCT probes are often equipped with simple optics that generate A-scans, i.e., single-column depth images. A sequence of A-scans can be considered as M-scan. However, the lack of spatial correlation complicates OCT image interpretation. Especially during in-vivo imaging with natural tissue motion and movements of the laryngoscope. However, spatial information can be derived from endoscopic image data, e.g., by estimating tissue and instrument motion.⁸ For example, the Minimum Output Sum of Squared Error (MOSSE) filter⁹ has been shown to be an efficient approach to track soft tissue features.

Further author information: (Send correspondence to S. Latus)

S. Latus: E-mail: sarah.latus@tuhh.de, Telephone: +49 40 42878 3389

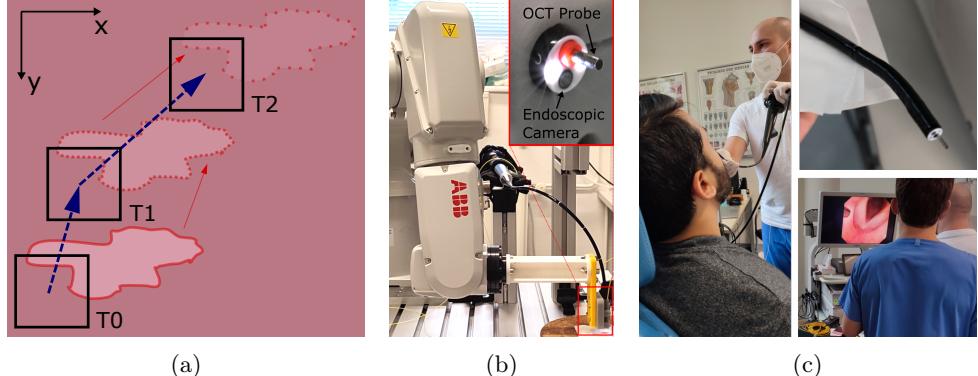


Figure 1: (a) An initial tracking template (T_0) is selected in the first endoscopic image for training the MOSSE filter. For subsequent images with a moving tissue structure, we obtain the updated tracking templates (T_1 , T_2) and corresponding relative motion vectors (blue). (b) Experimental setup to perform different motion profiles with a robot. Laryngoscope with inserted OCT probe is shown on top. (c) Impressions from the in-vivo study.

In this study, we introduce an approach for motion-compensated OCT imaging in the laryngeal region using a miniature OCT probe inserted in the working channel of a flexible laryngoscope. Utilizing a MOSSE filter, we track tissue features in the endoscopic images to estimate the relative motion caused by both tissue displacements and movements of the laryngoscope. Based on this motion trajectory, we spatially align the OCT-A scan and enable a three-dimensional analysis of sub-surface tissue structures. We evaluate our approach in two experimental studies: First, we use a robot to move the laryngoscope and quantify the tracking performance on bovine tissue. Second, we conduct an in-vivo study on laryngeal tissue to illustrate the benefit of spatially meaningful alignment of OCT A-scans.

2. MATERIAL AND METHODS

2.1 Relative motion tracking

Our goal is to estimate the relative motion between the laryngoscope and tissue by employing a MOSSE filter⁹ for two-dimensional tracking of tissue features in the endoscopic images. We train the MOSSE filter with an initial template f_0 chosen in the first image (T_0 in Fig. 1a) and adapt the template position with each new image as the feature appearance changes. Thus, we apply the MOSSE filter to search for a template position (T_1) with maximum correlation to the previous template (T_0). In general, the MOSSE filter H is defined as

$$H_i^* = \frac{A_i}{B_i}, \quad (1)$$

with

$$A_i = \eta G \odot F_i^* + (1 - \eta) A_{i-1}, \quad B_i = \eta F_i \odot F_i^* + (1 - \eta) B_{i-1}. \quad (2)$$

The operator \odot indicates element-wise multiplication in the Fourier Domain and $*$ denotes the complex conjugate. The training output G results after Fourier transformation of a 2D Gaussian-shaped peak fitted to the initial template position. We chose a learning rate $\eta = 0.01$ and $\sigma = 1$ as the standard deviation of the Gaussian peak to control the template adaption to new images. Consequently, the template shift $(\Delta x_i, \Delta y_i)$ with maximum correlation follows as

$$(\Delta x_i, \Delta y_i) = \underset{x,y}{\operatorname{argmax}} \{ \mathcal{F}^{-1}(F_i \odot H_{i-1}^*) \}. \quad (3)$$

To optimize the correlation results, we preprocess the templates f_i as follows. We first convert the pixel values to grayscale, apply logarithmic scaling, normalize the pixel values, and finally multiply the template with a two-dimensional Hann window. Maximizing the correlation by adapting the position of the tracking template is performed iteratively for all successive images. By analyzing the vectors between the tracking templates, we obtain a relative motion trajectory. Due to the reduced endoscopic frame rate, we linearly interpolate the trajectory to obtain a two-dimensional position for each A-scan.

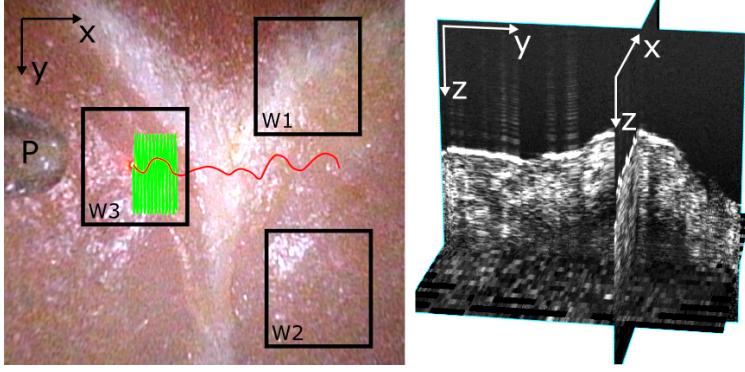


Figure 2: (Left) Endoscopic image from laboratory study with visible OCT probe (P), initial tracking templates (W1, W2, W3), and freehand motion trajectory (E2, red). (Right) Spatially aligned OCT A-scans for sinusoidal motion trajectory (E1, green).

Table 1: Quantitative tracking performance for lab experiments (E1, E2) and initial tracking window positions (W1, W2, W3). The RMSEs in x and y are listed in mm, which indicate the difference between the ground truth and the tracked trajectory.

Traj.	Wind.	x	y
E1	W1	0.080	0.082
	W2	0.276	0.817
	W3	0.081	0.082
E2	W1	0.179	0.113
	W2	3.114	0.179
	W3	1.343	0.284

2.2 Imaging setup

We insert a forward-viewing OCT probe in the working channel of a flexible laryngoscope (ENF-VT3, Olympus). The laryngoscope enables imaging rates of up to 30 Hz. To correct distortions of the endoscopic images and obtain scaling factors we conduct a camera calibration. An endoscopic image size of 1080×1080 pixels corresponding to approximately 9×9 mm for a working distance of 90 mm follows. The OCT probe consists of a single-mode fiber (F-SMF28, Newport Corp.) with a spliced GRIN lens (F25896, OFS Fitel) to achieve a working distance of approximately 5 mm. A stainless steel capillary tube of 1.8 mm outer diameter and 1 cm length stabilizes the optics and enables wipe disinfection of the probe for in-vivo use. We acquire A-scans with a frequency of 91 kHz using an OCT system (Telesto I, Thorlabs). In addition, a guide laser (KLS635, Thorlabs) with a limited output power of 0.1 mW is coupled to visualize the OCT imaging position.

2.3 Experimental evaluation

In our experimental study on bovine tissue, we use a serial robotic arm (IRB 120, ABB, Fig. 1b) to perform two motion trajectories with the laryngoscope: a sinusoidal course (E1) and a simulated freehand motion (E2) with motion ranges of 2×1.5 mm and 8×1 mm, respectively. The motion trajectories are visualized in red and green in Fig. 2. We synchronize the robot motion and image acquisition by recording trigger signals from all systems.

In our in-vivo study, we acquired image data from five healthy volunteers. For each participant, the physician navigates the laryngoscope to the vocal cords manually (Fig. 1c). Once a target region is visible, images were acquired synchronously while the endoscope was held stable. We crop the in-vivo data with large motion amplitudes and velocities for comparison with robot experiments.

3. RESULTS

To quantitatively analyze our study on bovine tissue, we compare the image based motion trajectories with the ground truth of robot motion. We chose three initial tracking templates: W1) multiple distinct features, W2) fewer features, and W3) laser dot in the center as shown in Fig. 2. We register the tracked and ground-truth trajectories using the iterative closest point algorithm¹⁰ and calculate the root mean square error (RMSE) in x and y direction. The respective registration errors are listed in Tab. 1. For a template with several features (W1), we obtain the best results on both trajectories. We achieve comparable results for a template with fewer features (W2), as long as small motion amplitudes in x and y direction are applied (E1). With laser dot in the template (W3), the largest RMSEs of up to 3.11 mm follow, since the template does not adapt to the relative movement and remains in the same position in the image. Spatially aligned OCT A-scans for motion trajectory E1 and tracking window position W1 are exemplary shown in Fig. 2.

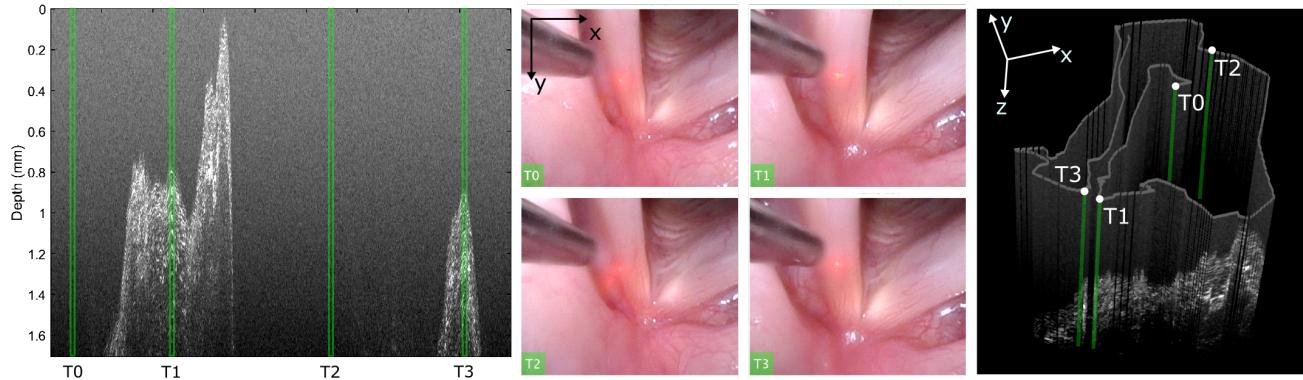


Figure 3: Exemplary OCT data acquired from vocal cords. Time points T0, T1, T2, and T3 of the OCT M-scan (left) are correlated to the respective endoscopic frames (middle). Spatial arrangement of OCT A-scans (right).

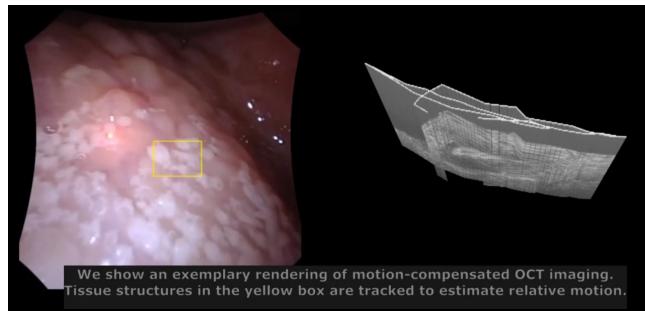


Figure 4: Video 1 - Graphical abstract demonstrating our motion-compensated OCT imaging approach on the bottom of the tongue. Access here: <http://dx.doi.org/10.1117/12.3006729.1>

Our in-vivo study demonstrates the benefit of the spatial alignment of OCT A-scans. Image data acquired from vocal cords that tend to spontaneous motion are visualized in Fig. 3. Using the OCT M-scan, no spatial information is available for image interpretation. After applying the MOSSE filter, the movement of the vocal cords can be considered, enabling a three-dimensional OCT image representation. OCT-A scans acquired spatially close to each other but with a large temporal difference (T1, T3) are now shown with morphological relationship. In addition to this example, our graphical abstract (Fig. 4) demonstrates the application of our motion-compensated OCT imaging approach on tissue structures at the bottom of the tongue.

4. DISCUSSION AND CONCLUSION

Our quantitative results demonstrate tracking performances with RMSEs of up to 0.08 mm, corresponding to a few pixels in the endoscopic image. However, the tracking performance depends on the number of features in the initial tracking window and its position relative to the guide laser. In future studies, the tracking performance could be further improved by using narrow band images that contain more extensive features. Furthermore, different localization approaches could be investigated such as the average of synthetic exact filters¹¹ or machine learning methods. In our in-vivo study, we demonstrate the benefit of motion-compensated OCT imaging. Tissue structures that have been imaged multiple times, e.g., due to reflective motion, can be spatially correlated for an improved three-dimensional interpretation of OCT data. Classification of malignant tissue studies based on spatially aligned OCT images will be the focus of future studies in patients with relevant pathologies.

ACKNOWLEDGMENTS

The authors state that they have no conflict of interest. This work was partially supported by the i³ initiative of TUHH, the Forschungszentrum Medizintechnik (fmthh), and by the Interdisciplinary Competence Center for Interface Research (ICCIR) on behalf of UKE and TUHH.

REFERENCES

- [1] Nocini, R., Molteni, G., Mattiuzzi, C., and Lippi, G., "Updates on larynx cancer epidemiology," *Chinese Journal of Cancer Research* **32**(1), 18 (2020).
- [2] Gale, N., Gnepp, D. R., Poljak, M., Strojan, P., Cardesa, A., Helliwell, T., Šifrer, R., Volavšek, M., Sandison, A., and Zidar, N., "Laryngeal squamous intraepithelial lesions: an updated review on etiology, classification, molecular changes, and treatment," *Advances in anatomic pathology* **23**(2), 84–91 (2016).
- [3] Wittig, L., Betz, C., and Eggert, D., "Optical coherence tomography for tissue classification of the larynx in an outpatient setting-a translational challenge on the verge of a resolution?," *Translational Biophotonics* **3**(1), e202000013 (2021).
- [4] Ellebrecht, D. B., Latus, S., Schlaefer, A., Keck, T., and Gessert, N., "Towards an optical biopsy during visceral surgical interventions," *Visceral medicine* **36**(2), 70–79 (2020).
- [5] Fu, Z., Jin, Z., Zhang, C., He, Z., Zha, Z., Hu, C., Gan, T., Yan, Q., Wang, P., and Ye, X., "The future of endoscopic navigation: A review of advanced endoscopic vision technology," *IEEE Access* **9**, 41144–41167 (2021).
- [6] Otte, S., Otte, C., Schlaefer, A., Wittig, L., Hüttmann, G., Drömann, D., and Zell, A., "Oct a-scan based lung tumor tissue classification with bidirectional long short term memory networks," in [*2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*], 1–6 (2013).
- [7] Latus, S., Sprenger, J., Neidhardt, M., Schädler, J., Ron, A., Fitzek, A., Schlüter, M., Breitfeld, P., Heinemann, A., Püschel, K., et al., "Rupture detection during needle insertion using complex oct data and cnns," *IEEE Transactions on Biomedical Engineering* **68**(10), 3059–3067 (2021).
- [8] Li, Y., Richter, F., Lu, J., Funk, E. K., Orosco, R. K., Zhu, J., and Yip, M. C., "Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics," *IEEE Robotics and Automation Letters* **5**(2), 2294–2301 (2020).
- [9] Bolme, D. S., Beveridge, J. R., Draper, B. A., and Lui, Y. M., "Visual object tracking using adaptive correlation filters," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], 2544–2550 (2010).
- [10] Rusinkiewicz, S. and Levoy, M., "Efficient variants of the icp algorithm," in [*Proceedings third international conference on 3-D digital imaging and modeling*], 145–152, IEEE (2001).
- [11] Bolme, D., Draper, B., and Beveridge, J., "Average of synthetic exact filters," *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009* , 2105–2112 (06 2009).