Medical Image Analysis 40 (2017) 80–95

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

# Stereo vision-based tracking of soft tissue motion with application to online ablation control in laser microsurgery

Andreas Schoob*, Dennis Kundrat, Lüder A. Kahrs, Tobias Ortmaier

*Leibniz Universität Hannover, Institute of Mechatronic Systems, Appelstr. 11a, 30167 Hanover, Germany*

ARTICLE INFO

*Article history:*
Received 8 November 2016
Revised 11 April 2017
Accepted 6 June 2017
Available online 8 June 2017

*Keywords:*
Non-rigid tracking
Stereo vision
Epipolar constraint
Motion compensation

ABSTRACT

Recent research has revealed that image-based methods can enhance accuracy and safety in laser microsurgery. In this study, non-rigid tracking using surgical stereo imaging and its application to laser ablation is discussed. A recently developed motion estimation framework based on piecewise affine deformation modeling is extended by a mesh refinement step and considering texture information. This compensates for tracking inaccuracies potentially caused by inconsistent feature matches or drift. To facilitate online application of the method, computational load is reduced by concurrent processing and affine-invariant fusion of tracking and refinement results. The residual latency-dependent tracking error is further minimized by Kalman filter-based upsampling, considering a motion model in disparity space. Accuracy is assessed in laparoscopic, beating heart, and laryngeal sequences with challenging conditions, such as partial occlusions and significant deformation. Performance is compared with that of state-of-the-art methods. In addition, the online capability of the method is evaluated by tracking two motion patterns performed by a high-precision parallel-kinematic platform. Related experiments are discussed for tissue substitute and porcine soft tissue in order to compare performances in an ideal scenario and in a setup mimicking clinical conditions. Regarding the soft tissue trial, the tracking error can be significantly reduced from 0.72 mm to below 0.05 mm with mesh refinement. To demonstrate online laser path adaptation during ablation, the non-rigid tracking framework is integrated into a setup consisting of a surgical Er:YAG laser, a three-axis scanning unit, and a low-noise stereo camera. Regardless of the error source, such as laser-to-camera registration, camera calibration, image-based tracking, and scanning latency, the ablation root mean square error is kept below 0.21 mm when the sample moves according to the aforementioned patterns. Final experiments regarding motion-compensated laser ablation of structurally deforming tissue highlight the potential of the method for vision-guided laser surgery.

## 1. Introduction

Surgery on delicate anatomical structures often demands high-resolution imaging and microinstruments for precise soft tissue manipulation. More advanced tools, such as medical lasers, facilitate contactless treatment of the pathology and minimize tissue trauma. A state-of-the-art clinical application is transoral laser microsurgery (TLM) for resection of benign or cancerous tissue on vocal cords (Rubinstein and Armstrong, 2011). Regarding the surgical treatment, a direct line-of-sight is established by inserting a laryngoscope in the throat of the patient. Precise resection of the lesion is achieved using a stereo microscope providing a magnified view of the surgical site and an ablation laser manually steered with a micromanipulator attached to the setup. Since the surgeon operates at a large distance from the patient, long and intensive training is required to master this task. Furthermore, soft tissue deformation induced by respiration artifacts and manipulation strongly affects the accuracy of laser ablation. Furthermore, misalignment of the laser path and loss of focus are evoked by the non-stiff mechanical fastening of the laser system to the patient; thus, motion externally applied to the microscope head most likely results in positional deviation of the laser spot. Deformations and camera motion are difficult to cope with, especially when the aim is function preservation with resection margins of less than 1 mm. To overcome this limitation, vision-based tracking of tissue motion and its application to motion-compensated laser ablation is addressed in this study as a continuation of our recently discussed method (Schoob et al., 2016) for image stabilization during incision planning. Moreover, the proposed method is not solely restricted to laser microsurgery. Further vision-guided, robot-assisted interventions or augmented reality concepts involving

---

* Corresponding author.
*E-mail address:* andreas.schoob@imes.uni-hannover.de (A. Schoob).

http://dx.doi.org/10.1016/j.media.2017.06.004

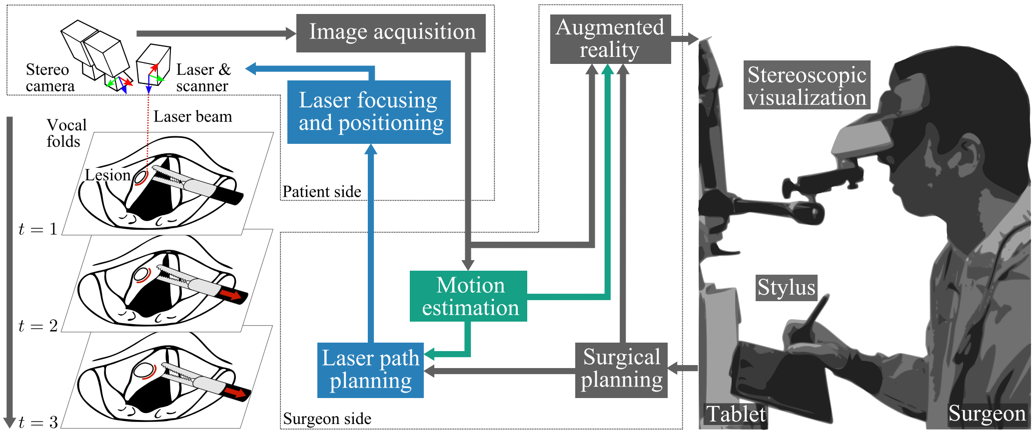

**Fig. 1.** Workflow for application of motion compensation in laser phonomicrosurgery. While exposing the vocal fold lesion (oval structure) by pulling with the grasping forceps, tissue motion is tracked to adapt online the ablation scan pattern (red line). Vision-guided laser control, as considered in this study, is intended to be integrated into a surgical framework with intuitive, stylus-tablet-based planning and augmented reality visualization as developed in the µRALP-project. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

superimposing tracking results directly to stereoscopic displays are conceivable.

Advances in laser surgery have been achieved regarding tablet-based planning interfaces (Tang et al., 2006; Mattos et al., 2014; Schoob et al., 2015c) and vision-guided laser control and micro-robotic scanning units for beam deflection (Dagnino et al., 2015; Renevier et al., 2016). Except for recent developments in the field of laser photocoagulation in retinal surgery, where tissue motion tracking performs adequately when rigid, affine or similarity transforms are used (Yang et al., 2015; Prokopetc and Bartoli, 2016), online estimation of larger tissue deformation during ablation has not been addressed so far. In particular, respiratory motion artifacts and tissue manipulation with surgical forceps as well as camera movements can lead to unintended injury of risk structures surrounding the lesion. To overcome this limitation, deformation tracking for online adaptation of laser spot positioning on the tissue surface is required (see Fig. 1).

Recently, vision-based tracking has been focused on minimally invasive surgery due to advances in medical imaging, augmented reality and robotics. Early studies discussed motion tracking, particularly in the field of beating heart surgery, considering matching of salient feature points (Ortmaier et al., 2005; Stoyanov et al., 2005; Sauvée et al., 2006). Intensive research has been conducted to improve robustness of feature-based tracking by including geometrical constraints for spatial consistency (Yip et al., 2012), multi-affine clustering of the target region (Puerto-Souza and Mariottini, 2013), affine-invariant feature descriptors (Giannarou et al., 2013), or online tracking-by-detection for surgical site retargeting (Ye et al., 2016).

By contrast, physical or geometric models can be incorporated into the non-rigid tracking framework. Associated optimization then aims at minimizing the shape bending energy and the matching error between the current frame and its template model. If shape priors are available or acquired preoperatively, organ deformation can be accurately estimated in real time depending on the complexity of the mechanical model and its intraoperative registration (Suwelack et al., 2014; Haouchine et al., 2015; Collins et al., 2016). If tracking of local deformations without knowledge of the anatomical shape is intended, stereo-based methods considering Free-Form Deformation (FFD) (e.g., piecewise bi-linear maps or B-splines) or Radial Basis Functions (RBF) (e.g., Thin Plate Splines (TPS)) have been shown to perform well for beating heart motion estimation (Lau et al., 2004; Stoyanov et al., 2004; Richa et al., 2010). In order to reduce the computational load when TPS is used, tracking can be split into intra-frame shape registration and inter-frame motion estimation (Yang et al., 2014). If a deformation is small, primitive models, such as quasi-spherical triangles, can perform as accurately as TPS-based methods (Wong et al., 2013). To further accelerate tracking, inverse compositional optimization (Brunet et al., 2011) or learning of non-linear template transformation provide promising solutions (Tan et al., 2014).

In contrast to RBF-based models, which are mainly limited to smooth and continuous deformations, alignment to local geometric changes can be efficiently achieved with piecewise warps providing local support and invertibility (Sotiras et al., 2013). A noteworthy method in the field of vision-based, non-rigid tracking (Pilet et al., 2008) estimates deformations with a triangular mesh of hexagonal elements. In this case, a quadratic energy term is formulated penalizing local surface curvature, whereas outliers are determined with a coarse-to-fine robust estimator function. In addition to considering progressive finite Newton (PFN) optimization (Zhu et al., 2009b), application to soft tissue motion estimation for white light and multispectral imaging has been recently discussed (Stoyanov and Yang, 2009; Stoyanov et al., 2012; Du et al., 2015). Piecewise affine warps have been considered not only for endoscopic vision but also for online ultrasound image registration, due to their reduced computational complexity (Preiswerk et al., 2014; Royer et al., 2017).

Most model-based, non-rigid tracking methods cannot operate at image-acquisition rates of 30 Hz and higher. In particular, direct methods often require a non-deterministic, Gauss-Newton-like optimization scheme. Thus, convergence is not ensured until the camera acquires the next frame. In this regard, tracking with a fixed number of iterations or using general purpose graphics processing units (GPGPU) provide only a limited solution to the problem. In particular, if the image alignment error is large and the minimization process requires several frames, tracking might fail if the tissue concurrently undergoes significant motion or deformation.

This study presents a novel method for stereoscopic tracking of soft tissue motion. Extending our recent work (Schoob et al., 2016), we follow the idea of splitting the optimization into (1) robust, quasi-deterministic tracking and (2) appearance-based mesh refinement to compensate for tracking inaccuracies such as drift. Instead of sequential processing, as described in the original work (Zhu et al., 2009b), concurrent computation of both steps is proposed. Once convergence is reached for the mesh refinement, affine-invariant fusion with respect to the current tracking