Robust endoscopic pose estimation for Intraoperative Organ-Mosaicking

Daniel Reichard*[a], Sebastian Bodenstedt[a], Stefan Suwelack[a], Martin Wagner[b], Hannes Kenngott[b], Beat Peter Müller-Stich[b], Rüdiger Dillmann[a], Stefanie Speidel[a]
[a]Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Adenauerring 2, D-76131 Karlsruhe, Germany;
[b]Department of General, Abdominal and Transplant Surgery, University of Heidelberg, Im Neuenheimer Feld 110, D-69120 Heidelberg

ABSTRACT

The number of minimally invasive procedures is growing every year. These procedures are highly complex and very demanding for the surgeons. It is therefore important to provide intraoperative assistance to alleviate these difficulties. For most computer-assistance systems, like visualizing target structures with augmented reality, a registration step is required to map preoperative data (e.g. CT images) to the ongoing intraoperative scene. Without additional hardware, the (stereo-) endoscope is the prime intraoperative data source and with it, stereo reconstruction methods can be used to obtain 3D models from target structures. To link reconstructed parts from different frames (mosaicking), the endoscope movement has to be known. In this paper, we present a camera tracking method that uses dense depth and feature registration which are combined with a Kalman Filter scheme. It provides a robust position estimation that shows promising results in ex vivo and in silico experiments.

Keywords: endoscopic pose estimation, feature tracking, organ-mosaicking, endoscope pose, intraoperative registration, projective data association

1. INTRODUCTION

Minimally invasive surgery (MIS) is a challenging field of work. To assist the surgeons with the difficulties tied to it, computer assisted intraoperative support is becoming more and more important. For intraoperative assistance often a live target model out of the operating room (OR) is needed. The problem with stereo reconstruction of endoscopic images is that only a small field of the scene is reconstructed per frame. Therefore a mosaicking is needed, which assembles multiple frames together. In our previous work [1], we used a voxel based method that integrates each depth frame with the scene already stored in the voxel volume. The first depth frame and the associated camera position is set as the origin of the camera coordinate system. Every following camera movement is computed relative to the first image. This is possible as the depth frames are integrated dynamically before the next frame is computed. This frame-to-origin registration counteracts the camera drifting problem that occurs if frame-to-frame registration is used.

The incremental movements between frames are estimated through Projective Data Association (PDA). The PDA allows us to use all depth pixels for the registration process in real time. This advantage over the common Iterative Closest Point (ICP) algorithm is achieved at the cost of a limitation described in section 2.1.

The homogenous structures and challenging image characteristics in MIS can lead to failure in one registration step, which often results in a possibly unrecoverable system state. To prevent this, an additional tracking approach can be included. Feature based methods like [2] and [3] are presenting a promising addition, as they rely on texture information, which is not used by the PDA approach.

*daniel.reichard@kit.edu; fax +49 (0)721 608-48270; his.anthropomatik.kit.edu

2. METHODS

2.1 Mosaicking and Projective Data Association

The Projective Data Association approach is based on the assumption, that the movement between incremental time steps is small [4]. Simple Iterative Closest Point algorithms usually compare every point in one set with all of the points from the other set. This would be too expensive for dense depth map comparisons. Using the PDA method, only points that are projected in proximity of the same camera pixel have to be taken into account because of the movement restriction.

When the new camera position is computed, the current point cloud R_i can be transformed into the global world coordinate system. To store and refine the single partial maps, they are integrated into a voxel volume that represents a Truncated Signed Distance Function (TSDF) [5]. Every voxel p contains the distance to the next object surface point F(p) and a weight w(p) that corresponds to the certainty of this value (i.e. how often it was observed). F(p) and w(p) are computed as suggested by Newcombe et al. [6].

$$S(p) = \{F(p), w(p)\} \tag{1}$$

2.2 Feature Matching and Tracking

To compute a 3-dimensional pose, at least three corresponding 3D points in the reference and current frame are needed. However, dealing with image noise and matching errors, more points are sensible for a stable estimation. The first step to create 3D feature-points is detecting 2D features in the left and right image of the stereo-endoscope. We are using a SIFT [7] feature detector and descriptor, which is rotation and scale invariant.

Stereo Feature Matching To compute 3D features, correspondences between 2D features detected in the stereo-images have to be found. For every feature descriptor in the left image, a set of nearest neighbors in the other image is computed using a kd-tree search. The resulting set is then checked against outliers using

- the Lucas-Kanade optical flow implementation [8]
- epipolar line restrictions
- the matching quality given by the feature descriptor distance

Feature Tracking The features between different time frames are matched similar to the stereo matching. Optical flow leads to possible correspondences. It thereby estimates the feature positons in the new image and restricts the search to an uncertainty ellipsoid. To reliably compute the endoscope pose, matching features in every frame are integrated into the system state of the extended Kalman Filter (EKF).

2.3 Endoscopic Pose Estimation

An extended Kalman Filter (EKF) is used for pose estimation. The Kalman Filter works with a prediction and measurement model [2, 9]. First, the position is predicted through the system state and is then corrected by a measurement step. In our case the measurement consists of the 3D positions from the detected features and the endoscope position computed with the Projective Data Association approach (figure 1).

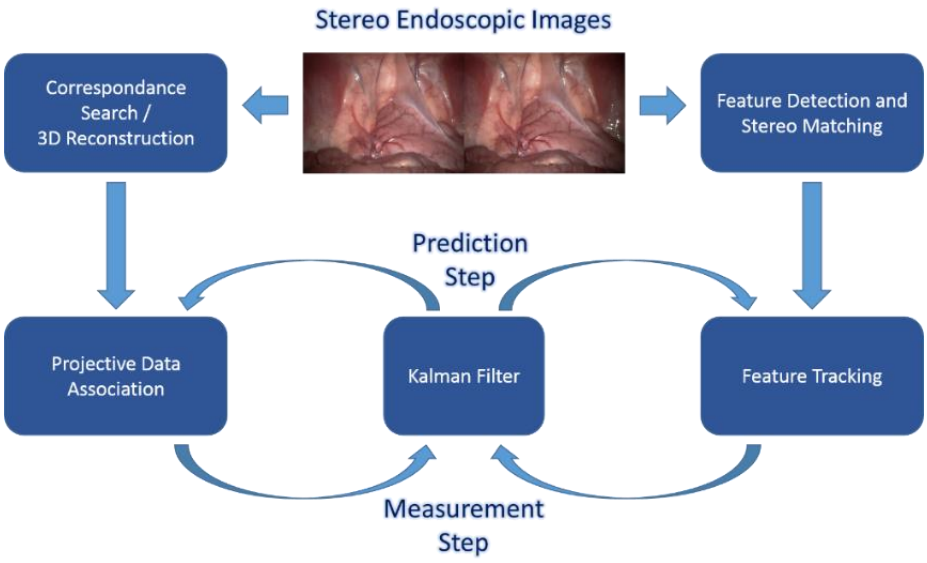

Figure 1. Pose estimation process

The System State s_t is composed of the endoscope pose position y_t^W, the endoscope orientation quaternion q_t^{WC}, the velocity v_t^W, the acceleration a_t^W, the angular velocity ω_t and the x_i 3D positions of the landmarks.

$$s_t = \left(y_t^W, q_t^{WC}, v_t^W, a_t^W, \omega_t, x_{1,t}^W, \ldots, x_{N,t}^W\right) \tag{2}$$

Process Model The translational movement of the camera system in the time interval *dt* is described with the state transition equation

$$t_{t+1} = t_t + v_t dt + \frac{1}{2} a_t (dt)^2 . \tag{3}$$

The rotational change is represented with an incremental orientation quaternion as proposed in [10]. Assuming $\{C\}$ is the camera coordinate system at the time *t*, $\{\tilde{C}\}$ is the camera coordinate system at *t*+*dt* and $\{W\}$ the world coordinate system. The transformation from $\{\tilde{C}\}$ to $\{C\}$ can be computed approximately with the angular velocitys of the rotation axes:

$$\begin{aligned} q^{C\tilde{C}} &\approx \left(\sqrt{1-\varepsilon}, \frac{w_x dt}{2}, \frac{w_y dt}{2}, \frac{w_z dt}{2}\right)^T \\ \varepsilon &= \left(\frac{w_x^2}{4} + \frac{w_y^2}{4} + \frac{w_z^2}{4}\right)(dt)^2 \end{aligned} . \tag{4}$$

Measurement Model The current system state vector s_t is mapped onto the measurement with

$$x_i^C = R^{CW}\left(x_i^W - t^W\right). \tag{5}$$

The measurement equation is used for every landmark integrated into the Kalman Filter.

3. RESULTS

The evaluation consists of an in silico and an ex vivo data set (figure 2). The in silico liver sequence was made with a simulation framework and contains 100 images of a circular motion. The ex vivo data (porcine tissue) was created with a tracked Wolf PAL stereo endoscope and contains 134 images. The reference endoscope pose was tracked with a Polaris tracking system.

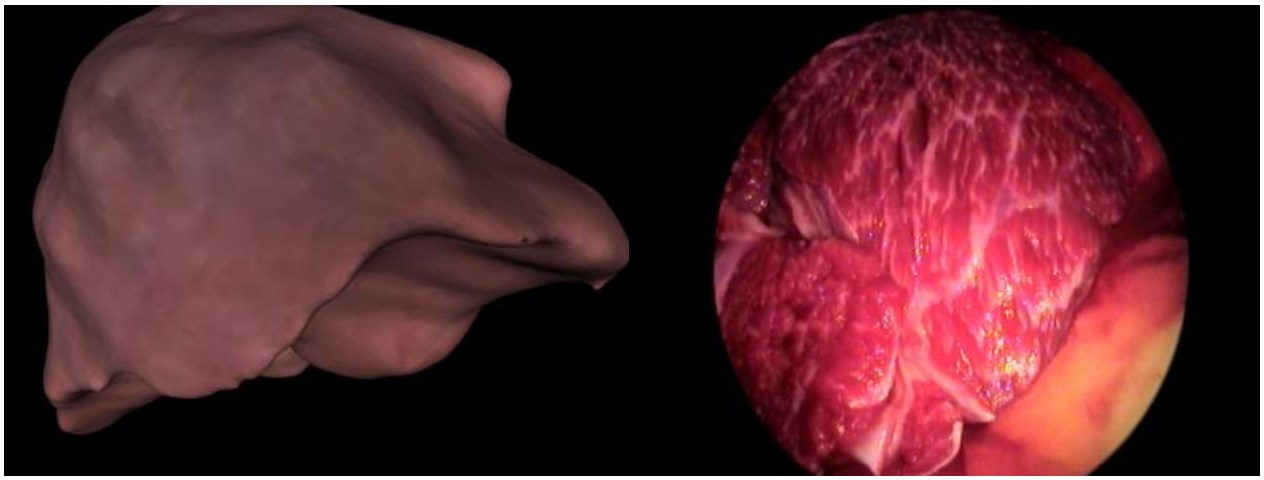

Figure 2. sample images from in silico (left) and ex vivo (right) experiment

3.1 Simulation Data

The simulation results (table 1) are showing that the combined approach is working reliable. The PDA approach is also delivering good results while the feature tracking results are slightly worse than with the other methods. Within the controlled environment of the simulation every method is working without major failures. This explains why the combined method does not clearly surpass the individual methods.

Method	Error X-Axis	Error Y-Axis	Error Z-Axis	Error Orientation
Combined	**4.37** (±2.72)	**1.43** (±0.93)	**7.41** (±2.88)	**2.51** (±0.43)
Feature Tracking	**6.65** (±4.03)	**1.28** (±1.00)	**9.56** (±3.51)	**3.31** (±0.69)
PDA depth map	**4.72** (±3.44)	**1.34** (±1.09)	**7.57** (±3.43)	**1.72** (±0.46)

Table 1. The absolute mean error and standard deviation for simulated data. Computed for each Axis and for orientation.

3.2 Ex Vivo Data

As with the simulated data, the ex vivo experiment (table 2 and figure 3) is showing promising results. The combined approach presents the overall smallest error. With the more realistic and challenging environment the combined method is proving the advantage of integrating different image qualities like structure and texture.

Method	Error X-Axis	Error Y-Axis	Error Z-Axis	Error Orientation
Combined	**7.40** (±4.37)	**1.77** (±0.95)	**2.12** (±1.78)	**5.56** (±3.31)
Feature Tracking	**8.85** (±5.88)	**20.10** (±12.86)	**2.87** (±2.85)	**18.52** (±10.58)
PDA depth map	**7.35** (±4.27)	**3.65** (±1.95)	**3.32** (±3.02)	**4.86** (±2.60)

Table 2. The absolute mean error and standard deviation for simulated data. Computed for each Axis and for orientation.

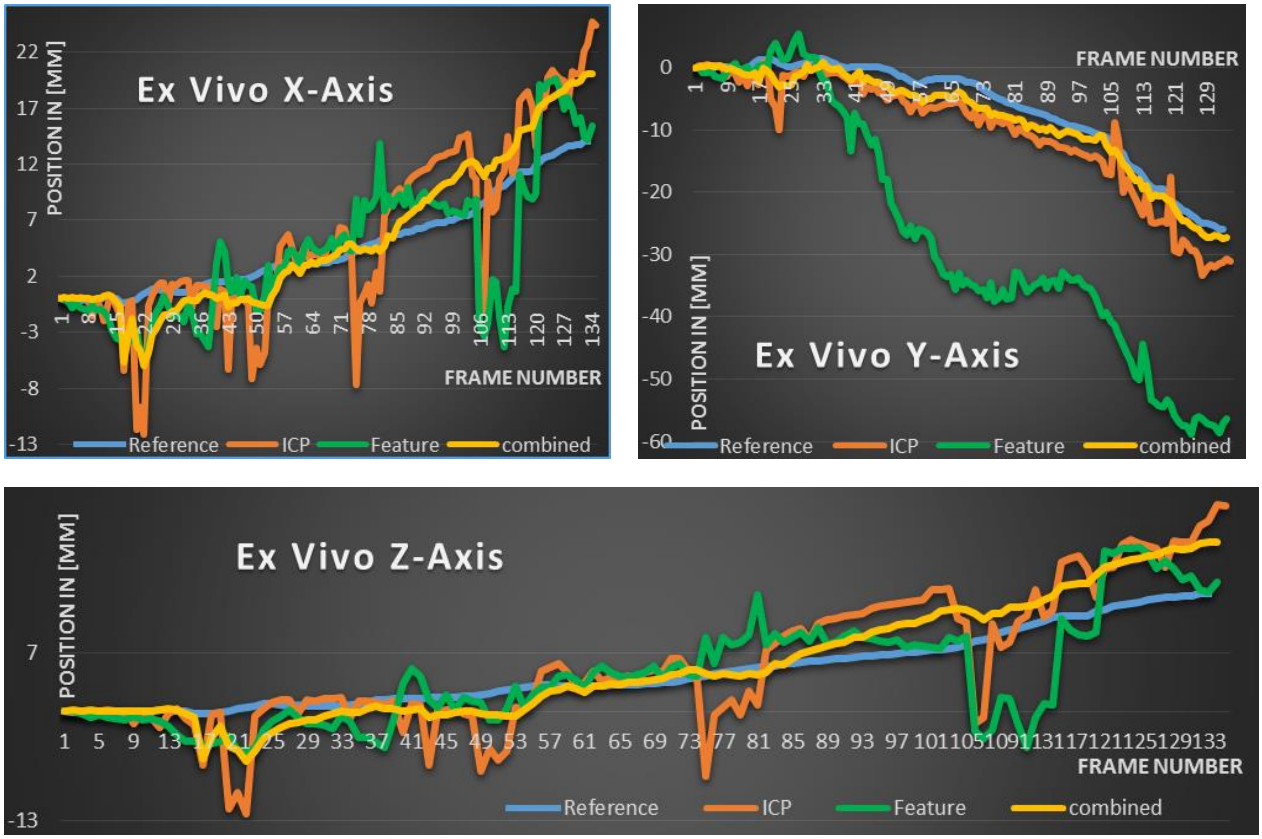

Figure 3. Estimated Axis values for ex vivo experiment.

4. CONCLUSION

We present a novel approach for robust endoscope tracking used for 3D mosaicking in laparoscopic interventions. The method is based on a Kalman Filter scheme that combines measurement inputs from dense depth and sparse stereo image feature tracking. The incorporation of two different tracking methods and the movement prediction model of the Kalman Filter is presenting a more robust way to deal with the challenging image scenes in laparoscopic surgery. The method shows promising results. Moving forward to the direction of real surgery, the visually more challenging in vivo datasets will show the benefit of using complementary image information for robust pose estimation.

In our future work, we will focus on features particularly suited for laparoscopic image scenes. Classical feature methods like SIFT have shown a semi-optimal performance in the laparoscopic domain. A feature detector and descriptor specifically designed for the task could lead to great improvements in stability and accuracy. The soft tissue deformation during surgery is another big challenge that has great impact on the pose estimation accuracy. Biomechanical models and simulation are presenting themselves as a promising part of the solution.

ACKNOWLEDGMENTS

The present research is sponsored by the Klaus Tschira Foundation and was conducted within the setting of the SFB/Transregio 125, Project A01 funded by the German Research Foundation.

REFERENCES

[1] Reichard, Daniel, et al. "Intraoperative on-the-fly organ-mosaicking for laparoscopic surgery." *Journal of Medical Imaging* 2(4), 045001-045001 (2015)

[2] Mountney, Peter, et al. "Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006. Springer Berlin Heidelberg, 347-354 (2006)

[3] Speidel, S., et al. "Robust feature tracking for endoscopic pose estimation and structure recovery." *SPIE Medical Imaging* International Society for Optics and Photonics, 2013. (2013)

[4] Blais, Gérard, and Martin D. Levine. "Registering multiview range data to create 3D computer objects." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 17(8) 820-824 (1995)

[5] Curless, Brian, and Marc Levoy. "A volumetric method for building complex models from range images." *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* ACM, (1996)

[6] Newcombe, Richard A., et al. "KinectFusion: Real-time dense surface mapping and tracking." *Mixed and augmented reality (ISMAR) 2011 10th IEEE international symposium on* IEEE, (2011)

[7] Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60(2), 91-110 (2004)

[8] Lucas, Bruce D., and Takeo Kanade. "An iterative image registration technique with an application to stereo vision." *IJCAI* 81, (1981)

[9] Davison, Andrew J., et al. "MonoSLAM: Real-time single camera SLAM." *Pattern Analysis and Machine Intelligence IEEE Transactions on* 29(6), 1052-1067 (2007)

[10] Azarbayejani, Ali, and Alex P. Pentland. "Recursive estimation of motion, structure, and focal length." *Pattern Analysis and Machine Intelligence IEEE Transactions on* 17(6), 562-575 (1995)