Landmark-based Feature Tracking for Endoscopic Motion Analysis

S. Friedl[1,2], B. Morgus[1], A. Kage[1], C. Münzenmayer[1], T. Wittenberg[1], T. Bergen[1]

[1] Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany
[2] University Hospital Erlangen, Germany

Contact: sven.friedl@iis.fraunhofer.de

Abstract:
Automated image analysis and interpretation within computer assisted minimally invasive surgery (MIS) most often depend and rely on manually defined landmarks, visible in endoscopic views. More specific, within many types of applications, such landmarks must be tracked automatically during the intervention. Typical feature tracking approaches are able to track slightly changing landmarks over time, as they occur in endoscopic image sequences, but are originally most often designed to track automatically detected salient points. In this contribution an approach is presented, where the advantages of feature descriptors and corresponding matchers can be used to track manually defined landmarks. Based on such initiated landmark points, local feature detection and tracking utilizing SURF or KLT features as descriptors, is executed. Within the region of interest as a constraint, movements of the detected features can be used to approximate the original landmark movements.

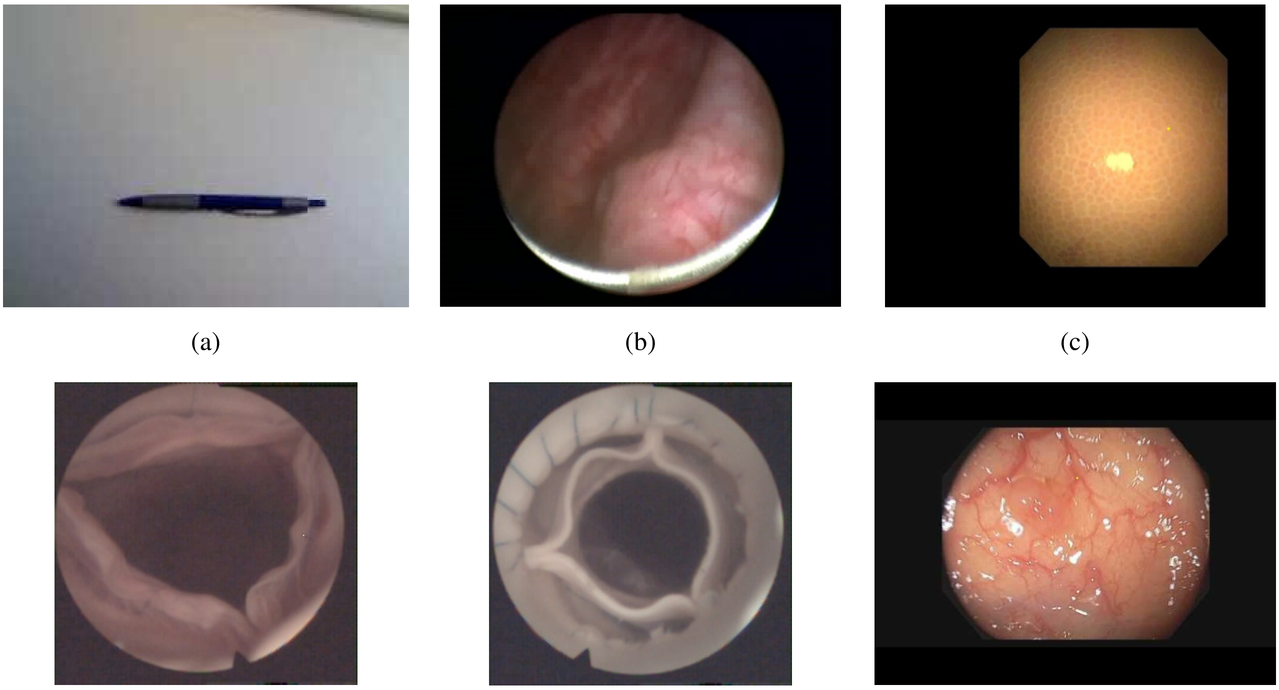
Keywords: Feature Tracking, Anatomical Landmarks, SURF, KLT

1 Problem

In various clinical scenarios, as e.g. minimally invasive surgery (MIS), endoscopy-based diagnostics in the upper and lower GI tract or clinical motion analysis of organs, the view onto the site is usually restricted by endoscopic apertures. Nevertheless, computer based (surgical) assistance systems often rely on manually defined and initiated landmarks, which must be tracked during the intervention. A typical application of this type is for example the tracking of landmarks on the beating heart within open or minimal invasive heart surgery to enable the augmentation of the view with matched and warped pre-operative image data [1,2,3,4]. Also for clinical motion analysis as the determination of movements of heart valves, manually defined landmarks have to be tracked [5,6]. Nevertheless, due to the slight but constant change of the underlying anatomical structure as well as changes of endoscopic orientation and illumination, classic template matching approaches such as correlation or SSD are neither stable nor sufficient [1,2] for such applications. Alternatively, feature tracking methods are used to determine movements in similar image data [7,8]. The literature dealing with tracking of medical image data from monocular image sequences describes such approaches, which are designed to automatically detect newly appearing local features in the scene and are hence not designed to deal with externally provided landmarks. To solve this problem and to exploit the advantages of combining robust feature descriptors with manually defined landmark points, a local feature tracking approach for endoscopic motion analysis is presented and evaluated on various types of endoscopic imagery.

2 Methods

Robust feature descriptors and the corresponding matching algorithms are well known and established methods to identify and track prominent identical points in sequences of consecutive image frames. These methods have been optimized to track feature points with specific characteristics in monocular image sequences. In contrast, manually initiated landmarks for tracking are unlikely to coincide with optimal (in the sense of feature tracking) features. Thus, depending on the visibility of interesting anatomical structures or landmarks and their movement and velocity within an endoscopic image sequence, a region of interest (ROI) can be defined around a manually initiated landmark point. As one side condition the ROIs must be defined sufficiently small and in such a way that clearly visible landmarks and the interrelated anatomy lies within the region and independent movements of adjacent ROIs do not interfere with each other. Additionally, the ROIs have to be large enough to cover the possible displacements of consecutive image frames based on the occurring movements. Now, within each manually initiated region, local feature detection and tracking is applied using well-established feature tracking approaches, which are promising for clinical (real-time) applications: the *Kanade-Lucas-Tomasi* (KLT) feature tracker [9,10] and *speeded up robust features* (SURF) [11]. SURF tracking is realized by descriptor comparison within the ROI.



(a) (b) (c) (d) (e) (f)

Figure 1: Typical images from experimental monocular data sets: ball-pen test sequence (a), view inside a bladder (b), surface of a liver (c), native heart valve (d), artificial heart valve (e), and colon tissue (f).

Within our evaluation framework, these matching algorithms are applied to the manually selected and initiated ROIs to recognize corresponding feature points in consecutive frames. Assuming that detected feature points vary from the given landmarks, but still describe the surrounding anatomic region of interest, the resulting movements can be used to estimate the movement of the original landmark.

3 Results

To evaluate the proposed approach, it has been applied to five monocular endoscopic video image sequences of different organs. These image sequences cover the surface of a human liver, the view into the human bladder, tissue of the colon, a native and as well as an artificial heart valve, cf. Figure 1. The movements within those sequences vary from almost pure translation of the endoscope (e.g. liver) up to complex deformations of the organ (e.g. heart valves). To prove the principal concept and the correctness of the implementation, a well-behaving test sequence with only one moving item (a ball-pen) was recorded. Specifically, this rigid object being different from the organic structures and being displaced over time with a constant and stable background has been chosen to be independent from application related restrictions in the tracking task, where several movements of landmarks, organs and tissue background are overlaying each other. As ground truth data for comparison and evaluation, the selected landmarks have been labeled manually in all frames of the sequences and the corresponding coordinates have been stored. Due to the intention to estimate the movement of a certain anatomical structure and not to track the exact landmark, the spatial distance between the ground truth points and the tracked features has not been considered as a meaningful error measure. Instead, the movement *m* of the tracked Points *Tr* in frame *i*, relative to the initial key frame, is determined as

$$m = \sqrt{(Tr_{xi} - Tr_{x0})^2 + (Tr_{yi} - Tr_{y0})^2}\ [Pixel].$$

For each of the endoscopic image sequences, the movement of the ground truth (GT) was compared to the detected motion using both the SURF as well as the KLT descriptors. The resulting trajectories of all six sequences are shown in Figure 2. Ideal tracking with respect to the manually labeled ground truth data would result in identical trajectories. The more parallel the trajectories of the tracking approaches are, compared to the trajectory of the ground truth, the better the result can be regarded.

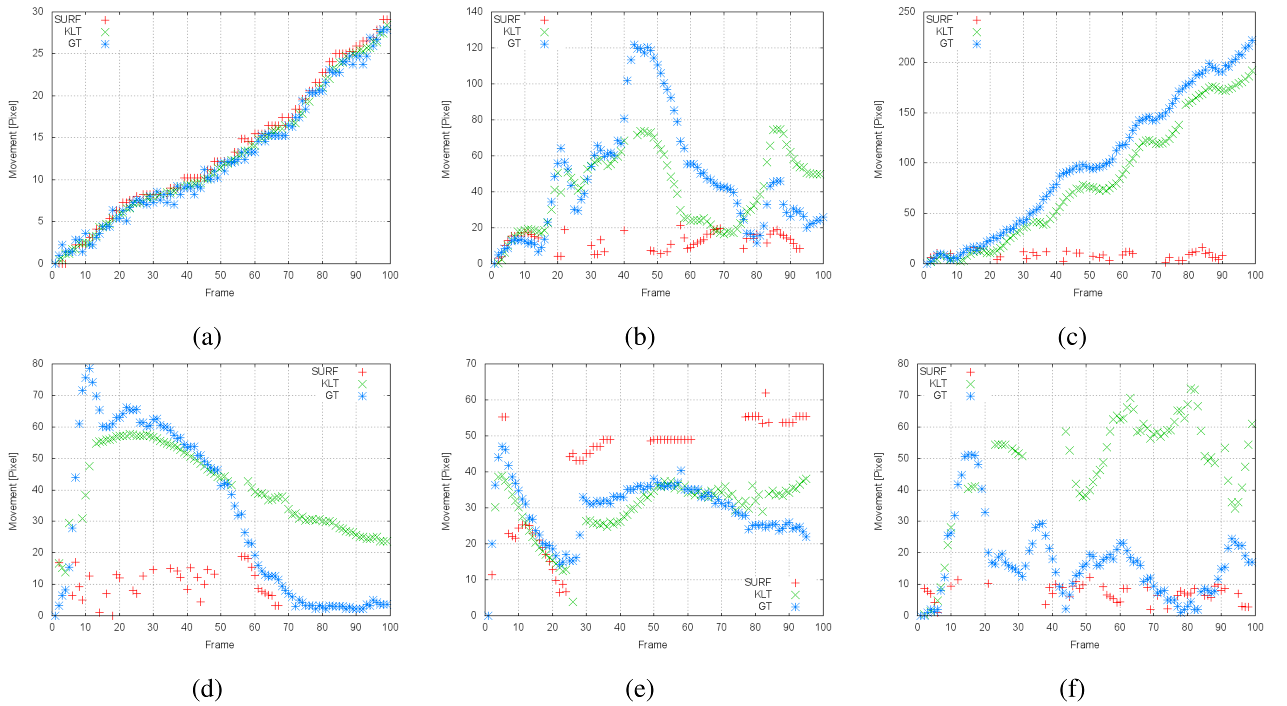


Figure 2: Trajectories of the movements for the six evaluated recordings for *Speeded Up Robust Features* (SURF, +), *Kanade-Lucas-Tomasi* (KLT, ×), and the annotated ground truth (GT, *): ball-pen test sequence (a), view inside a bladder (b), surface of a liver (c), native heart valve (d), artificial heart valve (e), and colon tissue (f).

	KLT	SURF
Ball pen	100 %	100 %
Bladder	32 %	14 %
Liver	75 %	5 %
Native heart valve	4 %	9 %
Artificial heart valve	25 %	0 %
Colon	13 %	29 %

Table 1: Manually rating of the consistency of the evaluated feature tracking methods

In addition, both tracking approaches were evaluated by manually rating the consistency of feature tracking. For each sequence, we counted the number of frames, in which the organic structure (i.e. initially selected feature) was tracked correctly, independently of the pixel based ground-truth distance. Table 1 depicts the results obtained.

4 Discussion

As can bee seen in Figure 2(a), as well as in Table 1, upper row, the movements within the reference ball-pen test sequence could be tracked with only small deviations. Thus, the principle approach of exploiting robust feature descriptors for manually defined and initiated landmarks seems to be promising. However, applying the proposed tracking method to real medical endoscopic image data, the results differ. For the liver (Fig. 2(c)) and both heart valve sequences (Fig. 2(d+e)), the movements of the manually selected landmarks can only be roughly estimated. In the case of the bladder (Fig. 2(b)), the difference between the ground truth and the tracked movement is increasing, while for the colon (Fig. 2(f)), the original movement is hardly identified in the tracking trajectories. Interestingly, in some sequences (bladder, native heart valve) the KLT tracker yields closer results to the ground truth while in other sequences (colon, artificial heart valve) the SURF features seem better. The rating depicted in Table 1 confirms these results. Although none of the approaches yielded better results throughout all image sequences, the KLT approach showed better stability than the SURF approach in most cases. Dealing with medical image data leads to various possible sources of error, e.g. low contrast images with significant noise, substantial organ deformations, as well as structured surfaces and specular highlights.

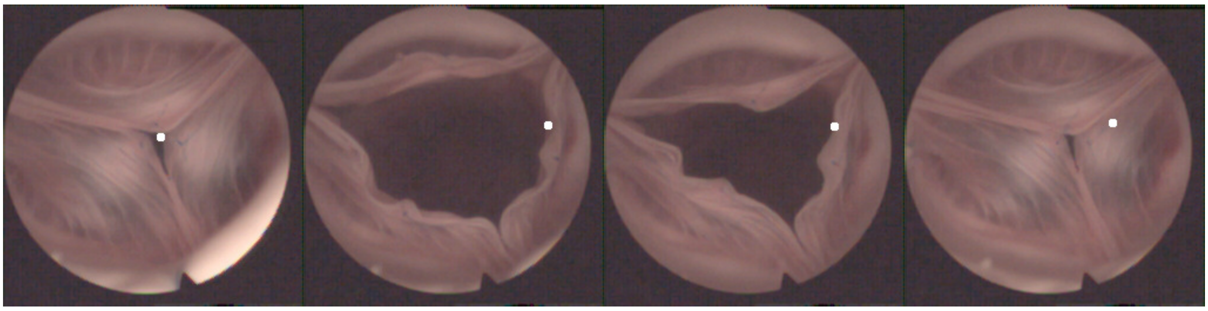


Figure 3: Example frames for a tracking result of native heart valves applying the KLT approach. Shown is each 25th frame of a consecutive sequence. The total distance of movement is approx. 125 pixel and approx. 5 to 10 pixel between each frame of the sequence.

Figure 3 shows an example for a tracking result applying the KLT tracker to a recording of a native heart valve. As can be seen, the cusps are tracked during the opening phase of the heart valve but lost while closing the orifice. Even though the proposed approach is a promising concept and a good start for further research, a clinical use of landmark based tracking in medical monocular endoscopic image data will demand substantial further improvements. Besides applying and evaluating enhanced tracking and correspondence methos for landmark tracking in monocular endoscopic images, a technical alternative for such scenarios could be the use of stereoscopic endosopes [12]. Nevertheless, such stereoscopic imaging have much larger diameters and can thus not be applied for all applications, and are currently only available as rigid endoscopes.

References

[1] S. Friedl, T. Wittenberg, M. Kondruweit. Interactive registration and visualization of cardiac video and angiography. In IFMBE Proc. Vol. 25/IV, World Congress on Med. Physics & Biomedical Engineering, pp. 468-471, 2009

[2] T. Ortmaier, M. Gröger, and G. Hirzinger. Multisensorielle Schätzung der Herzbewegung in der minimal invasiven Chirurgie CURAC-Jahrestagung, October 4-5, 2002, Leipzig – Germany

[3] T. Ortmaier, M. Groeger, and G. Hirzinger: Robust Motion Estimation in Robotic Surgery on the Beating Heart. Proc's Computer Assisted Radiology & Surgery (CARS) , June 26-29, 2002, Paris – France, pp. 206-211.

[4] T. Wittenberg, K. Drechsler, D. Kaltenbacher, S. Friedl, C. Reis, G. Sakas, J. Stallkamp, C. Rotinat, Y. Perrot, M. Kondruweit. 'MISS heart': Assisting systems for minimal invasive smart suturing in cardiac surgery? A conceptually closed loop approach. In IFMBE Proc. Vol. 25/IV, World Congress Med. Physics & Biomed. Eng., pp. 445-448, 2009

[5] A.P. Condurache, T. Hahn, U.G. Hofmann, M. Scharfschwerdt, Martin Misfeld, Til Aach. Automatic measuring of quality criteria for heart valves. Med. Imaging 2007: Image Processing, SPIE, San Diego, CA,, 17-22.2.2007

[6] T. Wittenberg, R. Cesnjevar, S. Rupp, M. Weyand, M. Kondruweit. High-Speed-Camera Recordings and Image Sequence Analysis of Moving Heart-Valves: Experiments and First Results. In T. Buzug, D. Holz, S. Weber, J. Bongartz, M. Kohl-Bareis, U. Hartmann (Eds), Advances in Med. Engineering, Springer Proc's in Physics 114, pp. 169-174. Workshop, 7.-9.3.2007 in Remagen, Springer, Heidelberg, 2007.

[7] T. Bergen, S. Ruthotto, C. Münzenmayer, S. Rupp, D. Paulus, C. Winter. Feature-based real-time endoscopic mosaicking. In Proc. 6th International Symposium on Image and Signal Processing and Analysis, pp. 695-700, 2009

[8] T. Bergen, A. Schneider, C. Münzenmayer, F. Knödgen, H. Feussner, T. Wittenberg, C. Winter. Echtzeit-Stitching endoskopischer Bilder für eine erweiterte Sicht in chirurgischen Eingriffen. Endoskopie Heute, 24(1):60-61, 2011

[9] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In Proc. 7th International Joint Conference on Artificial Intelligence, pp. 674-679, 1981

[10 J. Shi and C. Tomasi. Good features to track. In Proc IEEE Conference on Computer Vision and Pattern Recognition, pp. 593-600, 1994

[11] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool. SURF: Speeded Up Robust Features. Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-359, 2008

[12] W. Lau, N. Ramey, J. Corso, N. Thakor, G. Hager. Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation. In C. Barillot, D.R. Haynor, and P. Hellier (Eds.): MICCAI 2004, LNCS 3217, pp. 494–501, 2004.