A Real-Time Target Tracking Algorithm for a Robotic Flexible Endoscopy Platform

Chapter · February 2016

DOI: 10.1007/978-3-319-29965-5

A Real-Time Target Tracking Algorithm for a Robotic Flexible Endoscopy Platform

Nanda van der Stap[1(✉)], Luuk Voskuilen[1], Guido de Jong[1], Hendrikus J.M. Pullens[2], Matthijs P. Schwartz[2], Ivo Broeders[3], and Ferdi van der Heijden[1]

[1] MIRA Institute for BMT and TM, Carré 3.526, PO box 217, 7500 AE Enschede, The Netherlands
{n.stap,f.vanderheijden}@utwente.nl

[2] Department of Gastroenterology and Hepatology, Meander Medical Center, Amersfoort, The Netherlands

[3] Department of Surgery, Meander Medical Center, PO Box 1502, 3800 BM Amersfoort, The Netherlands

Abstract. Complex endoscopic interventions require a new generation of devices and instruments. A robotic platform for flexible endoscopy through telemanipulation was developed to meet this demand. The concept of telemanipulation allows the development of software for computer-aided surgery. Intelligent navigation such as automated target centralization could assist the endoscopist during procedures.

A real-time algorithm was designed for tracking a target region that is of specific interest for the surgeon. Therefore, the physician needs to indicate the region to be tracked, which then will be centralized (locked). The goal of this research is to investigate the robustness and accuracy of the tracking algorithm during endoscopic interventions. The region of interest can be a polyp for polypectomy, Vater's ampulla for Endoscopic Retrograde CholangioPancreatography (ERCP), Barrett's epithelia for gastroscopic biopsy or any area in more complex procedures. The algorithm was tested in vitro on image sequences obtained during real endoscopic interventions.

The indicated area of interest could be tracked in all image sequences, with an accuracy of 91.6 % (Q1–Q3 77.7 %–99.0 %, intraclass correlation). The algorithm was robust against instruments or smoke in the field of view. Tracking was less robust against very large camera movements.

The developed target lock worked robustly, in real-time and was found to be accurate. Improvements include improving the robustness of the algorithm against motion blur and drift.

1 Introduction

A trend towards minimizing the invasiveness of surgical procedures exists. Instead of open surgery, endoscopic or keyhole surgery is more often performed. Endoscopic surgery, where rigid cameras or endoscopes are used, decreases the amount of scarring and blood loss in the patient, leading to less pain and faster

X. Luo et al. (Eds.): CARE 2015, LNCS 9515, pp. 81–89, 2016.
DOI: 10.1007/978-3-319-29965-5_8

recovery times. With the transition from open to endoscopic surgery, flexible endoscopes are looked at for more complex procedures as well. Originally, only diagnostics and small therapeutic interventions (the removal of polyps in the colon for example) were performed with flexible endoscopes. Nowadays, increasingly large tumors are being removed, with the aim to spare the patient from a more invasive surgical procedure.

However, flexible endoscopes have vital drawbacks that make them difficult to handle, especially during complex procedures [1]. Endoscope handling is not intuitive and far from ergonomic. The instruments that can be inserted through the working channel of the endoscope are not capable of triangulation. Mostly, only one instrument channel is present, causing many instrument changes and a significant loss of dexterity.

Robotization is thought to improve the handling properties and dexterity of flexible endoscopes. Additionally, navigation can be automated (e.g. [2–4]). Our research is specifically aimed at automating endoscope navigation and fixation during interventions. Ultimate aim is to use image-based control to correct the endoscope tip once the endoscope is at the intervention site. The endoscopist can indicate a focus area manually, and the system described here will track this area continuously in the image. The tracked area will be kept in the center of the screen as much as possible, resulting in a so-called 'target lock'. The target should be centralized despite movement of the endoscope or the environment.

Others have investigated re-targeting of endoscopes. A useful application is for instance the optical biopsy, a visualization of cellular structures using optical instruments in the working channel of an endoscope. SLAM (Simultaneous Localization and Mapping) techniques combined with probe tracking, video manifolds for the patient-specific clustering of images and epipolar geometry recovery are examples of solutions for the retargeting problem in this application [5–8]. Chu et al. describe a flexible tip for a rigid endoscope and target tracking, but it is unclear which tracking approach they use [9].

We are interested in developing a clinically successful target locking system for robotized flexible endoscopy. This leads to the following requirements:

1. Real clinical added value: the procedure of interest should not be hindered or take longer due to the automation.
2. Robust to low texture frames, large movements, varying illumination conditions, instrument interference and occlusions (fluids, tissue deformation).
3. Accurate enough to enable small tip corrections in the order of a few millimeters.
4. Real-time functionality, such that the endoscopist can direct all his/her attention towards the surgical target instead of controlling the camera.
5. Easy correction functionality: if the target changes, for instance if tissue is taken from it, it should be easy to re-localize the target region.

Feedback for the control will be obtained from the images by visual motion tracking and correcting for this. Motion tracking has been employed in flexible endoscopy [2]. A key issue is robustness and accuracy of tracking, implying an accurate outlier detection mechanism. Visual motion tracking is challenging due to the nature of the images (often low in texture and suffering from the artifacts named above in 2.). There is a trade-off between accuracy and computational effort with feature tracking algorithms. For our application, the algorithm must run fast enough to accurately correct the tip before the endoscopic image has changed too much.

The contribution of this paper lies in the real-time, accurate, feature-based target tracking aimed at optimal system performance with real clinical value. Therefore, real clinical data is used for evaluation. The described application is developed for the robotized flexible endoscopy system Teleflex [10], but minor changes can make it suitable for other robotized endoscopes.

2 Materials and Methods

System requirements were established in close collaboration with several expert endoscopists. A thorough clinical evaluation among the endoscopists led to the conclusion that routine colonoscopies, ERCPs and EMRs (Endoscopic Mucosal Resections) were the interventions most likely to benefit from a target lock. These procedures are known for their clinical complexity with respect to specific sub-interventions. The most difficult part during an ERCP is the insertion of a probe in Vater's papilla. Once the papilla is in the proper position in the view, the endoscope should remain fixed so that the probe can be manipulated properly. Similar situations were indicated for colonoscopies and EMRs. Twelve image sequences were selected and contained the various sub-interventions of interest for the target lock (Table 1). A sub-intervention was estimated to last for 4 s on average; this is therefore the length of the sequences. An exception forms the papilla insertion during an ERCP. Therefore, the length of sequence 3 was doubled.

To use images as control feedback, the *bandwidth of the motion* should be an order of magnitude less than the *sample frequency*. The Nyquist frequency for 25 fps is 12.5 Hz. For control purposes, the rule of thumb is to have a 5–10 times higher sample frequency than the motion bandwidth. In this case, most motions have a frequency of 0–5 Hz. With an effective frame rate (sample frequency) of 24–25 fps this requirement was met in our system.

Real-time optimization can be done by cropping or down-scaling of the images, frame-skipping, code- and platform-optimizations or heterogeneous computing. The latter will result in the most significant improvement without data loss. To enable heterogeneous computing, the feature tracking algorithms were implemented using OpenCV with OpenCL.

Image sequences had a resolution of 768×576 and a frame rate of 25 frames per second (fps). All results were generated using a HP Elitebook 8570 w mobile workstation running on a 64 bit operating system (Windows 8.1) with an Intel Core i7-3630QM processor, 8 GB DDR3 RAM and an AMD FirePro M4000 graphics card. Programming was done using Microsoft Visual Studio Express 2013, with libraries from OpenCV 2.4.9 and OpenCL 1.1. A colonoscope, an ERCP-scope and a pediatric colonoscope (for EMR) were used to record the procedures. The properties of each of them are listed in Table 2.

Table 1. Image sequence properties. APC: Argon Plasma Coagulation. Note: in all sequences, instruments are present in the field of view (FOV).

Number	Procedure	Intervention	Target	Disturbing factor
1	Colonoscopy	Polypectomy	Polyp	Coarse movements
2	Colonoscopy	Polypectomy	Polyp	Poor illumination; Occlusions
3	ERCP	Cannulation	Vater's papilla	Target near edge of the FOV; Occlusions
4	ERCP	Sphincterotomy	Vater's papilla	Target near edge of FOV; Smoke; Occlusions
5	EMR	Injection	Primary tumor	Large target; Color change due to dye injection
6	EMR	APC	Residual lesion	Large area of removed mucosa; Small target; Sparks
7	Colonoscopy	Injection	Polyp	Color change due to dye
8	Colonoscopy	Polypectomy	Polyp	Large polyp; Dirt on lens
9	ERCP	Cannulation	Vaters ampulla	Endoscope motion
10	ERCP	Stent removal	Vaters ampulla	Multiple instruments in view
11	EMR	Partial resection	Primary tumour	Large target; Coarse movements; Dye injection
12	Colonoscopy	APC	Polyp	Sparks; Bubbles; Small target

Table 2. Properties of each endoscope used to record the image sequences. FOV: Field of View. DoF: Depth of Field. DoV: Direction of View.

Procedure	Endoscope	Properties
Colonoscopy	Olympus CF-H180AL	FOV: 170°; DoF: 2–100 mm; Length: 1680 mm
ERCP	Olympus TJF-160VR	FOV: 100°; DoF: 5–60 mm; Length: 1240 mm
EMR	Olympus PCF-PH190L	FOV: 140°; DoF: 2–100 mm; Length: 1680 mm

2.1 Algorithm

For accurate and robust feature tracking, SIFT (Scale Invariant Feature Transform [11]) will be suitable, because blob-like features are abundantly present in the image sequences that were used (Fig. 1). However, detecting and matching these features takes considerable computational effort. For our real-time application, we therefore chose to use SURF features (Speeded-Up Robust Features [12]). These are nearly as accurate as SIFT features, but decrease computational effort considerably [13].

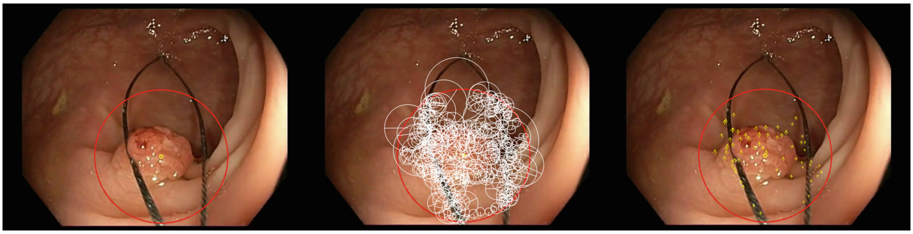

Fig. 1. Example image of polypectomy with polyp and instrument present. Left: scene with ROI indication. Middle: SURF features withing ROI (indicated as white circles). Right: final motion vectors (indicated as arrows).

As stated in the Introduction, an accurate outlier detection mechanism is key to robust system performance. In the original SURF algorithm brute force matching was selected for accuracy reasons. In an attempt to reduce computational time and improve reliability of the matches, we added the second-to-nearest-neighbor (SNN) distance ratio check, as proposed by Lowe [11]. To increase the number of features, even in low-textured areas, preprocessing (grayscale conversion and histogram equalization) was applied on the images. Outlier removal based on vector magnitude was added to the algorithm to improve robustness. The complete algorithm works as follows:

1. Initialization:
   (a) Acquire and visualize the first image, reference image I_{ref}.
   (b) Select the target ROI in I_{ref}: a circle with position $\mathbf{x}_{ref}$ and radius R.
   (c) Set $\mathbf{x}_{target} = \mathbf{x}_{ref}$.
   (d) $\{\mathbf{y}_{ref}(n), \mathbf{f}_{ref}(n)\} = \text{get\_surf\_features}(I_{ref}, \mathbf{x}_{ref})$.
2. Acquire current image I_{cur}.
3. $\{\mathbf{y}_{cur}(m), \mathbf{f}_{cur}(m)\} = \text{get\_surf\_features}(I_{cur}, \mathbf{x}_{target})$.
4. Match $\{\mathbf{f}_{ref}(n)\}$ to $\{\mathbf{f}_{cur}(m)\}$ with SNN distance ratio check, yielding matched indices $\{n(k), m(k)\}$, with $k = 1, ..., K$.
5. Get displacement vectors $\{\mathbf{d}(k) = \mathbf{y}_{cur}(m(k)) - \mathbf{y}_{ref}(n(k))\}$
6. Remove outliers. Condition: $\|\mathbf{d}(k)\| > 2 * median(\{\|\mathbf{d}(k)\|\})$.
7. $\mathbf{x}_{target} = \mathbf{x}_{ref} + mean(\{\mathbf{d}(k)\})$.
8. Repeat till end from 2.

Procedure $\{\mathbf{y}(n), \mathbf{f}(n)\} = \text{get\_surf\_features}(I, \mathbf{x})$

1. Convert image I from RGB to grayscale.
2. Apply histogram equalization to I to increase feature number.
3. Detect SURF key point positions $\{\mathbf{y}(n)\}$ and key point descriptors $\{\mathbf{f}(n)\}$.

2.2 Analysis

Targets to be tracked (Table 1) were manually annotated throughout the image sequences by an expert interventional endoscopist (>2000 endoscopies). The automatically found location was compared for accuracy to the manual results using intra-class correlation analysis (ICC, [14]). Tracking error was given by the Root Mean Square Error (RMSE) of the distance in pixels between the two targets. Computational times were recorded to measure real-time performance of the system. Robustness was measured by counting the number of feature matches and inlying matches per frame.

3 Results

In all sequences, the manually indicated target could be tracked with an high accuracy of 91.6 % (Q1–Q3 77.7 %–99.0 %, see Table 3). Sequence 7 had the best tracking results with a correlation of 99.9 % and a RMSE of 9.4 pixels. Median tracking error was 38.9 pixels (Q1–Q3: 28.7–76.3, see Table 3). The algorithm proves robust against smoke, fluid and instrument interference, color changes, occlusions and poor illumination. Sequences that suffered from these artifacts nonetheless led to the best results (Fig. 2). The text boxes and show that large motions and motion blur are the cause for the biggest tracking errors.

Table 3. Results per image sequence.

Sequence	ICC (%)	Median matches	Median inliers	RMSE (pixels)
1	81.9	77	34	149.1
2	89.4	165	56	32.4
3	99.7	154	67	25.5
4	98.8	123	32	34.2
5	96.6	201	73	27.4
6	93.8	239	73	28.0
7	99.9	184	62	9.4
8	73.6	205	61	116.0
9	99.8	88	32	16.6
10	75.7	236	38	43.7
11	78.4	131	24	123.0
12	63.5	223	58	57.1

The median number of matches and inliers per frame was at most 239 and 73, respectively (Table 3). Note that the lowest number of matches and inliers correspond to the lowest ICC. Our matching approach was more accurate and slightly faster than the original brute force matching, with an average of 43.79

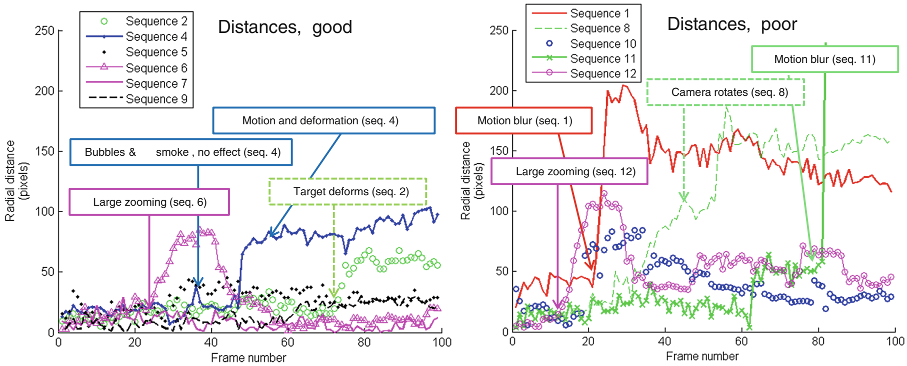

Fig. 2. Left: sequences with best ICC and smallest RMSE. Graphs show distance in pixels between the center of the two (automatic and annotated) targets. Note the text boxes that explain the larger shifts. Right: sequences with the poorest outcome. All tracking errors are caused by large motions.

($\pm$4.37) ms against 48.55 ($\pm$4.33) ms per frame. We have uploaded the image sequences in the additional conference materials to illustrate the disturbing factors more clearly.

4 Discussion

In this study, the accuracy and robustness of a real-time tracking algorithm for automated target centralization in a robotized flexible endoscopy system was evaluated. This algorithm can be used for a variety of interventions. Here, we evaluated the algorithm using six image sequences of three different interventions. The achieved median accuracy was 91.6 %, which is an excellent result.

Robustness of the algorithm was shown by the continuous ability to track the target throughout the sequences, independent from procedure or tissue type, although several disturbing factors were present (Table 1). Inlying vectors mostly remained present and tracking was kept accurate, even with occlusions, color and illumination changes, surgical instruments, smoke and fluids present. Large and fast movements still form a problem; this caused most errors. If such an error occurred, the tracking was disturbed. For longer tracking periods of the same region this means re-initialization of the algorithm in its current form is sometimes necessary. However, we expect robustness to be improved with system implementation (see below).

Computations took a mean total time of 43.79 ms per frame. The algorithm could theoretically track the target every $\frac{1sec}{25frames}$ = 40 ms. However, when using a newer computer with a better CPU (Intel Core i5-3570K) and a better GPU (AMD Sapphire Tri-X R9 290), computational time was below 40 ms and real-time system performance was ensured.

A limitation of this research is that zooming motion is assumed to be absent, although small zooming motions were present in the used sequences. Therefore, our current focus is on the implementation of the algorithm in the robotic system, complete with zooming functionality.

Current research further includes optimizing the robotic control based on the feedback that is generated from this algorithm. When combining the algorithm with the robotic control, large tracking errors (such as these over 100 pixels) will be diminished by the integral action that is present in the robotic controller. This action effectively smoothes the feedback signal because of the limited displacement possibility of the motors within a certain time frame. If this smoothing will not be enough to obtain the desired robustness, smart filtering with which previous information (key frames) is employed will be added to the system.

Finally, we will focus on establishing clinical relevance and patient safety. The algorithms are integrated in the Teleflex system, which is currently being evaluated in a phase II clinical trial, and we expect good results from this evaluation.

5 Conclusion

A target lock was designed for complex flexible endoscopic interventions. The algorithm performed accurately, robustly and worked in real-time. Intelligent navigation in robotized systems could assist the endoscopist during complex and time-consuming procedures. Clinical added value for the patient still needs to be objectively evaluated, but preliminary evaluation results seem promising.

References

1. Valdastri, P., Simi, M., Webster III, R.J.: Advanced technologies for gastrointestinal endoscopy. Ann. Rev. Biomed. Eng. **14**, 397–429 (2012). http://www.ncbi.nlm.nih.gov/pubmed/22655598
2. Van der Stap, N., Reilink, R., Misra, S., Broeders, I., van der Heijden, F.: The use of the focus of expansion for automated steering of flexible endoscopes. In: IEEE BioRob, pp. 13–18. IEEE, Rome (2012). http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6290804
3. van der Stap, N., Slump, C.H., Broeders, I.A.M.J., van der Heijden, F.: Image-based navigation for a robotized flexible endoscope. In: Luo, X., Reich, T., Mirota, D., Soper, T. (eds.) CARE 2014. LNCS, vol. 8899, pp. 77–87. Springer, Heidelberg (2014)
4. Ott, L., Nageotte, F., Zanne, P., Mathelin, M.D., Member, S.: Robotic assistance to flexible endoscopy by physiological-motion tracking. IEEE Trans. Robot. **27**(2), 346–359 (2011)
5. Atasoy, S., Mateus, D., Meining, A., Yang, G.-Z., Navab, N.: Endoscopic video manifolds for targeted optical biopsy. IEEE Trans. Med. Imaging **31**(3), 637–653 (2012). http://www.ncbi.nlm.nih.gov/pubmed/22057050
6. Allain, B., Hu, M., Lovat, L., Cook, R., Vercauteren, T., Ourselin, S., Hawkes, D.: Re-localisation of a biopsy site in endoscopic images and characterisation of its uncertainty. Med. Image Anal. **16**(2), 482–496 (2012). http://www.ncbi.nlm.nih.gov/pubmed/22197442

7. Ye, M., Giannarou, S., Patel, N., Teare, J., Yang, G.-Z.: Pathological site retargeting under tissue deformation using geometrical association and tracking. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 67–74. Springer, Heidelberg (2013)
8. Liu, J., Wang, B., Hu, W., Zong, Y., Si, J., Duan, H.: A non-invasive navigation system for retargeting gastroscopic lesions. Biomed. Mater. Eng. **24**(6), 2673–2679 (2014). http://www.ncbi.nlm.nih.gov/pubmed/25226971
9. Chu, Y.-J., Liu, S.-P., Luo, R.C., Hu, R.-H., Yeh, C.-C., Peng, Y.W., Yen, P.-L.: Dynamic tracking of anatomical object for a steerable endoscope. In: International Conference on Advanced Intelligent Mechatronics (2012)
10. Ruiter, J., Rozeboom, E., Voort, M.V.D., Bonnema, M., Broeders, I.: Design and evaluation of robotic steering of a flexible endoscope. In: IEEE BioRob, pp. 761–767. IEEE, Roma (2012)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 1–28 (2004)
12. Bay, H., Ess, A., Tuytelaars, T., Vangool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008). http://linkinghub.elsevier.com/retrieve/pii/S1077314207001555
13. Speidel, S., Krappe, S.: Robust feature tracking for endoscopic pose estimation and structure recovery. In: Proceedings of the SPIE. SPIE digital library, vol. 8671, pp. 1–7 (2013). http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1663315
14. Jellinek, E.M.: On the use of the intra-class correlation coefficient in the testing of the difference of certain variance ratios. J. Educ. Psychol. **31**(1), 60–63 (1940). http://dx.doi.org/10.1037/h0062703