Image-Based Navigation for a Robotized Flexible Endoscope

Nanda van der Stap[1(✉)], C.H. Slump[1], Ivo A.M.J. Broeders[1,2], and Ferdi van der Heijden[1]

[1] Robotics and Mechatronics group, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands
n.stap@utwente.nl
[2] Department of Surgery, Meander Medical Center, P.O. Box 1502, 3800 BM Amersfoort, The Netherlands

Abstract. Robotizing flexible endoscopy enables image-based control of endoscopes. Especially during high-throughput procedures, such as a colonoscopy, navigation support algorithms could improve procedure turnaround and ergonomics for the endoscopist.

In this study, we have developed and implemented a navigation algorithm that is based on image classification followed by dark region segmentation. Robustness and accuracy were evaluated on real images obtained from human colonoscopy exams. Comparison was done using manual annotation as a reference. Intraclass correlation (ICC) was employed as a measure for similarity between automated and manual results.

The discrimination of the developed classifier was 6.8, making it a reliable classifier. In the experiments, the developed algorithm gave an ICC of 93 % (range 84.7–98.8 %) over the test image sequences on average. If images were classified as 'uninformative', which led to re-initialization of the algorithm, this was predictive for the result of dark region segmentation accuracy.

In conclusion, the developed target detection algorithm provided accurate results and is thought to provide reliable assistance in the clinic. The clinical relevance of this kind of navigation and control is currently being investigated.

Keywords: Flexible endoscopy · Robotized endoscopy · Vision-based steering

1 Introduction

Robotizing flexible endoscopes has potential advantages for routine medical diagnostic as well as for complex therapeutic interventions [1]. Flexible endoscopes are 1.8 m long flexible tubes, controlled by steering knobs and with a lens system and CCD or CMOS chip in the tip. This configuration makes them suitable for minimally invasive procedures. Demand for these kinds of procedures is growing with the trends of more preventive medicine and of reducing the amount of scarring due to surgical procedures. The design of flexible endoscopes has sufficed for decades, but the increased demand now leads to a need for more efficient procedures in terms of turnaround. Additionally, one instrument with two Degrees of Freedom (DOF) that protrudes from the tip does not meet the dexterity that is required for complex interventions. Control of the

X. Luo et al. (Eds.): CARE 2014, LNCS 8899, pp. 77–87, 2014.
DOI: 10.1007/978-3-319-13410-9_8

endoscope is far from intuitive; high physical workload and injuries among endoscopists have been reported (e.g. [2]).

A telemanipulation system for currently used flexible endoscopes has been developed [3]. This system, called TeleFlex, aims at improving procedure efficiency and reducing endoscopist workload at relatively low cost. Image-based navigation and control of the robotic system may help to improve the intuitiveness of the system and reduce procedure time [4]. During routine diagnostic 'high-throughput' procedures, specifically colonoscopies (endoscopy of the colon), the endoscope is first fully inserted in the colon and then retracted slowly to interpret possible lesions. The retraction phase has a recommended duration of 6 min [5] while the mean insertion phase of a colonoscope is 10–20 min [6], so efficiency improvement can be achieved during the insertion phase.

Centralization of the target direction by automatic endoscope tip correction is thought to lead to a smooth insertion (e.g. [7]), but this application has never been brought into clinical practice. The target direction in images of the colon can easily be found by segmentation of the darkest region in the image (Fig. 1, point A). This region in most cases represents the deepest part of the inside of the organ (lumen) and thus corresponds to the target direction for steering. The center of this region will be called 'target' from now on.

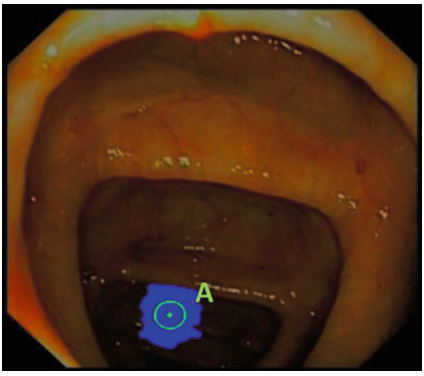

Fig. 1 Example of dark region segmentation (target direction determination) in a real colonoscopy image.

A problem with image-based navigation in colonoscopy is that the lumen often is not visible. This situation occurs when the tip of the endoscope is (almost) touching the wall of the colon, a so-called 'red-out'. In addition, fluids like blood or water frequently occlude the field of view and influence image analysis. Bubbles and motion blur are other artifacts that occur often during these kinds of exams (see Sect. 2.1).

At our department an attempt was made to develop image-based endoscope navigation earlier [8]. The center of gravity (CoG) of the image was used as the target area towards which the endoscope tip was steered. While the algorithm performs well on a plastic model of the esophagus and stomach, in real colonoscopic image sequences it suffers from the artifacts mentioned above and it drifts quickly.

Probably, one of the main reasons automated navigation algorithms for colonoscopy are currently not working in a clinical setting is that the colon is too unpredictable for complete autonomous navigation. During colonoscopy, the colon is insufflated with air, which changes the diameter and folds of the colon significantly. In addition, the endoscope itself causes a considerable shape change of the colon, limiting the use of pre-procedural imaging information (e.g. CT images). Therefore, we have chosen for a semi-autonomous approach of endoscope navigation and have designed and realized a support system. The navigation algorithm that was developed for this system was tested on real colonoscopy images. Main requirements for such an algorithm are that it should be able to perform robustly, accurately and in real-time in clinical practice. The focus in our design lies on clinical applicability. Goal of this study is to evaluate the accuracy and robustness of the designed algorithm.

2 Materials and Methods

A target centralization algorithm was designed for and implemented in a robotized flexible endoscopy system. Two main requirements were imposed on system functionality. First, the complete system needs to work in real-time without significant hardware requirements to become applicable in a clinical setting. Second, the endoscopist needs to be able to overrule the algorithm instantaneously and the complete functionality needs to be intuitive.

In the case of Teleflex, control of the system (Fig. 3) can either be established through remote user input (e.g. joystick or touchpad device), or through the image-based navigation algorithm. This algorithm's main task is to detect the target of the endoscope through image analysis. The navigation algorithm should not be influenced by frequently occurring artifacts like lumen occlusion, fluids on the lens or bubbles in the field of view.

In this section, first the system will be described in more detail (Sects. 2.1 and 2.2), and then the workings of the algorithm (Sect. 2.3). Section 2.4 contains the details of the analyses we have performed.

2.1 Materials

Endoscope systems typically produce RGB image sequences ranging from 25 to 30 frames per second of varying resolution, depending on vendor and model. In the case of this experiment, an Olympus CF180 colonoscope was used to record 576 × 768 pixel images with 25 frames per second. This endoscope has a field of view of 170° and a field depth of 3–100 mm. An endoscope can be advanced up to tens of mms per frame in some areas of the colon, making motion blur artifacts in images relatively common.

2.2 System design

The complete system setup and control is displayed in Figs. 2 and 3. Only the tip of the endoscope is controlled by the steering knobs, and consequently the robotized system

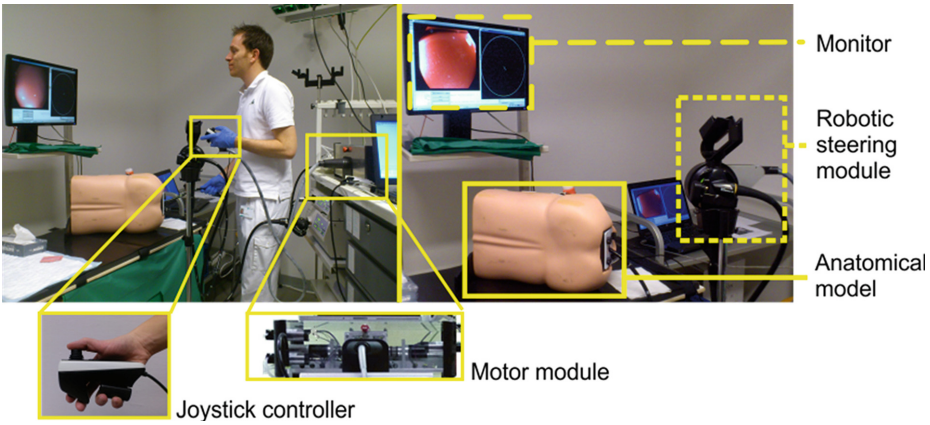

Fig. 2 Overview of the experimental setup. On the left, an endoscopist working with the system using a joystick controller. If a button is pressed on this controller, the navigation algorithm is enabled. The anatomical model is used to represent patient location. The robotic steering module is clicked on a current flexible endoscope and motors in the motor module turn the endoscopic steering knobs, using information from joystick or navigation algorithm input.

only influences tip motion in the currently used setup. The computer that is used for navigation and control is a standard laptop.

The user can enable the automated steering by keeping a button on the interface pressed. If the button is pressed, the navigation algorithm computes a steer command by finding the target direction. The algorithm can be overruled by using the joystick or by releasing the button. If the algorithm is enabled, the system attempts to centralize the target, i.e. turn the motors in the motor unit to rotate the endoscope's steering knobs until the tip is in the correct position. However, constraints are imposed in the movements of the motors to prevent instability of the system. Therefore it is possible that the tip position is not reached before the next target location update (analysis of the next image). Nonetheless, by using previous, reliable information, the target ultimately will be reached with the navigation algorithm.

2.3 Navigation algorithm

Detection and correction of all possible artifacts separately is thought to lead to considerable computational effort and to still not lead to a robust solution for endoscope navigation because of the unpredictable environment. However, not in every frame a new target needs to be determined. Directional changes during colonoscopy occur gradually, which means previous accurate information can be used in the current frame. Yet, determining accuracy of target direction estimation automatically is not straightforward. Dark region segmentation leads to a large spread in results for many possible accuracy features. Therefore, we have designed an image classifier based on image variance and entropy that will *predict* whether a frame contains reliable information

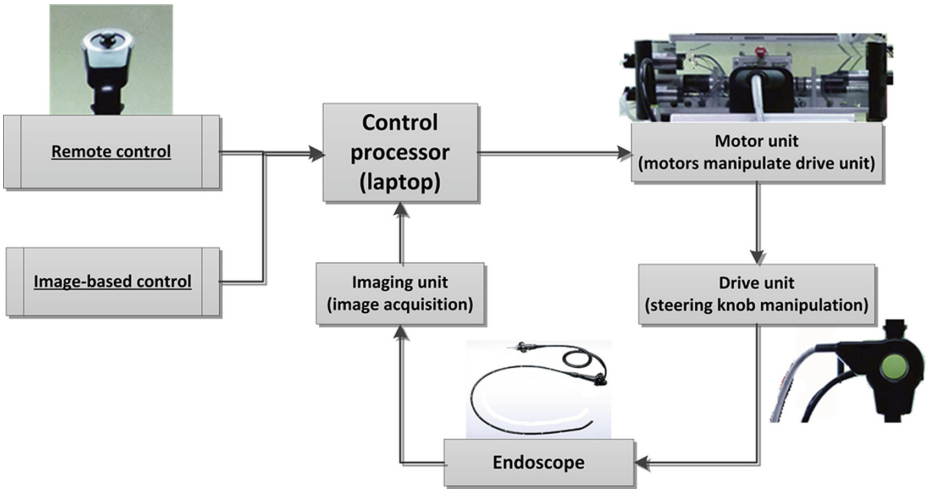

Fig. 3 Flowchart of components in the robotized endoscopy system. Two types of control input can be discerned: remote (user) controlled steering and image-based steering. The motor unit controls the endoscopic steering knobs through the drive unit that is attached at the endoscope handle. Images are acquired from the video processor of the endoscope system.

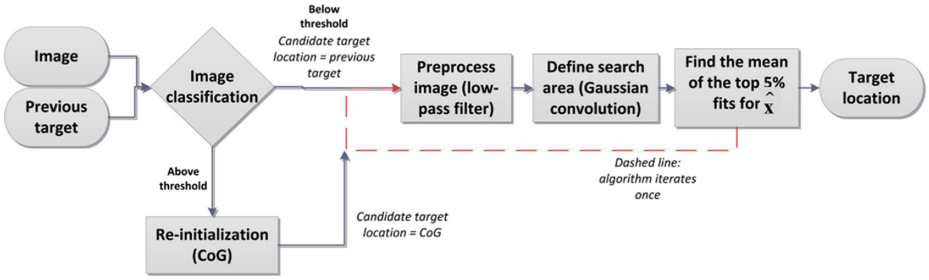

Fig. 4 Stepwise approach of the algorithm with an image and the previous target as inputs. If the image is ruled 'uninformative', it does not resemble the previous frame at all. Therefore, previous target information is then discarded and the CoG is recomputed. This location will then be used as candidate location for the target.

and thus whether the segmentation can be expected to be reliable (Fig. 4). If it is expected that the information is unreliable, the algorithm will re-initialize by computing the Center of Gravity (CoG) and using this as input location for the Gauss filter.

Let i be the frame index of an image sequence $f_i(\mathbf{x})$ with $\mathbf{x}$ the 2D pixel positions, and let $\mathbf{y}(i)$ be the 2D position of the center of the lumen. $\hat{\mathbf{y}}(i)$ is the estimate of $\mathbf{y}(i)$. The algorithm is initialized in the first frame of the sequence ($i = 1$) by the CoG computation described in [8]. This provides the estimate $\hat{\mathbf{y}}(1)$. All other frames of the sequence are processed according to:

(1) invert and low pass filter the frame: $g(\mathbf{x}) = (1 - f_i(\mathbf{x})) * h(\mathbf{x})$
(2) window the result at the previous found position: $g_{win}(\mathbf{x}) = w(\mathbf{x} - \hat{\mathbf{y}}(i))g(\mathbf{x})$
(3) estimate the lumen at the position of global maximum: $\hat{\mathbf{y}}(i) = \arg\max g_{win}(\mathbf{x})$

Gaussian low pass filtering is applied to suppress the influence of noise. The inversion of the frame turns the search for a minimum into a search for a maximum. The window function is a Gaussian with the peak located at the previously estimated lumen position. So, it restricts the search area to a vicinity of that previous position, and also biases the current estimate towards this previous one. If a frame is classified as 'uninformative' (see below), the sequence is regarded as terminated, and the algorithm is re-initialized with the following frame.

Image classifier. Rationale behind the image classifier is that wall view frames, or red-outs, contain the least information. These frames are recognized from three features: image variance after high- and low-pass filtering and image entropy. We have established a training data set consisting of two classes: wall view images and other images. These images were obtained from colonoscopy exams in several patients. The classifier was then designed as follows. Suppose the data is organized in a three-dimensional (3D) vector $\mathbf{z}$: high-pass filtered variance, low-pass filtered variance and entropy of two classes, ω_1 and ω_2. Differentiation between the classes occurs by transforming the data into a log-likelihood ratio Λ, defined as:

$$\Lambda = \log \frac{p(\mathbf{z}|\omega_1)}{p(\mathbf{z}|\omega_2)} \tag{1}$$

with $p(\mathbf{z}|\omega_k)$ the likelihood function [9]. Under the assumptions of normally distributed data sets with equal covariance matrices, Λ becomes a linear mapping of $\mathbf{z}$:

$$\Lambda = \mathbf{a}^T\mathbf{z} + b \tag{2}$$

The likelihood ratios Λ, conditioned on the two classes, are two Gaussians with equal variance V and means $\mu_1 = -\frac{1}{2}V$ and $\mu_2 = \frac{1}{2}V$, respectively. If the densities are normalized with respect to variance, the distance between the means becomes $\frac{\mu_2 - \mu_1}{\sqrt{V}} = \frac{V}{\sqrt{V}} = \sqrt{V}$. This normalized distance is called the discrimination of the classifier. A value above 4 makes the classifier reliable [9].

2.4 Experiments

For the complete algorithm to work properly, the classifier had to be trained first. 265 images of colonoscopy exams were collected; 172 were labeled in class uninformative and the remaining 93 were labeled informative. A Linear Bayes Normal Classifier mapping was applied from the training set using the 'PRtools' toolbox [10] in Matlab R2011b (The Mathworks Inc., Natick, MA, USA). The mean values of the three features for both classes, 3 × 1 vectors $\mathbf{m}_1$ and $\mathbf{m}_2$, and the covariance C result from this mapping. For the log likelihood ratio Λ then:

$$\mathbf{a} = \mathbf{C}^{-1}(\mathbf{m}_1 - \mathbf{m}_2), \text{ and } b = -0.5 \cdot (\mathbf{m}_1 + \mathbf{m}_2)' \cdot \mathbf{a}. \quad (3)$$

After the classifier was trained successfully, it was implemented in the algorithm. Next, the algorithm was tested for robustness on five image sequences from human colonoscopy exams. In Table 1, properties of the test image sequences are listed. We have categorized the sequences in Easy, Medium or Hard. We expect the sequence of category 'Hard' to provide poorer results than the sequence of category 'Easy'. For Medium (Hard) better results are expected than for (Medium) Hard.

For each sequence, the steering target was manually annotated per frame. This target was defined as 'the point to which a human user would steer the endoscope'. The algorithm resulted in several measurement parameters for each sequence. The target location was detected and compared to the manual results using intra-class correlation analysis (ICC, [11, 12]). The ICC is the main measure to determine accuracy of the algorithm. Additionally, the number of useful frames according to our classifier provides information on the quality of the frames in a sequence. This is expected to clarify whether a poor result in target direction over a sequence is due to frame quality or has another cause and therefore provides information on algorithm robustness as well.

Table 1. Properties and categorization of the test set of image sequences.

Sequence	Artifact description	No. of frames	Category
1	No artifacts, clean bowel	77	Easy
2	Bubbles, motion blur, wall contact	100	Hard
3	Out of focus, motion blur, wall contact	174	(Medium) Hard
4	Motion blur	31	Medium (Hard)
5	Colitis (red, no folds in the wall)	175	Medium

3 Results

The best result was obtained with σ, the peak width of the Gaussian window function, set at 3. The discrimination of our classifier then was 6.8 using all three features, making it excellent. Because of the excellent result a feature reduction was attempted. With the two best features, the maximum discrimination was 6.1. Yet, the error rate then still was 1.5 % due to not fully normally distributed data. To minimize the error rate all three features were kept and the classifier was implemented in the navigation algorithm.

Figure 5 displays plots of target location in X and Y coordinates for a good result with many target location changes (A) and the poorest result (B) in terms of accuracy. In Table 2, accuracy and robustness results are summarized.

In the top graphs of Fig. 5 can be seen that automatic target detection follows manually annotated locations really well. The dashed vertical lines indicate areas for which $\Lambda\mathbf{z}$ (zRes, bottom graphs) was far below the threshold, shown for reference as a

horizontal line. In the top graphs it is seen that for these regions, the automatic detection corresponds poorly to the manual gold standard.

From the results in Table 2 can be seen that the ICC between the results of our algorithm and the manual results is high with an average of 93 % (range 84.7–98.8 %). A high overall ICC corresponds to a high number of useful frames according to the classifier. The best result indeed was obtained with sequence 1, which we categorized as 'Easy'. The poorest result came from sequence 5, which we categorized as 'Medium'.

Sequence 5 only contained 6 (3.43 %) useful frames, leading to the question whether the classifier worked properly in this sequence. In Fig. 6, a representative image from this sequence is displayed. The colon of this patient was inflamed, causing a red appearance without the usual characteristic folds in the wall. The variance and entropy of these images thus deviate from images of patients without this disease, which is why the classifier ruled so many images as 'uninformative'. On the one hand, these frames do contain relevant information and should thus not be classified as such.

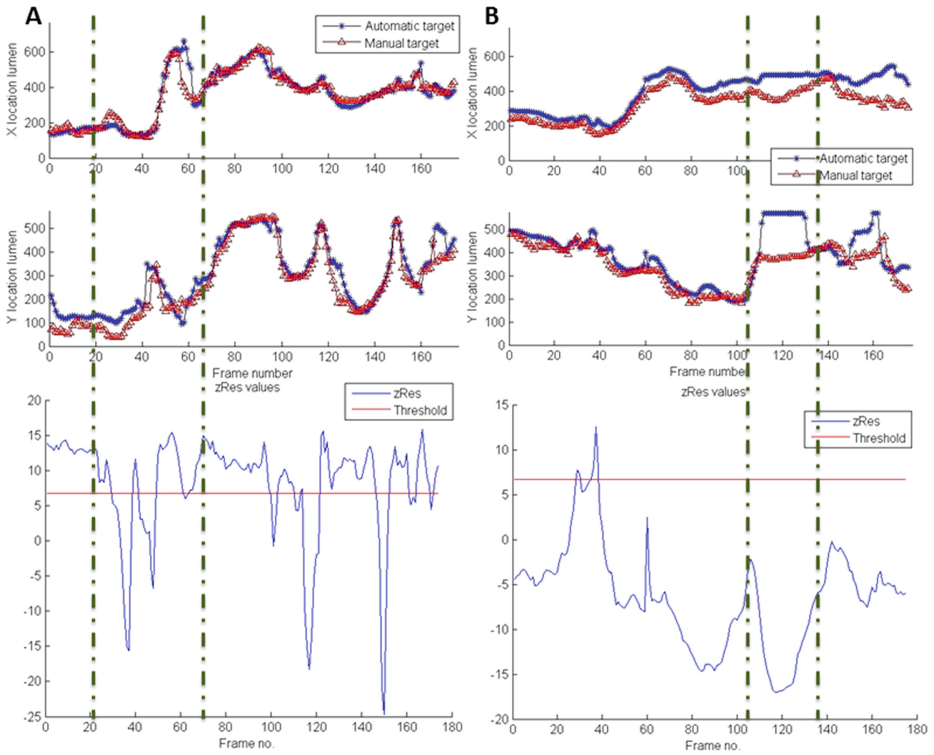

Fig. 5 The top graphs illustrate the target location results in x (top) and y (bottom) coordinates in pixels. The bottom graphs display the log likelihood of z, resulting from the classifier. Plots A correspond to sequence 3, plots B to sequence 5. The dashed vertical lines show areas where automatic and manual target locations do not correspond very well. In the bottom graphs, low zRes values can be seen in these areas.

Table 2 Results of robustness and accuracy of the target detection.

Sequence	ICC in %	No. of frames	No. of useful frames (% of total)
1	98.8	77	64 (83.12)
2	89.4	100	72 (72)
3	97.1	174	131 (75.29)
4	95.4	31	25 (80.65)
5	84.7	175	6 (3.43)

However, from the results of the target localization algorithm can be concluded that a reliable target is difficult to find in images from this sequence. From a navigation standpoint, the classification is therefore still correct.

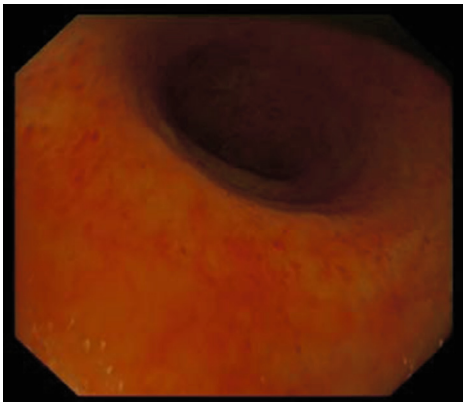

Fig. 6 Representative image from image sequence 5. The characteristic folds in the wall are missing and the colon wall is unusually vascularized and red. These features have impact on the variance and entropy of the image, and therefore influence the classifier results (Color figure online).

4 Conclusion

In terms of ergonomics and procedure efficiency, clinical relevance of robotized flexible endoscopy is beyond question. Once a robotized system is being considered, image-based navigation is a logical addition. In this research, an assistive algorithm for endoscope insertion was developed, implemented and evaluated for accuracy and robustness in real human colonoscopy images.

The algorithm was based on dark region segmentation after image classification. Earlier research has been done in the field of dark region segmentation in endoscopy. Chettaoui et al. [13] developed an algorithm that distinguishes the normal lumen from so-called ‘diverticulas’, bulges in the colonic wall that can be mistaken for lumen. Because the algorithm we developed takes the mean of the 5 % darkest regions, and furthermore because the location is updated 25 times per second, we do not expect

significant influence in control performance in case of diverticulas. Moreover, our algorithm is meant as an assistive navigation tool, which means the endoscopist can turn it off at any moment.

Zhiyu [14] describes an extensive analysis on lumen boundary extraction. In Sect. 2 was stated that the main requirement of an endoscopic navigation algorithm is real-time, intuitive functionality. The method proposed by Zhiyu is complicated and computationally intensive, making it not very applicable to clinical practice.

In the experiments it was found that the developed target centralization algorithm always detected a suitable target or lumen direction, with an ICC with manual annotation of the target of 93 %. The categorization we established beforehand was not predictive for accuracy of the results, but the image classifier that was developed was an excellent predictor of images in which a lumen could be detected reliably. In the sequence from a patient with an inflamed colon, the classifier ruled many frames as 'uninformative', making the algorithm less accurate. However, an ICC of 84.7 % was still obtained in this case. It is therefore expected that this kind of patients can still be treated using the described system. The algorithm further works in real-time on standard equipment, making it suitable for clinical practice.

Although these results are promising, future work will be needed to improve navigation functionality further. Possible improvements are to incorporate heading direction information [11]. Furthermore, user feedback on the system's functionality during experiments in a realistic test environment will be obtained to adapt to clinical practice and increase possible acceptance in the clinic. Finally, clinical relevance of the complete robotized endoscopy system needs to be established. Therefore, the first patient study has started and our goal is to include 32 patients.

References

1. Taylor, R.H.: A perspective on medical robotics. Proc. IEEE **94**(9), 1652–1664 (2006)
2. Pedrosa, M.C., Farraye, F.A., et al.: Minimizing occupational hazards in endoscopy: personal protective equipment, radiation safety, and ergonomics. Gastrointest. Endosc. **72** (2), 227–235 (2010)
3. Ruiter, J., Rozeboom, E., et al.: Design and evaluation of robotic steering of a flexible endoscope. In: IEEE BioRob, pp. 761–767 (2012)
4. Van Der Stap, N., Van Der Heijden, F., Broeders, I.A.M.J.: Towards automated visual flexible endoscope navigation. Surg. Endosc. **27**(10), 1–13 (2013)
5. Waye, J.D., Rex, D.K., Williams, C.B.: Colonoscopy: Principles and Practice, 2nd edn, pp. 267–345. Blackwell Publishing Ltd, Chichester (2009)
6. Lee, S.-H., Chung, I.-K., et al.: An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. Gastrointest. Endosc. **67**(4), 683–689 (2008)
7. Gillies, D., Khan, G.: Vision based navigation system for an endoscope. Image Vis. Comput. **14**, 763–772 (1996)
8. Reilink, R., Stramigioli, S., Misra, S.: Image-based flexible endoscope steering. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. i, p. 6 (2010)
9. Van Der Heijden, F., Duin, R.P.W., et al.: Classification, Parameter Estimation and State Estimation, pp. 45–79. Wiley, Chichester (2004)

10. Duin, R.P.W., Tax, D.M.J.: PRTools – a matlab toolbox for pattern recognition, 14 July, 2013. http://prtools.org. Accessed 15 May 2014
11. van der Stap, N., Reilink, R., et al.: The use of the focus of expansion for automated steering of flexible endoscopes. In: IEEE BioRob, pp. 13–18 (2012)
12. 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)
13. Chettaoui, H., Thomann, G., et al.: Extracting and tracking Colon's 'Pattern' from Colonoscopic Images. In: IEEE Canadian Conference on Computer Robot Vision, pp. 65–71 (2006)
14. Zhiyun, X.: Computerized detection of abnormalities in endoscopic oesophageal images. Ph.D. thesis, Nanyang Technological University (2000)