Image-Based Flexible Endoscope Steering

Rob Reilink, Stefano Stramigioli, and Sarthak Misra
University of Twente, Enschede, The Netherlands

Abstract— **Manually steering the tip of a flexible endoscope to navigate through an endoluminal path relies on the physician's dexterity and experience. In this paper we present the realization of a robotic flexible endoscope steering system that uses the endoscopic images to control the tip orientation towards the direction of the lumen. Two image-based control algorithms are investigated, one is based on the optical flow and the other is based on the image intensity. Both are evaluated using simulations in which the endoscope was steered through the lumen. The RMS distance to the lumen center was less than 25% of the lumen width. An experimental setup was built using a standard flexible endoscope, and the image-based control algorithms were used to actuate the wheels of the endoscope for tip steering. Experiments were conducted in an anatomical model to simulate gastroscopy. The image intensity-based algorithm was capable of steering the endoscope tip through an endoluminal path from the mouth to the duodenum accurately. Compared to manual control, the robotically steered endoscope performed 68% better in terms of keeping the lumen centered in the image.**

I. INTRODUCTION

Flexible endoscopy is a minimally invasive medical procedure to examine the internal body cavities. Common procedures include gastroscopy, the inspection of the esophagus and the stomach via the mouth (Fig. 1), and colonoscopy which involves the inspection of the colon via the rectum. During endoscopy, the physician holds the proximal end of the flexible endoscope that contains the control wheels, and uses this to steer the tip. The tip contains a camera and a light source that allow the physician to investigate the gastrointestinal (GI) tract via his/her monitor.

During clinical procedures, the endoscope is first inserted up to the required length and then the inspection is performed while slowly retracting the endoscope. The insertion requires spatial reasoning and dexterity, therefore it may take significant time and effort. The physician needs one hand to maneuver the flexible tube, while his/her other hand has to operate the control wheels that steer the tip. The control is not very intuitive, since the two degrees of freedom (left-right and up-down) are controlled by two concentric wheels. Experience is required to manipulate the controls to steer the endoscope in the appropriate direction [1]. This makes the steering difficult, especially for less experienced physicians. A robotic system can be used to improve the performance of the physician and the clinical outcome of
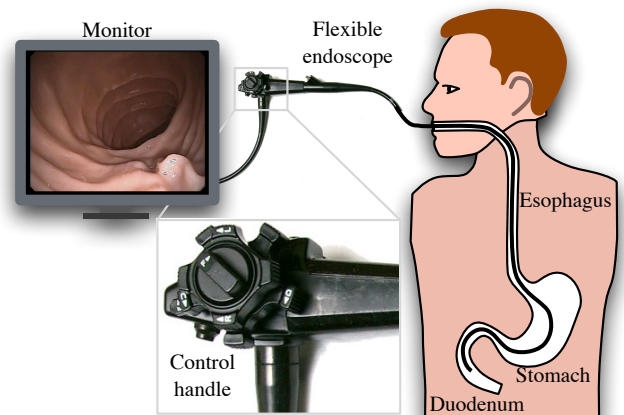
Fig. 1. **Conventional gastroscopy:** The physician uses the endoscope control handle to steer the endoscope through the patient's gastrointestinal tract while observing the endoscopic images on the monitor.

the procedure [2]. Controlling a robotic flexible endoscopy system will require computing the desired tip orientation. Using a purely mechanics-based approach to compute the required tip orientation to steer the endoscope through the GI tract would require an accurate model of the endoscope as it interacts with the soft tissue. This is realistically not possible since the *in vivo* patient-specific elastic properties of the soft tissue are not known *a priori*. An alternative approach to compute the required tip orientation is to use the endoscopic images. An overview of vision algorithms that process endoscopic images is given by Liedlgruber [3]. Related work in flexible endoscopy includes lumen detection [4], [5] and polyp detection [6], [7]. However, these algorithms were not designed for use in the feedback of a control loop. As such, their performance in terms of robustness and latency may not be sufficient under all conditions. Another approach based on the image gradient is proposed by Gomez *et al.* [8]. Although they claim this approach is suitable for real-time processing, they do not show experiments where the algorithm is actually used in a feedback control loop.

This research presents a method to robotically steer the endoscope using the endoscopic images i.e., a visual servoing approach [9]. In order to provide the physician with a clear image, our goal is to keep the furthest part of the lumen centered in the endoscopic image. We investigate two vision algorithms to find the preferred endoscope direction: (i) An optical flow-based and (ii) an image intensity-based method. Both algorithms were first implemented and tested in simulation. Subsequently, an experimental setup was constructed that allowed a commercially available endoscope to be controlled by the vision algorithm. This setup was used

to steer the endoscope through the GI tract of an anatomical model up to the duodenum (Fig. 1).

This paper is structured as follows: In Section II, the use of optical flow to infer depth information from the images will be discussed, while in Section III, the use of the intensity distribution of the images to acquire this information will be described. Section IV will provide simulation results that indicate that both approaches can be used to control an endoscope through the lumen in a rendered environment. Section V will describe the experimental setup that was developed to test these approaches on an anatomical model. Section VI will discuss the experiments that were done with this setup and the results. Finally, Section VII concludes and provides possible directions for future work.

II. OPTICAL FLOW-BASED IMAGE PROCESSING

The first approach to finding the direction of the lumen that we have investigated is based on optical flow, which is the perceived motion of the environment as observed by a camera [10]. By comparing two subsequent images taken from a camera, it is possible to estimate the optical flow. The resulting optical flow field has been used to steer mobile robots away from obstacles and through corridors [11],[12], and for the control of aerial robots [13]. We have investigated this approach since the task of steering the endoscope through the lumen is similar to steering a mobile robot through a corridor. In the following two subsections we will discuss the theory of depth estimation from optical flow and an implementation that we have used to process endoluminal image sequences.

A. Depth estimation from optical flow

The key feature in the optical flow-based approach is dependency of the optical flow on the distance of the perceived features. If the depth of the scene can be estimated, we can steer away from nearby objects. The dependency of the optical flow on the distance of the environment is most easily described using a spherical camera model, M, that projects points, $\mathbf{p} \in \mathbb{R}^3$, onto a sphere, S^2 (Fig. 2(a)), [13]:

$$M : \mathbb{R}^3 \to S^2; \quad \mathbf{p} \mapsto \frac{\mathbf{p}}{|\mathbf{p}|} \quad . \tag{1}$$

Note that the actual camera can have a 'usual' flat image plane though, its perceived image may be mapped onto a sphere. For each point, $\mathbf{q} := M(\mathbf{p})$, we define

$$\lambda(\mathbf{q}) : S^2 \to \mathbb{R} \quad , \tag{2}$$

as the distance from $\mathbf{p}$ to the camera optical center. If we have a camera moving within a static environment, the optical flow, $\theta(\mathbf{q})$, is the sum of a rotational part, $\theta_{\mathbf{R}}(\mathbf{q})$, and a translational part, $\theta_{\mathbf{T}}(\mathbf{q})$:

$$\theta(\mathbf{q}) := \underbrace{-\mathbf{\Omega} \times \mathbf{q}}_{\theta_{\mathbf{R}}(\mathbf{q})} + \underbrace{\frac{-1}{\lambda(\mathbf{q})} \left(\mathbf{I} - \mathbf{q}\mathbf{q}^{\mathrm{T}}\right) \mathbf{V}}_{\theta_{\mathbf{T}}(\mathbf{q})} \,, \tag{3}$$

where $\mathbf{V}$ denotes the translational velocity of the optical center and $\mathbf{\Omega}$ the rotational velocity around the optical center. The depth information is contained in $\theta_{\mathbf{T}}(\mathbf{q})$. A common approach to obtain $\theta_{\mathbf{T}}(\mathbf{q})$ is to cancel the rotational optical

(a) Point $\mathbf{q} \in S^2$ is obtained by projecting point $\mathbf{p} \in \mathbb{R}^3$ onto a unit sphere. The optical flow $\theta(\mathbf{q})$ is defined in the tangent space of S^2. The central part C is used to estimate camera rotation.

(b) Optical flow balancing uses the difference between mean optical flow in the left (L) and right (R) sections to control rotation around the x axis. I shows the part of S^2 covered by the camera image.
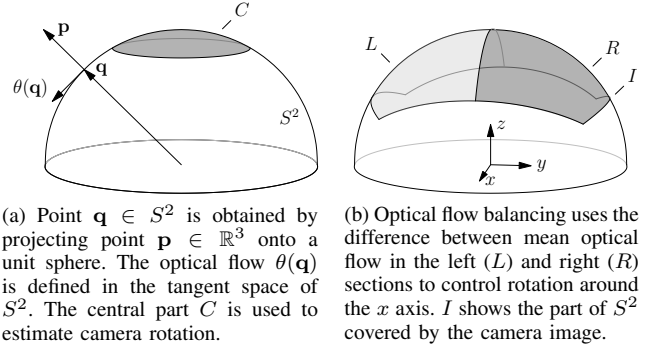
Fig. 2. Schematic representation of the spherical camera model used for optical flow processing.

flow, $\theta_{\mathbf{R}}(\mathbf{q})$, by estimating $\mathbf{\Omega}$ using odometry (e.g., [11],[12]) or inertial motion sensor data (e.g., [13]). However, there does not exist a sensor to obtain this data in an endoscope. Estimating $\mathbf{\Omega}$ from the control inputs using a kinematics model is not considered feasible, since the kinematics of the tip are highly dependent of the overall shape of the endoscope. Therefore, a method was chosen that solely uses the optical flow as its input.

In order to estimate the camera motion we use the central region of the image as a reference (C in Fig. 2(a)). In this region, the translational velocity, $\mathbf{V}$, is approximately in the direction of the camera optical axis. Therefore, in C, $\mathbf{q}$ will be approximately in the same direction as $\mathbf{V}$. This makes $(\mathbf{I} - \mathbf{q}\mathbf{q}^{\mathrm{T}})\mathbf{V}$ small since it is the projection of $\mathbf{V}$ on the plane orthogonal to $\mathbf{q}$. Aditionally, since the environment in the center of the image will be the far away from the camera, $1/\lambda(\mathbf{q})$ will be small. Therefore, in C, $\theta_{\mathbf{T}}(\mathbf{q}) \approx 0$ hence $\theta(\mathbf{q}) \approx \theta_{\mathbf{R}}(\mathbf{q}) = -\mathbf{\Omega} \times \mathbf{q}$. So, $\theta(\mathbf{q})$ can be used to estimate $\mathbf{\Omega}$.

The $\mathbf{\Omega}$ that is obtained can be used to calculate only the translational part, $\theta_{\mathbf{T}}(\mathbf{q})$, from the optical flow in the entire camera image. This can then be used to steer the camera away from points that are near i.e., whose observed $\theta_{\mathbf{T}}(\mathbf{q})$ is large. We will now describe the implementation of this method for the processing of a sparse optical flow field acquired from an endoscopic image sequence.

B. Implementation

A Lucas-Kanade optical flow algorithm [14] was used to obtain the sparse optical flow field from two subsequent camera images. Each of the n flow vectors that are found is represented as a vector pair $(\mathbf{u}_i, \mathbf{v}_i) \in S^2 \times S^2$, where subscript i denotes the i-th vector pair ($i = 1 \ldots n$). $\mathbf{u}_i$ and $\mathbf{v}_i$ represent approximately the same physical point in the previous and the current frame, respectively. For the pairs where $\mathbf{v}_i$ falls within central region C, the rotation from $\mathbf{u}_i$ to $\mathbf{v}_i$ is computed. These rotations are represented as quaternions and averaged, resulting in an estimation of the camera rotation between the frames [15]. This rotation is expressed as rotation matrix $\mathbf{R}$. Using $\mathbf{R}$, the translational flow vector belonging to each vector pair $(\mathbf{u}_i, \mathbf{v}_i)$ can be computed, which we define as

$$\theta_{\mathbf{T}i} := \frac{1}{\Delta t}(\mathbf{R}^{-1}\mathbf{v}_i - \mathbf{u}_i) \,, \tag{4}$$
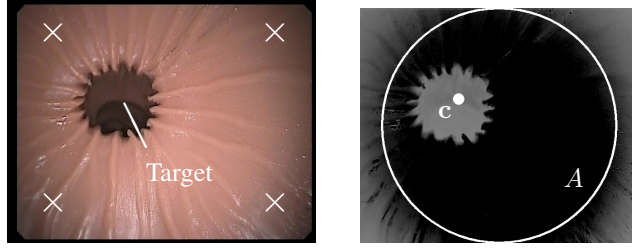
(a) Image from the stomach showing the target as a dark area and the dark corners (indicated by ×) caused by the inhomogeneous lighting.

(b) Image (a) equalized and inverted showing the circular ROI A and the corresponding centeroid **c**.

Fig. 3. The intensity-based algorithm finds the dark target in the image by computing the centroid of the equalized and inverted image over the circular region of interest (ROI).

where Δt is the frame time. In (4), $\mathbf{R}^{-1}\mathbf{v}_i$ is $\mathbf{v}_i$ compensated for the camera rotation.

An optical flow balancing controller [16] is used to control the camera orientation. This controller uses the computed optical flow vectors, $\theta_{\mathbf{T}i}$, to compute the desired camera rotational velocity, ω. Separate controllers are implemented for the left/right (pan) and the up/down (tilt) motion. For the pan control, the image is separated in a left (L) and a right (R) region (Fig. 2(b)). The mean left flow, ϕ_L, and the mean right flow, ϕ_R, are defined as

$$\phi_L := \text{mean}(\{||\theta_{\mathbf{T}i}||_2 \quad | \mathbf{v}_i \in L\}) , \qquad (5)$$

$$\phi_R := \text{mean}(\{||\theta_{\mathbf{T}i}||_2 \quad | \mathbf{v}_i \in R\}) . \qquad (6)$$

The desired rotational velocity of the camera around the x-(pan) axis is ω_x and is computed from (5) and (6) as

$$\omega_x = K(\phi_R - \phi_L) , \qquad (7)$$

where K is a constant gain. Rotational velocity around the y-(tilt) axis is ω_y and is computed similar to (7) using the image separated in top and bottom regions.

III. INTENSITY-BASED IMAGE PROCESSING

Endoscopic images from the GI tract might have insufficient texture in some regions. Texture is required for the optical flow-based endoscope steering algorithm to work appropriately [10]. In order to provide reliable steering in the presence of limited image texture, we also considered an intensity-based approach, which is described in this section.

The arrangement of the light source and the camera in the tip of the endoscope cause the part of the lumen that is furthest away to appear as a dark area in the image (Fig. 3(a)). This has been exploited to extract an accurate description of the lumen wall contour, which may be used to find polyps e.g., [6], [7], [17]. In this body of research, adaptive thresholding was used to obtain a binary image which was then processed to obtain the lumen wall shape. In [4] and [5] the dark area in the image is used to find the lumen position. This is also done by processing a binary image that is obtained by thresholding.

For our purpose of finding the direction of the lumen, we are not so much interested in the actual shape of the wall, but more in a robust estimation of the center of the lumen. This robustness is required since the results will be used as feedback in the control loop. Therefore, we

propose to use a method based on the intensity centroid. This algorithm will be described in the remainder of this section. Its implementation in simulation and experiments are described in Section IV and V.

The input to our system is the grey-scale image, $I(x,y)$, that is captured from the endoscope video system, where x and y are the horizontal and vertical pixel positions, respectively. $x=0$, $y=0$ is the center of the image. $I(x,y)$ is an 8-bit image with 0 representing black and 255 representing white. In order to obtain robustness against contrast and intensity variations, histogram equalization is applied which normalizes the brightness and increases the contrast of the image. This is done using the OpenCV function `cvEqualizeHist` [18].

In order to find the direction of the lumen we will calculate center of the dark region in the image, which we define as the centeroid **c** of the inverted image. This inverted image, $I''(x,y)$, is defined as

$$I''(x,y) := 255 - I'(x,y) , \qquad (8)$$

where $I'(x,y)$ is $I(x,y)$ with histogram equalization applied. The centroid is computed over a circular region of interest (ROI) A. This ROI is centered in the image and has a diameter equal to the image height (Fig. 3(b)). Using a circular ROI makes the algorithm invariant to camera rotations around the optical axis. It also prevents undesired influence from the dark corners that may appear in the image due to inhomogeneous lighting (Fig. 3(a)). The centroid **c**, is computed as

$$\mathbf{c} = \begin{bmatrix} c_x \\ c_y \end{bmatrix} = \frac{\sum_A \begin{bmatrix} x \\ y \end{bmatrix} \cdot I''(x,y)}{\sum_A I''(x,y)} , \qquad (9)$$

where $\sum_A$ denotes summation over the area A.

The desired rotational velocity of the camera, ω_x around the x-axis and ω_y around the y-axis, are computed as

$$\begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} = -K \begin{bmatrix} c_x \\ c_y \end{bmatrix} , \qquad (10)$$

where K is a constant gain. This rotates the camera such that the center of the dark region will be in the center of the image. Thus, the camera is rotated to look in the direction of the lumen.

IV. SIMULATION OF FLEXIBLE ENDOSCOPIC PROCEDURE

The optical flow-based and the intensity-based vision algorithms were tested in simulation before applying them in an experimental setup. For both vision algorithms, the closed loop behavior of the system was verified in a simulation of a flexible endoscopic procedure. This was done using a custom-built simulation environment using Blender [19]. Blender is used to render an image of the virtual environment. This image is processed using the vision processing algorithm under consideration. The results are used to update the virtual camera position. The interaction between the lumen and the endoscope was not considered in this simulation. The lumen was modeled as a rigid body.

Fig. 4 depicts the virtual environment and the path that the virtual camera followed in the simulation. In order to
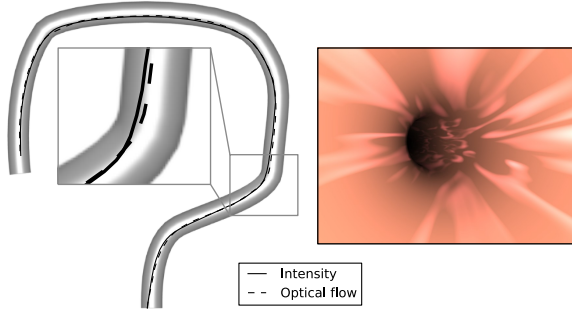
Fig. 4. **Simulated robotically steered flexible endoscopy:** The camera follows the path through the lumen using two different vision algorithms: the intensity- and the optical flow-based methods. **Right:** View from the camera inside the lumen.
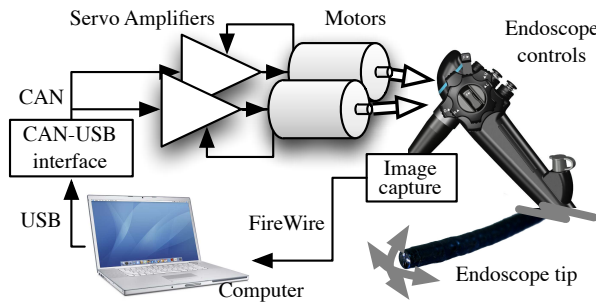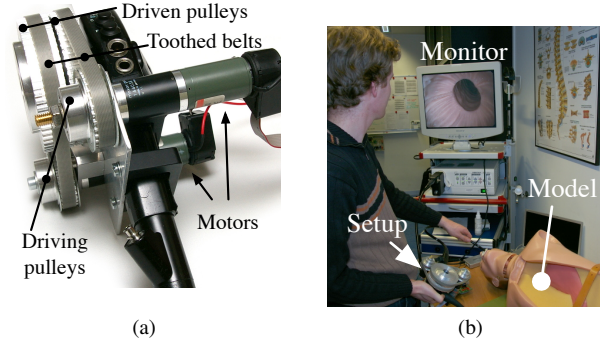

(a)   (b)

Fig. 6. **Realization of the setup:**(a) Various mechanical parts are attached to the endoscope, including motors, pulleys, and a toothed belt drive. (b) A robotically steered gastroscopy was performed on an anatomical model to evaluate the setup.
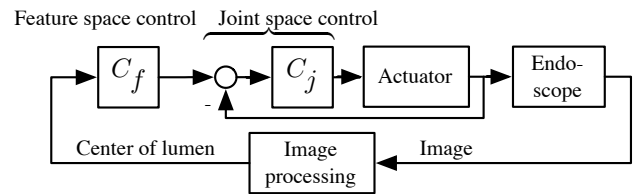

Fig. 5. **Experimental setup used for steering the endoscope:** The steering is based on images which are captured through the endoscope camera.

Fig. 7. The image-based look-and-move structure is used to control the center of the lumen to be in the center of the image.

assess the performance of the two algorithms, we consider the root mean square (RMS) distance between the camera position and the center line of the lumen. This is 21% of the lumen width for the optical flow-based algorithm, and 24% of the lumen width for the intensity-based algorithm. Fig. 4 also shows that the deviations from the lumen center were largest in the bends, where the camera trajectory 'cuts the corner'. This is due to the fact that the camera has a limited field of view i.e., the algorithm only perceives the environment in front of the camera and will therefore keep the lumen ahead of the camera centered.

Using this simulation setup, the vision algorithms were tested with varying light conditions. The light intensity was increased to up to 400% of the intensity as used for the simulations shown in Fig. 4. The RMS difference between these trajectories was less than 5% of the lumen width. This indicates that for the simulated conditions, both algorithms are capable of steering the camera through the lumen.

V. EXPERIMENTAL SETUP

A motorized endoscope setup was developed to test the endoscope steering algorithms. Fig. 5 shows an overview of this setup. Except for the mechanical connection to the endoscope, all components are common of-the-shelf products. The endoscope that is used is a EG-2930K gastroscope (Pentax, Tokyo, Japan). Images from the endoscope are captured by an ADVC55 video capture device (GrassValley, Conflans St. Honorine, France) and transferred via FireWire™ into the computer (MacBook Pro, Apple, Cupertino, CA, USA) that does the image processing. The control algorithm uses

the data obtained from the images to compute setpoints for the motor positions. These setpoints are sent to the servo amplifiers via a CANUSB interface (Lawicel, Tyringe, Sweden). The Elmo Whistle servo amplifiers (Elmo Motion Control, Petach-Tikva, Israel) control the motors, which are fitted with encoders for position feedback. The S2326 motors (Maxon, Sachseln, Switzerland) are coupled to the controls of the endoscope, allowing control of the orientation of the tip in two dimensions. The following subsections will discuss the mechanical interface that connects to the endoscope and the control architecture that is used to control the system.

A. Design

A mechanical interface was constructed such that it can be fitted to the proximal end of a commercially available flexible endoscope (Fig. 6(a)). The base that supports the two motors is mounted to the shaft of the endoscope. A toothed belt drive couples the motors to the endoscope. The driven pulleys are press-fitted over the control wheels of the endoscope.

B. Control architecture

The implemented control method is of a dynamic image-based look-and-move structure [9], as shown in Fig. 7. This structure has a joint space control loop inside an image-based control loop. The feature to be controlled is the position of the center of the lumen. Since the task of steering the endoscope does not require a high bandwidth, a simple integral controller $C_f = K/s$ was implemented as feature space control law, with K a constant gain. More sophisticated controllers could be used as well (e.g., incorporating friction and backlash compensation). The gain was tuned manually on the actual setup.

VI. EXPERIMENTS AND EVALUATION

In order to test the robotically steered endoscope, an anatomical gastroscopy model (OGI Phantom CLA 4, Coburger Lehrmittelanstalt, Coburg, Germany) was used. Fig. 6(b) shows the experimental setup in use. During initial experiments, it appeared that the endoscopic images from some regions of the GI tract did not contain enough texture for the optical flow algorithm to work reliably. Therefore, only the intensity-based algorithm was evaluated on the anatomical model.

During the experiment, the endoscope was manually fed into the model. When the end of the duodenum was reached, the endoscope was retracted while the image-based steering system ensured that the lumen was kept centered.

In order to evaluate the performance of the system, we have compared the output of the vision algorithm, denoted c_a, with a reference, denoted c_r. c_r was obtained by manually analyzing the images that were recorded. For each image, the center of the lumen was marked.

The performance of the robotically steered endoscope was compared against ten manual gastroscopies. These were performed on the same anatomical model by five Technical Medicine[1] students (1 male, age 22 years and 4 female, average age 22 years). All subjects had done flexible endoscopy training and had previous experience on the anatomical model that was used. The subjects were asked to try to keep the lumen well centered and to focus on accuracy rather than speed. They manipulated the endoscope control wheels while an assistant fed the endoscope into the anatomical model. The following subsections will discuss the performance of the lumen detection, the overall system performance, and the comparison of the overall system performance against the manually performed gastroscopies.

A. Intensity-based vision algorithm performance

In fig. 8(a) and (b) we compare the position of the center of the lumen as determined by the algorithm, c_a, and the reference, c_r. c_a and c_r are expressed in mm, and computed as

$$\mathbf{p}_{\mathrm{mm}} = \mathbf{p}_{\mathrm{pix}} \cdot \frac{w_{\mathrm{mm}}}{w_{\mathrm{pix}}} , \qquad (11)$$

where w_{mm} and w_{pix} are the width of the monitor in mm and in pixels, respectively, and $\mathbf{p}_{\mathrm{mm}}$ and $\mathbf{p}_{\mathrm{pix}}$ denote a point expressed in mm and in pixels, respectively. Fig. 8 shows that in the mouth and the throat, the deviations between c_a and c_r are larger than in the other sections. These deviations are caused by the fact that there are more irregularities like the palate and the tongue (Fig. 9(a)).

As a performance measure for the vision algorithm, we define for every frame

$$e_v := ||\mathbf{c_r} - \mathbf{c_a}||_2 , \qquad (12)$$

as the error of the vision algorithm. This measure is shown in Fig. 8(c). This graph also shows clearly that the performance of the algorithm is better in the esophagus, stomach, and

[1]Technical Medicine is a Master's level program at the University of Twente where students study to integrate advanced technologies within the medical sciences to improve patient care.
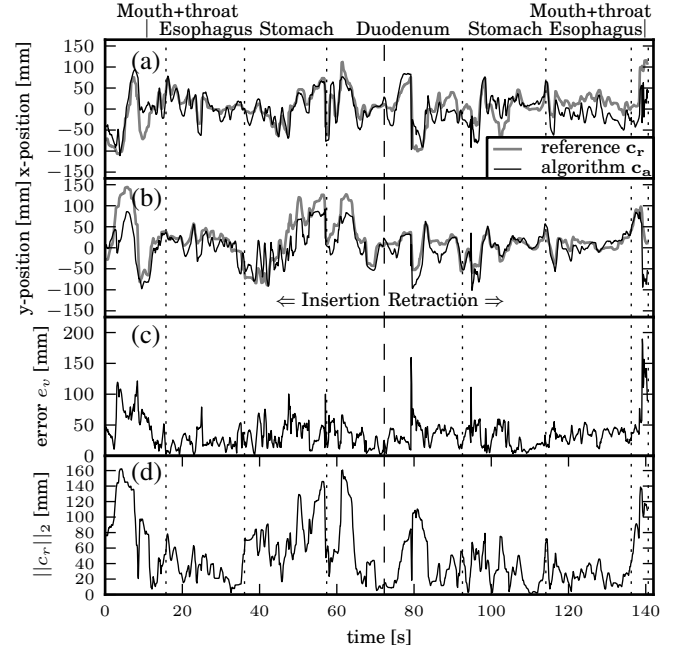

Fig. 8. **Evaluation of the intensity-based vision algorithm and the overall performance:** The algorithm is well capable of tracking the lumen center in the esophagus, the stomach, and the duodenum. In the mouth and throat the deviations are larger.
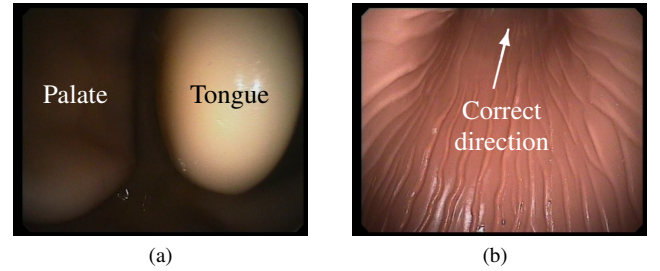

Fig. 9. **Endoscopic images during gastroscopy:** (a) In the mouth, the target is less clear which decreases the performance of the vision algorithm. (b) In the stomach, the target is not always completely in view, but the intensity-based algorithm finds the correct direction.

duodenum than in the mouth and the throat. In the stomach, the intensity-based vision algorithm is able to find the correct direction, even when the exit point is not visible (Fig. 9b). Over the entire experiment, the RMS of the error, e_v, was 42 mm, which equates to 10% of the width of the image.

B. Overall system performance

In order to assess the overall system performance, we define for every frame the position error

$$e_p := ||\mathbf{c_r}||_2 , \qquad (13)$$

as the overall system performance measure. This is the Euclidian distance between the center of the lumen, c_r, and the center of the image. This measure is shown in Fig. 8(d). The system preforms best in the esophagus and during the retraction in the stomach. In the other sections of the GI tract, the system performance reduces due to the following:

- In the stomach during insertion, the endoscope is in a large open space. Therefore, the endoscope position is not well constrained and the endoscope may curl inside the stomach. This makes the control more challenging.

This is an inherent difficulty in gastroscopy, which is also observed in manually conducted gastroscopies.

- In the mouth and the throat, the reduced performance of the vision algorithm degrades the overall system performance. The system performance may be improved by adapting the vision algorithm to cope with the structures found in the mouth and the throat.
- In the duodenum, the endoscope needs to make a sharp turn. Since there is no feed-forward path in the controller, an error in the feature space is required to get the required motion output. Adding a feed-forward path in the feature space controller may improve this.

Over the entire experiment, the RMS of the position error, e_p, was 66 mm, which is 16% of the width of the image.

C. Comparison of robotic steering with manual steering

Like in the robotic steering experiment, for each manually conducted gastroscopy the recorded endoscopic images were analyzed manually. Again, for every image the position of the center of the lumen was marked as a reference, $\mathbf{c_r}$. The same performance measure (13) was used. For each experiment, the RMS of the position error was computed. The average RMS position error over all ten experiments was 110 mm (standard deviation 10 mm). This equates to 27% of the monitor width. This error is 68% higher than in the robotically steered experiment.

VII. Conclusions and future work

This study presented a system that is capable of controlling a flexible endoscope through an endoluminal path from the mouth to the duodenum. Two approaches for detecting the lumen position from endoscopic images were investigated, one was based on optical flow, the other on image intensity. Using both approaches, an endoluminal path was followed during a simulated endoscopy. The RMS distance to the lumen center was 21% of the lumen width for the optical flow-based algorithm and 24% for the intensity-based algorithm.

The intensity-based algorithm was used to steer a conventional flexible endoscope robotically in an experimental setup. This setup was evaluated using an anatomical model. In this experiment, gastroscopy was performed where the GI tract was followed with an RMS error of 16% of the screen width. As a comparison, the same experiment was done using a manually steered endoscope by five Technical Medicine students. The robotically steered endoscope performed 68% better than the manually steered endoscope in terms of keeping the lumen centered in the image. The results indicate that the intensity-based vision algorithm has the potential to improve flexible endoscopy over conventional manual control.

Future directions within this research project will focus on improving system performance. This could be accomplished by using the intensity-based vision algorithm in a shared control system that uses virtual fixtures [20]. A physician and the vision algorithm would share the control of the endoscope. The physician would use a haptic device to steer the endoscope, while being guided by a force that directs the endoscope towards the virtual fixture (i.e., center of the lumen). This way, the physician stays in full control, while his/her performance is improved by the haptic guidance.

Overall system performance could be also improved by incorporating a more sophisticated feature space controller. This controller, that uses the output of the vision algorithm to steer the endoscope, should account for the endoscope's inherent properties (e.g., friction, backlash, and joint compliance). Simulation studies have shown that the optical flow-based algorithm can be used to estimate depth in endoluminal images, and to steer an endoscope. As part of our future work, we will investigate using an optical-flow algorithm that is adapted to work under insufficient texture conditions. Future studies will also include evaluating the performance of the system under more realistic operating conditions.

References

[1] G. C. Harewood, "Relationship of colonoscopy completion rates and endoscopist features," *Digestive Diseases and Sciences*, vol. 50, no. 1, pp. 47–51, 2005.

[2] P. Dario *et al.*, "Smart surgical tools and augmenting devices," *IEEE Trans. Robot. Autom.*, vol. 19, no. 5, pp. 782 – 792, 2003.

[3] A. U. M. Liedlgruber, "Endoscopic image processing - an overview," in *Proc. 6th Int'l Symp. on Image and Signal Processing and Analysis*, 2009.

[4] C. K. Kwoh *et al.*, "Automated endoscope navigation and advisory system from medical imaging," in *Physiology and function from multidimensional images*, San Diego, CA, USA, Feb 1999.

[5] P. Wang *et al.*, "An adaptive segmentation technique for clinical endoscopic image processing," in *Proc. 24th Annual Conf. and the Annual Fall Meeting of the Biomedical Engineering Society*, vol. 2, 2002, pp. 1084 – 1085.

[6] K. V. Asari, "A fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images." *Medical Engineering & Physics*, vol. 22, no. 2, pp. 89–96, 2000.

[7] S. Xia *et al.*, "A novel methodology for extracting colon's lumen from colonoscopic images," *Journal of Systemics, Cybernetics and Informatics*, vol. 1, pp. 7–12, 2003.

[8] G. Gomez *et al.*, "The pq-histogram as a navigation clue," in *Proc. IEEE Int'l. Conf. on Robotics and Automation*, 2002.

[9] S. Hutchinson *et al.*, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 651–670, 1996.

[10] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.

[11] D. Coombs *et al.*, "Real-time obstacle avoidance using central flow divergence, and peripheral flow," *IEEE Trans. Robot. Autom.*, vol. 14, no. 1, pp. 49–59, 1998.

[12] C. McCarthy and N. Barnes, "Performance of optical flow techniques for indoor navigation with a mobile robot," in *Proc. IEEE Int'l. Conf. on Robotics and Automation*, vol. 5, New Orleans, LA, USA, April-May 2004, pp. 5093–5098.

[13] R. Mahony *et al.*, "A new framework for force feedback teleoperation of robotic vehicles based on optical flow," in *Proc. IEEE Int'l. Conf. on Robotics and Automation*, Kobe, Japan, May 2009, pp. 1079–1085.

[14] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int'l. Joint Conf. on Artificial Intelligence*, 1981, pp. 674–679.

[15] C. Gramkow, "On averaging rotations," *Journal of Mathematical Imaging and Vision*, vol. 15, no. 1-2, pp. 7–16, 2001.

[16] J. Santos-Victor *et al.*, "Divergent stereo for robot navigation: learning from bees," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun 1993, pp. 434 –439.

[17] S. J. Phee *et al.*, "Automation of colonoscopy. ii. visual control aspects," *IEEE Eng. Med. Biol. Mag.*, vol. 17, no. 3, pp. 81–88, 1998.

[18] "OpenCV," Willow Garage, Inc., Menlo Park, CA, USA. [Online]. Available: opencv.willowgarage.com

[19] "Blender," Stichting Blender Foundation, Amsterdam, the Netherlands. [Online]. Available: www.blender.org

[20] A. Bettini *et al.*, "Vision-assisted control for manipulation using virtual fixtures," *IEEE Trans. Robot.*, vol. 20, no. 6, pp. 953 – 966, 2004.