2011 IEEE/RSJ International Conference on
Intelligent Robots and Systems
September 25-30, 2011. San Francisco, CA, USA

Three-Dimensional Pose Reconstruction of Flexible Instruments from Endoscopic Images

Rob Reilink, Stefano Stramigioli, and Sarthak Misra

***Abstract*— A position and orientation sensing system is developed for the feedback control of endoscopic instruments in advanced flexible endoscopes. The images that are taken by the endoscopic camera are used to match a kinematic model to the observed instrument. Using the pseudo-inverse of the Jacobian of the forward kinematics, the estimated state of the model is continuously updated so as to match feature points from the images to the model. An experiment was performed inside a colon model, in which reference markers with known locations were touched with the instrument. The root mean square position estimation errors were 1.7 mm, 1.2 mm and 3.6 mm in the horizontal (x), vertical (y), and away-from-camera (z) directions, respectively.**

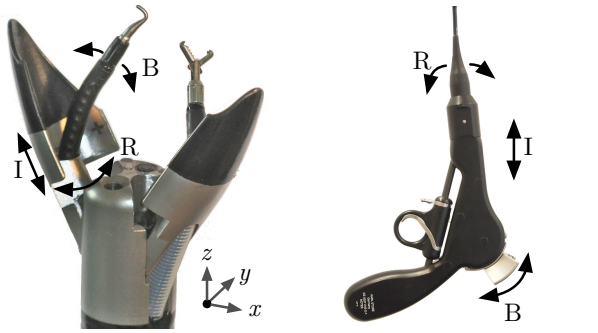
I. Introduction

Endoscopy is a minimally invasive procedure that allows the physician to examine the internal body cavities. The physician uses a flexible endoscope to conduct this procedure. In addition to examination, interventions can be performed using instruments that emerge from the endoscope tip. Recently, Natural Orifice Transluminal Endoscopic Surgery (NOTES) and Single Port Access (SPA) surgery have emerged as new procedures, where the physician performs surgery using a flexible endoscope. These procedures are supposed to result in less trauma than conventional minimally invasive surgical procedures [1]. Advanced endoscopes, that will enable these surgical procedures, are currently being developed. These include the EndoSAMURAI (Olympus Corp., Tokyo, Japan) and the ANUBIS (Karl Storz GmbH & Co. KG, Tuttlingen, Germany; Fig. 1). However, steering these advanced endoscopes is difficult since the controls are not very ergonomic, and multiple physicians are required to operate the endoscope [2]. The latter is undesirable since it requires optimal coordination between the physicians involved, and because of the procedural costs.

A solution would be to employ a robotic telemanipulation system that enables a single physician to control an advanced endoscope in an intuitive way. For laparoscopic surgery, the DaVinci system (Intuitive Surgical, Inc., Sunnyvale, USA) allows this. In this system, the physician operates two 7-Degree of Freedom (DOF) joysticks, which can be coupled to the endoscopic camera or the instruments. A similar approach could be applied to flexible endoscopes, by robotically actuating the endoscope and the instruments.

This research is conducted within the TeleFLEX project, which is funded by the Dutch Ministry of Economic Affairs and the Province of Overijssel, within the Pieken in de Delta (PIDON) initiative. The ANUBIS endoscopic instrument was provided by Karl Storz GmbH & Co. KG.

The authors are affiliated with MIRA - Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands. {*r.reilink, s.stramigioli, s.misra*}@utwente.nl



(a) Endoscope tip with instruments (b) Control handle

Fig. 1. The endoscopic instruments of the ANUBIS endoscope have three degrees of freedom: insertion (I,q_1), rotation (R,q_2), and bending (B,q_3). They are operated by rotating and translating the control handle (R and I) and by moving a lever on the handle (B).

In previous work, we have studied intuitive steering of the endoscope using haptic guidance [3]. As a next step, our goal is to implement intuitive and accurate control of the endoscopic instruments. However, this is difficult since there exists significant flexibility between the instrument at the tip and its actuation points at the control handle (Fig. 1b). Combined with the internal friction within the endoscope, this causes a hysteresis effect, making the steering of the instrument very difficult [4]. Using a feed-forward compensation, the effect may be reduced. Reduction is however only possible up to a limited extent, since the parameters of the flexibility and the friction change with the (unknown) shape and force loading of the instrument [5]. Therefore, we consider a feedback approach. For this purpose, our aim is to develop a sensor system that can measure the position and orientation of the endoscopic instrument.

For a feedback approach, sensing of the actual position and orientation of the endoscope instruments is required. Adding extra sensors to the endoscope is difficult and expensive, since the available space is limited, and the sensors will need to be sterilizable. Therefore, we will investigate the use of the endoscopic camera images to determine the instrument position and orientation.

In the proposed approach, we use image processing techniques to extract feature points from the endoscopic images. These points are compared with the estimated positions of these feature points, obtained from a kinematic model. Based on the deviations between the model and the observations, the state of the model is updated, such that the model will converge to the observations.

(a) (b)

Fig. 2. **Instrument feature points to be tracked:** (a) The points that are tracked are marked with arrows. (b) Using the position of point $\mathbf{c}$ and the orientation of the instrument $\mathbf{d}$, the position of the top feature point $\mathbf{p}_1$ is approximated.

In this paper, we will describe a system that considers only a single instrument. However, the work is extensible to tracking multiple instruments, which are used in advanced flexible endoscopes. This paper is structured as follows: In Section II, the kinematics model of the endoscope instrument and the model of the endoscopic camera are described. Section III describes the computer vision algorithm that is used to extract features from the endoscopic images. In Section IV, the use of these features to estimate the state of the instrument is discussed. In order to evaluate the proposed approach, an experiment was done using an actual endoscope in a colon model. This experiment, and the results, are described in Section V. Finally, Section VI concludes and provides directions for future work.

II. Model of the Endoscopic Instrument

In order to be able to estimate the state of the endoscopic instrument, we will use a kinematics model of the instrument and a model of the endoscope camera. These models will be discussed in this section. We will use $\mathbf{q} := \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}^{\mathrm{T}}$ to denote the state of the instrument actuation in the insertion (q_1), rotation (q_2) and bending (q_3) directions (Fig. 1).

For the state estimation, we will use the two-dimensional (2D) coordinates of two distinct feature points in the image as the input. These points are shown in Fig. 2a. These points are chosen because they are clearly distinguishable in the image. Furthermore, they are unlikely to be occluded if the gripper is used to grab tissue. We will denote the position of the k-th feature point in three-dimensional (3D) cartesian space as $\mathbf{p}_k$ ($k = 1, 2$). We will denote its projection on the 2D camera image plane as $\mathbf{x}_k$. For the state estimation, the relation between $\mathbf{q}$ and $\mathbf{p}_k$ (forward kinematics) is required, this will be denoted as the function f_k:

$$f_k : \mathbb{R}^3 \rightarrow \mathbb{R}^3; \mathbf{q} \mapsto \mathbf{p}_k \quad . \tag{1}$$

Furthermore, the relation between the derivatives $\dot{\mathbf{q}}$ and $\dot{\mathbf{p}}_k$ are required for the state estimation. These relations are derived in the remainder of this section.

A. Kinematics model of the instrument

For the kinematics of the instrument, we use a model similar to that of Bardou et al. [4]. This model assumes a constant radius of curvature within the bending section of the instrument. This assumption is considered to be valid as long as there are no external forces on the instrument, and the friction of the cables inside the bending section is low, which is the case. We will describe the kinematics for the Karl Storz ANUBIS instrument. However, the proposed approach could be made to work with other instruments as well, by adapting the kinematics model appropriately.

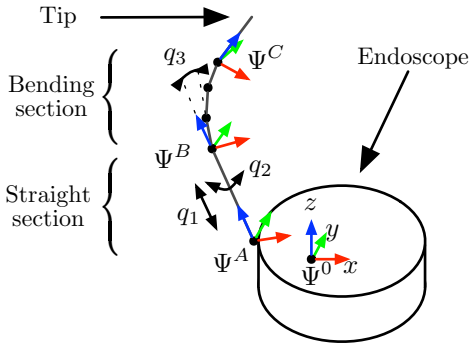


Fig. 3. **Kinematics model of the instrument:** The optical center of the camera is the world frame Ψ^0. Base frame Ψ^A is where the instrument emerges from the endoscope. The straight section is described by the insertion (q_1) and rotation (q_2) DOFs. The bending section consists of multiple links, connected by joints (only 3 links shown for clarity)

The instrument is modeled as a straight section, which can be inserted/retracted (q_1) and rotated (q_2), followed by a bending section (q_3), and a tip (Fig. 3). In the case of the ANUBIS instrument, the bending section of the physical instrument consists of 9 interconnected rigid bodies. This section is modeled as a series of 10 parallel rotational joints, which are interconnected by 9 links.

For the analysis, the world frame Ψ^0 is chosen at the camera optical center, with the z-axis pointing in the camera viewing direction (Fig. 3). Note that the camera is not located exactly in the center of the endoscope tip. The instrument base frame Ψ^A is located at the point where the instrument emerges from the endoscope, with the z-axis in the direction of the instrument shaft. The mapping between Ψ^0 and Ψ^A is described by the homogeneous transformation matrix[1]

$${}^0_A\mathbf{H} = \begin{bmatrix} & & -12 \\ \mathbf{R}_y(-15°) & & 3 \\ & & 0 \\ 0\ 0\ 0 & & 1 \end{bmatrix} \quad , \tag{2}$$

where $\mathbf{R}_y(\cdot)$ denotes a matrix describing a rotation around the y-axis

$$\mathbf{R}_y(\cdot) := \begin{bmatrix} \cos(\cdot) & 0 & \sin(\cdot) \\ 0 & 1 & 0 \\ -\sin(\cdot) & 0 & \cos(\cdot) \end{bmatrix} \quad . \tag{3}$$

The straight section is a translation along the z-axis at a distance given by q_1, combined with a rotation around the z-axis, at an angle given by q_2. We define Ψ^B at the end of the straight section (Fig. 3). The homogeneous transformation between Ψ^A and Ψ^B is described by

$${}^A_B\mathbf{H} = \begin{bmatrix} & & 0 \\ \mathbf{R}_z(q_2) & & 0 \\ & & q_1 \\ 0\ 0\ 0 & & 1 \end{bmatrix} \quad , \tag{4}$$

where $\mathbf{R}_z(\cdot)$ denotes a matrix describing a rotation around

[1] All distances are expressed in mm

the z-axis

$$\mathbf{R}_z(\cdot) := \begin{bmatrix} \cos(\cdot) & -\sin(\cdot) & 0 \\ \sin(\cdot) & \cos(\cdot) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

The bending section consists of a series of joints, rigidly interconnected by links. The homogeneous transformation of each individual joint is described by $\mathbf{H}_\mathrm{J}$, and the homogeneous transformation of each individual link is described by $\mathbf{H}_\mathrm{L}$

$$\mathbf{H}_\mathrm{J} = \begin{bmatrix} & & & 0 \\ & \mathbf{R}_y(q_3) & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{H}_\mathrm{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \ell \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad , \tag{6}$$

where $\ell = 1.6$ mm is the length of each link.

We define Ψ^C at the end of the bending section (Fig. 3). The homogeneous transformation matrix ${}^B_C\mathbf{H}$ describing the complete bending section is given by

$${}^B_C\mathbf{H} = (\mathbf{H}_\mathrm{J}\mathbf{H}_\mathrm{L})^9\mathbf{H}_\mathrm{J} \quad . \tag{7}$$

For the computation of the position of the feature points, the point at the centerline of the gripper, at the border between the dark and the light region is required (denoted $\mathbf{c}$ in Fig. 2a and b). This point is located at $[0, 0, 5]^\mathrm{T}$ in Ψ^C. Furthermore, the direction of the instrument is required ($\mathbf{d}$ in Fig. 2b). These are computed as

$$\mathbf{c} = {}^0_C\mathbf{H} \begin{bmatrix} 0 \\ 0 \\ 5 \\ 1 \end{bmatrix}, \ \mathbf{d} = {}^0_C\mathbf{H} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \text{ where } {}^0_C\mathbf{H} = {}^0_A\mathbf{H}\,{}^A_B\mathbf{H}\,{}^B_C\mathbf{H} \,. \tag{8}$$

Depending on the instrument position and orientation, different points on the circumference of the instrument will be projected to the top or bottom feature points. $\mathbf{d}$ denotes the direction of the instrument. The approximation of the top feature point is illustrated in Fig. 2. In this figure, $\mathbf{o}$ is the camera optical center, and $\mathbf{c} - \mathbf{o}$ is the vector from $\mathbf{o}$ to the instrument center $\mathbf{c}$. The top feature point $\mathbf{p}_1$ is on the circumference of the instrument, in the direction perpendicular to $\mathbf{d}$ and $\mathbf{c} - \mathbf{o}$:

$$\mathbf{p}_1 = \mathbf{c} + r\frac{(\mathbf{c}-\mathbf{o}) \times \mathbf{d}}{||(\mathbf{c}-\mathbf{o}) \times \mathbf{d}||} \quad , \tag{9}$$

with r denoting the radius of the instrument, $\times$ denoting the vector cross product (disregarding the 4th homogeneous coordinate) and $||\cdot||$ denoting the Euclidian norm. Similar to (9), the position of the bottom feature point is computed as

$$\mathbf{p}_2 = \mathbf{c} - r\frac{(\mathbf{c}-\mathbf{o}) \times \mathbf{d}}{||(\mathbf{c}-\mathbf{o}) \times \mathbf{d}||} \quad . \tag{10}$$

The estimated feature points $\mathbf{p}_1$ and $\mathbf{p}_2$ describe the forward kinematics f_k in (1).

B. Differential kinematics of the instrument

For the state estimation, it is required to know the relation between the change of the state ($\dot{\mathbf{q}}$) and the change of the positions of the feature points ($\dot{\mathbf{p}}_k$). This relation is given by the Jacobian $\mathbf{F}_k$ as

$$\dot{\mathbf{p}}_k = \mathbf{F}_k(\mathbf{q})\dot{\mathbf{q}} \tag{11}$$

This relation will be used to find the change of states that is required to move the estimated feature point positions. In order to find this relation, the geometric Jacobian will be used [6]. We will use $\mathbf{T}_c^{a,b}$ to denote the twist of frame Ψ^b with respect to frame Ψ^c expressed in frame Ψ^a. We will use $\hat{\mathbf{T}}_i$ to denote the unit twist of Ψ^C with respect to Ψ^0, associated with joint q_i expressed in Ψ^0.

Given the unit twists $\hat{\mathbf{T}}_1$, $\hat{\mathbf{T}}_2$ and $\hat{\mathbf{T}}_3$, and the joint velocities $\dot{\mathbf{q}}$, the twist $\mathbf{T}_0^{0,C}$ can be computed as

$$\mathbf{T}_0^{0,C} = \hat{\mathbf{T}}_1\dot{q}_1 + \hat{\mathbf{T}}_2\dot{q}_2 + \hat{\mathbf{T}}_3\dot{q}_3 \quad . \tag{12}$$

Using the matrix form of the twist, denoted $\tilde{\mathbf{T}}$, the motion of the points $\mathbf{p}_k$ can be written as

$$\dot{\mathbf{p}}_k = \tilde{\mathbf{T}}_0^{0,C}\mathbf{p}_k = \underbrace{\begin{bmatrix} \tilde{\hat{\mathbf{T}}}_1\mathbf{p}_k & \tilde{\hat{\mathbf{T}}}_2\mathbf{p}_k & \tilde{\hat{\mathbf{T}}}_3\mathbf{p}_k \end{bmatrix}}_{\mathbf{F}_k(\mathbf{q})}\dot{\mathbf{q}} \quad , \tag{13}$$

where $\dot{\mathbf{p}}_k$ is expressed in frame Ψ^0. The computation of $\hat{\mathbf{T}}_1$, $\hat{\mathbf{T}}_2$, and $\hat{\mathbf{T}}_3$ is described in the remainder of this section.

q_1 and q_2 represent the insertion and rotation of the straight section, which are a translation along the z-axis and a rotation around the z-axis of frame Ψ^A, respectively. Thus, the combined twist of the straight section is

$$\begin{aligned} \mathbf{T}_A^{A,B} = &\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^\mathrm{T}\dot{q}_1 \\ &+ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^\mathrm{T}\dot{q}_2 \quad . \end{aligned} \tag{14}$$

This can be expressed in frame Ψ^0 using the Adjoint operator (denoted $\mathrm{Ad}_\mathbf{H}$):

$$\mathbf{T}_A^{0,B} = \mathrm{Ad}_{{}^0_A\mathbf{H}}\mathbf{T}_A^{A,B} \tag{15}$$

Combining (14) and (15) shows the unit twists associated with q_1 and q_2:

$$\hat{\mathbf{T}}_1 = \mathrm{Ad}_{{}^0_A\mathbf{H}}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^\mathrm{T} \tag{16}$$

$$\hat{\mathbf{T}}_2 = \mathrm{Ad}_{{}^0_A\mathbf{H}}\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^\mathrm{T} \tag{17}$$

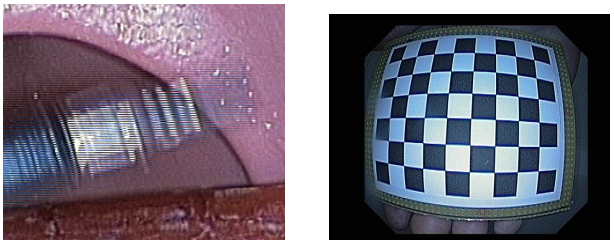
For the bending section, we fix a frame Ψ^i to the previous link for every rotational joint i ($i = 1, \ldots, 10$). These are oriented such, that Ψ^1 coincides with Ψ^B. For each joint i, the associated twist represents a rotation around the y-axis of the frame Ψ^i:

$$\mathbf{T}_i^{i,i+1} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^\mathrm{T}\dot{q}_3 \tag{18}$$

This can be expressed in frame $\Psi^B (= \Psi^1)$:

$$\mathbf{T}_i^{B,i+1} = \mathrm{Ad}_{{}^B_i\mathbf{H}}\mathbf{T}_i^{i,i+1} \quad , \tag{19}$$

in which the homogeneous transformation matrix ${}^B_i\mathbf{H}$ is



(a) (b)

Fig. 4. **Image distortions to be compensated by the pre-processing:** (a) Jaggered edges are caused by the interlaced video signal. (b) The fish-eye lens causes severe barrel distortion.

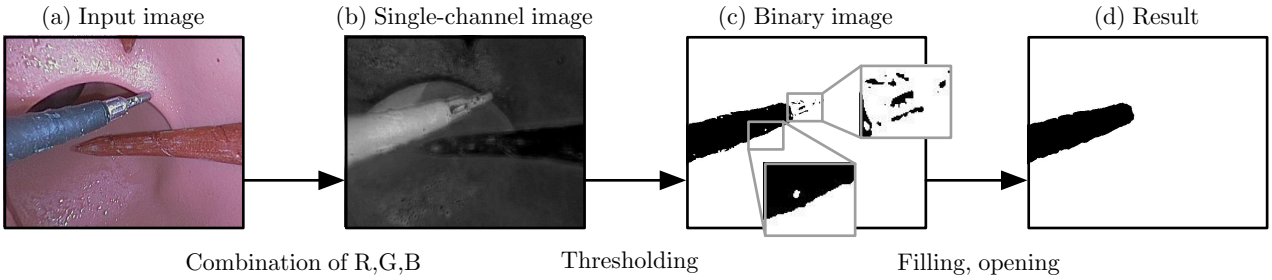


Fig. 5. **Instrument detection:** From the color input image (a), the color channels are combined in an optimal way to get a single-channel image (b). This image is thresholded to get a binary image (c). Finally, holes in this region are filled and small regions are removed using the binary opening operation, resulting in image (d).

composed of all proceeding joints and links:

$$ {}^{B}_{i}\mathbf{H} = \left(\mathbf{H}_{\mathrm{J}}\mathbf{H}_{\mathrm{L}}\right)^{(i-1)} \quad . \tag{20} $$

The total twist associated with the bending section is the sum of the twists of each of the 10 joints, expressed in Ψ^B as

$$ \mathbf{T}_B^{B,C} = \sum_{i=1}^{10} \mathbf{T}_i^{B,i+1} \tag{21} $$

This can be expressed in Ψ^0 as

$$ \mathbf{T}_B^{0,C} = \mathrm{Ad}_{{}^{0}_{B}\mathbf{H}} \mathbf{T}_B^{B,C} \tag{22} $$

Combining (18-22) gives the unit twist for q_3 as

$$ \hat{\mathbf{T}}_3 = \mathrm{Ad}_{{}^{0}_{B}\mathbf{H}} \left(\sum_{i=1}^{10} \mathrm{Ad}_{{}^{B}_{i}\mathbf{H}} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}} \right) \tag{23} $$

Given unit twists $\hat{\mathbf{T}}_1$, $\hat{\mathbf{T}}_2$ and $\hat{\mathbf{T}}_3$, as computed in (16), (17), and (23), respectively, the Jacobian $\mathbf{F}_k(\mathbf{q})$ defined in (11) that describes the differential kinematics can be determined according to (13).

C. Camera model

A pinhole camera model is used. The points $\mathbf{p}_k$ in the 3D cartesian space are mapped to points $\mathbf{x}_k$ in the 2D image plane according to

$$ \mathbf{x}_k = g(\mathbf{p}_k) \quad . \tag{24} $$

In (28), g denotes the pinhole projection function:

$$ g : \mathbb{R}^3 \to \mathbb{R}^2; \mathbf{p_k} \mapsto \frac{f}{p_z} \begin{bmatrix} p_x \\ p_y \end{bmatrix} \quad , \tag{25} $$

where f is the focal distance and p_x, p_y, and p_z are the x-, y- and z-components of $\mathbf{p}_k$ (for notational simplicity, the subscript k was left out in (25)). Taking the derivative of (25) yields the relation between $\mathbf{p}_k$ and $\mathbf{x}_k$, given by the Jacobian $\mathbf{G}(\mathbf{p}_k)$ [7]:

$$ \dot{\mathbf{x}}_\mathbf{k} = \mathbf{G}(\mathbf{p}_k)\dot{\mathbf{p}}_k, \quad \text{where} \tag{26} $$

$$ \mathbf{G}(\mathbf{p}_k) := \frac{\partial g}{\partial \mathbf{p}_k} = f \begin{bmatrix} \frac{1}{p_z} & 0 & -\frac{f p_x}{{p_z}^2} \\ 0 & \frac{1}{p_z} & -\frac{f p_y}{{p_z}^2} \end{bmatrix} \tag{27} $$

D. Combined model of the instrument and the camera

The kinematics model of the instrument and the model of the camera can be combined in order to find the relation between the state ($\mathbf{q}$) and the positions of feature points in the camera image ($\mathbf{x}_k$). This is the composition of the camera model g from (25) and the kinematics model f_k from (1):

$$ \mathbf{x}_k = (g \circ f_k)(\mathbf{q}) \quad . \tag{28} $$

Therefore, the relation between the actuator velocities $\dot{\mathbf{q}}$ and the velocities of the feature points in the camera image $\dot{\mathbf{x}}_k$ is the product of the Jacobians $\mathbf{G}(\mathbf{p}_k)$ and $\mathbf{F}_k(\mathbf{q})$, combining (11) and (26):

$$ \dot{\mathbf{x}}_k = \mathbf{J}_k(\mathbf{q})\dot{\mathbf{q}} \text{ , where } \mathbf{J}_k(\mathbf{q}) := \mathbf{G}(f_k(\mathbf{q}))\mathbf{F}_k(\mathbf{q}) \ . \tag{29} $$

This relation is used to perform the state estimation for the instrument, as will be described in Section IV.

III. Computer vision algorithm

A computer vision algorithm has been developed to find the observations, denoted $\tilde{\mathbf{x}}$, that are input to the state estimator which will be described in the next section. The computer vision algorithm processes the color image stream originating from the endoscope camera in three steps. Firstly, the images are pre-processed. Secondly, the endoscopic instrument is detected. Finally, the features that represent the observations ($\tilde{\mathbf{x}}$) are extracted. These steps will be described in the remainder of this section.

A. Pre-processing

The image stream, coming from the endoscope camera, is first pre-processed by a de-interlacer and a lens calibration algorithm [8], [9]. The output of the endoscope is an interlaced video stream, meaning that the odd and even numbered lines of each image are sampled at different time instants. When these are combined into a single image, artifacts may be introduced (Fig. 4a). These artifacts may adversely affect the vision algorithm. Therefore, gstreamer was used to de-interlace the images using the 'greedyh' algorithm [8]. Further, since the endoscope has a quite severe lens distortion (Fig. 4b), the resulting images were corrected using the Camera Calibration Toolbox for Matlab [9].

B. Instrument detection

The instrument detection process is depicted in Figure 5. The pre-processed RGB color image is first converted to a single-channel image. The red, green, and blue channels are combined linearly. The weights of this linear combination are determined using Fishers Linear Discriminant method [10], in order to maximize the contrast between the dark section of the instrument, and the background. The resulting image is thresholded using a global threshold. It was found during experiments, that the exact level of this threshold did not influence the results significantly. This suggests that the algorithm is relatively insensitive to changes in lighting conditions. However, a fixed global threshold algorithm may

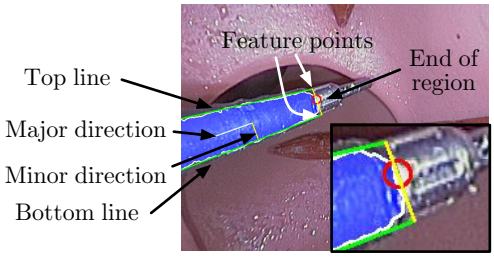


Fig. 6. **Feature extraction**: The major and minor directions are determined from the instrument region. Subsequently, the top and bottom lines are fitted to the top and the bottom edges of the region.

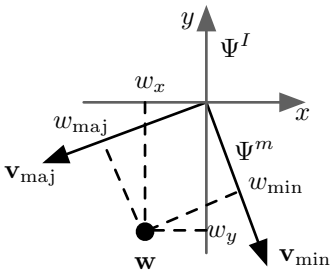


Fig. 7. Using the eigenvectors that point in the major and minor directions $\mathbf{v}_{\text{maj}}$ and $\mathbf{v}_{\text{maj}}$, coordinates of point $\mathbf{w}$ can be transformed from the image coordinate system Ψ^I to major/minor coordinate system Ψ^m.

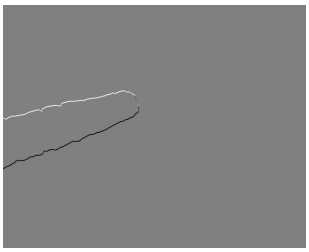


Fig. 8. Using a directional filter, the top and bottom edges of the instrument region are found. The top edge is brighter than the background, while the bottom edge is darker.

be unsuitable for robust instrument detection in clinical images. In this case, a more sophisticated (adaptive) approach may be required.

Morphological operations were applied on the resulting binary image [11]. In particular, holes of the regions were filled, and small regions, present at the tip of the instrument, were removed using a binary opening operation (Fig. 5c). The resulting image contains a single region representing the dark section of the endoscopic instrument.

C. Feature extraction

Given the region representing the dark section of the endoscopic instrument, first the major and minor directions are found (Fig. 6). For this purpose, the edge of the region is considered as a point cloud, with a point at each pixel position belonging to the edge. Of this point cloud, the eigenvectors of the covariance matrix are computed. The eigenvector belonging to the largest eigenvalue represents the major axis. The eigenvector belonging to the smallest eigenvalue represents the minor axis. The eigenvectors are each negated as necessary in order to ensure that the major axis always points towards the left-hand side of the image, and the minor eigenvector always points towards the bottom of the image. The eigenvectors are normalized and combined into matrix $\mathbf{V}$, which can be used as a coordinate transform between image coordinate system Ψ^I, and major and minor axis-coordinate system Ψ^m (Fig. 7):

$$\mathbf{V} = \begin{bmatrix}\mathbf{v}_{\text{maj}} & \mathbf{v}_{\text{min}}\end{bmatrix} \quad , \quad \begin{bmatrix}w_{\text{maj}} \\ w_{\text{min}}\end{bmatrix} = \mathbf{V}^{-1}\begin{bmatrix}w_x \\ w_y\end{bmatrix} \quad , \quad (30)$$

where $\mathbf{v}_{\text{maj}}$ and $\mathbf{v}_{\text{min}}$ represent the unit-length eigenvectors associated with the largest and the smallest eigenvalue, respectively. w_x and w_y are the x- and y-coordinates of a point $\mathbf{w}$ expressed in Ψ^I, while w_{maj} and w_{min} are its coordinates expressed in Ψ^m .

In order to extract the top and bottom edges of the region, a gradient filter is applied on the image of the region. The gradient is computed in the direction of the minor axis using a Sobel filter [11]. The directional gradient is computed using a convolution kernel that is a linear combination of the Sobel filter in the x-direction ($\mathbf{S}_x$) and in the y-direction ($\mathbf{S}_y$):

$$\mathbf{S}_x = \begin{bmatrix}-1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1\end{bmatrix}, \quad \mathbf{S}_y = \begin{bmatrix}-1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1\end{bmatrix} \quad . \quad (31)$$

The directional filter kernel $\mathbf{S}_d$ is

$$\mathbf{S}_d = v_{\text{min}_x}\mathbf{S}_x + v_{\text{min}_y}\mathbf{S}_y \quad , \quad (32)$$

where v_{min_x} and v_{min_y} are the x- and y-component of $\mathbf{v}_{\text{min}}$, respectively. The resulting gradient image is shown in Fig. 8.

In the gradient image, the points belonging to the top edge are found by selecting points with positive values (white in Fig. 8). Similarly, the points belonging to the bottom edge are found by selecting points with negative values.

Then, two lines are fitted to the top and bottom points using a linear least squares fit (Fig. 6). Furthermore, the point of the region with the lowest coordinate in the major direction, w_{maj}, (i.e. most towards the tip) is selected as the end of the region. This is shown as the red circle in Fig. 6. From this point, a line is defined in the minor direction. The intersections between this line and the top and bottom lines are the feature points.

IV. State Estimation

Using the model of the endoscopic instrument, we aim to estimate the state $\mathbf{q}$ from the endoscopic images. In the following sections, we will use $\mathbf{u}$ to denote the vector that combines the estimated positions of the two feature points in the 2D image plane:

$$\mathbf{u} := \begin{bmatrix}\mathbf{x}_1 \\ \mathbf{x}_2\end{bmatrix} \quad , \quad (33)$$

with $\mathbf{x}_1$ and $\mathbf{x}_2$ the estimated positions of the top and bottom feature points in the image plane, respectively, as computed using (28).

A block diagram of the estimation algorithm is shown in Figure 9. For a given current state $\mathbf{q}$, the estimated observations $\mathbf{u}$ are computed using the kinematics and camera models. From the estimated observations $\mathbf{u}$, and the actual observations (denoted $\tilde{\mathbf{u}}$), the errors $\mathbf{e} := \tilde{\mathbf{u}} - \mathbf{u}$ are computed.

The controller C determines the error dynamics. The desired change of error (denoted $\bar{\mathbf{e}}$) is computed based on the current error $\mathbf{e}$ according to

$$\bar{\mathbf{e}} = K\mathbf{e} \quad , \quad (34)$$

with K a positive constant gain.

The dependence of $\dot{\mathbf{e}}$ on the change of states $\dot{\mathbf{q}}$ is given by

$$\dot{\mathbf{e}} = \dot{\tilde{\mathbf{u}}} - \dot{\mathbf{u}} = \dot{\tilde{\mathbf{u}}} - \mathbf{J}\dot{\mathbf{q}} \text{ , with } \mathbf{J} := \begin{bmatrix}\mathbf{J}_1(\mathbf{q}) \\ \mathbf{J}_2(\mathbf{q})\end{bmatrix} \quad , \quad (35)$$

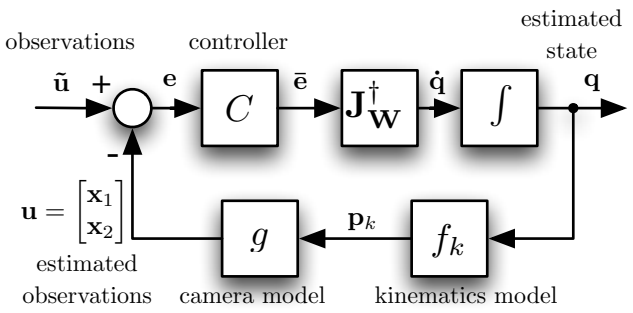


Fig. 9. **State estimation:** From the current estimated state $\mathbf{q}$, the estimated observations $\mathbf{x}_k$ are computed using the kinematics and camera models. These are compared to the actual observations $\tilde{\mathbf{x}}$, and the error $\mathbf{e}$ is used to update the estimated state.

where $\mathbf{J}_1$ and $\mathbf{J}_2$ are the Jacobians for the individual feature points as given in (29). Because the dimension of $\mathbf{q}$ is smaller than the dimension of $\mathbf{e}$, there is in general no $\mathbf{q}$ that realizes a perfect match between the estimated and the actual observations (i.e., a $\mathbf{q}$ such that $\mathbf{e} = 0$). Similarly, the desired change of error $\bar{\mathbf{e}}$ is in general not realizable (i.e., there is in general no $\dot{\mathbf{q}}$ such that $\dot{\mathbf{e}} = \bar{\mathbf{e}}$).

Therefore, the pseudo-inverse of the Jacobian, $\mathbf{J}^{\dagger}$, is used to compute the change in estimated state $\dot{\mathbf{q}}$ that realizes $\dot{\mathbf{e}} = \bar{\mathbf{e}}$ as 'good' as possible. More specifically, we seek the $\dot{\mathbf{q}}$ that minimizes

$$||\mathbf{W}(\dot{\mathbf{e}} - \bar{\mathbf{e}})||_2 \quad , \tag{36}$$

with $\mathbf{W}$ a weighting matrix. This is obtained by taking

$$\dot{\mathbf{q}} = \mathbf{J}^{\dagger}_{\mathbf{W}}\bar{\mathbf{e}} \quad , \tag{37}$$

with $\mathbf{J}^{\dagger}_{\mathbf{W}}$ the weighted pseudo-inverse given by [12]

$$\mathbf{J}^{\dagger}_{\mathbf{W}} := (\mathbf{J}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{J})^{-1}\mathbf{J}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W} \quad . \tag{38}$$

The weighting matrix $\mathbf{W}$ is used to influence the minimization of (36). For the state estimation, an accurate estimation of the direction of the tip is more important than the position of the feature points, since small errors in the estimation of the direction will result in large errors of the tip position. This can be expressed by specifying $\mathbf{W}$ in Ψ^m: errors in the major direction will be penalized more than errors in the minor direction. This is obtained by taking

$$\mathbf{W} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{V}^{-1} & 0 \\ 0 & \mathbf{V}^{-1} \end{bmatrix} \tag{39}$$

The first term of (39) is a diagonal matrix that specifies the weights for the components of $\dot{\mathbf{e}} - \bar{\mathbf{e}}$ expressed in Ψ^m. The second term transforms the components of $\dot{\mathbf{e}} - \bar{\mathbf{e}}$ from Ψ^I to Ψ^m.

V. Experimental results

A test setup has been constructed in order to evaluate the performance of the image processing and the state estimation described in the previous sections. It involves a tip attachment that fits onto the tip of a conventional colonoscope (Fig. 10). This tip attachment contains a channel that allows a flexible instrument to be passed through, and ensures that this instrument will emerge at a fixed orientation

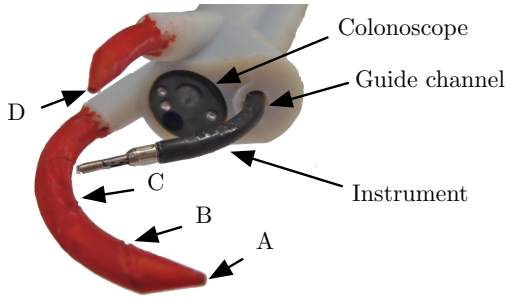


Fig. 10. A tip attachment was constructed to fit onto the tip of a colonoscope. A guide channel, which guides the instrument, ensures that the instrument emerges at a fixed position. Furthermore, there are four reference points (labelled A-D) that can be touched with the instrument.

with respect to the endoscope tip. Furthermore, there are several reference points, whose 3D location with respect to the endoscope tip are known (labelled A-D in Fig. 10). These reference points provide ground truth values for the position of the instrument. The reference is red, so as to minimize the influence on the instrument detection algorithm.

An Olympus colonoscope system (Exera II CV-180/CLV-180, Olympus Corp., Tokyo, Japan) was used for the experiment. For the instrument, a gripper of the Karl Storz ANUBIS system was used (Karl Storz GmbH & Co. KG, Tuttlingen, Germany). The endoscope, the tip attachment and the instrument were inserted in the colon of a colonoscopy model (KKM40, Kyoto Kagaku, Kyoto, Japan). The model was coated with a transparent lubricant internally as per the manufacturers instructions, in order to create the specular reflections similar to those that are present in clinical colonoscopic images. The instrument was operated manually, and was positioned to touch the reference points. The images were recorded for off-line processing.

The results of the state estimator are shown in Figure 11. These graphs show the horizontal (x), vertical (y), and away-from-camera (z) coordinates of the position of the tip expressed in Ψ^0, computed based on the estimated state $\mathbf{q}$. Using the images of the experiment, the periods where the instrument touches a reference point where manually marked. The true position of each reference point, $\mathbf{p}_R$, is known from the design of the tip attachment. During the periods where the tip touches a reference point, the true position of that point is also shown in Fig. 11.

During the periods where the tip touches a reference point, the estimation error $\mathbf{\Delta}$ is computed, which is defined as the difference between the estimated tip position $\mathbf{p}_T$ and the reference point position $\mathbf{p}_R$:

$$\mathbf{\Delta} = \mathbf{p}_T - \mathbf{p}_R, \quad \text{with} \quad \mathbf{p}_T = {}^0_C\mathbf{H} \begin{bmatrix} 0 \\ 0 \\ 20 \\ 1 \end{bmatrix} \quad , \tag{40}$$

where $\mathbf{\Delta}$, $\mathbf{p}_T$ and $\mathbf{p}_R$ are expressed in frame Ψ^0. The orientation of the gripper, which is contained in ${}^0_C\mathbf{H}$, is not evaluated explicitly, but only through $\mathbf{p}_T$, which is dependent on the rotation component of ${}^0_C\mathbf{H}$.

For each of the points, the root mean square (RMS) value is computed for the components Δ_x, Δ_y and Δ_z. These are shown in Table I. The table also shows the RMS value of the Euclidian norm of $\mathbf{\Delta}$, $||\mathbf{\Delta}||_2$.

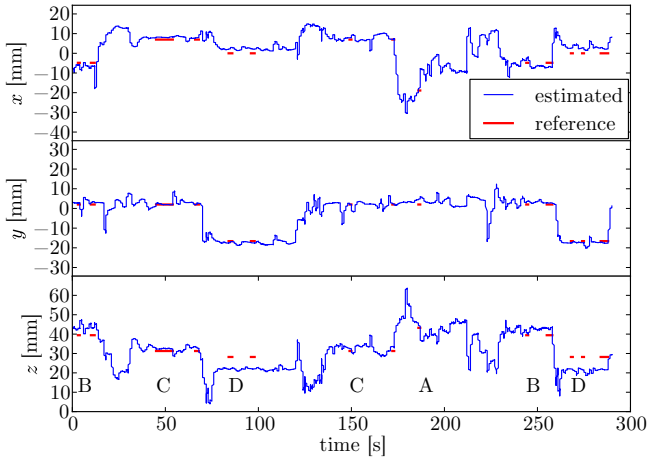


Fig. 11. The plots show the x-, y- and z- coordinate of the estimated tip position. The positions of the reference points are indicated at those times where a reference point was touched.

It can be noted that the errors in the z-direction are in general larger than the errors in the x- and y- directions. This is due to the fact that the motions in the x- and y-direction are better observable from the camera viewpoint. The deviation in the z-direction for reference D is a systematic (i.e. repeatable) error: at both times where reference D is touched (around 90 s and 270 s) a constant deviation in the z-direction is present. A difference between the parameters of the kinematic model (e.g. the lengths of the links) and the actual instrument is the most likely cause here. However, for the purpose of controlling the instrument in a feedback loop, with the set-point given by a physician, repeatability and low noise are more important than absolute accuracy. Systematic deviations from the correct position can be compensated by the physician as long as the deviations are repeatable and have a low noise.

VI. Conclusions and Future Work

A state estimator has been designed that can estimate the orientation and the tip position of a 3-DOF endoscopic instrument, based on the endoscopic images. Using this estimator, the position of the tip of the instrument was estimated, and compared to four reference points with known positions. The RMS position estimation error, averaged over the four reference points was 1.7, 1.2 and 3.6 mm for the x-, y- and z- directions, respectively.

For future work, our aim is to use the developed estimator in a feedback loop to control the 3-DOF position of the endoscopic instrument accurately. Accurate control of the endoscopic instrument is essential for a physician to use the instrument effectively. In order to use the estimator in this setting, several points need to be addressed.

Currently, the lens distortion is compensated off-line, by pre-processing the individual images. However, if the lens distortion is incorporated in the camera model, this step would no longer be required. The image processing algorithm would then process the 'raw' images, and the estimator would compensate for the distortion.

Furthermore, the detection of the feature points by the image processing algorithm could be improved by using

TABLE I

For each reference point, the RMS estimation error of the tip position in the x-, y- and z-direction are given.

Reference	Δ_x RMS [mm]	Δ_y RMS [mm]	Δ_z RMS [mm]	$\|\|\boldsymbol{\Delta}\|\|_2$ RMS [mm]
A	1.3	2.4	2.6	3.8
B	2.0	0.9	3.8	4.4
C	1.0	0.6	1.5	1.9
D	2.4	0.8	6.3	6.8
Mean	1.7	1.2	3.6	4.2

dedicated markers on the instrument. Since markers can be chosen to be of a contrasting color, their detection is easier. This can result in faster, more accurate and more robust measurements of the feature points.

With the current estimator, a systematic error exists when touching reference point D. It is suspected that a difference between the parameters of the kinematic model and the actual instrument causes this systematic error. This is not a critical problem for the purpose of controlling the instrument, since repeatability and low noise are more important than absolute accuracy. However, methods to reduce the systematic error and improve the model need to be studied. Possible improvements include identifying the kinematic parameters, and modeling the internal friction and flexibilities.

Additionally, when the estimator is used for the purpose of controlling the instrument, the dynamic performance of the estimator is of importance. In order to measure the dynamic performance, the approach with fixed reference points will no longer be viable. Using a magnetic position tracking system (e.g. Aurora, NDI, Waterloo, Canada), the dynamic performance of the estimator could be studied and optimized so as to make the estimator suitable for use in a feedback control system.

References

[1] Kalloo *et al.*, "Flexible transgastric peritoneoscopy: a novel approach to diagnostic and therapeutic interventions in the peritoneal cavity," *Gastrointestinal Endoscopy*, vol. 60, no. 1, pp. 114–117, 2004.

[2] Marescaux *et al.*, "Surgery without scars: report of transluminal cholecystectomy in a human being," *Archives of Surgery*, vol. 142, no. 9, pp. 823–826, 2007.

[3] Reilink *et al.*, "Evaluation of flexible endoscope steering using haptic guidance," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 7, no. 2, pp. 178–186, 2011.

[4] Bardou *et al.*, "Control of a multiple sections flexible endoscopic system," in *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, 2010, pp. 2345–2350.

[5] Abbott *et al.*, "Design of an endoluminal notes robotic system," in *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS)*, San Diego, CA, USA, 2007, pp. 410–416.

[6] S. Stramigioli and H. Bruyninckx, *Geometry and Screw Theory for Robotics*, Tutorial from IEEE Int'l. Conf. on Robotics and Automation, Seoul, Korea, 2001.

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

[8] "Gstreamer: open source multimedia framework." [Online]. Available: http://gstreamer.freedesktop.org/

[9] J.-Y. Bouguet, "Camera calibration toolbox for Matlab." [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc

[10] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, no. 7, pp. 179–188, 1936.

[11] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2002.

[12] Y. Nakamura, *Advanced Robotics, Redundancy and Optimization*. Addison-Wesley, 1991.