2012 IEEE International Conference on Robotics and Automation
RiverCentre, Saint Paul, Minnesota, USA
May 14-18, 2012

Pose Reconstruction of Flexible Instruments from Endoscopic Images using Markers

Rob Reilink, Stefano Stramigioli, and Sarthak Misra

***Abstract*— A system is developed that can reconstruct the pose of flexible endoscopic instruments that are used in advanced flexible endoscopes using solely the endoscopic images. Four markers are placed on the instrument, whose positions are measured in the image. These measurements are compared to a three-dimensional rendered model of the instrument. The pseudo-inverse of the interaction matrix between the state of the model and the marker positions in the image is used to update the state such that the model will track the real instrument. An experiment was performed in which the instrument was moved inside a colon model, while the tip position was simultaneously measured with an electromagnetic tracking system. The root mean square errors of the position estimation were 2.3 mm, 2.2 mm and 1.7 mm in the horizontal (x), vertical (y) and away-from-camera (z) directions, respectively.**

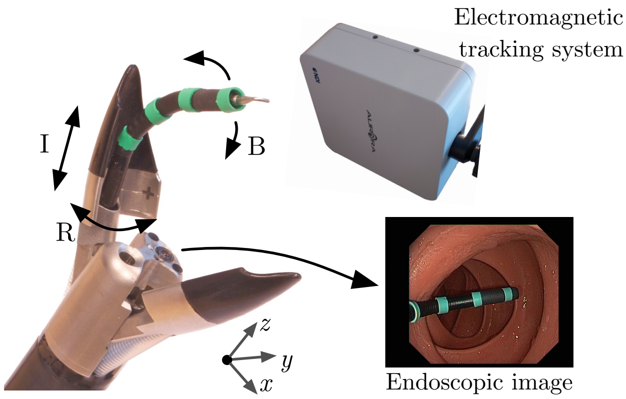

Fig. 1. The endoscopic instruments of the ANUBIS endoscope have three degrees of freedom: insertion (I,q_1), rotation (R,q_2), and bending (B,q_3). The pose of the endoscopic instrument is reconstructed from the endoscopic image. An electromagnetic tracking system is used as a reference to evaluate the performance of the pose estimation.

I. Introduction

Flexible endoscopy is a procedure in which a physician examines the internal body cavities of the patient in a minimally invasive way. Using small instruments that emerge from the endoscope tip, the physician can also execute minor interventions e.g., polyp removal or mucosal resection. Recently, Single Port Access (SPA) and Natural Orifice Transluminal Endoscopic Surgery (NOTES) have emerged as new surgical procedures which are performed using an advanced flexible endoscope [1]. These endoscopes feature multiple instruments, allowing the physician to perform complex minimally invasive interventions. They include the ANUBIS (Karl Storz GmbH & Co. KG, Tuttlingen, Germany, Fig. 1) and the EndoSAMURAI (Olympus Corp., Tokyo, Japan). However, multiple physicians are required to operate these endoscopes [2]. This is undesirable because of procedural costs, and because optimal coordination between the physicians is difficult.

In order to allow a single physician to control the endoscope and the instruments, a robotic teleoperation solution could be employed. The physician could then operate the complete system from a console similar to the daVinci system that is used for laparoscopic surgery (Intuitive Surgical, Inc., Sunnyvale, USA). This solution would require actuating the endoscope and the instruments. However, this is difficult, since significant flexibility is present between the control handle and the instrument at the tip. This causes hysteresis, making the steering of the instrument difficult [3]. Since this effect depends on the (unknown) shape of the endoscope, feed-forward compensation of this effect is only possible up to a limited extent.

Therefore, we consider a feedback approach. This requires measuring the current pose of the endoscopic instrument accurately. Adding extra sensors to the endoscope in order to measure the position and orientation of the instrument is difficult because of space constraints and sterility issues. Therefore, we aim to determine the instrument pose using solely the endoscopic camera images. In previous work, we have studied the use of endoscopic camera images to estimate the instrument position using a marker-less approach [4]. The resulting pose reconstruction algorithm was able to estimate the position of the instrument with an RMS error of 1.7 mm, 1.2 mm, and 3.6 mm in the horizontal, vertical, and away-from-camera directions, respectively. However, due to the limited contrast between the instrument and the background, a computationally intensive image processing algorithm was required to detect the instrument. The limited contrast also caused measurement inaccuracies and reduced robustness of the algorithm. Furthermore, lens-distortion was corrected by pre-processing all images, again resulting in a high computational load.

In this paper, we will describe a marker-based approach. It is hypothesized that a marker-based approach will result in a more robust and accurate estimation, since the markers are relatively easy to detect. Also, we have included the lens-distortion correction within the estimation algorithm, elimi-

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

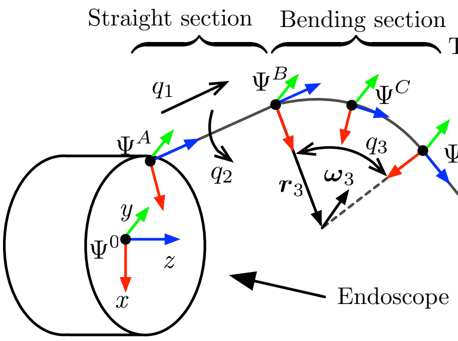

Fig. 2. The kinematic model of the endoscopic instrument describes the tip motion as a function of the states q_1, q_2, and q_3. Radius $\mathbf{r}_3$ and axis of rotation $\boldsymbol{\omega}_3$ define the curvature of the bending section.

nating the computationally intensive pre-processing step of undistorting the images. We have evaluated the algorithm in a silicone phantom colon. We have used an electromagnetic tracking system as a reference, enabling evaluation of the performance over the complete workspace.

Our approach is as follows: in order to find the instrument position and orientation, we will first find the positions of the markers in the endoscopic image. Based on these observations, the state of the kinematic model of the instrument is updated so as to match the observations as close as possible. Then, using the kinematic model, the position and orientation of the instrument tip is computed.

This paper is organized as follows: Section II describes the kinematic model of the endoscopic instrument. Section III discusses the detection of the markers in the endoscopic image and the extraction of feature points. Section IV shows how the detected feature points are used to estimate the pose of the instrument. The experimental validation of the proposed method is described in Section V. Section VI concludes and provides possible directions for future work.

II. KINEMATIC MODEL OF THE INSTRUMENT

We use a model similar to that of Bardou et al. [3] to describe the kinematics of the endoscopic instrument. The model assumes a constant curvature along the bending section. This assumption is valid as long as there are no significant external forces acting on the instrument, and there is limited friction of the cables inside the bending section, which is the case in our experiments. For application in actual surgical procedures, the effects of external forces may need to be compensated for.

The kinematic model is depicted in Fig. 2. The instrument is modeled as a translation (q_1) and a rotation (q_2) in the straight section, followed by a bending section (q_3) and a straight tip. We will use the geometric Jacobian [5] to describe the motion of the tip with respect to the endoscope camera as a function of q_1, q_2, and q_3. We will use $\mathbf{T}_l^{k,m}$ to denote the motion of frame Ψ^l with respect to Ψ^m expressed in Ψ^k. We will use $\hat{\mathbf{T}}_{l,j}$ to denote the unit twist of frame Ψ^l with respect to Ψ^0 associated with joint q_j, expressed in frame Ψ^0.

Fig. 2 shows the coordinate frame Ψ^0 which is fixed to the endoscope, located at the camera optical center, with the z-axis pointing in the camera viewing direction. Frame Ψ^A is located at the point where the instrument emerges from the endoscope, and oriented such that the z-axis points in the instrument direction. Frame Ψ^B is fixed to the end of the straight section which extends and rotates (q_1 and q_2). Frame Ψ^C is fixed half way the bending section and frame Ψ^D is fixed at the end of the bending section. The markers on the instrument are located at the tip and at the origins of Ψ^B, Ψ^C, and Ψ^D.

A. Kinematics of the straight section

The motion of the end of Ψ^B, described in terms of infinitesimal twists, with respect to the endoscope is

$$\mathbf{T}_B^{0,0} = \mathbf{T}_A^{0,0} + \mathbf{T}_B^{0,A} \ . \tag{1}$$

Ψ^A is fixed with respect to Ψ^B, hence $\mathbf{T}_A^{0,0} = \mathbf{0}$. $\mathbf{T}_B^{0,A}$ describes the motion of the straight section, induced by q_1 and q_2. Thus,

$$\mathbf{T}_B^{0,A} = \hat{\mathbf{T}}_{B,1}\dot{q}_1 + \hat{\mathbf{T}}_{B,2}\dot{q}_2 \ , \tag{2}$$

in which unit twists $\hat{\mathbf{T}}_{B,1}$ and $\hat{\mathbf{T}}_{B,2}$ represent a translation along the z-axis of Ψ^A and a rotation around the z-axis of Ψ^A, respectively

$$\hat{\mathbf{T}}_{B,1} = \mathrm{Ad}_{{}^0_A\mathbf{H}} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}} \ , \tag{3}$$

$$\hat{\mathbf{T}}_{B,2} = \mathrm{Ad}_{{}^0_A\mathbf{H}} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}} \ , \tag{4}$$

where $\mathrm{Ad}_{{}^0_A\mathbf{H}}$ denotes the Adjoint operator that changes twist coordinates from Ψ^A to Ψ^0.

From the differential kinematics, the forward kinematics can be derived using the Brockett's Product of Exponentials [5]. The pose of Ψ^A with respect to Ψ^0 is fixed by the geometry of the endoscope tip i.e., the point and direction at which the instrument leaves the endoscope. The pose of frame Ψ^B with respect to Ψ^0 is

$${}^0_B\mathbf{H} = \exp(\tilde{\hat{\mathbf{T}}}_{B,1}\ q_1 + \tilde{\hat{\mathbf{T}}}_{B,2}\ q_2)\ {}^0_B\mathbf{H}(0) \ , \tag{5}$$

where $\tilde{\hat{\mathbf{T}}}$ is the unit twist expressed as a 4×4 matrix and ${}^0_B\mathbf{H}(0) = {}^0_A\mathbf{H}$ is the initial configuration for $q_1 = q_2 = 0$.

B. Kinematics of the bending section

The bending section has a constant curvature, and can be described by a finite twist around axis $\boldsymbol{\omega}_3$. $\boldsymbol{\omega}_3$ is in the y-direction of frame Ψ^B, located at a position $\boldsymbol{r}_3$ from the origin of Ψ^B. $\boldsymbol{r}_3$ is in the x-direction of frame Ψ^B (Fig. 2). Thus, $\boldsymbol{\omega}_3 = \begin{bmatrix} 0 & \omega_3 & 0 \end{bmatrix}^{\mathrm{T}}$ and $\boldsymbol{r}_3 = \begin{bmatrix} r_3 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$, where both are expressed in frame Ψ^B. ω_3 and r_3 are related through the constant chord length of the bending section ℓ, where $\ell = \omega_3 r_3$. The length of $\boldsymbol{\omega}_3$ represents the angle of the arc which is defined as q_3. Thus, the finite twist describing the bending section is

$$\mathbf{S}_D^{B,B} = \begin{bmatrix} \boldsymbol{\omega}_3 \\ \boldsymbol{r}_3 \wedge \boldsymbol{\omega}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ q_3 \\ 0 \\ 0 \\ 0 \\ \ell \end{bmatrix} \ , \tag{6}$$

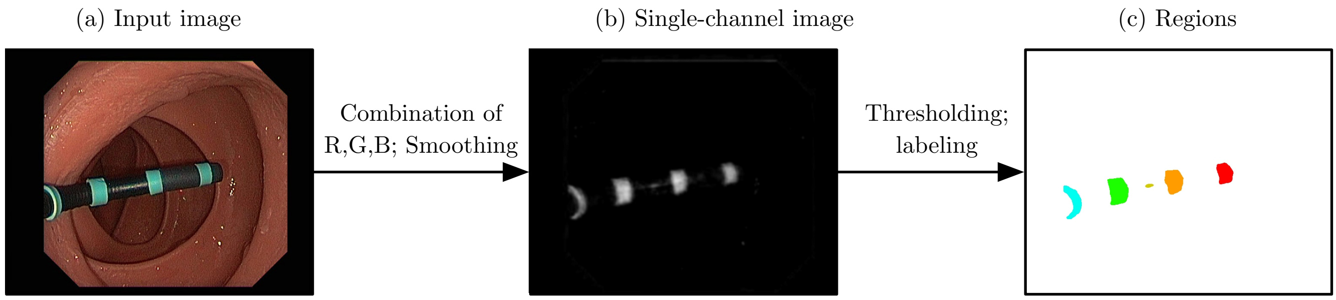

Fig. 3. The input image (a) is converted to a single-channel image and smoothed (b). This image is thresholded and labelled, resulting in regions representing the markers (c).

where $\mathbf{S}_D^{B,B}$ denotes the finite twist of frame Ψ^D with respect to frame Ψ^B, expressed in frame Ψ^B. This finite twist describes the position of frame Ψ^D with respect to frame Ψ^B.

The infinitesimal twist $\mathbf{T}_D^{B,B}$ can be derived from the finite twist $\mathbf{S}_D^{B,B}$. The infinitesimal twist (in matrix form) that describes the motion of frame Ψ^D with respect to frame Ψ^B is

$$\tilde{\mathbf{T}}_D^{B,B} := {}^B_D\dot{\mathbf{H}}\ {}^D_B\mathbf{H} \text{ , where } {}^D_B\mathbf{H} = \exp(\tilde{\mathbf{S}}_D^{B,B}) \tag{7}$$

$$= \frac{\partial\ {}^B_D\mathbf{H}}{\partial q_3}\dot{q}_3\ {}^D_B\mathbf{H} \tag{8}$$

$$= \begin{bmatrix} 0 & 0 & 1 & \frac{\ell}{{q_3}^2}(-1+\cos q_3) \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & \frac{\ell}{{q_3}^2}(q_3 - \sin q_3) \\ 0 & 0 & 0 & 0 \end{bmatrix} \dot{q}_3 \ . \tag{9}$$

Rewriting (9) in vector form, and transforming it to frame Ψ_0, we find the unit twist

$$\hat{\mathbf{T}}_{D,3} = \mathrm{Ad}_{{}^0_B\mathbf{H}} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \frac{\ell}{{q_3}^2}(-1+\cos q_3) \\ 0 \\ \frac{\ell}{{q_3}^2}(q_3 - \sin q_3) \end{bmatrix} . \tag{10}$$

The unit twist of frame Ψ^C is found similarly, substituting ℓ by $\frac{\ell}{2}$ in (10), since Ψ^C is located halfway the bending section

$$\hat{\mathbf{T}}_{C,3} = \mathrm{Ad}_{{}^0_B\mathbf{H}} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \frac{\ell}{2{q_3}^2}(-1+\cos q_3) \\ 0 \\ \frac{\ell}{2{q_3}^2}(q_3 - \sin q_3) \end{bmatrix} . \tag{11}$$

III. Marker Detection

In order to track the pose of the instrument robustly and accurately, markers were placed on the instrument (Fig. 3a). The markers were colored green, since this color shows a good contrast against the colon, which is colored red. The input to the marker tracking algorithm are the endoscopic images, which are captured via the FireWire output of the endoscope system. The output of the endoscope system is an interlaced video stream i.e., the odd and even numbered

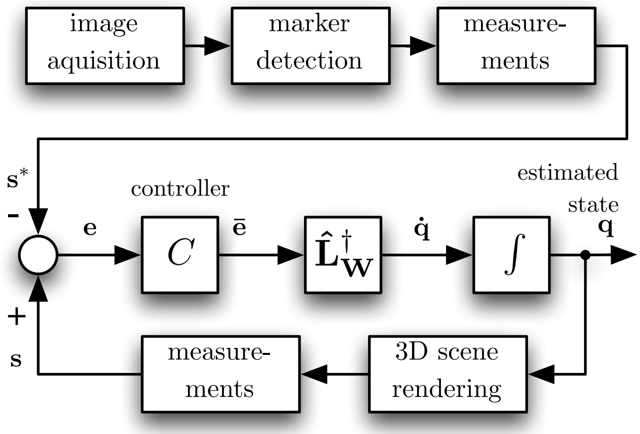

Fig. 4. Measurements $\mathbf{s}^*$ are obtained from the markers in the acquired image. Using a visual servo control loop, the estimated state $\mathbf{q}$ is updated so as to control the measurements from the 3D scene rendering $\mathbf{s}$ to match $\mathbf{s}^*$. This way, the 3D scene rendering will match the acquired image.

lines in the image are not sampled simultaneously. In order to reduce artifacts that this may cause, gstreamer [6] was used to de-interlace the images.

Subsequently, the red, green, and blue channels of the resulting color image are combined linearly into a single-channel image, and smoothed using a Gaussian filter (Fig. 3b). The weights of this linear combination are chosen using Fishers Linear Discriminant method [7] so as to get a maximum contrast between the markers and the rest of the image. The resulting image is thresholded in order to get a binary image. From the resulting binary image, the separate connected regions are labeled using the scipy ndimage module [8]. The four largest regions are considered to represent the four markers. Of each of these regions, the centroid is measured. These measurements are the inputs to the state estimation algorithm which is described in the next section.

IV. State Estimation

Fig. 4 shows a block diagram of the estimator, which implements a gradient descent optimization. The estimator is a virtual visual servo control loop [9], [10]. As such, in this case the controller does not control a physical system, but it controls a three-dimensional rendering of the scene.

The control loop has as its input a set of measurements $\mathbf{s}^*$ that are obtained from the endoscopic image. From the scene rendering, measurements $\mathbf{s}$ are obtained. The visual servo control loop controls the state $\mathbf{q}$ such, that the error $\mathbf{e}$ between $\mathbf{s}^*$ and $\mathbf{s}$ is minimized. The measurement vector $\mathbf{s}$ comprises the two-dimensional position of the centroid $\mathbf{c}_i$ of each marker region i ($i = 1 \dots 4$)

$$\mathbf{s} := \begin{bmatrix} c_{1_x} & c_{1_y} & \dots & c_{4_x} & c_{4_y} \end{bmatrix}^{\mathrm{T}} , \qquad (12)$$

where c_{i_x} and c_{i_y} denote the x- and y- coordinate of the centroid of marker region i, respectively.

A. 3D scene rendering

Due to occlusion effects, the centroid of each marker in the observed endoscopic image is not necessarily equal to the projection of its geometrical center. In order to duplicate these effects in the feed-back path of the estimator, a 3D scene rendering was implemented. The model of the endoscopic instrument is rendered using OpenGL [11]. The rendering uses a model of the endoscopic camera that was obtained using the Camera Calibration Toolbox for Matlab [12]. The lens distortion and the movement of the instrument are implemented to run on the Graphics Processing Unit (GPU), reducing the computational load on the Central Processing Unit (CPU) of the computer. From the resulting rendered scene, the measurements $\mathbf{s}$ are obtained.

B. Approximation of the interaction matrix

The relationship between the change of the error $\dot{\mathbf{e}}$ and the change of the state $\dot{\mathbf{q}}$ is given by the interaction matrix $\mathbf{L}(\mathbf{q})$ [10]

$$\dot{\mathbf{e}} = \mathbf{L}(\mathbf{q})\dot{\mathbf{q}} \ , \ \text{where } \mathbf{L}(\mathbf{q}) := \frac{\partial \mathbf{s}}{\partial \mathbf{q}} = \frac{\partial \mathbf{e}}{\partial \mathbf{q}} \qquad (13)$$

Usually, in visual servo applications, $\mathbf{L}(\mathbf{q})$ cannot be computed directly, because it depends on the (unknown) depth of objects in the scene. Because in our application, a 3D rendering is being controlled, the depth of all objects in the scene is known, and therefore $\mathbf{L}(\mathbf{q})$ could be computed accurately. However, this would be computationally intensive. Therefore, we approximate $\mathbf{L}(\mathbf{q})$ (denoted $\hat{\mathbf{L}}(\mathbf{q})$) using the following simplifications:

- For each marker, the motion of the centroid $\mathbf{c}_i$ is assumed to be equal to the motion of the projection of the center onto the image plane.
- Occlusion effects are ignored.
- Lens distortion effects are ignored.

Note that these approximations are only used when computing $\hat{\mathbf{L}}(\mathbf{q})$. In the 3D rendering of the scene the aforementioned effects are taken into account.

Next, we define $\mathbf{p}_i$ as the 3D position of the center of marker i, expressed in frame Ψ^0. The geometric Jacobian $\mathbf{J_g}$, which relates the velocities $\dot{\mathbf{p}}_i$ to the state change $\dot{\mathbf{q}}$ is

$$\begin{bmatrix} \dot{\mathbf{p}}_1 \\ \vdots \\ \dot{\mathbf{p}}_4 \end{bmatrix} = \mathbf{J_g}\dot{\mathbf{q}} \ , \ \text{where } \mathbf{J_g} := \begin{bmatrix} \frac{\partial \mathbf{p}_1}{\partial \mathbf{q}} \\ \vdots \\ \frac{\partial \mathbf{p}_4}{\partial \mathbf{q}} \end{bmatrix} . \qquad (14)$$

$\mathbf{J_g}$ can be computed using the unit twists as given in (3), (4), (10) and (11). The twist of a frame Ψ^l with respect to frame Ψ^0, expressed in frame Ψ^0 is composed of the contributions of $\dot{q}_1$, $\dot{q}_2$, and $\dot{q}_3$

$$\mathbf{T}_l^{0,0} = \hat{\mathbf{T}}_{l,1}\dot{q}_1 + \hat{\mathbf{T}}_{l,2}\dot{q}_2 + \hat{\mathbf{T}}_{l,3}\dot{q}_3 \ . \qquad (15)$$

The velocity of point $\mathbf{p}_i$, which is fixed to frame Ψ^l, expressed in frame Ψ^0 is [5]

$$\dot{\mathbf{p}}_i = \tilde{\mathbf{T}}_l^{0,0}\mathbf{p}_i \ . \qquad (16)$$

Matrix $\mathbf{J_g}$ in (14) can be obtained by combining (15) and (16) for each point $\mathbf{p}_1 \cdots \mathbf{p}_4$. Since marker 1 is fixed to frame Ψ^B, marker 2 is fixed to frame Ψ^C and markers 3 and 4 are fixed to frame Ψ^D, the resulting Jacobian matrix is

$$\mathbf{J_g} = \begin{bmatrix} \tilde{\hat{\mathbf{T}}}_{B,1}\ \mathbf{p}_1 & \tilde{\hat{\mathbf{T}}}_{B,2}\ \mathbf{p}_1 & 0 \\ \tilde{\hat{\mathbf{T}}}_{C,1}\ \mathbf{p}_2 & \tilde{\hat{\mathbf{T}}}_{C,2}\ \mathbf{p}_2 & \tilde{\hat{\mathbf{T}}}_{C,3}\ \mathbf{p}_2 \\ \tilde{\hat{\mathbf{T}}}_{D,1}\ \mathbf{p}_3 & \tilde{\hat{\mathbf{T}}}_{D,2}\ \mathbf{p}_3 & \tilde{\hat{\mathbf{T}}}_{D,3}\ \mathbf{p}_3 \\ \tilde{\hat{\mathbf{T}}}_{D,1}\ \mathbf{p}_4 & \tilde{\hat{\mathbf{T}}}_{D,2}\ \mathbf{p}_4 & \tilde{\hat{\mathbf{T}}}_{D,3}\ \mathbf{p}_4 \end{bmatrix} . \qquad (17)$$

Note that since the poses of frames Ψ^C and Ψ^D with respect to Ψ^B are independent of q_1 and q_2, $\tilde{\hat{\mathbf{T}}}_{B,1} = \tilde{\hat{\mathbf{T}}}_{C,1} = \tilde{\hat{\mathbf{T}}}_{D,1}$, and $\tilde{\hat{\mathbf{T}}}_{B,2} = \tilde{\hat{\mathbf{T}}}_{C,2} = \tilde{\hat{\mathbf{T}}}_{D,2}$.

Given the 3D motions of the markers $\dot{\mathbf{p}}_i$, the approximate observed motion of the centroid in the image plane $\mathbf{c}_i$ is given by the image Jacobian $\mathbf{J_I}$ [10]

$$\dot{\mathbf{c}}_i \approx \mathbf{J_I}(\mathbf{p}_i)\dot{\mathbf{p}}_i \ , \text{ for } \mathbf{J_I}(\mathbf{p}_i) = \frac{f}{p_z}\begin{bmatrix} 1 & 0 & -\frac{p_x}{p_z} & 0 \\ 0 & 1 & -\frac{p_y}{p_z} & 0 \end{bmatrix} , \qquad (18)$$

where f is the focal distance and p_x, p_y, and p_z are the x-, y- and z-components of $\mathbf{p}_i$, respectively. This approximation comprises the aforementioned simplifications. Combining this for all four markers yields the combined image Jacobian $\mathbf{J_C}$

$$\dot{\mathbf{s}} = \begin{bmatrix} \dot{\mathbf{c}}_1 \\ \vdots \\ \dot{\mathbf{c}}_4 \end{bmatrix} \approx \mathbf{J_C} \begin{bmatrix} \dot{\mathbf{p}}_1 \\ \vdots \\ \dot{\mathbf{p}}_4 \end{bmatrix} , \ \text{where} \qquad (19)$$

$$\mathbf{J_C} = \begin{bmatrix} \mathbf{J}_I(\mathbf{p}_1) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_I(\mathbf{p}_2) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_I(\mathbf{p}_3) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_I(\mathbf{p}_4) \end{bmatrix} \qquad (20)$$

Combining (17) and (20) yields the approximate relation between the change of the state $\dot{\mathbf{q}}$ and the change of the measurements $\dot{\mathbf{s}}$

$$\dot{\mathbf{s}} \approx \underbrace{\mathbf{J_C}\mathbf{J_g}}_{\hat{\mathbf{L}}(\mathbf{q})}\dot{\mathbf{q}} \qquad (21)$$

The resulting $\hat{\mathbf{L}}(\mathbf{q})$ is the approximation of the interaction matrix, which is used in the controller that is described in the following section.

C. Controller

Controller C (Fig. 4) implements a proportional control law

$$\bar{\mathbf{e}} = -K\mathbf{e} \ , \qquad (22)$$

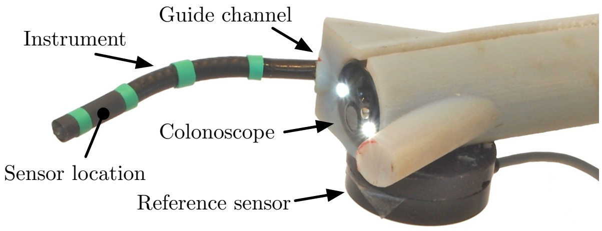

Fig. 5. An endoscope attachement was created that fits onto a conventional endoscope, and contains a guide channel that positions the instrument next to the endoscope. A 6-DOF electromagnetic reference sensor was fixed to the endoscope attachment. A 5-DOF electromagnetic position and orientation sensor was positioned in the instrument to evaluate the state estimation accuracy.

where K is a positive constant gain. $1/K$ is the time constant of the proportional controller. The resulting $\bar{\mathbf{e}}$ is the required change of the measurements $\mathbf{s}$ that will decrease the error $\mathbf{e}$. However, since the dimension of state $\mathbf{q}$ is less than the dimension of $\bar{\mathbf{e}}$, this change is in general not realizable. Therefore, we use the weighted Moore-Penrose pseudo-inverse [13]

$$\hat{\mathbf{L}}^{\dagger}_{\mathbf{W}} := \left(\hat{\mathbf{L}}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\hat{\mathbf{L}}\right)^{-1}\hat{\mathbf{L}}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W} \ , \tag{23}$$

where $\mathbf{W}$ is the weighting matrix. $\mathbf{W}$ can be chosen to influence the optimization. In the case that a perfect match between $\mathbf{s}$ and $\mathbf{s}^*$ is not realizable (due to model inaccuracies), the choice of $\mathbf{W}$ determines the resulting $\mathbf{q}$ and thus, the resulting $\mathbf{s}$. By taking

$$\dot{\mathbf{q}} = \hat{\mathbf{L}}^{\dagger}_{\mathbf{W}}\bar{\mathbf{e}} \ , \tag{24}$$

we obtain the $\dot{\mathbf{q}}$ that minimizes $||\mathbf{W}(\dot{\mathbf{e}} - \bar{\mathbf{e}})||_2$. It should be noted that the inverse in (23) cannot be computed in case the rank of $\hat{\mathbf{L}}$ is less than 3. This is the case when $q_3 = 0$ i.e., when the instrument is straight. In this case, rotation around the instrument axis (q_2) is unobservable since the instrument is circular symmetric. Thus, this situation should be avoided.

For the experiments, weighting matrix $\mathbf{W}$ was chosen to be a diagonal matrix

$$\mathbf{W} = \mathrm{diag}(1,1,1,1,1,1,4,4) \ . \tag{25}$$

This way, the error in c_{4_x} and c_{4_y} (the position error of the marker near the tip) is weighted four times more than the other markers. This is done because a position error of that marker directly results in an error in the estimated x- and y-coordinates of the tip position. It was empirically found that increasing the weighting factor for c_4 greater than 4 did not significantly influence the resulting estimation errors.

V. Experimental results

In order to evaluate the performance of the tracking and state estimation algorithms that were described in the previous sections, an experimental setup was built. A conventional colonoscope system (Exera II CV-180/CLV-180, Olympus Corp., Tokyo, Japan) was used. An endoscope attachment was created that fits onto the tip of this colonoscope (Fig. 5). This attachment adds a guide channel to the endoscope, which guides the endoscopic instrument, such that it emerges from the tip in a fixed direction. The guide channel was designed such that this direction is similar to the way the instruments are positioned in the Karl Storz Anubis endoscope.

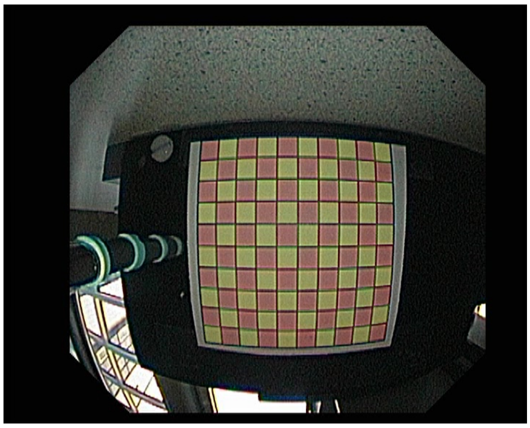

Fig. 6. **Verification of the camera model:** An image is taken of a 5 mm line grid and the camera model is used to overlay a 5 mm checkerboard. The the line grid and the checkerboard match with a maximum deviation of 2 pixels, showing that the camera model is accurate.

The camera model that is used for the rendering of the 3D scene was verified by taking an image of a 5 mm line grid, and overlaying a 5 mm checkerboard which was transformed using the camera model (Fig. 6). The line grid and the checkerboard pattern match with a maximum error of 2 pixels.

A electromagnetic tracking system (Aurora®, NDI, Waterloo, Canada) was used to track the position and orientation of the endoscopic instrument. Using this system allows evaluation of the estimator performance over the full workspace of the instrument while it is inserted in the colon model (unlike e.g. a visual tracking solution using external cameras). It also enables evaluation of the dynamic performance i.e., how well is the instrument tracked when it moves quickly. A disadvantage of the electromagnetic tracking system is that the accuracy is degraded by the metal which is present in the endoscopic instrument. The accuracy of the tracking system was found to be in the order of 2 mm.

A 5-degree-of-freedom (DOF) electromagnetic sensor was attached to the tip of the instrument as shown in Fig. 5. A 6-DOF reference sensor was fixed to the tip of the endoscope. The electromagnetic field generator was positioned approximately 300 mm above the setup. The measurement data was recorded at 40 Hz, the fixed sample rate of the tracking system. The measurements were resampled to 25 Hz using linear interpolation, in order to match the frame rate of the endoscope system.

In order to create a realistic environment, the setup was placed inside the colon of a colonoscopy model (KKM40, Kyoto Kagaku, Kyoto, Japan). The model was coated with a viscous fluid according to the manufacturers instructions. This way, the lighting conditions are similar to those in clinical images, including the specular reflections as can be seen in Fig. 3a. The endoscopic instrument was manually operated. It was moved around within the workspace, while

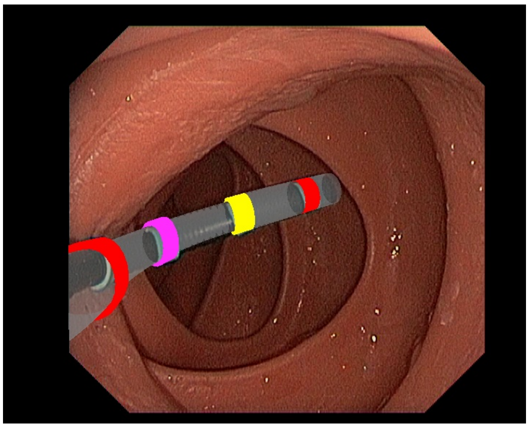

Fig. 7. The 3D rendering of the model of the instrument is overlaid on the endoscopic image. The image shows that the model is able to track the observed instrument accurately. The accompanying video demonstrates the tracking of the instrument by the model.

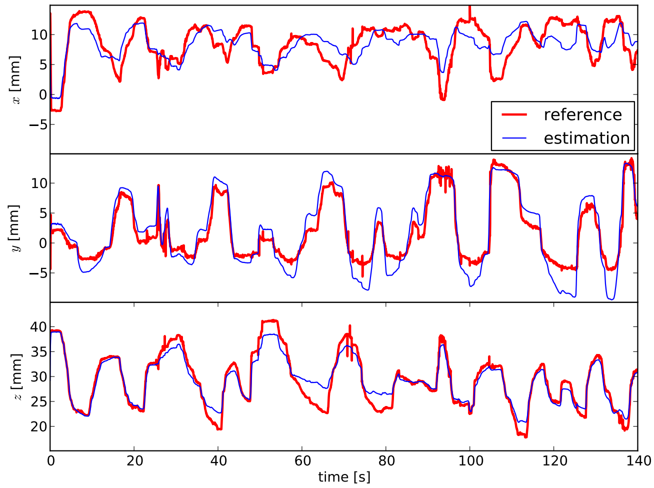

Fig. 8. The graphs show the x-, y- and z-coordinate of the estimated position of the tip based on the endoscopic images, as well as the reference position as measured by the electromagnetic tracking system.

the endoscopic camera images and the reference position given by the electromagnetic tracking system were recorded.

As shown in Fig. 7, the state estimator was able to track the markers of the instrument. The model of the instrument matches the image of the actual instrument accurately. The results of the experiments are shown in Fig. 8. The root-mean-square (RMS) error between the reference and the estimation are 2.3 mm, 2.2 mm, and 1.7 mm in the x-, y- and z-directions, respectively. This is in the same order of magnitude as the accuracy of the electromagnetic tracking system. The orientation accuracy was not evaluated.

Compared to our previous work [4], in which no markers were used, the accuracy in the z-direction is significantly improved (from 3.6 mm to 1.7 mm RMS error). The accuracy in the x- and y-directions has degraded (from 1.7 mm and 1.2 mm to 2.3 mm and 2.2 mm, respectively), due to the limited accuracy of the reference measurements as described above.

VI. Conclusion and Future Work

An estimator has been designed that uses endoscopic images to estimate the tip pose of a 3-DOF endoscopic instrument. An experiment was performed in which the endoscopic instrument was operated in a colon model. The estimated tip position was compared against reference measurements performed using an electromagnetic tracking system. The RMS errors were 2.3 mm, 2.2 mm, and 1.6 mm in the x-, y- and z-directions, respectively. Part of these errors are due to errors in the reference measurements, most likely caused by the interference of the metal parts of the instruments on the electromagnetic tracking system. By using the markers on the instrument, the estimation error in the z-direction was reduced from 3.6 mm to 1.6 mm.

For future work, we will evaluate the designed estimator using a more accurate reference measurement. An X-ray system will be used to create a top-view of the instrument, thus allowing us to measure the distance from the instrument tip to the camera (the z-direction) accurately.

Finally, our goal is to use the developed estimator to control the position of the endoscopic instrument accurately. The estimated tip pose will be used as the feedback to the controller. Using this control system, a physician could control the instrument using a multi-DOF input device in an intuitive way, enabling easier and faster execution of the endoscopic procedure.

References

[1] A. N. 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] J. 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] B. 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.

[4] R. Reilink, *et al.*, "Three-dimensional pose reconstruction of flexible instruments from endoscopic images," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, San Francisco, USA, 2011, pp. 2076–2082.

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

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

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

[8] "SciPy: scientific tools for Python." [Online]. Available: www.scipy.org

[9] É. Marchand and F. Chaumette, "Virtual visual servoing: a framework for real-time augmented reality," *Eurographics*, vol. 21, no. 3, 2002.

[10] F. Chaumette and S. Hutchinson, "Visual servo control. I. basic approaches," *Robotics Automation Magazine, IEEE*, vol. 13, no. 4, pp. 82 –90, dec. 2006.

[11] "OpenGL: The industry standard for high performance graphics." [Online]. Available: www.opengl.org

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

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