

Segmentation and Guidance of Multiple Rigid Objects for Intra-operative Endoscopic Vision

C. Doignon, F. Nageotte, and M. de Mathelin

Control, Vision and Robotic Group, LSIIT (UMR ULP-CNRS 7005)
University Louis Pasteur of Strasbourg, Pole API, Bd. Brant, 67412 Illkirch, France
{doignon,nageotte,demathelin}@lsiit.u-strasbg.fr

Abstract. This paper presents an endoscopic vision framework for model-based 3D guidance of surgical instruments used in robotized laparoscopic surgery. In order to develop such a system, a variety of challenging segmentation, tracking and reconstruction problems must be solved. With this minimally invasive surgical technique, every single instrument has to pass through an insertion point in the abdominal wall and is mounted on the end-effector of a surgical robot which can be controlled by automatic visual feedback. The motion of any laparoscopic instrument is then constrained and the goal of the automated task is to safely bring instruments at desired locations while avoiding undesirable contact with internal organs. For this "eye-to-hands" configuration with a stationary camera, most control strategies require the knowledge of the out-of-field of view insertion points location and we demonstrate it can be achieved *in vivo* thanks to a sequence of (instrument) motions without markers and without the need of an external measurement device. In so doing, we firstly present a real-time region-based color segmentation which integrates this motion constraint to initiate the search for region seeds. Secondly, a novel pose algorithm for the wide class of cylindrical-shaped instruments is developed which can handle partial occlusions as it is often the case in the abdominal cavity. The foreseen application is a good training ground to evaluate the robustness of segmentation algorithms and positioning techniques since main difficulties came from the scene understanding and its dynamical variations. Experiments in the lab and in real surgical conditions have been conducted. The experimental validation is demonstrated through the 3D positioning of instruments' axes (4 DOFs) which must lead to motionless insertion points disturbed by the breathing motion.

1 Introduction

One may observe since few years a growing spectrum of computer vision applications to surgery, particularly to intra-operative guidance [1, 2]. On the one hand computer vision techniques bring a lot of improvements and gain in reliability in the use of visual information, on the other hand medical robots provide a significant help in surgery, particularly for the minimally invasive surgery, as it is for the laparoscopic surgery. Minimally invasive surgery is a very attractive technique since it provides position accuracy, it avoids surgical opening and then it reduces the recovery time for the patient. In counterpart, motions of surgical instruments are constrained to by the insertion point locations in the abdominal wall, reducing the dexterity since only four degrees of freedom are available.

Our research in this field aims at expanding the potentialities of such robotic systems by developing visual tracking and servoing techniques to realize semi-autonomous tasks [3, 4]. Endoscopic vision systems are used for that purpose, however many obstacles remain to be overcome to achieve an accurate positioning of laparoscopic instruments inside the abdominal cavity by visual feedback. Many difficulties are emanating from the scene understanding, the time-varying lighting conditions, the presence of specularities and bloodstained parts, and a non-uniform and moving background due to patient breathing and heart beating. But, for this "eye-to-hands" robotic vision system, one of the most tricky problem is the unknown relative position/orientation of robot arms holding the instruments w.r.t. the camera frame [3]. This transformation mainly depends on the insertion points location which must be recovered to express the relative velocity screw in the appropriate frame.

The outline of the paper is as follows. In the next section, we review some existing endoscopic vision systems used in robotized laparoscopy. In section three, we describe the fast region-based color segmentation of surgical instruments. We present the laparoscopic kinematic constraint together with the 3D pose estimation of surgical instruments in section four. Throughout the paper, results are provided and a conclusion is given in section five.

2 Related Work on Vision-based Robotic Guidance for Minimally Invasive Abdominal Surgery

Prior researches have been conducted to process laparoscopic images for the development of 3D navigation systems in the human body. One of the pioneered work was that of Casals *et al.* [5] which used a TV camera microoptics mounted on a 4 DOFs industrial robot (with 2 passive joints) to realize a 2D tracking of a surgical instrument with markers. Projections of markers were approximated by straight lines in the image segmentation process and the tracking task was to keep the imaged markers close to the image center. This guidance system worked at a sampling rate of 5 Hz with the aid of an assistant. Wei *et al.* [1] have used a stereoscopic laparoscope mounted on a robot arm and have designed a color

marker to realize a tracking task. By means of a color histogram, the color bin with the lowest value is selected to mark the instrument. This spectral mark was then utilized to control the robot motion at a sampling rate of 15 Hz. An interesting feature of the proposed technique is the choice of HSV color space for segmentation, leading to a good robustness with respect to lighting variations. Wang *et al.* [6] have proposed to enhance laparoscope manoeuvring capabilities. In so doing, they have conceived a general framework that uses visual modelling and servoing methods to assist the surgeon in manipulating a laparoscope mounted on a robot end-effector. Color signatures are used in a Bayesian classifier to segment endoscopic images into two classes (organ and markerless instrument). Finally, this framework has been applied to the instrument localization (the 2D position of the imaged tip of instrument) and 2D tracking with 3 DOFs of the AESOP robot in a way to follow the laparoscope. Like for the two previous related works, it's a visual tracking system with active vision guidance in order to keep the instrument close to the image center, that is there is no need to the estimate of the 3D motion of the instrument.

For these related works, it is assumed that the endoscopic camera is mounted on a robot (eye-in-hand). Other more recent works are rather related to the tracking of free-hand or robotized instruments with respect to the internal organs with the aid of a stationary camera. Hayashibe *et al.* [7] have designed an active scanning system with structured lighting for the reconstruction of 3D intraoperative local geometry of pointed organs. With a 2D galvano scanner and two cameras (one of the two is a high speed one), a real-time registration of the scene of interest is performed via the triangulation principle in order to alleviate the surgeon to mentally estimating the depth. An external device equipped with leds (the Optotrak system from Northern Digital Inc.) was used to calibrate the laser and the cameras coordinate frames. The authors have reported a total measuring time of 0.5 s to provide the 3D geometry of the liver under laparoscopic surgery conditions and have realized non-master-slave operation for the AESOP surgical robot guided by the surgeon.

A robot vision system that automatically positions a single laparoscopic instrument with a stationary camera is described by Krupa *et al.* [3]. Laser pointers are designed to emit markers on the organ. A visual servoing algorithm is carried out to position a marked instrument by combining pixel coordinates of the laser spots and the estimated distance between the pointed organ surface and the tip of the instrument thanks to the projective invariance of the cross-ratio. Successful experiments using this system were done on living pigs. In this work, 3 DOFs of the instrument were tracked (pan/tilt/penetration depth) thanks to a two-stage visual servoing scheme that partly decouple the control of the pointed direction (given in the image) and the control of the depth. It is worth noticing that a on-line identification of the Jacobian matrix for pan/tilt control (first stage) was realized with appropriate robot joint motions to directly get expressions of the velocity screw components in the instrument frame. At the Center for Computer Integrated Surgical Systems and Technology (CISST), several techniques for assisting surgeons in manipulating the 3D space within the

human body have been developed not only for the abdominal cavity but also for eye, sinus and thoracic surgery. Some of them involve (mono- and stereo-) vision-based robot control and articulated instruments [2] and in order to obtain the robot(fixed frame)-to-camera transformation, the Optotrak system is used in a preliminary setup. Burschka *et al.* have noticed an offset of approximately 5 mm (compared to the stereovision tracking) which is due to an error in the cameras-Optotrak calibration because of the difficulty of segmenting led centers.

Our objectives are to bring solutions of the previously mentioned problems in this complex environment including dynamical changes, with landmark-free approaches. No previous work is directly related to the 3D location recovery of insertion points with respect to the endoscopic camera. However, some solutions have been provided by Krupa *et al.* [3] and also by Ortmaier *et al.* [8] but with respect to the robot frame, which inherently introduces errors of the robot model. Moreover, these methods need markers on the instruments. Robotic tasks may require interactions with tissues, instruments must be autoclavable before a surgical operation and since several one may alternatively be used (depending on the subtask addressed), it is not convenient to always use artificial landmarks placed on endoscopic tools. In this paper, techniques related to image processing and computer vision have been specially designed so as to be dedicated to the interpretation of visual data coming from the abdominal cavity for robotic purposes. In particular, we investigate the on-line localization recovery of the out-of-field of view insertion points in the abdominal wall which is useful for image regions classification and for the temporal consistency of instruments motion.

3 Segmentation inside the Abdominal Cavity

For applications involving robots, image segmentation as well as classification and recognition must be fast and fully automatized. Moreover, since we deal with color images, it's suitable to analyze the multispectral aspect of the information to identify regions of interest. In laparoscopic surgery, many surgical instruments have cylindrical metallic parts leading to grey regions with many specularities in the image. In [9], the detection of a single laparoscopic instrument has achieved by means of the Hough transform but it requires the knowledge of the 3D position of insertion point while in Doignon *et al.* [10], we addressed the detection of boundaries of grey regions in color endoscopic images accounting for laparoscopic instruments. It was based on a recursive thresholding of histograms of color purity attribute S (saturation) and it works at half the video rate. The color image segmentation we designed here is based on chromatic HS (Hue-Saturation) color attributes when HSI is chosen as the color space representation. The joint color feature $\mathcal{H} = S H$ from which the first derivative is closely related to the shadow-shading-specular quasi-invariant $|\mathbf{H}_x^c| = S \cdot H_x$ [11] seems to be an appropriate discriminant cue and is shown in Fig. 1 (right). H_x denotes the spatial differentiation of hue H (a change of \mathcal{H} may also occur with a change of the color purity S). A well-known drawback of hue is its

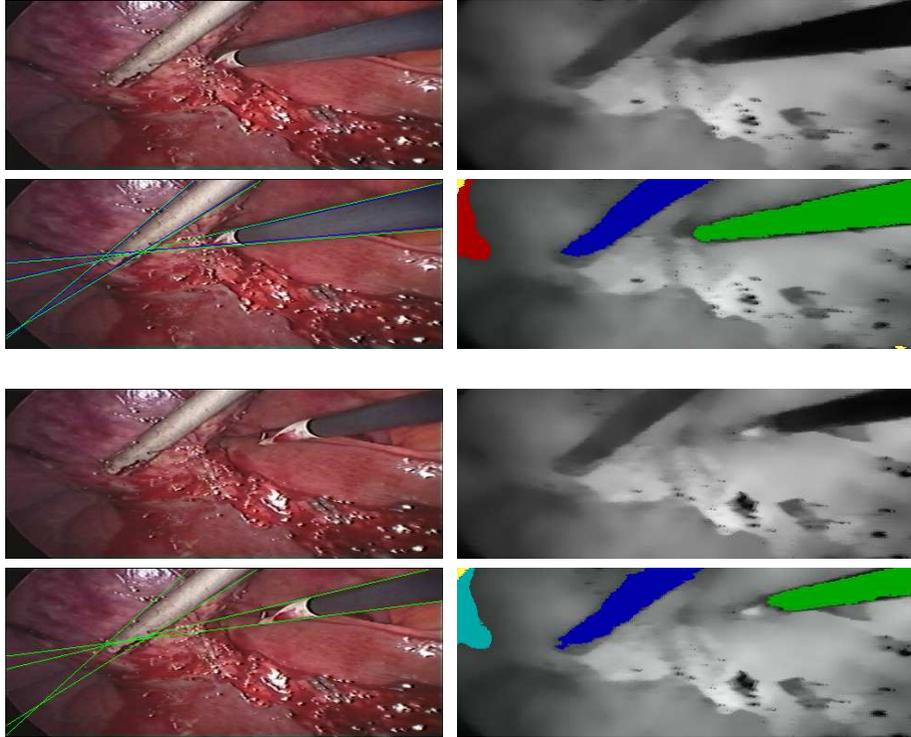


Fig. 1. The results of the region-based color (hue-saturation) segmentation for frames 74 (the 4 top images) and 578 (bottom). In right, the \mathcal{H} (filtered) images and the selected (coloured) image regions. The apparent contour of instruments is delineated with a pair of straight lines (in green).

undefinedness for achromatic pixels, i.e., for small S and small changes round the grey axis result in large changes of the direction of that quasi-invariant and therefore the derivative of hue is unbounded. However, van de Weijer *et al* [11] have shown that the norm of \mathbf{H}_x^c remains bounded. It follows that its integral is also bounded, and hence, \mathcal{H} is bounded. As noticed by van de Weijer *et al*, the discriminance of this quasi-invariant of photometric color feature is efficient and suitable to deal with specularities. To get out an oversegmentation, a fast Sigma filter algorithm has been performed on the \mathcal{H} image. This is a non-linear filtering which has the capability to either smooth pixel attributes inside region and to equally preserve the topological properties of edges. Results are very similar to the well-known anisotropic diffusion process [12]. However, it is very fast and in [13], we have presented the real-time implementation of this filtering.

We have followed a region-based segmentation approach, and, since any instrument is constrained to pass through the insertion point, the automatic detection of seeds to initiate the region growing process is reduced to a one-dimensional

search of low \mathcal{H} values along the image boundaries. Once regions have been segmented, the region boundaries are ordered and used to perform a robust two-class line fitting. It first consists in a contour classification algorithm which determines the farthest edge from the seed in the list of boundaries as a discriminant class separator. Then, a least-median of squares method is carried out to each class for modelling the apparent contour with a pair of line parameters, \mathbf{I}^+ and \mathbf{I}^- , or to reject the region if the euclidean distance between pixels and corresponding line is too large (see for example the red and lightblue labeled regions in Fig. 1). Nevertheless, it is yet possible that a region which does not correspond to an instrument may be selected with this method. Then, we will see in the next section how the motion constraint can help to solve this problem.

4 Model-based Pose Approach with Motion Constraint

The aim of this section is to formalize the motion constraint. First of all, a scene structure from motion approach is developed to get the location of the insertion points. For this purpose, a two-step algorithm with closed-form solution of the pose parameters is presented.

4.1 The Motion Constraint in Minimally Invasive Surgery

As previously mentioned, any laparoscopic instrument is constrained to pass through the incision point. Usually, the structure of the scene from motion involves multiple views and the well-known factorization method exploits geometric constraints between views acquired by one or several cameras, in motion (e.g. see [14–17]). In opposite, the main feature of the multiview approach presented here is that it properly exploits existing motion constraints of the robotized instruments observed by a stationary camera.

At a first approximation, let consider the patient breathing being no impact on the abdominal wall deformation, that is any insertion point is assumed to be motionless. We denote with $(R_c) = (C, \mathbf{x}^c, \mathbf{y}^c, \mathbf{z}^c)$ the reference frame attached to the camera with projection centre C , $(R_I) = (O^I, \mathbf{x}^I, \mathbf{y}^I, \mathbf{z}^I)$ the reference frame attached to a laparoscopic instrument with an arbitrary origin O_I . Without loss of generality, we assume vector \mathbf{z}^I with the same orientation as the instrument axis. The small incision area in the abdominal wall for an instrument is represented with a geometrical point I and that of the endoscope with the geometrical point E . Under these assumptions and with these notations, the position vector \mathbf{EI} is constant, and for a stationary camera, vector \mathbf{CI} is also constant. If the position and orientation of the instrument frame (R_I) are respectively the vector \mathbf{t} and the rotation matrix $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ expressed in the camera frame (R_c) , it comes:

$$\mathbf{CI} = \mathbf{t} + \mathbf{R} \mathbf{O}^I \mathbf{I} = \mathbf{t} + \lambda \mathbf{R} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\top = \mathbf{t} + \lambda \mathbf{r}_3, \quad \lambda \in \mathbb{R} \quad (1)$$

Since most instruments exhibit a surface of revolution (SOR), with few exceptions, the attitude of the axis of revolution may conveniently be represented with

the Plücker coordinates as it is for any 3D straight line. Plücker coordinates are a couple of algebraically dependent vectors (\mathbf{v}, \mathbf{w}) such that $\mathbf{w} = \mathbf{v} \times \mathbf{t}$. They may alternatively be gathered in the following matrix \mathbf{L} or its dual \mathbf{L}^* :

$$\mathbf{L} = \begin{bmatrix} [\mathbf{w}]_{\times} & -\mathbf{v} \\ \mathbf{v}^T & 0 \end{bmatrix}, \quad \mathbf{L}^* = \begin{bmatrix} [\mathbf{v}]_{\times} & -\mathbf{w} \\ \mathbf{w}^T & 0 \end{bmatrix}. \quad (2)$$

This is a suitable representation since one may easily deal with geometrical transformations [18] including the perspective projection [19]. This (4×4) matrix is defined up to a scale, skew-symmetric, singular and the rank value (2) is expressing the orthogonality constraint between the two vectors \mathbf{v} and \mathbf{w} . With this representation, the laparoscopic kinematic constraint may be expressed for $\mathbf{v} = \mathbf{r}_3$ as the common intersection of multiple convergent lines. Since any (homogeneous) point \mathbf{X} is on \mathbf{L} if $\mathbf{L}^* \mathbf{X} = 0$, given n displacements $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$ corresponding to the set of dual Plücker matrices $\{\mathbf{L}_1^*, \mathbf{L}_2^*, \dots, \mathbf{L}_n^*\}$, a unique intersection of lines is obtained with a rank-3 $(4n \times 4)$ matrix \mathbf{G}_n^T such that

$$\mathbf{G}_n = [\mathbf{L}_1^*, \mathbf{L}_2^*, \dots, \mathbf{L}_n^*]. \quad (3)$$

That is, the null-space of \mathbf{G}_n^T must be a one-dimensional subspace and the intersection may be computed with n ($n \geq 2$) 3D displacements of the instrument. By computing the SVD of \mathbf{G}_n^T , one obtains the common intersection by taking the singular vector associated with the null singular value (or the smallest one in presence of noisy data). The sign ambiguity of the solution is dispelled as the only valid one is corresponding to an intersection $I = (I_x, I_y, I_z)$ occurring in front of the camera ($I_z > 0$).

The perspective projection of the 3D line L_j is the image line \mathbf{l}_j defined by

$$[\mathbf{l}_j]_{\times} = \mathbf{K}^c \mathbf{P}^c L_j (\mathbf{K}^c \mathbf{P}^c)^T = [(\mathbf{K}^c)^{-T} \mathbf{w}_j]_{\times} \quad (4)$$

where \mathbf{K}^c is the matrix of camera parameters, \mathbf{P}^c is the (3×4) projection matrix and $[\mathbf{l}]_{\times}$ is the skew-symmetric matrix of vector \mathbf{l} . Since the intersection is preserved by projective transformation, the n corresponding convergent image lines $\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_n$ must satisfy

$$(\mathbf{l}_1 \ \mathbf{l}_2 \ \dots \ \mathbf{l}_n)^T i = \underbrace{(\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_n)^T}_{\mathbf{W}_n} (\mathbf{K}^c)^{-1} i = \mathbf{0} \quad (5)$$

where i is the image of the insertion point I . It follows that a set of n 3D straight lines is projecting to n convergent image lines if the above $(n \times 3)$ matrix \mathbf{W}_n is of rank 2. It's only a necessary condition which does not ensure the convergence of the 3D lines, but which makes so important the accurate estimation of the imaged axis of revolution (any line \mathbf{l}_j) which requires the recovery of the Plücker coordinates presented in the next paragraph. Once the pose estimation is done with the measurements $(\mathbf{I}_p^-, \mathbf{I}_p^+)$ of a putative image region p , the following criterion is used as a discriminant classification parameter

$$\min_j |\mathbf{I}_p^T i_j| < \tau, \text{ for } j = 1, \dots, m \quad (6)$$

to attach the region to one of the m insertion points, otherwise it is rejected.

4.2 Pose computation of a right circular cylinder

We present a novel algorithm for the pose estimation of a cylinder. As a close related work, Wong *et al.* [20] exploit the invariance of surfaces of revolution (SOR) to harmonic homology and have proposed an algorithm which is able to recover the orientation and the depth (or the focal length of the lens) while an image rectification is performed to coincide the imaged revolution axis of a SOR with one image axis and when the image of a latitude circle is available (assuming that the principal point is located at the image center and that the camera has unit aspect ratio) from the resulting silhouette which exhibits a bilateral symmetry after a rectification which brings the revolution axis to coincide with one image axis. With this method, an initial guess of the imaged symmetry axis is found by numerical minimization of a cost function. and if the image of a latitude circle in the SOR is also available, the depth can be estimated. The method we propose here is especially designed for cylindrical objects. It's a direct method (all components are computed in one stage), it does not need any image transformation and no latitude circle, hence it can deal with partial occlusion of the apparent contour as it is for this application area.

Given the matrix K^c , the cylinder radius r_c and the image of its contour generator (the apparent contour), we look for the determination of the Plücker coordinates (\mathbf{r}, \mathbf{w}) of the cylinder's rotation axis satisfying the non-linear equation $\mathbf{r}^T \mathbf{w} = 0$. It can be easily shown (from [21]) that the apparent contour is a set of two straight lines represented with the pair of vectors \mathbf{I}^- and \mathbf{I}^+ satisfying

$$\begin{aligned} (\mathbf{I}^-)^T m &\equiv \{(K^c)^{-T} (\mathbf{I} - \alpha[\mathbf{r}]_{\times}) \mathbf{w}\}^T m = 0 \\ (\mathbf{I}^+)^T m &\equiv \{(K^c)^{-T} (\mathbf{I} + \alpha[\mathbf{r}]_{\times}) \mathbf{w}\}^T m = 0 \quad , \end{aligned} \quad (7)$$

for any point m lying on the apparent contour and $\alpha = r_c / \sqrt{\|\mathbf{w}\|^2 - r_c^2}$.

To compute the pose parameters, we define the three vectors $\mathbf{y} = \alpha[\mathbf{r}]_{\times} \mathbf{w}$, $\boldsymbol{\rho}^- = K^c \mathbf{I}^-$ and $\boldsymbol{\rho}^+ = K^c \mathbf{I}^+$. With these notations, (7) can be written as follows

$$\mu_1 \boldsymbol{\rho}^- = \mathbf{w} - \mathbf{y} \quad ; \quad \mu_2 \boldsymbol{\rho}^+ = \mathbf{w} + \mathbf{y} \quad (8)$$

where μ_1 and μ_2 are two non-null scale factors. Vectors \mathbf{y} and \mathbf{w} are algebraically dependent (but not linearly) since they satisfy $\mathbf{y}^T \mathbf{w} = 0$ and $\|\mathbf{y}\| = |\alpha| \|\mathbf{w}\|$. The latter one is developed so as to take into account the expression for α

$$r_c^2 (\|\mathbf{w}\|^2 + \|\mathbf{y}\|^2) = \|\mathbf{w}\|^2 \|\mathbf{y}\|^2 \quad (9)$$

To summarize, what we have to do is to solve the following homogeneous deficient-rank system

$$\begin{bmatrix} -\mathbf{I} & \mathbf{I} & -\boldsymbol{\rho}^- & \mathbf{0} \\ \mathbf{I} & \mathbf{I} & \mathbf{0} & -\boldsymbol{\rho}^+ \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{w} \\ \mu_1 \\ \mu_2 \end{bmatrix} = \mathbf{A}_{6 \times 8} \mathbf{x} = \mathbf{0} \quad (10)$$

for the unknown vector $\mathbf{x} = (\mathbf{y}^T, \mathbf{w}^T, \mu_1, \mu_2)^T$, subject to $\mathbf{y}^T \mathbf{w} = 0$ and (9). Since $\mathbf{A}_{6 \times 8}$ has a rank equal to 6, the SVD $\mathbf{U}_{6 \times 8} \mathbf{D} (\mathbf{v}_1, \dots, \mathbf{v}_8)^T$ has two null singular

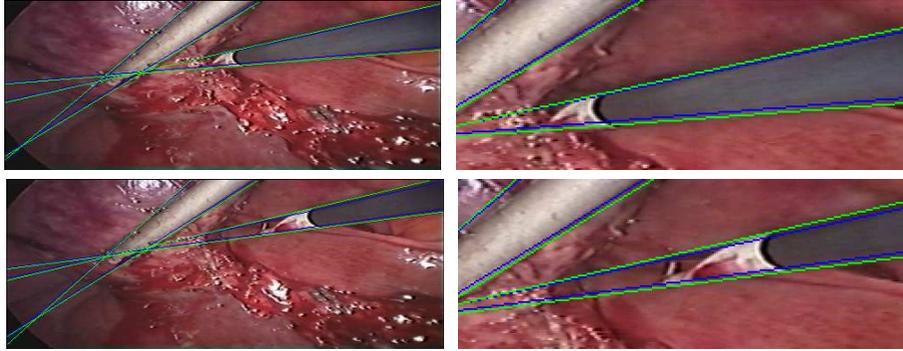


Fig. 2. Results of the pose for two frames picked up from the sequence. Blue curves are the perspective projections of the contour generator of the cylindrical-shaped instrument with the estimated pose, whereas blue ones are those corresponding to the two-class fitting of the apparent contour. (right) magnification of left images.

values and the null-space of $\mathbf{A}_{6 \times 8}$ is spanned by the right singular vectors \mathbf{v}_7 and \mathbf{v}_8 and provides a 2-parameter family of solutions as a linear combination of the two last columns of \mathbf{V} as

$$\mathbf{x} = \lambda \mathbf{v}_7 + \tau \mathbf{v}_8, \quad \text{for } \lambda, \tau \in \mathbb{R}. \quad (11)$$

The second step consists in the introduction of non-linear constraints. Substituting $\mathbf{y} = (x_1, x_2, x_3)^\top$ and $\mathbf{w} = (x_4, x_5, x_6)^\top$ from (11) in $\mathbf{y}^\top \mathbf{w} = 0$ gives the following homogeneous quadratic equation in λ and τ

$$a_1 \lambda^2 + a_2 \lambda \tau + a_3 \tau^2 = 0 \quad (12)$$

where a_i are scalar functions of \mathbf{v}_7 and \mathbf{v}_8 . Two real solutions for $s = \tau/\lambda$, s^- and s^+ , can be computed from (12). Then, reporting these solutions in (9) with substitutions from (11) gives an homogeneous quadratic equation in τ^2 :

$$c_1(s) \tau^2 + c_2(s) \tau^4 = 0 \quad (13)$$

and the solutions are $\tau = 0$ (double) and $\tau = \pm \sqrt{-\frac{c_1(s)}{c_2(s)}}$. The two null solutions for τ are those corresponding to the trivial solution $\mathbf{x} = 0$ since $\mathbf{y}^\top \mathbf{w} = 0$ and (9) are both satisfied with null vectors. Moreover, the sign of the non-null solutions for τ can not be determined since both \mathbf{x} and $-\mathbf{x}$ are solutions. As one can notice, since $\tau = s \lambda$ with $s^- = -1/s^+$, the solution for the pair of vectors (\mathbf{y}, \mathbf{w}) with s^+ is also the solution for the pair of vectors $(-\mathbf{w}, -\mathbf{y})$ with s^- .

4.3 Experimental results

Results concerning the pose are shown (sketched) in Fig. 2. In this figure, blue curves are the perspective projections of the contour generator of the cylindrical-

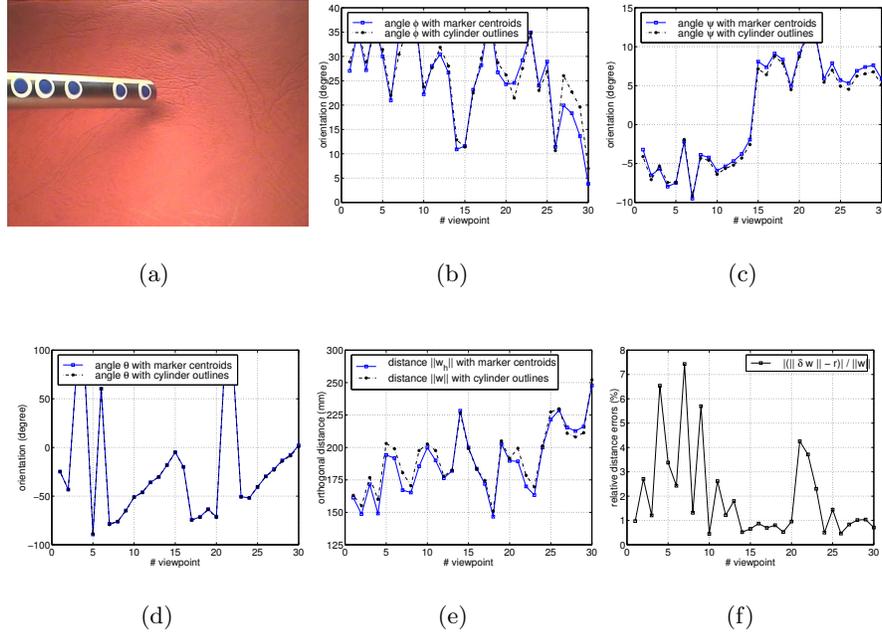


Fig. 3. (a) Image of the laparoscope with blue markers. (b-f) Comparison of the markers-based Haralick’s method and the method based on apparent contours of a right circular cylinder for the 4 DOFs: angles (b-d) and orthogonal distances (e). Whereas the orientation of the cylinder should be equal with and without markers, the norm of the vector $\mathbf{w} - \mathbf{w}_h$ must be equal to the radius of the cylinder $r_c = 5$ mm (f).

shaped instrument with the estimated pose, whereas blue ones are those corresponding to the two-class fitting of the apparent contour. With the proposed method, the curves should be perfectly superimposed, however the small residual error (1.2 pixels in average) is probably due to a mis-identification of lens distortion parameters. A cylindrical laparoscope with blue markers stucked on its surface has been used for primary experiments. Centroids of these markers are such that we get a set of 5 collinear object points in the axis direction. A set of endoscopic images has been captured with 30 viewpoints (see Fig. 3-a). With this equipment, we have compared the pose computation from apparent contours of the cylinder (\mathbf{r}, \mathbf{w}) with the proposed method and the Haralick’s method for the pose of a set of collinear points [22]. The latter method determines the orientation \mathbf{r}_h of the straight line supporting the points as well as a position vector \mathbf{t}_h (given the interpoint distances and an arbitrary origin for the points reference). We then compute the following cross-product $\mathbf{w}_h = \mathbf{r}_h \times \mathbf{t}_h$ to get the Plücker coordinates. Due to the relative position of these markers w.r.t. the cylinder axis, vectors \mathbf{r} and \mathbf{r}_h should coincide whereas the Euclidean

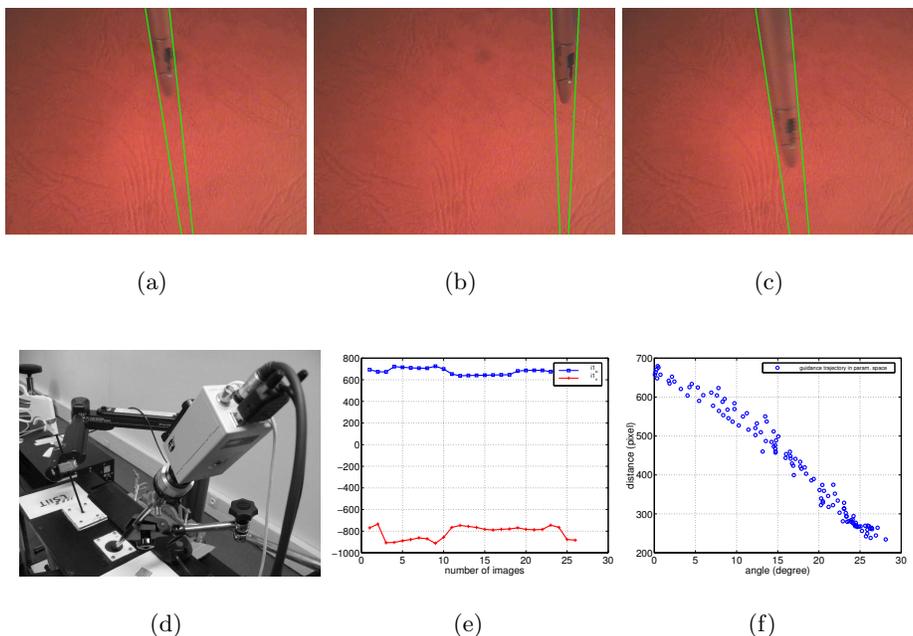


Fig. 4. Experiments in the lab to validate the proposed method. (a-c) Three endoscopic images with the segmentation of a single surgical instrument. The image lines resulting from the two-class fitting of the apparent contours are drawn in green. (d) A training box is used together with the endoscope fixed onto a monoCCD camera. The instrument is mounted onto the end-effector of the AESOP3000 surgical robot. (e) Temporal variations of i_1 coordinates in the image plane while moving the surgical instrument in front of the camera. (f) The dual parameter space of convergent lines (θ, ρ) (imaged instrument axis), "points" (blue bullet) must be collinear with a perfectly motionless insertion point.

norm of vector $\delta\mathbf{w} = \mathbf{w} - \mathbf{w}_h$ should be equal to the cylinder radius $r_c = 5$ mm whatever is the camera viewpoint. This experimental validation is depicted in Fig. 3-b:c for the orientation (angles ϕ and ψ) of the rotation axis, in Fig. 3-d for the inclination of the interpretation plane w.r.t. to the optical axis (angle θ) and in Fig. 3-e for the orthogonal distances w.r.t. camera centre. The results show a good agreement and consistency for the orientation of the instrument axis. However, results about relative distance error are not as good as expected. This error is 3.1 % in average, but for several viewpoints, there are significant differences (up to 7.6 %) between $\|\mathbf{w} - \mathbf{w}_h\|$ and the cylinder radius (Fig. 3-f). With a training box at the lab and a motionless insertion point (I_1), displacements and pose estimation of a surgical instrument has carried out (see Fig. 4) with the AESOP surgical robot. During the guidance of the instrument, we no-

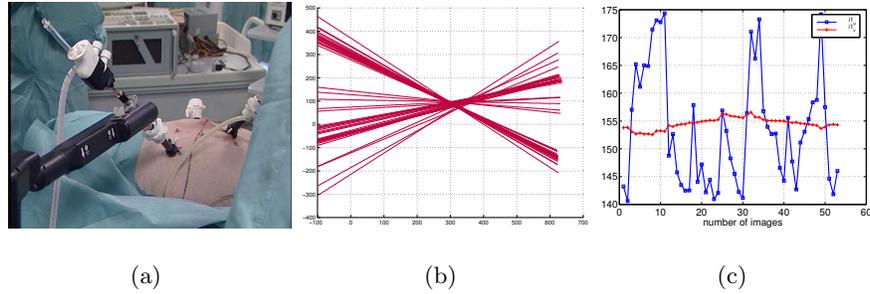


Fig. 5. (a) The Aesop surgical robot in the operating room. Trocars are inserted to incision points to guide the laparoscopic instruments or to hold the stationary camera. (b) The (I_x, I_y) coordinates of the convergent point I_1 during the guidance of an instrument. (c) Temporal variations of the perspective projection of I_1 (i_1) as the intersection of imaged symmetric axes $\mathbf{1}$ for a sequence of 52 images.

ticed some small temporal variations of the image (i_1) of the insertion point due to error in the overall segmentation (Fig. 4-e) and pose estimation. In Fig. 4-f, we have reported the dual parameter space of convergent lines (distance from the origin versus angle of line direction), since a unique intersection of lines must lead to perfectly collinear points (blue bullets).

We have depicted in Fig. 5-a the experimental setup used in the operating room and we have also reported the first two coordinates of the first insertion point $I_1 = (304; 88; 224)$ found with the proposed method in Fig. 5-b. The precision of the imaged point $i_1 = (157.5; 154.2)$ (Fig. 5-c) is given by the standard deviations which are $\sigma_u = 10.4$ and $\sigma_v = 1.2$ pixels respectively in the horizontal and vertical directions, and with 52 images (about 2 s). Results exhibit a significantly better precision found in the vertical direction. This can be explained either by the breathing motion or by a no sufficient spread of orientation motions in one direction while the robot is guiding the instrument. Another experimentation has been done to validate the convergence of the imaged instrument axes of cylindrical instruments. Fig. 6 shows the location of the insertion point location in the image with the least mean square method (Fig. 6-a:b) and with a robust (least median of squares) estimation method (Fig. 6-c:d). The latter method is able to cope with outliers, that is it keep only the salient endoscopic views with the more accurate 3D pose estimations.

5 Conclusion

In this paper, we have tackled a set of problems to solve for the 3D guidance of surgical instruments in minimally invasive surgery inside the abdomen. For this complex environment with dynamical changes, we have presented the automatic

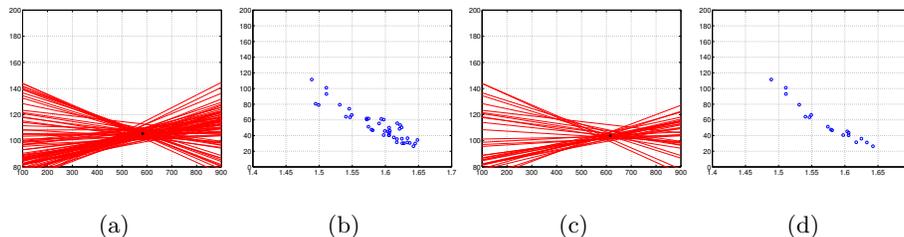


Fig. 6. (a) The convergent imaged symmetric axes and the estimated image of the insertion point i_1 at (593.4; 105.5) (black cross) computed with the least mean squares during the guidance of an instrument. (b) In the dual parameter space of convergent lines (θ, ρ) , "points" (blue bullet) must be collinear. (c) The estimated image of the insertion point i_1 at (615.5; 103.9) (black cross) and the parameter space (d) with the robust estimation when 50 % of data (outliers) are rejected.

detection and positioning of cylindrical-shaped objects with endoscopic views of the human body and we have brought some solutions especially in the context of the robotized laparoscopic surgery. Then, in the first part of the paper, we briefly present a fast segmentation of grey regions and, in the second part, the 3D pose and constrained motion of surgical instruments is described with details. With this article, we have addressed some issues with a non-uniform and moving background with time-varying lighting conditions, offer some generic and context-based solutions with landmark-free approaches. The representation of the instrument axis motion with the Plücker coordinates (4 DOFs) has been shown to be suited to deal with partial occlusions and also for the decoupling of the pan/tilt control, the penetration depth and the rotation axis of instruments. This is an important practical contribution for the achievement of vision-based semi-autonomous tasks with robots in minimally invasive surgery. In particular, the on-line localization of out-of-field of view insertion points (and their images) is an important issue to drive the image segmentation, the regions selection process and finally to improve the reliability while tracking the surgical instruments.

References

1. Wei, G.Q., Arbter, K., Hirzinger, G.: Real-time visual servoing for laparoscopic surgery. *IEEE Engineering in Medicine and Biology* **16** (1997) 40–45
2. Burschka, D., Corso, J.J., Dewan, M., Hager, G.D., Lau, W., Li, M., Lin, H., Marayong, P., Ramey, N.: Navigating inner space: 3-d assistance for minimally invasive surgery. In: *Workshop Advances in Robot Vision, IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, Sendai, Japan* (2004) 67–78
3. Krupa, A., Gangloff, J., Doignon, C., de Mathelin, M., Morel, G., Leroy, J., Soler, L., Marescaux, J.: Autonomous 3-d positioning of surgical instruments in robo-

- tized laparoscopic surgery using visual servoing. *IEEE Trans. on Robotics and Automation* **10** (2003) 842–853
4. Nageotte, F., Zanne, P., de Mathelin, M., Doignon, C.: A circular needle path planning method for suturing in laparoscopic surgery. In: *Proceedings of the IEEE Int'l Conf. on Robotics and Automation*, Barcelona, Spain (2005) 516–521
 5. Casals, A., Amat, J., Prats, D., Laporte, E.: Vision guided robotic system for laparoscopic surgery. In: *Proc. of the IFAC Int. Congress on Advanced Robotics*, Barcelona, Spain (1995) 33–36
 6. Wang, Y.F., Uecker, D.R., Wang, Y.: A new framework for vision-enabled and robotically assisted minimally invasive surgery. *Journal of Computerized Medical Imaging and Graphics* **22** (1998) 429–437
 7. Hayashibe, M., Nakamura, Y.: Laser-pointing endoscope system for intra-operative geometric registration. In: *Proc. of the 2001 IEEE International Conference on Robotics and Automation*, Seoul, South Korea (2001)
 8. Ortmaier, T., Hirzinger, G.: Cartesian control issues for minimally invasive robot surgery. In: *Proceedings of the IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, Takamatsu, Japan (2000)
 9. Voros, S., Orvain, E., Cinquin, P., Long, J.A.: Automatic detection of instruments in laparoscopic images: a first step towards high level command of robotized endoscopic holders. In: *IEEE Conf. on Biomed. Robotics and Biomechanics*. (2006)
 10. Doignon, C., Nageotte, F., de Mathelin, M.: Detection of grey regions in color images: application to the segmentation of a surgical instrument in robotized laparoscopy. In: *Proceedings of the IEEE/RSJ Int'l Conference on Intelligent Robots and Systems*, Sendai, Japan (2004)
 11. van de Weijer, J., Gevers, T., Geusebroek, J.M.: Color edge detection by photometric quasi-invariants. In: *Proc. of ICCV*, Nice, France (2003) 1520–1526
 12. Perona, P., Shiota, T., Malik, J.: Anisotropic diffusion. In: *Geometry-driven diffusion in Computer Vision*. Kluwer Academic Publisher (1994) 73–92
 13. Doignon, C., Graebling, P., de Mathelin, M.: Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. *Real-Time Imaging* **11** (2005) 429–442
 14. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography. *Int'l Journal of Computer Vision* **9** (1992) 137–154
 15. Weinshall, D., Tomasi, C.: Linear an incremental acquisition of invariant shape models from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17** (1995) 512–517
 16. Sturm, P., Triggs, W.: A factorization based algorithm for multi-image projective structure and motion. In: *In Proceedings of The European Conference on Computer Vision*. (1996) 709–720
 17. Ma, Y., Soatto, S., Košeká, J., Sastry, S.: *Invitation to 3D Vision: From Images to Geometric Models*. Springer-Verlag (1003)
 18. Bartoli, A., Sturm, P.: The 3d line motion matrix and alignment of line reconstructions. In: *Proceedings of CVPR*, Hawaii, USA (2001) 287–292
 19. Hartley, R., Zisserman, A.: *Multiple view geometry in computer vision*. Cambridge Univ. Press (2000)
 20. Wong, K.Y., Mendonça, P.R.S., Cipolla, R.: Reconstruction of surfaces of revolution from single uncalibrated views. *Image and Vis. Computing* **22** (2004) 829–836
 21. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation* **8** (1992) 313–326
 22. Haralick, R.M., Shapiro, L.G.: *Computer and Robot Vision*. Volume 2. Addison-Wesley Publishing (1992)