Pose estimation and feature tracking for robot assisted surgery with medical imaging

C. Doignon, F. Nageotte, B. Maurin, A. Krupa

▶ **To cite this version:**

C. Doignon, F. Nageotte, B. Maurin, A. Krupa. Pose estimation and feature tracking for robot assisted surgery with medical imaging. Kragic, D. and Kyrki, V. Unifying Perspectives in Computational and Robot Vision, Springer Verlag, 2007. inria-00350651

HAL Id: inria-00350651
https://inria.hal.science/inria-00350651

Submitted on 7 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapter 1
Pose Estimation and Feature Tracking for Robot Assisted Surgery with Medical Imaging

Christophe Doignon,
Florent Nageotte,
Benjamin Maurin
and Alexandre Krupa

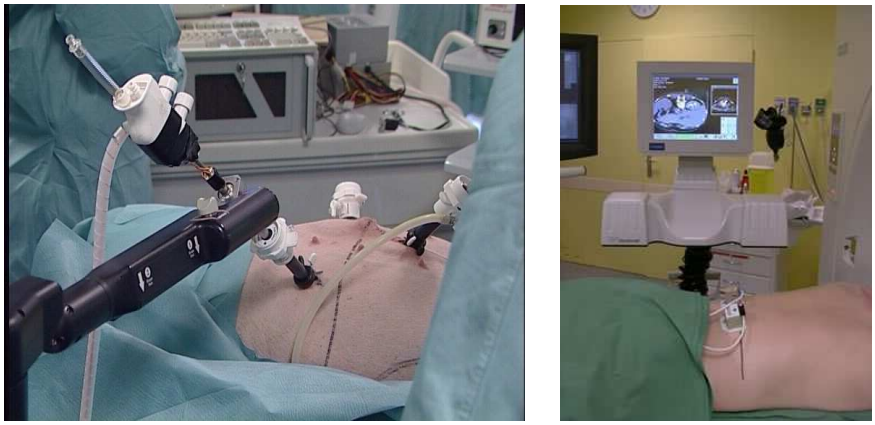
1.1 Introduction

The field of vision-based robotics has been widely growing for more than three decades, and more and more complex 3-D scenes are within robot vision capabilities thanks to better understanding of the scenes, improvement of computer capabilities and control theory. The achievement of applications like medical robotics, mobile robotics, micro-robotic manipulation, agricultural automation or the observation by aerial or underwater robots needs the integration of several research areas in computer vision and automatic control ([32, 19]).
For the past two decades, medical robot and computer-assisted surgery have gained increasing popularity. They have expanded the capabilities and comfort for both patients and surgeons in many kinds of interventions such as local therapy, biopsies, tumors detection and removal with techniques like multi-modal registration, online visualization, simulators for specific interventions and tracking. Medical robots provide a significant help in surgery, mainly for the improvement of positioning accuracy and particularly for intra-operative image guidance [36]. The main challenge in visual 3-D tracking for medical robotic purposes is to catch the relevant video information from images acquired with endoscopes [5], ultra-sound probes [17, 21] or scanners [35, 26] so as to evaluate the position and the velocity of objects of interest which usually are natural or artificial landmarks attached to a surgical instrument.

Christophe Doignon and Florent Nageotte
University of Strasbourg, LSIIT (UMR ULP-CNRS 7005), Control, Vision and Robotics Team, Boulevard Brant, 67412 Illkirch, France. e-mail: {doignon,nageotte}@lsiit.u-strasbg.fr

Benjamin Maurin
Cerebellum Automation Company, 178, route de Cran Gevrier, 74650 Chavanod, France.
e-mail: maurin@eavr.u-strasbg.fr

Alexandre Krupa
IRISA/INRIA Rennes (Lagadic Team), Campus de Beaulieu, 35042 Rennes, France.
e-mail: alexandre.krupa@irisa.fr



(a) (b)

Fig. 1.1 (a) The laparoscopic experimental setup. The instrument is mounted on the end effector of a surgical robot and inserted through the abdominal wall while the laparoscope is inserted through another insertion point. (b) An image-guidance with CT scanners feedback control during the percutaneous insertion of a radio-frequency needle.

This chapter presents several 3-D pose estimation algorithms and visual servoing-based tracking with monocular vision systems such as endoscopes and CT scanners (see Fig. 1.1) developped in an attempt to improve the guidance accuracy. These are intended for the 3-D positioning and guidance of surgical instruments in the human body. The efficiency of most of model-based visual servoing approaches relies on correspondences between the position of tracked visual features in the current image and their 3-D attitude in the world space. If these correspondences contain errors then the servoing usually fails or converges towards a wrong position. Overcoming these errors is often achieved by improving the quality of tracking algorithms and features selection methods ([37, 20]). Following this purpose, the work integrates several issues where computational vision can play a role:

1. estimating the distance between the tip of a laparoscopic instrument and the targeted organ with projected collinear feature points,
2. estimating the 3-D pose of an instrument using a multiple features tracking and a virtual visual servoing,
3. positioning a cylindrical-shaped instrument,
4. registering the instantaneous position of a robot using stereotaxy.

The chapter is organized as follows. In the next Section, the problem of the pose estimation of surgical instruments with markers is stated and solved for some degrees of freedom. In Section 3, we focus on the positioning of the symmetry axis of a cylindrical-shaped instrument. Applications of both Sections use endoscopic vision in laparoscopy. The stereotactic registration with a single view (2-D/3-D registration) is studied as a pose estimation problem in Section 4. Finally, a conclusion with some perspectives is drawn in Section 5.

1.2 Pose estimation of a laparoscopic instrument with landmarks

1.2.1 Pose estimation with collinear markers

There exist several difficulties when tackling the problem of estimating the 3-D position of a laparoscopic surgical instruments with a single endoscopic view. One difficulty is the use of monocular vision which gives poor depth information. Another one relies on the highly unstructured nature of the scene with varying lighting conditions and with a background moving due to breathing or heart beating. To solve these problems, we conceived five years ago a special instrument which projects a laser pattern onto the organ surface in order to provide the relative orientation of the instrument with respect to the organ, even if the instrument is not in the camera field of view. Optical markers have been added on the tip of the surgical instrument. These markers (composed of three circular LEDs) were directly projected onto the image and in conjunction with images of the laser pattern, they were used to guide the instrument (see Fig. 1.2). We combined image feature points (spots center coordinates) and depth information for positioning the instrument with respect to the pointed organ [22].

1.2.1.1 Pose estimation with a calibrated endoscope

Recovering the relative orientation (2 degrees of freedom - a unit vector **r**) and position (3 degrees of freedom - a vector **t**) of a set of n collinear points such as the optical markers and laser projections in Fig. 1.2 with respect to the camera has been

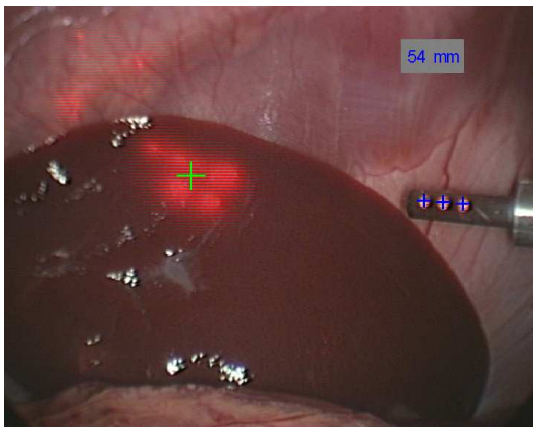


Fig. 1.2 A surgical instrument with a laser pointing device (laser beam - big cross) and three optical markers (small crosses). The cross-ratio is computed with cross centres and controlled in successive images so as to estimate the distance and its variations between the tip and the pointed organ surface.

previously investigated by Haralick fifteen years ago [15]. The interpoint distances (structure) and a focal length f of the camera are assumed to be known. Haralick solved this problem with a linear algorithm. Let $P_0 = \mathbf{t}$, $P_1 = \mathbf{t} + \lambda_1 \mathbf{r}$,..., $P_{n-1} = \mathbf{t} + \lambda_{n-1}\mathbf{r}$ be n discriminated points where λ_i represents the distance between the $(i+1)$th and ith points. The first point P_0 is arbitrarly chosen as the origin ($\lambda_0 = 0$), hence the perspective projection $Q_i = (u_i, v_i, 1)^\mathsf{T}$ of the ith point is given by

$$[0\ 0\ 1]\ (\lambda_i \mathbf{r} + \mathbf{t}) \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathtt{K}^c\ (\lambda_i \mathbf{r} + \mathbf{t}) \tag{1.1}$$

where $\mathtt{K}^c$ is a (3×3) upper diagonal matrix containing the internal parameters of the camera. From the above equation, Haralick built a homogeneous linear system with a uni-variate matrix $\mathtt{K}^c = \mathtt{diag}(f, f, 1)$ and vectors $\mathbf{t}$ and $\mathbf{r}$ as unknowns

$$\underbrace{\begin{bmatrix} \mathbf{A}_r & \mathbf{A}_t \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} \mathbf{r} \\ \mathbf{t} \end{bmatrix} = \mathbf{0}\ . \tag{1.2}$$

$\mathbf{A}$ is a $(2n \times 6)$ real matrix and a closed-form solution can be found with $n \geq 3$ discriminated points. This system may be reformulated as a classical optimization problem with an equality constraint:

$$\min \|\mathbf{A}_r\ \mathbf{r}\ +\ \mathbf{A}_t\ \mathbf{t}\| \quad \text{subject to} \quad \mathbf{r}^\mathsf{T}\mathbf{r}\ = 1\ , \tag{1.3}$$

where $\mathbf{A}_r$ and $\mathbf{A}_t$ are two $(2n \times 3)$ real matrices defined as:

$$\mathbf{A}_r = \underbrace{\begin{bmatrix} \lambda_0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & \lambda_0 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \lambda_{n-1} & 0 \\ 0 & 0 & \cdots & \cdots & 0 & \lambda_{n-1} \end{bmatrix}}_{\Lambda} \mathbf{A}_t\ ,\ \mathbf{A}_t = \begin{bmatrix} \begin{bmatrix} 1\ 0\ 0 \\ 0\ 1\ 0 \end{bmatrix} \left(\mathtt{K}^c - \begin{bmatrix} u_0 \\ v_0 \\ 1 \end{bmatrix} [0\ 0\ 1] \right) \\ \vdots \\ \begin{bmatrix} 1\ 0\ 0 \\ 0\ 1\ 0 \end{bmatrix} \left(\mathtt{K}^c - \begin{bmatrix} u_{n-1} \\ v_{n-1} \\ 1 \end{bmatrix} [0\ 0\ 1] \right) \end{bmatrix} \tag{1.4}$$

The solution for $\mathbf{r}$ is given by the eigenvector associated with the smallest eigenvalue of the following symmetric matrix

$$\mathbf{E}\ =\ \mathbf{A}_r^\mathsf{T} \left(\mathbf{I}\ -\ \mathbf{A}_t \left(\mathbf{A}_t^\mathsf{T} \mathbf{A}_t \right)^{-1} \mathbf{A}_t^\mathsf{T} \right)\ \mathbf{A}_r \tag{1.5}$$

and the position vector $\mathbf{t}$ is given by $\mathbf{t} = -\left(\mathbf{A}_t^\mathsf{T}\mathbf{A}_t\right)^{-1}\mathbf{A}_t^\mathsf{T}\ \mathbf{A}_r\ \mathbf{r}$. We end up with two different estimates for $\mathbf{r}$ (a twofold ambiguity in the sign of $\mathbf{r}$). However, for real objects placed in front of the camera, the third component of vector $\mathbf{t}$ must be strictly positive assuming that the camera z-axis (usually, the optical axis) is pointed towards the scene. This leads to the uniqueness of the solution for the pose.

It is worth pointing out that collinearity is a projective invariant property which is not fully exploited in this technique for pose recovery. Moreover, in presence of both noisy data and close points in the object pattern, matrices $\mathbf{A}_r$ and $\mathbf{A}_t$ are ill-conditioned, which introduces some significant bias in the results. The use of the least mean squares and the lack of data normalization in the original algorithm tend the solution to be sensitive to the condition number. One has to pay attention to data normalization since the pose estimation may be computed with points not always well scattered. This may also lead to numerical problems. To lower the condition number, it seems advisable to normalize data coordinates with an affine transformation as in [16].

1.2.1.2 Distance with Collinear landmarks

To perform 3-D positioning of an instrument with respect to an organ [22], we need to estimate the distance between the instrument and the targeted organ (depth d_0 in Fig. 1.3). Since the three optical markers centers P_1, P_2 and P_3 are placed along the instrument axis, we assumed they are collinear with the laser spot's barycentre P.

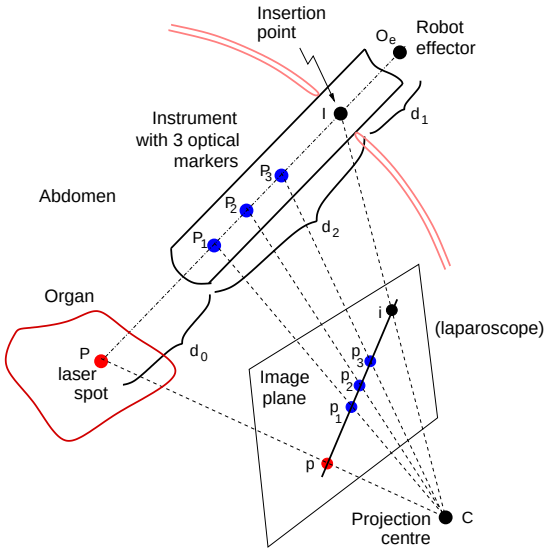


Fig. 1.3 The basic geometry involved for the relative instrument positioning with the laser spot aligned with three collinear LED centres.

Under this assumption, a cross-ratio can be computed from these four points [28]. This projective invariant can also be computed in the image using their respective projections p_1, p_2, p_3 and p (see Fig. 1.2 and 1.3) and can be used to estimate the

depth d_0. In other words, since a 1-D projective basis can be defined either with $\{P_1, P_2, P_3\}$ or their respective images $\{p_1, p_2, p_3\}$, the-cross ratio is a projective invariant built with the fourth point (P or p). Consequently, a 1-D homography H exists between these two bases, so that the straight line Δ corresponding to the instrument axis is transformed, in the image, into a line $\delta = H(\Delta)$ as shown in Fig. 1.3. The cross-ratio τ is given by:

$$\tau = \frac{\left(\frac{\overline{pp_2}}{\overline{p_1p_2}}\right)}{\left(\frac{\overline{pp_3}}{\overline{p_1p_3}}\right)} = \frac{\left(\frac{\overline{PP_2}}{\overline{P_1P_2}}\right)}{\left(\frac{\overline{PP_3}}{\overline{P_1P_3}}\right)} \tag{1.6}$$

and d_0 is obtained as

$$d_0 = \overline{PP_1} = (1-\tau)\frac{\overline{P_1P_3}}{\tau - \frac{\overline{P_1P_3}}{\overline{P_1P_2}}} = \alpha\,\frac{1-\tau}{\tau-\beta} \tag{1.7}$$

where α and β depend only on the known relative position of P_1, P_2 and P_3. To simplify the computation of the cross-ratio in the image plane, it's necessary to characterize the straight line δ in order to relate the pixels coordinates of an image point $p = (u, v, 1)^\mathsf{T}$ and its projective coordinates $(s\lambda, s)^\mathsf{T}$ on δ. Let $(-b, a)^\mathsf{T}$ be the normalized cosine direction of δ and $p_k = (u_k, v_k, 1)^\mathsf{T}$ a point on δ. This gives:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} -b & u_k \\ a & v_k \\ 0 & 1 \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} \lambda \\ 1 \end{bmatrix} \tag{1.8}$$

or

$$\lambda = \begin{bmatrix} -b & a \end{bmatrix} \begin{bmatrix} u - u_k \\ v - v_k \end{bmatrix} \tag{1.9}$$

with :

$$\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0 \tag{1.10}$$

where $(-c)$ is the orthogonal distance from δ to the image origin. The computation of the cross-ratio is then:

$$\tau = \frac{\lambda_0 + \overline{p_1p_2}}{\overline{p_1p_2}} \frac{\overline{p_1p_3}}{\lambda_0 + \overline{p_1p_3}} \tag{1.11}$$

From equation (1.7), it is straightforward that d_0 is a function of τ which, in turn, is a function of λ_0, $\overline{p_1p_2}$ and $\overline{p_1p_3}$. Similar computations lead to the same relationship between d_2 and another cross-ratio μ defined with the points P_1, P_2, P_3, I and their respective projections provided that i, the perspective projection of the incision point I (see Fig. 1.4), can be recovered [10]. Since I is generally not in the camera field of view, this can be achieved by considering a displacement of the surgical instrument

between two configurations yielding straight lines δ and δ' in the image. Then, i is the intersection of these lines given that I is invariant. Finally :

$$\mu = \frac{\left(\frac{\overline{p_1p_3}}{\overline{p_2p_3}}\right)}{\left(\frac{\overline{p_1i}}{\overline{p_2i}}\right)} = \frac{\left(\frac{\overline{P_1P_3}}{\overline{P_2P_3}}\right)}{\left(\frac{\overline{P_1I}}{\overline{P_2I}}\right)} , \tag{1.12}$$

$$d_2 = \overline{P_1I} = \frac{\frac{\alpha}{1-\beta}}{\mu + \frac{\beta}{1-\beta}} . \tag{1.13}$$

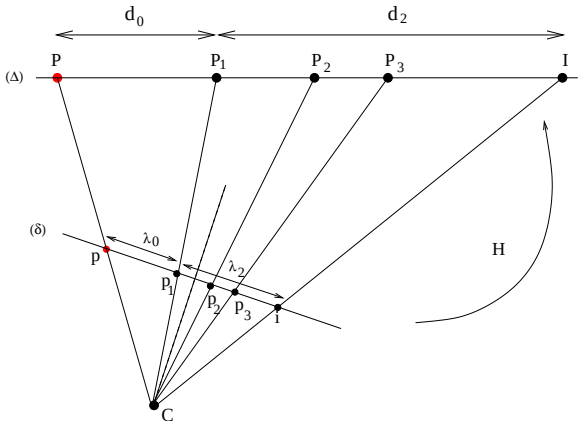


Fig. 1.4 Markers P_1,P_2,P_3 on the tool axis Δ and their images p_1, p_2, p_3 on line δ. Note that $i = H(I)$ is invariant during surgical procedures.

1.2.2 Pose estimation with Multiple Features

1.2.2.1 Objectives and Related work

In many early works, the images of laparoscopic instruments are segmented, in order to control the position of the endoscopic camera. These methods are based on the structure, mainly the apparent lines of the instrument [1, 6], or on its frequential features [8, 41]. In order to make the detection more robust and accurate, instruments can be marked with structuring markers as described before ([1, 43, 22]) or frequential (color) markers ([42, 38]). Most of these works use only the 2-D position of the instrument in the image and the accuracy of the features extraction is not so important for aimed applications. On the contrary, the 3-D pose estimation requires a very accurate feature extraction step.

None of the earlier works have focused on the complete 6 degrees of freedom detection. To determine the six degrees of freedom, the non symmetric part of the instrument has to be used or the instrument has to be marked. The second solution avoids using a CAD model and can be applied to any kind of instrument. We use the marker presented in Fig. 1.5, which is composed of twelve black spots on a white area, building four "marker lines" which can be discriminated from each other by using the cross-ratio invariant. For more information on the choice of this marker, the interested reader can refer to [30]. The detection of this marker is mainly based on the intensity of the white area and the black spots in the endoscopic images. It can thus be used in grey level images as well as color endoscopic images. The results of the detection technique are shown in Fig. 1.5 for laboratory endoscopic images. The image features noted $\mathbf{s}$ are n points corresponding to the centres of the visible spots (generally $n \in [0,6]$) and two lines $\mathbf{l}^+$ and $\mathbf{l}^-$ corresponding to the apparent contours of the shaft in the endoscopic image and represented by their distance ρ to the origin of the image and their orientation θ:

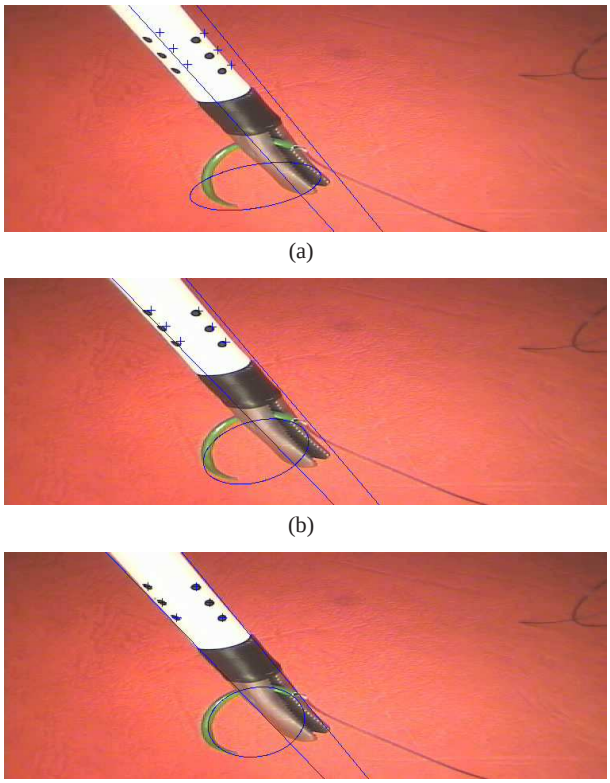
$$\mathbf{s} = \left(\mathbf{l}^+, \mathbf{l}^-, p_1, \cdots, p_n\right)^\mathsf{T} . \tag{1.14}$$

Depending on the size of the white marker area in the image, the complete extraction process can take up to 200 ms. In order to track the instrument at higher rate, we have developed techniques based on the moving edges method due to Bouthemy [3]. The main difficulty is to track the black spots which can appear and disappear, due to the rotation of the instrument around its own axis, and possibly due to occlusions. We have proposed a method based on the prediction of the spots appearance and disappearance which allows to track the markers without the need to register the images of the spots with the real positions [31]. Thus, the tracking of the instrument can be handled at a rate of 20 Hz as long as at least one spot is visible in the endoscopic images.

1.2.2.2 Pose Estimation of a Tagged Instrument

Our model-based pose estimation process requires a calibrated camera. Endoscopic cameras have a large field of view and include large radial distortion. As a consequence, the calibration method must estimate the distortion parameter [2, 39].

Only four degrees of freedom are necessary to estimate the attitude of the instrument axis (see next Section). Theoretically, the 4 degrees of freedom of the pose can be determined using the contour generator and its image (the apparent contour) of the cylinder [9]. However, the positions of the marking spots not only define the proper rotations and translations, but also give information on the orientation and position of the axis of the shaft. We then chose to estimate all the degrees of freedom of the instrument. This can be done with analytical methods using both the apparent contours and one known point at the cylinder's surface [29]. Other methods, like those proposed by Horaud [18], Haralick [14], DeMenthon [7] or Quan [33] can also be



(a)

(b)

(c)

Fig. 1.5 (a-c) The pose estimation as a virtual visual servoing process with multiple geometric features (apparent lines, marker-lines and eventually the circular needle). (a) The straight lines are the projections with the initial virtual camera position. (c) The projections when the error vector $\mathbf{s} - \mathbf{s}_d$ tends to 0.

used. However, the full pose estimation is interesting for robustness considerations only if all the available information given by the apparent lines and all the spots is used. To this purpose, the Virtual Visual Servoing (VVS) due to Marchand and Sundareswaran ([24, 34]) may handle the information redundancy. VVS is a numerical iterative method for minimizing the error between the extracted features and the forward projection of the object in the images, based on the image-based visual servoing (IBVS) schemes. This process needs the computation of an interaction matrix which relates the variations of each image feature and the the camera velocity screw τ. With the image features we use, the interaction matrix L_s has the following form:

$$\begin{pmatrix} \dot{\mathbf{l}}^+ \\ \dot{\mathbf{l}}^- \\ \dot{p}_1 \\ \vdots \\ \dot{p}_n \end{pmatrix} = \underbrace{\begin{pmatrix} L_{line}(\mathbf{l}^+) \\ L_{line}(\mathbf{l}^-) \\ L_{pt}(p_1) \\ \vdots \\ L_{pt}(p_n) \end{pmatrix}}_{L_s} \underbrace{\begin{pmatrix} {}^cV^C_{c/i} \\ {}^c\omega_{c/i} \end{pmatrix}}_{\tau} . \tag{1.15}$$

The interaction matrices associated to a point L_{pt} and to a line L_{line} can be found in the works of Chaumette [11]. In order to guarantee a fast convergence and a good stability of the VVS, it is useful to initialize the algorithm close enough to the real pose (from which a desired feature vector $\mathbf{s}_d$ can be defined). To this purpose, we use either the DeMenthon iterative method when at least four points are visible. From the obtained initial estimate, the following control law is applied to the virtual camera

$$\tau = -\lambda \tilde{L_s}^+ (\mathbf{s} - \mathbf{s}_d) \tag{1.16}$$

until the control vector becomes smaller than a specified value. The process converges quickly towards the real pose of the camera (see Fig. 1.5-c).

1.3 Pose estimation of a laparoscopic instrument without landmarks

1.3.1 Problem statement and perspective projection

The aim of this section is to briefly present a new algorithm for the determination of the pose of a straight homogeneous circular cylinders (SHCC) without markers, that is to say directly from the apparent contour. More details are provided in [9]. The apparent contour (γ) of a cylinder is a set of points which intersect the viewline and the image plane. It is the projection of a 3-D curve on the cylinder's surface referred to as the contour generator (Γ).
Given the matrix $\mathtt{K}^c$ of camera intrinsic parameters, the cylinder radius r_c and the apparent lines $\{\mathbf{l}^-, \mathbf{l}^+\}$, we look for the estimation of the Plücker coordinates $(\mathbf{r}, \mathbf{w})$ of the cylinder axis (see Fig. 1.7) satisfying the non-linear equation $\mathbf{r}^\mathsf{T}\mathbf{w} = 0$. This means that one has to solve a polynomial equation for a unique (double) solution (see [9]) that is, for a null discriminant. This one equals $B^2 - AC = 0$ with

$$\begin{cases} A = m^\mathsf{T}(\mathtt{K}^c)^{-\mathsf{T}}[\mathbf{r}]_\times \, [\mathbf{r}]_\times^\mathsf{T} m \\ B = \quad m^\mathsf{T}(\mathtt{K}^c)^{-\mathsf{T}}[\mathbf{r}]_\times \mathbf{w} \\ C = \qquad \|\mathbf{w}\|^2 - r_c^2 \end{cases} \tag{1.17}$$

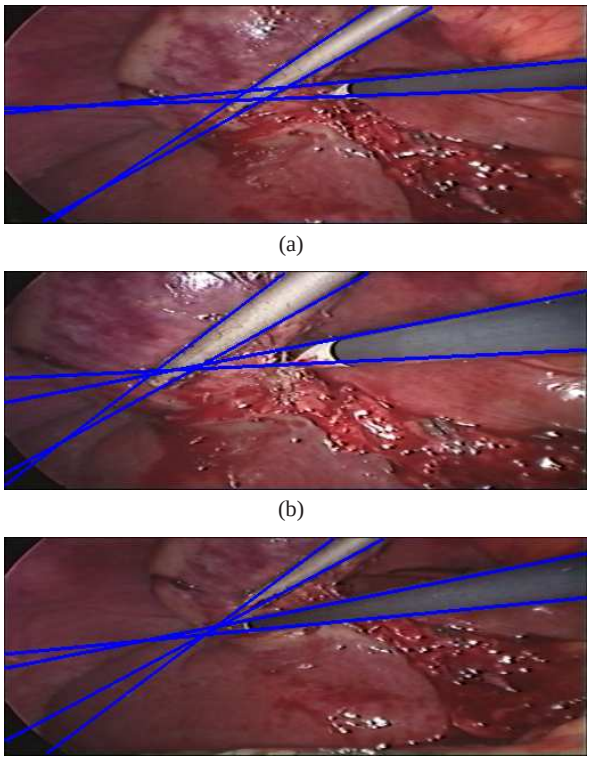
and after some computations, it can be expressed as

$$\left(\frac{r_c\, B}{\sqrt{C}} + \mathbf{w}^\mathsf{T}(\mathtt{K}^c)^{-1} m\right)^\mathsf{T} \left(\frac{r_c\, B}{\sqrt{C}} - \mathbf{w}^\mathsf{T}(\mathtt{K}^c)^{-1} m\right) \; = \; 0 \, . \tag{1.18}$$

If the scalar $C \leq 0$, the projection center $(0,0,0,1)^{\mathsf{T}}$ is located inside the cylinder (or on its surface, if $C = 0$) and the above equation yields no real solutions. We do not consider these special cases in the remainder, we rather focus this work on the more practical situation with real solutions. In the case of a circular cylinder with infinite height and a constant radius, equation (1.18) shows that the apparent contour is a set of two straight lines represented either with the pair of vectors $\mathbf{l}^{-}$ and $\mathbf{l}^{+}$ satisfying

$$\begin{cases} (\mathbf{l}^{-})^{\mathsf{T}}\, m \equiv \{(\mathtt{K}^{c})^{-\mathsf{T}}\,(\mathtt{I} - \alpha[\mathbf{r}]_{\times})\,\mathbf{w}\}^{\mathsf{T}}\, m = 0 \\ (\mathbf{l}^{+})^{\mathsf{T}}\, m \equiv \{(\mathtt{K}^{c})^{-\mathsf{T}}\,(\mathtt{I} + \alpha[\mathbf{r}]_{\times})\,\mathbf{w}\}^{\mathsf{T}}\, m = 0 \end{cases} \tag{1.19}$$

with $\alpha = r_c/\sqrt{\|\mathbf{w}\|^2 - r_c^2}$, or alternatively with the (3×3) matrix $\mathtt{C} = \mathbf{l}^{-}\,\mathbf{l}^{+\mathsf{T}} + \mathbf{l}^{+}\,\mathbf{l}^{-\mathsf{T}}$ satisfying $m^{\mathsf{T}}\,\mathtt{C}\,m = 0$. The latter is a rank-2 symmetrical matrix defined up to a scale factor and both representations are equivalent to model the apparent contour with 4 parameters.



(a) (b) (c)

Fig. 1.6 (a-c) Segmentation of three endoscopic images of surgical instruments in the abdominal cavity. The degenerate conic fitting superimposed with the overall apparent contour of the two detected cylindrical instruments.

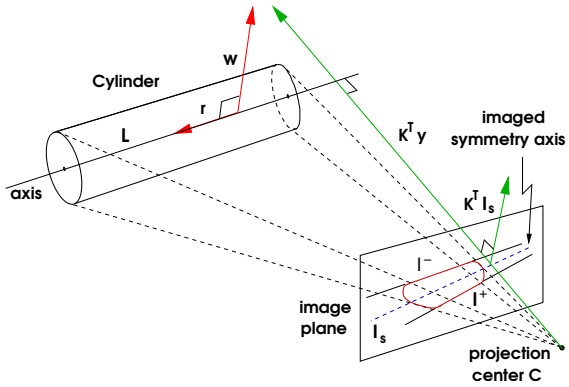


Fig. 1.7 A cylinder and its image with the perspective projection $\mathtt{P}$. The backprojection of the apparent lines $(\mathbf{l}^-,\mathbf{l}^+)$ is a pair of planes $\mathtt{P}^\mathsf{T}\mathbf{l}^-$ and $\mathtt{P}^\mathsf{T}\mathbf{l}^+$, passing through the projection center. The image of the cylinder axis Δ is the axis $\mathbf{l}_s$ of the harmonic homology $\mathtt{H}$ relating the apparent lines.

1.3.2 Direct pose computation

In this paragraph we present a linear algorithm for the pose estimation. Starting from (1.19), the matrix $\mathtt{C}$ can be related to the pose parameters since we have

$$\begin{aligned}
\mathtt{K}^{c\mathsf{T}}\,\mathtt{C}\,\mathtt{K}^c &\equiv \mathtt{K}^{c\mathsf{T}}\left(\mathbf{l}^-\,\mathbf{l}^{+\mathsf{T}} + \mathbf{l}^+\,\mathbf{l}^{-\mathsf{T}}\right)\mathtt{K}^c \\
&\equiv (\alpha[\mathbf{r}]_\times - \mathtt{I})\,\mathbf{w}\mathbf{w}^\mathsf{T}\,(\alpha[\mathbf{r}]_\times^\mathsf{T} + \mathtt{I}) + (\alpha[\mathbf{r}]_\times + \mathtt{I})\,\mathbf{w}\mathbf{w}^\mathsf{T}\,(\alpha[\mathbf{r}]_\times^\mathsf{T} - \mathtt{I}) \\
&\equiv \alpha^2[\mathbf{r}]_\times\mathbf{w}\mathbf{w}^\mathsf{T}[\mathbf{r}]_\times + \mathbf{w}\mathbf{w}^\mathsf{T} \\
&\equiv [\mathbf{r}]_\times\,(\alpha^2\frac{\mathbf{w}\mathbf{w}^\mathsf{T}}{\|\mathbf{w}\|^2} + \frac{[\mathbf{w}]_\times[\mathbf{w}]_\times^\mathsf{T}}{\|\mathbf{w}\|^2})\,[\mathbf{r}]_\times^\mathsf{T} \;=\; [\mathbf{r}]_\times\,(\mathtt{I} - (1-\alpha^2)\frac{\mathbf{w}\mathbf{w}^\mathsf{T}}{\|\mathbf{w}\|^2})\,[\mathbf{r}]_\times^\mathsf{T} \\
&\equiv \left(\mathtt{I} - \mathbf{r}\mathbf{r}^\mathsf{T} - \mathbf{z}\mathbf{z}^\mathsf{T}\right) = \left[\,\mathbf{a}\;\mathbf{z}_u\;\mathbf{r}\,\right]\begin{bmatrix}1 & 0 & 0\\ 0 & 1-\sigma^2 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{a}^\mathsf{T}\\ \mathbf{z}_u^\mathsf{T}\\ \mathbf{r}^\mathsf{T}\end{bmatrix}
\end{aligned} \tag{1.20}$$

with $\mathbf{z} = \frac{\sqrt{1-\alpha^2}}{\|\mathbf{w}\|}\,[\mathbf{r}]_\times\mathbf{w}$ and the unit vector $\mathbf{z}_u = \mathbf{z}/\sigma$. On the other hand, the SVD has the following expression $\mathtt{K}^{c\mathsf{T}}\,\mathtt{C}\,\mathtt{K}^c = \mathbf{U}\,\mathbf{D}\,\mathbf{U}^\mathsf{T} = \mathbf{U}\,\mathrm{diag}(\lambda_1,\lambda_2,0)\,\mathbf{U}^\mathsf{T}$. Then, it is easy to see that $\mathbf{U} = \left[\,\mathbf{a}\;\;\mathbf{z}_u\;\;\mathbf{r}\,\right]$ and

$$\sigma = \|\mathbf{z}\| = \sqrt{1-\alpha^2} = \sqrt{1-\frac{\lambda_2}{\lambda_1}}\,.$$

Finally, $\mathbf{w} \equiv \mathbf{z}_u \times \mathbf{r} = \mathbf{a}$ and

$$\|\mathbf{z}\| = \sqrt{1-\frac{r_c^2}{\|\mathbf{w}\|^2 - r_c^2}} \;\Rightarrow \|\mathbf{w}\| = r_c\,\sqrt{1+\frac{\lambda_1}{\lambda_2}}\,. \tag{1.21}$$

Many results are provided in [9]. In particular, we have compared the pose computation from the apparent contours of cylinders and the Haralick's method for the pose of a set of collinear points [15] with some artificial markers attached to the cylinder's surface. Here, we rather focus the discussion on the application of concern which is the image-guidance for intra-operative procedures in minimally invasive surgery (MIS). In laparoscopic surgery, most of surgical instruments have cylindrical parts and are metallic (see Fig. 1.6) leading to grey regions with many specularities in the image. Prior researchs involving such endoscopic images have been conducted in the field of color image segmentation [8]. Once regions have been segmented, the region boundaries are ordered and used to perform a degenerate conic-based contours fitting. With the calibrated and distortion-corrected endoscope used in the experiments, the 3-D localization of the two moving surgical instruments in Fig. 1.6 has been done with success for more than 300 successive images of the abdominal cavity of a pig.

With the proposed method, the location of each insertion point in laparoscopy can be recovered, on-line, with no marker, without any knowledge of robot kinematics and without an external measurement device [10]. Since any laparoscopic instrument is passsing through this point, the motion constraint in MIS can be expressed as the intersection of multiple convergent 3-D straight lines. Since any (homogeneous) point $\mathbf{X}$ is on L if $\mathtt{L}^\star\mathbf{X}=\mathbf{0}$, given n positions corresponding to the set of dual Plücker matrices $\{\mathtt{L}_1^\star,\mathtt{L}_2^\star,...,\mathtt{L}_n^\star\}$, the intersection of lines is obtained with a rank-3 $(4n\times 4)$ matrix $\mathtt{G}_n^\mathsf{T}$ such that

$$\mathtt{G}_n=[\mathtt{L}_1^\star,\mathtt{L}_2^\star,...,\mathtt{L}_n^\star]\ . \tag{1.22}$$

That is to say the null-space of $\mathtt{G}_n^\mathsf{T}$ must be a one-dimensional subspace and the intersection may be computed with n $(n\geq 2)$ 3-D positions. By computing the SVD of $\mathtt{G}_n^\mathsf{T}$, one obtains the common intersection with the singular vector associated to the null singular value (or the smallest one in presence of noisy data). Moreover, the perspective projection of the 3-D line L_j is the image line $\mathbf{l}_j$ defined by

$$[\mathbf{l}_j]_\times=\mathtt{K}^c\mathtt{P}^c\,\mathtt{L}_j\,(\mathtt{K}^c\mathtt{P}^c)^\mathsf{T}\ =\ [(\mathtt{K}^c)^{-\mathsf{T}}\,\mathbf{w}_j]_\times\ \Rightarrow\ \mathbf{l}_j\equiv(\mathtt{K}^c)^{-\mathsf{T}}\,\mathbf{w}_j\ , \tag{1.23}$$

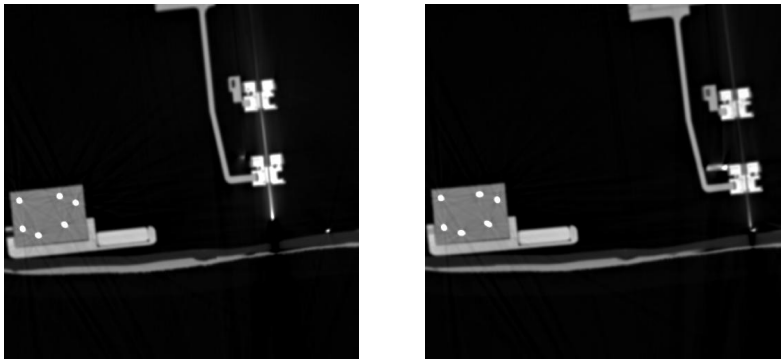
where $\mathtt{P}^c$ is the projection matrix. Since vector $\mathbf{l}_j$ is defined up to a scale, it does not depend on the magnitude of vector $\mathbf{w}_j$, hence the n convergent image lines $\mathbf{l}_1,\mathbf{l}_2,...,\mathbf{l}_n$ must satisfy

$$\big(\mathbf{l}_1\ \ldots\ \mathbf{l}_n\big)^\mathsf{T}\ i\ =\underbrace{\big(\mathbf{w}_1\ \ldots\ \mathbf{w}_n\big)^\mathsf{T}}_{\mathtt{W}_n}\ (\mathtt{K}^c)^{-1}\ i\ =\ \mathbf{0} \tag{1.24}$$

where i is the image of the insertion point I. It follows that a set of n 3-D straight lines is projected to n convergent image lines if the above $(n\times 3)$ matrix $\mathtt{W}_n$ is of rank 2. It is only a necessary condition which does not ensure the convergence of the 3-D lines, but it makes very important the accurate estimation of the imaged cylinder axis (lines $\mathbf{l}_j$), hence the estimation of its Plücker coordinates.

1.4 Pose estimation of stereotactic landmarks

This Section deals with the 2-D/3-D registration of a stereotactic frame from a single slice captured with a computed tomography (CT) scanner. A registration with a single image is very well suited for CT-guided robotic systems in interventional radiology, particularly to quickly correct the needle positioning (see Fig. 1.8) during percutaneous procedures [27].
In stereotaxy, line fiducials are usually used to produce a set of image points that are further employed in a pose estimation algorithm (see Fig. 1.8). To achieve the registration, the matching and pose estimation processes need to be robust and fast enough so as to be convenient in clinical conditions. To this end, a new formulation of the patient-to-modality stereotactic registration with a single image and for any arrangement of the fiducials has been proposed. It is worth pointing out that our solution requires very few fiducials in comparison with previous techniques.



(a) (b)

Fig. 1.8 "Look an move" with CT scanners. (a) The needle is maintained in the needle-holder jaws of a lightweight parallel robotic plateform (CT-Bot), which has to be moved to the target point (right). (b) A new acquisition to check for the final positioning.

1.4.1 The imaging model

Since most CT imaging devices execute some proprietary algorithms to generate image slices and since these algorithms usually are not in the public domain, we consider the imaging device as a black box. In other words, this work is focused on the delivery of a general framework for 2-D/3-D registration rather than a study of the physical properties of each step of image formation. A CT scanner provides slices of objects. It has internal parameters such as the thickness of a slice and scaling parameters that influence the tomographic reconstruction process from the projection

measurements. To take care of them, we propose an imaging model composed of an affine transformation accounting for intrinsic parameters, an Euclidean one for the rigid-body transformation that relates the scanner to the stereotactic frame and an orthographic projection that expresses the projection of a thin slice onto the image. To formulate the registration, we denote with $\mathscr{F}_0$ the reference frame attached to the fiducials and with $\mathscr{F}_{ct}$ the frame attached to the scanner. A scaled frame $\mathscr{F}_I$ is also attached to the CT image with pixel units instead of millimeters. A point in space like the origin of a reference frame is written in bold as **O**. The imaging model relates the coordinates of a 3-D point $\mathbf{P}_j$ expressed in $\mathscr{F}_{ct}$ and coordinates of the corresponding point ${}^I\mathbf{Q}_j = \begin{bmatrix} u_j & v_j \end{bmatrix}^\mathsf{T}$ in the image as

$$\begin{aligned} {}^{ct}\mathbf{OP}_j = \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} &= \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\text{orthographic}} \underbrace{\begin{bmatrix} s_x & g \\ 0 & s_y \end{bmatrix}}_{\text{intrinsic parameters}} \underbrace{\begin{bmatrix} u_j \\ v_j \end{bmatrix}}_{\text{pixel}} \\ &= {}^{ct}\Pi_\pi \, {}^\pi S_I \, {}^I\mathbf{Q}_j \end{aligned} \tag{1.25}$$

where ${}^{ct}\Pi_\pi$ is a (3×2) matrix accounting for the orthographic projection onto the cutting plane (π) (see Fig. 1.10) and the non-null entries of matrix ${}^\pi S_I$ are the intrinsic parameters of the scanner. They consist of two scaling factors s_x and s_y and a shearing parameter g accounting for a gantry tilt angle error or table bending during the scan. Usually, this parameter is very small and it is often neglected. However, it may be identified, since in some circumstances, it may decrease the registration accuracy as it is for MRI [4]. In the rest of the paper, this parameter will be neglected. Since there exists a rigid-body transformation between $\mathscr{F}_0$ and $\mathscr{F}_{ct}$, the expression for the vector $\mathbf{OP}_j$ in $\mathscr{F}_0$ is given by ${}^0\mathbf{OP}_j = R\,{}^{ct}\mathbf{OP}_j + \mathbf{t}$ where R is a rotation matrix and $\mathbf{t}$ is a position vector. Then, one may see the following expression

$${}^0\mathbf{OP}_j = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 \end{bmatrix} {}^\pi S_I \, {}^I\mathbf{Q}_j + \mathbf{t} \tag{1.26}$$

as a compact representation for the transformation $\mathscr{F}_I \rightarrow \mathscr{F}_0$ including the orthographic projection, where $\mathbf{r}_k$ is the k^{th} column of R. Therefore, the following (3×2) real matrix

$${}^0L_I = [\mathbf{r}_1 \; \mathbf{r}_2] \; {}^\pi S_I = [\mathbf{l}_1 \;\; \mathbf{l}_2] \tag{1.27}$$

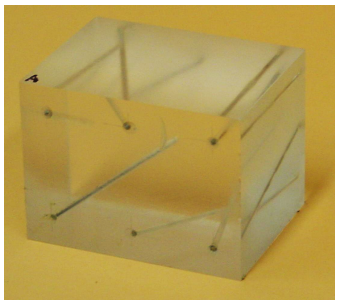
must satisfy the quadratic constraints coming from the orthonormality of any rotation matrix [1] :

$$\mathbf{l}_1{}^\mathsf{T}\,\mathbf{l}_1 = s_x^2 \;\;,\;\; \mathbf{l}_2{}^\mathsf{T}\,\mathbf{l}_2 = s_y^2 \;\text{ and }\; \mathbf{l}_1{}^\mathsf{T}\,\mathbf{l}_2 = 0\;. \tag{1.28}$$

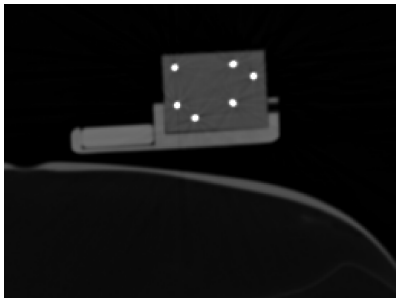
Finally, considering the notations for homogenous coordinates of ${}^I\mathbf{Q}_j$ as ${}^I\underline{\mathbf{Q}}_j = (u_j, v_j, 1)^\mathsf{T}$, equation (1.26) is rewritten as

$${}^0\mathbf{OP}_j = \begin{bmatrix} {}^0L_I & \mathbf{t} \end{bmatrix} {}^I\underline{\mathbf{Q}}_j\;. \tag{1.29}$$

[1] In accordance with (1.25), if the shearing parameter g is significant, equation (1.28) should be replaced by $\mathbf{l}_1^\mathsf{T}\mathbf{l}_1 = s_x^2$, $\mathbf{l}_1^\mathsf{T}\mathbf{l}_2 = s_x g$ and $\mathbf{l}_2^\mathsf{T}\mathbf{l}_2 = g^2 + s_y^2$.



(a)



(b)

Fig. 1.9 (a) The plastic cube with the line fiducials used for experiments. - (b) A CT scanner image (magnified) when the cube is placed on a phantom.

1.4.2 Modeling the fiducials

Fiducials used in stereotaxy are usually composed of rods (see Fig. 1.9-a) and are represented with straight lines [13, 35, 23]. Let Δ_j be the j^{th} line. This line may be represented with the origin $\mathbf{O}_j$ (3 dof) and a unit vector $\mathbf{y}_j$ (2 dof). Its intersection with the scanner plane (π) is (generally) a point $\mathbf{P}_j = \Delta_j \cap \pi$ (see Fig. 1.9-b), and substituting the expression of ${}^0\mathbf{OP}_j$ in (1.29), it can be expressed with

$$ {}^0\mathbf{OP}_j = {}^0\mathbf{OO}_j + \lambda_j \, {}^0\mathbf{y}_j = \left[{}^0L_I \; \mathbf{t} \right] {}^I\underline{\mathbf{Q}}_j , \quad \lambda_j \in \mathbb{R} \tag{1.30} $$

where $\mathbf{O}_j$ is the orthogonal projection of the origin of the frame $\mathscr{F}_0$ onto Δ_j, thus satisfying $\mathbf{O}_j\mathbf{P}_j \times \mathbf{y}_j = \mathbf{0}$. Therefore, to achieve the registration, one must solve (1.30) for ${}^0L_I$, $\mathbf{t}$ and the $\{\lambda_j\}$'s, that is for $(9+n)$ unknowns with n lines. Consequently, the size of the system to solve increases with the number of rods, leading to large

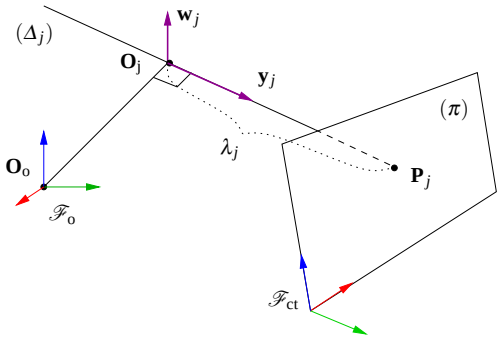


Fig. 1.10 A 3-D line Δ_j crossing the cutting plane (π). The pair of vectors $(\mathbf{y}_j, \mathbf{w}_j)$ is the Plückerian representation of the line.

matrices which must be precessed with many numerical operations [26]. To reduce the number of unknowns, we introduce the Plückerian representation [16]. The Plückerian coordinates of a 3-D line Δ_j are the pair of orthogonal vectors $(\mathbf{y}_j, \mathbf{w}_j)$ (see Fig. 1.10) where $\mathbf{w}_j$ is defined by $\mathbf{w}_j = \mathbf{y}_j \times \mathbf{OP}_j$. With this representation, the origin of the line as well as the $\{\lambda_j\}$'s are removed from the system. The above definition and the latter expression for ${}^0\mathbf{OP}_j$ can be gathered in the following equation, expressed in $\mathscr{F}_0$:

$$\left[{}^0\mathbf{y}_j\right]_\times \left[{}^0L_I \quad \mathbf{t}\right] {}^I\underline{\mathbf{Q}}_j = {}^0\mathbf{w}_j \,, \tag{1.31}$$

where $[\mathbf{y}_j]_\times$ is the (3×3) skew-symmetric (singular) matrix associated to $\mathbf{y}_j$. Equation (1.31) is the basis for our registration approach.

Generally, intersections of straight lines with the cutting plane should provide as many spots as there are lines (see Fig. 1.9-b). In practice, several spots may be missing in the image or in contrary some artifacts may appear [23]. In practice, line fiducials are bounded ($\lambda_j^{min} \leq \lambda_j \leq \lambda_j^{max}$). It is easy to compute these extremal values for any displacement (R,$\mathbf{t}$) and to check the relevance of the corresponding spot. To do so, a pre-multiplication with a unit vector ${}^0\mathbf{y}_j^\mathsf{T}$ in (1.30) gives the following expression

$$\lambda_j = {}^0\mathbf{y}_j{}^\mathsf{T} \left[{}^0L_I \quad \mathbf{t}\right] {}^I\underline{\mathbf{Q}}_j \,. \tag{1.32}$$

Given n lines/points correspondences, (1.31) can be expressed as a minimization problem with equality constraints:

$$\min_{\mathbf{x}} || A\,\mathbf{x} - \mathbf{b} ||^2 \quad \text{subject to} \quad \mathbf{x}^\mathsf{T} C\,\mathbf{x} = 0 \tag{1.33}$$

where $\mathbf{x} = \left[\mathbf{l}_1{}^\mathsf{T}\ \mathbf{l}_2{}^\mathsf{T}\ \mathbf{t}^\mathsf{T}\right]^\mathsf{T}$, C is a (9×9) symmetrical matrix with null entries except for $C_{14} = C_{25} = C_{36} = C_{41} = C_{52} = C_{63}$, A is a $(3n\times 9)$ and $\mathbf{b}$ is a $(3n\times 1)$ matrix, respectively defined as:

$$A = \begin{bmatrix} {}^I\underline{\mathbf{Q}}_1{}^\mathsf{T} \otimes [{}^0\mathbf{y}_1]_\times \\ \vdots \\ {}^I\underline{\mathbf{Q}}_n{}^\mathsf{T} \otimes [{}^0\mathbf{y}_n]_\times \end{bmatrix} \,,\ \mathbf{b} = \begin{bmatrix} {}^0\mathbf{w}_1{}^\mathsf{T} \\ \vdots \\ {}^0\mathbf{w}_n{}^\mathsf{T} \end{bmatrix} \,. \tag{1.34}$$

1.4.3 Registration as a pose estimation problem

This section aims at designing fast algorithms for estimating the parameters of the rigid registration, assuming a calibrated scanner is available (see [26] for uncalibrated scanners). We tackle this rigid registration problem (recovery of R and $\mathbf{t}$) with the minimum number of fiducials needed and by means of a linear algorithm. Given a single image, and considering the unknown vector $\xi = \left[\mathbf{r}_1{}^\mathsf{T}\ \mathbf{r}_2{}^\mathsf{T}\ \mathbf{t}^\mathsf{T}\right]^\mathsf{T}$, (1.31) becomes

$$\left[{}^I\underline{\mathbf{Q}}_j{}^\mathsf{T} \otimes [{}^0\mathbf{y}_j]_\times\right] S_9\,\xi = {}^0\mathbf{w}_j, \tag{1.35}$$

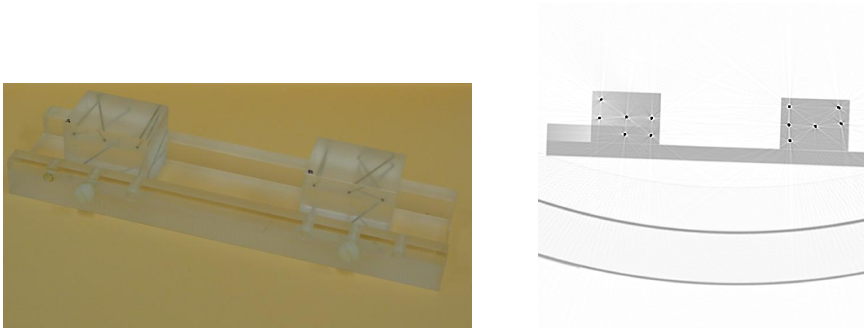
where $S_9 = \begin{bmatrix} {}^{\pi}S_I{}^{\mathsf{T}} \otimes I_3 & 0 \\ 0 & I_3 \end{bmatrix}$. With exactly 4 lines/points correspondences, (1.35) is a deficient-rank system which can be solved thanks to rotations properties. Except for some arrangements of the fiducials enumerated in [25], the matrix A has rank 8 when components contain uncorrupted data. However, with noisy data, the rank may be greater than 8. Hence, we wish to enforce the rank value because of the matrix structure (it is built with singular matrices). Therefore, there is a one-parameter family of solutions and (1.35) may be solved with the the Singular Value Decomposition (SVD). We summarize it as follows:

1. Find the SVD of A: $A = UDV^{\mathsf{T}}$, where the diagonal entries d_i of D are in descending numerical order,
2. Set $\mathbf{b}' = \begin{bmatrix} b'_1 & b'_2 & \dots & b'_9 \end{bmatrix}^{\mathsf{T}} = U^{\mathsf{T}}\,\mathbf{b}$ (see (1.34)),
3. Build the vector $\mathbf{z}$ defined by $z_i = b'_i/d_i$, for $i = 1, \cdots, 8$ and $z_9 = 0$,
4. The general solution is $\xi = S_9^{-1}\,(V\mathbf{z} + \gamma\,\mathbf{v}_9)$, where $\mathbf{v}_9$ is the last (rightmost) column of V.
5. Compute the value of scalar γ with the quadratic relations between $\mathbf{r}_1 = \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \end{bmatrix}^{\mathsf{T}}$ and $\mathbf{r}_2 = \begin{bmatrix} \xi_4 & \xi_5 & \xi_6 \end{bmatrix}^{\mathsf{T}}$.

Equation (1.35) can be solved provided that all combinations of triplets verify the conditions mentioned in previous section. Because of the presence of noise, R is not exactly a rotation matrix. One may enforce R to be a rotation by computing the SVD, $R = U\Sigma V^{\mathsf{T}}$ and by setting singular values to 1. If R' is the corrected rotation matrix, it is given by $R' = U\ \mathrm{diag}(1, 1, \det(UV))\ V^{\mathsf{T}}$ [12, 40].
A Newton-Raphson (N-R) numerical approach has also been carried out. It uses the initial guesses provided by the above least-squares method (LS) but we do not describe it here (see [26] for details).

1.4.4 Experimental validations

Experiments were conducted with a Siemens Somatom Plus CT scanner and with fiducials composed of two cubes with six rods each (see Fig. 1.11). The relative position of the cubes is constrained to by a guide rail on which the cubes are screwed. Each cube has been calibrated as well as their relative position with a Mitutoyo measuring machine which can achieve a precision better than 10 μm.
We have assessed the accuracy of the relative pose recovery by registering only one fiducial cube at a time, and by computing the relative position and rotation. By doing so, it is possible to verify the consistency of the pose estimation between two coordinate frames with a single image. To this end, a helicoidal sequence has been captured while a constant translation of the table is performed. In Fig. 1.12, we present the estimated position vector between the two cubes, as the orientation is approximately the identity matrix (it differs to the identity matrix by less than 10^{-5} on each component). The registration has been executed for each cube independently and once it has been done for both (with the LS method and the N-R min-



(a) (b)

Fig. 1.11 (a) The two cubes used for the experiments. - (b) image picked up from a CT helicoidal sequence with image size of (512×512) pixel.

imization), the relative position and orientation have been computed. The position $\mathbf{T} = \begin{bmatrix} 0 & -118.29 & 0 \end{bmatrix}$ mm was measured during a calibration procedure (orientations are equal for the two cubes' reference frames). Thus, this value (dotted line in Fig. 1.12) can be compared to the estimations.
As shown in Fig. 1.13, the 3-D pose algorithm works very well on experimental data, since the registration of each cube can serve to predict the position of the rods of the other cube. The estimated error bounding-boxes with an assumed spot location error of 0.25 pixel are also represented. As illustrated, all the detected spots are inside a box with boundaries corresponding to 3-D position errors always less than 2 mm.

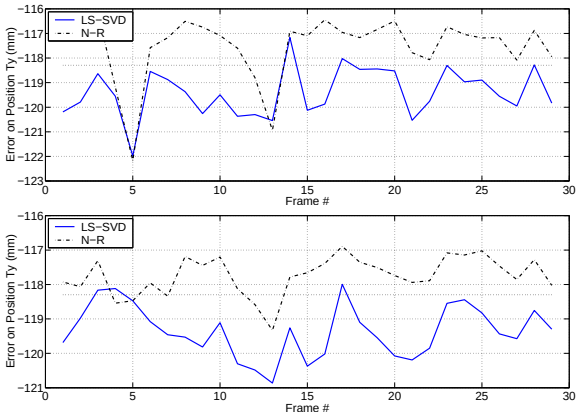


Fig. 1.12 Relative positions between the cubes during the acquisition (slice thickness is 0.5 mm) while translating the table. The first plot (up) is with 5 fiducials for the estimation while the second (down) is with 6 fiducials.

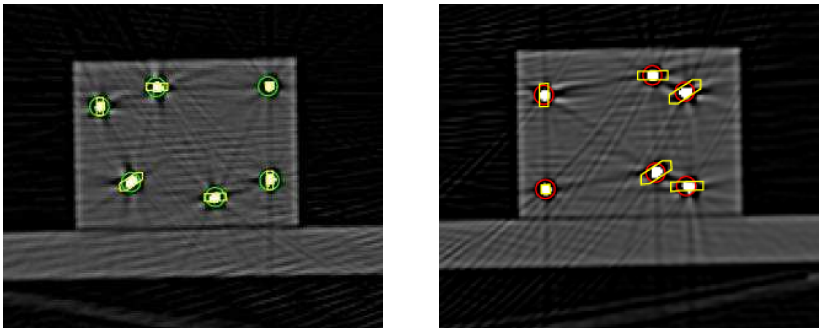


Fig. 1.13 CT images of the cubes. Each drawing box in one cube is corresponding to 2 mm of error bounds computed with the registration of the other cube.

1.5 Conclusion

In this chapter, we have described some pose estimation problems by means of intra-operative images. We focused the works on endoscopic views for assisted laparoscopy and CT images slices with X-ray scanners for the image guidance in interventional radiology. For vision-based integrated systems used in minimally invasive surgery, we have developed a set of techniques for assisting surgeons in navigating and manipulating the three-dimensional space within the human body. To that purpose, simple geometrical features have been attached to surgical instruments. Alternatively, when the task is sufficiently constrained by the shape of object of interest, we directly solve the pose without artificial markers: it is the case for the 4 degrees of freedom of a cylindrical needle-holder inside the human abdomen.

One path toward safety and reliability is to incorporate all the available video information. Following this issue, the virtual visual servoing has been used to combine both the apparent contour of the instrument and artificial markers in a numerical iterative process.
The recovery of out-of-field of view instrument in laparoscopy, the automatic suturing intervention demonstrated in vitro [31] and the positioning of a radio-frequency needle with CT scanners [25] are some applications we contributed for the aforementioned pose problems.

Finally, we believe that significant advances are possible when the geometric information is fused across time and across modality. Furthermore, pre-operative information like the insertion point's localization, the CAD model of instruments, the eye-to-hand calibration or the availability of several statistical atlases of organs can provide some strong constraints on the vision problem. These are crucial factors to achieve reliable dedicated vision systems while compensating small displacements due to patient breathing or any small disturbances which may occur during an image-guided surgical procedure.

Acknowledgements This work has been partially sponsored by the Region Alsace Council. The experimental part of this work has been made possible thanks to the collaboration with the Department of Radiology at the Strasbourg Hospital and the "Institut de Recherche contre les Cancers de l'Appareil Digestif". In particular, we would like to thank Prof. Marescaux, Leroy, Soler and Gangi for their advices, as well as for the use of their facilities.

References

1. A. Casals, J.A., Laporte, E.: Automatic guidance of an assistant robot in laparoscopic surgery. In: IEEE Int'l Conf. on Robotics and Automation, pp. 895–900. Minneapolis, USA (1996)
2. Bouguet, J.Y.: Camera calibration matlab toolbox. MRL - Intel Corp. URL http://*www.vision.caltech.edu/bouguetj/calib − doc/*
3. Bouthemy, P.: A maximum likelihood framework for determining moving edges. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(5), 499–511 (1989)
4. Breeuwer, M., Zylka, W., Wadley, J., Falk, A.: Detection and correction of geometric distortion in 3D CT/MR images. In: Proc. of Computer Assisted Radiology and Surgery, pp. 11–23. Paris (2002)
5. Burschka, D., Corso, J.J., Dewan, M., Hager, G., 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, in conjunction with the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 67–78. Sendai, Japan (2004)
6. Climent, J., Mars, P.: Automatic instrument localization in laparoscopic surgery. Electronic Letters on Computer Vision and Image Analysis **4**(1), 21–31 (2004)
7. Dementhon, D., Davis, L.S.: Model-based object pose in 25 lines of code. IJCV **15**(1), 123–141 (1995)
8. 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**, 429–442 (2005)
9. Doignon, C., de Mathelin, M.: A degenerate conic-based method for a direct fitting and 3-d pose of cylinders with a single perspective view. In: IEEE Int'l Conf. on Robotics and Automation, pp. 4220–4225. Roma, Italy (2007)
10. Doignon, C., Nageotte, F., de Mathelin, M.: The role of insertion points in the detection and positioning of instruments in laparoscopy for robotic tasks. In: In Proceedings of the Int'l. Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI, pp. 527–534 (Part I). Copenhaguen, Denmark (2006)
11. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. IEEE Trans. Robotics and Automation **8**(3), 313–326 (1992)
12. Goryn, D., Hein, S.: On the estimation of rigid body rotation from noisy data. IEEE Transactions on Pattern Analysis and Machine Intelligence **17**(12), 1219–1220 (1995)
13. Grunert, P., Maürer, J., Müller-Forell, W.: Accuracy of stereotactic coordinate transformation using a localisation frame and computed tomographic imaging. Neurosurgery **22**, 173–203 (1999)
14. Haralick, R., Lee, C., Ottenberg, K., Nlle, M.: Analysis and solutions of the three-point perspective pose estimation problem. In: IEEE Conf. Computer Vision and Pattern Recognition, pp. 592–598. Maui, Hawai, USA (1991)
15. Haralick, R.M., Shapiro, L.G.: Computer and Robot Vision, vol. 1. Addison-Wesley Publishing (1992)
16. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge Univ. Press (2000)
17. Hong, J., Dohi, T., Hashizume, M., Konishi, K., Hata, N.: An ultrasound-driven needle insertion robot for percutaneous cholecystostomy. Journal of Physics in Med. and Bio. pp. 441–455 (2004)

18. Horaud, R., Conio, B., Leboulleux, O., Lacolle, B.: An analytic solution for the perspective 4-point problem. Computer Vision, Graphics, and Image Processing **47**, 33–44 (1989)
19. Hutchinson, S., Hager, G., Corke, P.: A tutorial on visual servo control. IEEE Trans. Robotics and Automation **12**(5), 651–670 (1996)
20. Kragic, D., Christensen, C.: Cue integration for visual servoing. IEEE Trans. on Robotics and Autom. **17**(1), 19–26 (2001)
21. Krupa, A., Chaumette, F.: Control of an ultrasound probe by adaptive visual servoing. In: IEEE/RSJ Int'l. Conf. on Intelligent Robots and Systems, vol. 2, pp. 2007–2012. Edmonton, Canada (2005)
22. 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 robotized laparoscopic surgery using visual servoing. IEEE Trans. on Robotics and Automation, special issue on Medical Robotics **19**(5), 842–853 (2003)
23. Lee, S., Fichtinger, G., Chirikjian, G.S.: Numerical algorithms for spatial registration of line fiducials from cross-sectional images. American Ass. of Physicists in Medicine **29**(8), 1881–1891 (2002)
24. Marchand, E., Chaumette, F.: Virtual visual servoing: a framework for real-time augmented reality. In: Proceedings of the EUROGRAPHICS Conference, vol. 21 (3 of Computer Graphics Forum), pp. 289–298. Saarbrücken, Germany (2002)
25. Maurin, B.: Conception et réalisation d'un robot d'insertion d'aiguille pour les procédures percutanées sous imageur scanner. Ph.D. thesis, Louis Pasteur University, France (2005)
26. Maurin, B., Doignon, C., de Mathelin, M., Gangi, A.: Pose reconstruction from an uncalibrated computed tomography imaging device. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, WN, U.S.A. (2003)
27. Maurin, B., Gangloff, J., Bayle, B., de Mathelin, M., Piccin, O., Zanne, P., Doignon, C., Soler, L., Gangi, A.: A parallel robotic system with force sensors for percutaneous procedures under ct-guidance. In: Int'l Conf. on Medical Image Computing and Computer-Assisted Intervention. St Malo, France (2004)
28. Maybank, S.: The cross-ratio and the j-invariant. Geometric invariance in computer vision pp. 107–109 (1992)
29. Nageotte, F.: Contributions à la suture assistée par ordinateur en chirurgie mini-invasive. Ph.D. thesis, Louis Pasteur University, France (2005)
30. Nageotte, F., Doignon, C., de Mathelin, M., Zanne, P., Soler, L.: Circular needle and needle-holder localization for computer-aided suturing in laparoscopic surgery. In: SPIE Medical Imaging, pp. 87–98. San Diego, USA (2005)
31. Nageotte, F., Zanne, P., Doignon, C., de Mathelin, M.: Visual servoing-based endoscopic path following for robot-assisted laparoscopic surgery. In: IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, pp. 2364–2369. Beijin, China (2006)
32. Papanikolopoulos, N., Khosla, P., Kanade, T.: Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision. IEEE Trans. Rob. and Autom. **9**(1), 14–35 (1993)
33. Quan, L., Lan, Z.: Linear n-point camera pose determination. IEEE Transactions on PAMI **21**(8) (1999)
34. Sundareswaran, V., Behringer, R.: Visual-servoing-based augmented reality. In: In Proceedings of the IEEE Int'l. Workshop on Augmented Reality. San Francisco, USA (1998)
35. Susil, R.C., Anderson, J.H., Taylor, R.H.: A single image registration method for CT guided interventions. In: Proceedings of the Second Int'l Conf. on Medical Image Computing and Computer-Assisted Intervention, pp. 798–808. Cambridge, UK (1999)
36. Taylor, R., Funda, J., LaRose, D., Treat, M.: A telerobotic system for augmentation of endoscopic surgery. In: IEEE Int'l Conf. on Engineering in Med. and Bio., pp. 1054–1056. Paris, France (1992)
37. Tommasini, T., Fusiello, A., Trucco, E., Roberto, V.: Making good features track better. In: IEEE Int'l Conf. on Computer Vision and Pattern Recognition, pp. 178–183. Santa Barbara, USA (1998)

38. Tonet, O., Ramesh, T., Megali, G., Dario, P.: Image analysis-based approach for localization of endoscopic tools. In: Surgetica'04, pp. 221–228. Chambery, France (2005)
39. Tsai, R.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the shelf tv cameras and lenses. IEEE Journal of Robotics and Automation **3**(4), 323–344 (1987)
40. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on PAMI **13**(4), 376–380 (1991)
41. 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**, 429–437 (1998)
42. Wei, G.Q., Arbter, K., Hirzinger, G.: Automatic tracking of laparoscopic instruments by color-coding. In: S. Verlag (ed.) Proc. First Int. Joint Conf. CRVMed-MRCAS'97, pp. 357–366. Grenoble, France (1997)
43. Zhang, X., Payandeh, S.: Application of visual tracking for robot-assisted laparoscopic surgery. Journal of Robotic systems **19**(7), 315–328 (2002)