3D Tracking of Laparoscopic Instruments Using Statistical and Geometric Modeling*

Rémi Wolf[1], Josselin Duchateau[1], Philippe Cinquin[1], and Sandrine Voros[2]

[1] UJF-Grenoble 1, CNRS
[2] UJF-Grenoble 1, CNRS, INSERM
TIMC-IMAG UMR 5525, Grenoble, F-38041, France

Abstract. During a laparoscopic surgery, the endoscope can be manipulated by an assistant or a robot. Several teams have worked on the tracking of surgical instruments, based on methods ranging from the development of specific devices to image processing methods. We propose to exploit the instruments' insertion points, which are fixed on the patients abdominal cavity, as a geometric constraint for the localization of the instruments. A simple geometric model of a laparoscopic instrument is described, as well as a parametrization that exploits a spherical geometric grid, which offers attracting homogeneity and isotropy properties. The general architecture of our proposed approach is based on the probabilistic Condensation algorithm.

Keywords: laparoscopic surgery, image-based localization of surgical instruments, Condensation algorithm.

1 Introduction

Minimally invasive surgery (MIS) has become more and more popular with surgeons and the public in the recent years, although its superiority on open surgery has not been systematically proven in terms of per and post-operative complications ([8]). Furthermore, the mini-invasive approach is more expensive and more challenging technically for the surgeons than open surgery (loss of depth information, loss of tactile information, limited field of view, inverted movements of the instruments due to their insertion in the patient's abdominal cavity through "fixed" points on the abdominal wall).

These limitations highlight the need to develop further new methodologies dedicated to MIS. These methodologies can be divided in two categories: skills (learning, practicing and evaluating) and tools to perform these skills (end-effectors, endoscopic cameras). New approaches to learn MIS have already been developed, especially using virtual reality simulators [7] and video recordings of surgeries for the evaluation of the surgeon's performance [1]. Extensive research has also been performed on the optimization of laparoscopic tools, the most outstanding being the development of robots, either holding the endoscope inside

* This work has been supported by French National Research Agency (ANR) through TecSan program (project DEPORRA nANR-09-TECS-006).

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 203–210, 2011.

the abdominal cavity (AESOP®, ViKY®) or offering more degrees of freedom than the standard instruments (DaVinci®). Here we will concentrate on the enhancement of the endoscope manipulator. It is bothersome and distracting for the surgeons to control the positioning of the robot manually or by voice. Therefore there has been some research to visually track the tip of the surgeon's instruments and adapt the position of the endoscope to follow them[3,9].

In this paper, we present a new method to localize a laparoscopic instrument from video images in real time, using only a priori knowledge on the geometry of the instrument, the position of its insertion point and a statistical model for the displacement of the instrument based on the Condensation algorithm. Compared to other methods described until today, this algorithm allows us to automatically retrieve the 3D orientation and tip position of the instrument in the camera referential, offering new opportunities such as analysing the motion of the instruments to classify the operating task.

2 Materials and Methods

2.1 General Framework

The goal of our method is to track the motion of a laparoscopic instrument inside the abdominal cavity without modifying it (ie. without a physical tracking system such as infrared or magnetic localizers), using only information from the endoscopic image and a geometric model of the tool.

After an initial calibration procedure, we determine the insertion point of each instrument in 3D using a priori knowledge about its geometry. We then build a geode centered on the insertion point describing the different orientations that can be taken by the tool. A propagation algorithm using particles evolving over the geode's elements according to a determinist and a stochastic law is then computed (the **Con**ditional **Dens**ity Propag**ation** or Condensation algorithm [4]), with measures performed at each time step to guess the most likely orientation of the tool, after reconstruction and projection of the instrument model in the 2D image. Finally, we determine the tip depth along this orientation with an Otsu-based [5] segmentation method over a sliding window.

2.2 System Calibration

Our method requires system calibration: in particular it must be possible to calculate the 3D line in the fixed field of reference (FOR) that corresponds to a 2D image point on the screen. This requires an intrinsic calibration and distortion correction performed using Zhang's procedure[11] with a planar chessboard, which is validated in the medical imaging community.

If the endoscope is displaced by a robot, it is also necessary to know the camera's displacement in the robot's referential, which requires a "hand-eye" calibration of the robotic system. In this paper, we choose to work with a fixed camera, since we concentrate on the presentation of our localization method.

2.3 Instrument Model

Laparoscopic tools can have differents appearances, according to their color (which is not standardized) or to their tip which varies according to their fonction. However, they all have in common a cylindrical shape and a standardized radius, in order to fit in trocars. It is thus possible to model a tool as a finite cylinder of known diameter [2]. Our search, given a predefined insertion point (see section 2.4), is thus limited to its two angular parameters (pan and tilt) that define a direction axis, depicted by a 3D unit vector $\mathbf{e_1}$, and a translational parameter, γ, representing the instrument's depth along its axis (see section 2.6).

The knowledge of the axis of the tool and of its insertion point permits the reconstruction of its borders as seen by the camera (see fig.1, left): the vector $\mathbf{CT}$ is the translation vector representing the instrument insertion point T in the camera's FOR (where C is the camera's center of projection). We can construct a referential linked to the tool by defining two new vectors as follow:

$$\mathbf{e_2} = \frac{\mathbf{T} \otimes \mathbf{e_1}}{\|\mathbf{T} \otimes \mathbf{e_1}\|} \ , \quad \mathbf{e_3} = \mathbf{e_1} \otimes \mathbf{e_2}$$

The instrument axis can be defined by the line (T,$\mathbf{e_1}$) but also by the line (P,$\mathbf{e_1}$), where $\mathbf{CP} = \|\mathbf{T} \otimes \mathbf{e_1}\|\mathbf{e_3}$, as shown in fig. 1, left. The plane (P,$\mathbf{e2}$,$\mathbf{e3}$) is the only right section of the cylinder representing the tool that contains the camera's center of projection C. Working in this right section (fig.1, right), its is possible to define two points S_1 and S_2 that belong to the tangent planes to the instrument running through the camera center. S_1 and S_2 satisfy $\|\mathbf{PS}_{1,2}\| = \rho$, where ρ is the instrument's radius and $\mathbf{CS}_{1,2} \cdot \mathbf{PS}_{1,2} = 0$, i.e. the following system:

$$\begin{cases} \|\alpha\mathbf{e_2} + \beta\mathbf{e_3}\| = \rho \\ (\alpha\mathbf{e_2} + (\|\mathbf{T} \otimes \mathbf{e_1}\| + \beta)\mathbf{e_3}) \cdot (\alpha\mathbf{e_2} + \beta\mathbf{e_3}) = 0 \end{cases} \quad (\alpha, \beta) \in \mathbb{R}^2$$

This system admits two solutions (one for each tangent plane) :

$$(\alpha, \beta) = \left(\pm\rho\sqrt{1 - \frac{1}{\|\mathbf{T} \otimes \mathbf{e_1}\|}}, -\frac{\rho^2}{\|\mathbf{T} \otimes \mathbf{e_1}\|} \right)$$

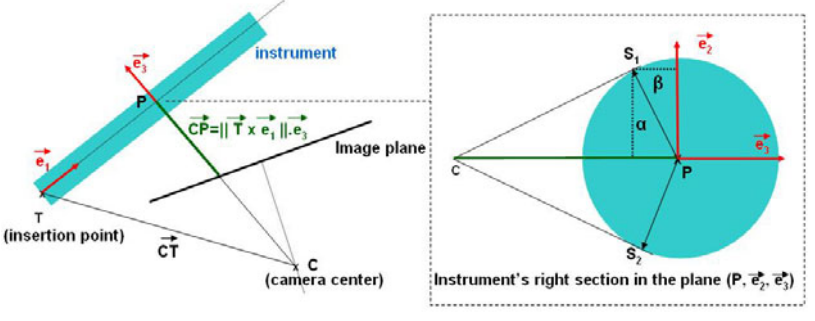

Fig. 1. Description of the tool model

Once these parameters have been extracted, one can easily build in 3D the lines defined by the tangent plane to the instrument passing through the camera center (corresponding to the instrument borders in the 2D image).

$$L = \mathbf{T} \pm |\alpha|\mathbf{e_2} + \beta\mathbf{e_3} + \gamma\mathbf{e_1} \quad \gamma \in \mathbb{R}$$

2.4 Determination of the Insertion Point

Several methods have already been developed in order to retrieve the lines representing the borders of a laparoscopic instrument in 2D images[9,3]. Here we use manual selection of these lines in a sequence of images with one camera position and a moving instrument, which allows us to find (thanks to the calibration and the geometric model of the tool) a beam of 3D lines each representing an instrument's axis. The intersection of these lines, which corresponds to the insertion point, is computed by solving a rank-3 system with a SVD decomposition. We will see in the discussion that this method could easily be automatized.

2.5 Instrument Localization

Our tracking system is based on the Condensation algorithm[4]: it uses factored sampling in which randomly generated sets represent the probability distribution of a certain feature. The random set is propagated using a dynamic model and visual observations, which is of particuliar interest in our situation in which background clutter can often cause direct observation to fail. We therefore need to build a specific dynamic model for our process, along with the corresponding observation model. Given our model construction, it is preferable to detect the instrument 3D position in two steps, determining the angular parameters before the translational parameter, since this parameter only influences the length of the apparent segments, and not their orientation.

Geodesic Geometry. Two parameters are needed to describe the angular orientation of the tool in a 3D space. The typical representation of these parameters would be standard latitude and longitude coordinates (pan and tilt), but these are not well suited to our problem as they are not isotropic (non-uniform distribution) and computational stability near the poles is non-trivial. A solution developed in the field of climate modeling is the use of quasi-uniform spherical geodesic grids [6] among which icosahedral grids give almost homogeneous and quasi-isotropic coverage of the sphere. These grids are obtained by the successive subdivisions of an isocahedron (fig. 2a). Parametrization is easily obtained by dividing the geodesic sphere into 5 panels (fig. 2b) composed almost exclusively of hexagonal cells. Each panel, once rotated and twisted, can be represented as a rectangular array (fig. 2c).

Each 3D unit vector that runs from the center of the sphere to a cell on its surface corresponds to a unique cell on one of the 5 rectangular arrays, as illustrated by the hexagonal cell marked in blue on the sphere in fig. 2a, which corresponds to an unique cell of the rectangular array in fig. 2c. This vector can easily be computed given the 2D position of the cell in the rectangular array, and in our algorithm

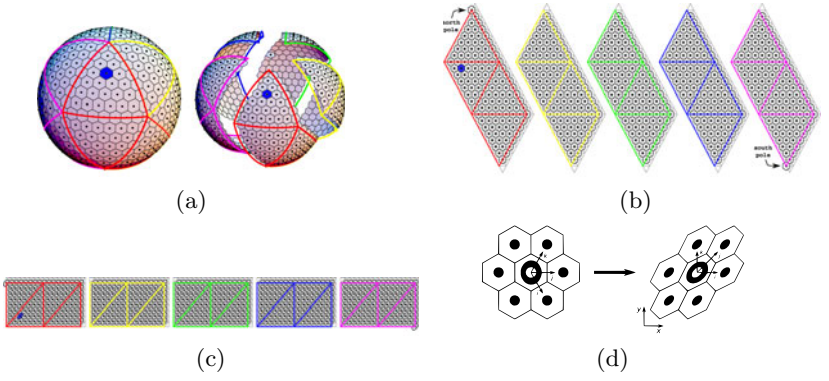

Fig. 2. Representation of the geodesic grid: (a) The spherical geodesic grid composed of hexagonal and pentagonal cells. (b) Cutting the geodesic sphere in 5 panels. (c) Twisted and rotated panels can be represented in rectangular arrays. (d) A particle evolves freely in three independent directions in the hexagonal tiling. Note that $j = i + k$. (images source http://kiwi.atmos.colostate.edu/BUGS/geodesic)

correspondences between 2D cells and 3D vectors are computed and stored in correspondence tables in order to increase speed for the rest of the execution.

Particle Model. In our model, each particle is described at a given time by five variables: the current and previous 2D coordinates of the particle $\mathbf{s_t} = [x_t, y_t, x_{t-1}, y_{t-1}]$ expressed in the panel variable p_t.

Particle dynamics are controlled by a linear stochastic equation, $\mathbf{s_t}$ evolving as:

$$\mathbf{s_t} = A\mathbf{s_{t-1}} + B\mathbf{w_t}, \ \textit{where}\ A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.8 & 0 & 1.8 & 0 \\ 0 & -0.8 & 0 & 1.8 \end{pmatrix}$$

Matrix A represents the deterministic evolution of the particle, $\mathbf{w_t}$ is a vector of standard normal variates and BB^T is the process noise covariance. The coefficients of A have been set empirically for now, describing an exponentially decreasing speed. The determination of B is achieved by noting that probabilities of random motion are isotropic: the correponding probability density function is described in Eq. 1, where i, j, k represent the three directions on the hexagonal grid (see fig. 2d). Fitting the hexagonal grid on the rectangular array parametrized by $x = i$ and $y = k$ coordinates, the probability density function can be expressed as Eq. 2.

$$p(i, j, k) \propto \exp(-(i^2 + j^2 + k^2)) \quad (1)$$
$$p(x, y) \propto \exp(-(x^2 + y^2 + (x + y)^2)) \quad (2)$$

which has a covariance matrix proportional to : $BB^T = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$

Particle dynamics are therefore almost defined for a single panel. The scale factor for the covariance matrix is however unknown and is determined empirically.

Measurements. At each time step, the different cell values are extracted using the instrument model. For each particle, the 3D vector $\mathbf{v}$ that corresponds to the 2D cell it occupies is extracted from the precomputed correspondence table. The segments that correspond to the projection in the 2D image of the borders of a cylinder of direction vector $\mathbf{v}$ passing through the insertion point $\mathbf{T}$ are then extracted according to the instrument model previously described.

The image measurements we use are based on contour detection: image derivatives along the $\mathbf{x}$ and $\mathbf{y}$ axis are approximated by convoluting the image with corresponding Sobel kernels. The two components being extracted, the gradient vector $\mathbf{G_i}$ is available in each image point. We define a score, calculated for each segment : $\sigma_j = \sum_{i=1}^{m_j}(\mathbf{G_i} \cdot \mathbf{n_j})$, where $j = (1, 2)$, $\mathbf{n_j}$ is the unit vector normal to the segment j and m_j is the number of points in segment j. Finally the score $\pi = |\sigma_1 - \sigma_2|$ is given to the cell and the particle (n_1 and n_2 have opposite directions). The instrument axis is then computed as the average of all the particle vectors weighted by their score.

2.6 Determination of the Instrument Depth

Once the instrument axis has been computed, we determine the tip position along this axis by looking for a transition point between the instrument and the background. We developed an algorithm using a sliding window moving along the projection of the instrument's axis in the 2D image. For each position of the sliding window, we find the optimal threshold value dividing the pixels into two distinct classes according to Otsu's method [5]. We consider that the transition between the intrument's body and its tip is the point where the inter-class variance will be maximal. Therefore we store the position of the sliding window for which the inter-class variance is maximal,the position of the middle of this window returning the position of the tip of the instrument.

3 Results

Our first evaluation parameter is the angular error between the theorical instrument's axis and the computed one, in 3D and in the projected image. First, we tested the validity of our method with simulated data. We built a fictive sequence using an abdominal background image on which we pasted a black mask, corresponding to known 3D positions and orientations of the instrument. Its dynamic law was governed by the same equation as the particles (addition of a deterministic and a stochastic evolution).

We then tested our method on a testbench modeling a laparoscopic surgery, and compared our results with 1) manual segmentation of the tool in the image followed by 3D reconstruction in the same fashion as in 2.4 and 2) 3D results obtained using an Optical localizer. Our testbench consisted of a fixed endoscope imaging a surgical instrument inserted through a 5mm trocart, with a background image of an abdominal cavity containing another tool (the blue tool in fig. 4). The angular error for each setup according to the number of particles and geode resolution can be found fig. 3. Fig.4 shows typical image results that we

Measure type	Simulated data		Testbench data			
			Manual segmentation		Optical localizer	
	Mean	Stdv	Mean	Stdv	Mean	Stdv
Best particle (3D)	**9.2°**	**3.5°**	**30.4°**	**14.7°**	**14.8°**	**8.20°**
Particles average (3D)	**8.3°**	**5.0°**	**34.0°**	**11.9°**	**8.6°**	**5.8°**
Best particle (2D)	**2.6°**	**1.7°**	**5.7°**	**3.8°**	**1.76°**	**1.47°**
Particles average (2D)	**2.5°**	**1.6°**	**6.6°**	**5.3°**	**1.77°**	**1.46°**

Fig. 3. Results for 1050 consecutive frames, using 2000 particles and a geode resolution of 163842 cells covering the 3D sphere

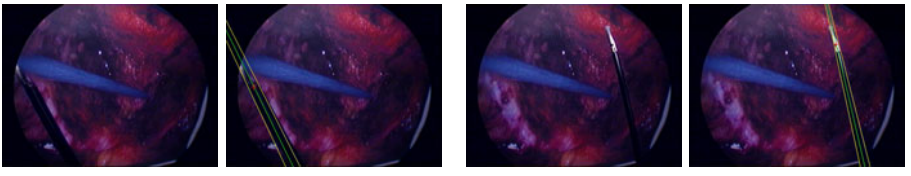

Fig. 4. Examples of the 2D results of an instrument's borders and tip detection in laparoscopic images

obtained on our testbench. Working with 768x576 images acquired by a clinical endoscope (no downsampling was performed), a geode resolution of 163842 cells and 500 or 2000 particles, our algorithm computes at the speed of respectively 16 and 8.6Hz with an Intel Xeon 2.67GHz, 3.48Go RAM PC.

Our second evaluation parameter was the 2D tip position in the image, which was compared to the tip position manually clicked on the images, with the error expressed in pixels. Our mean error is evaluated at 27.8 pixels for the same data set as in fig. 3.

4 Discussion - Conclusion

Our tracking system allows a surgeon to track the motion of his tools during an intervention, in order to either automatically control the movements of the robot, or to evaluate the quality of the surgeon's gesture. As shown by the results, our precision in the 2D orientation of the instrument inside the image is precise, robust and quick enough to allow the first anticipated use. The precision in the 3D orientation and the depth evaluation are still imperfect and should call for several improvements in our algorithm.

Increasing the number of cells in the geode (i.e. its resolution) does not slow down the algorithm as the correspondence table between the cells and 3D orientations is computed only once. However a limit is reached when the order is superior to 128 (163842 cells), as the surface of a cell at a distance of 35cm (the instrument's length) from the insertion point will be 3,13mm^2, which is approximatively the surface of the instrument tip.

Simulated data highlights the intrinsic error of the method, linked to the statistical and dynamical model, as the insertion point remains fixed, when testbench data shows as well the error due to the motion of the insertion point.

We now plan several improvements of our method to improve the 3D orientation detection and ensure its compliance with clinical conditions: first, we plan to allow the insertion point to move freely around its computed position , as it is the case in real surgery (in an amplitude range described in [10]), following a white noise or an evolution law which needs to be determined. Secondly, we plan to detect automatically the insertion point at the beginning of the procedure by extracting lines using the Hough transformation and the same pivot procedure as presented in section 2.4 (time-requesting computation, not adapted to real-time tracking).

Our next step will be to test our method on anatomic specimens mimicking true surgical interventions, in order to improve the deterministic and stochastic matrices describing the evolution of the particles - currently empirically evaluated - and to study their influence on the results. Improvements will also be brought to the tip detection in order to include an evolution model similar to that used for the axis orientation, thus smoothing aberrant measurements which pollute our mean result.

References

1. Aggarwal, R., Grantcharov, T., Moorthy, K., Milland, T., Darzi, A.: Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. Annals of Surgery 247(2), 372 (2008)
2. Cano, A.M., Gayá, F., Lamata, P., Sánchez-González, P., Gómez, E.J.: Laparoscopic tool tracking method for augmented reality surgical applications. In: Bello, F., Edwards, E. (eds.) ISBMS 2008. LNCS, vol. 5104, pp. 191–196. Springer, Heidelberg (2008)
3. 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(5-6), 429–442 (2005)
4. Isard, M., Blake, A.: Condensation—conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)
5. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (1979)
6. Randall, D.A., Ringler, T.D., Heikes, R.P., Jones, P., Baumgardner, J.: Climate modeling with spherical geodesic grids. Computing in Science and Engineering 4, 32–41 (2002)
7. Sherman, V., Feldman, L.S., Stanbridge, D., Kazmi, R., Fried, G.M.: Assessing the learning curve for the acquisition of laparoscopic skills on a virtual reality simulator. Surgical Endoscopy 19(5), 678–682 (2005)
8. Touijer, K., Guillonneau, B.: Laparoscopic radical prostatectomy: A critical analysis of surgical quality. European Urology 49(4), 625 (2006)
9. Uecker, D.R., Lee, C., Wang, Y.F., Wang, Y.: Automated instrument tracking in robotically assisted laparoscopic surgery. Journal of Image Guided Surgery 1(6), 308–325 (1995)
10. Voros, S., Long, J.A., Cinquin, P.: Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders. International Journal of Robotics Research 26(11-12), 1173–1190 (2007)
11. Zhang, Z.Y.: A flexible new technique for camera calibration. IEEE Transactions On Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)