

Automatic instrument localization in laparoscopic surgery¹

Joan Climent and Pere Marés

Computer Eng. & Automatic Control dept. (UPC). Pau Gargallo,5. 08028 Barcelona. Spain

Received 29 April 2004; revised 1 June 2004; accepted 13 July 2004

Abstract

This paper presents a tracking algorithm for automatic instrument localization in robotically assisted laparoscopic surgery. We present a simple and robust system that does not need the presence of artificial marks, or special colours to distinguish the instruments. So, the system enables the robot to track the usual instruments used in laparoscopic operations.

Since the instruments are normally the most structured objects in laparoscopic scenes, the algorithm uses the Hough transform to detect straight lines in the scene. In order to distinguish among different instruments or other structured elements present in the scene, motion information is also used.

We give in this paper a detailed description of all stages of the system.

Keywords – Laparoscopic surgery, Object recognition, Hough transform, Image processing system, real-time tracking.

1. Introduction

Laparoscopic surgery is a minimally invasive surgical procedure. The surgeon inserts instruments and a laparoscope into the patient's body through multiple incisions, and performs the operation viewing the images displayed on a video screen. The main problem of such a technique lies in the difficulties of mutual understanding between the surgeon and the camera assistant. The camera assistant also gets tired in long operations and the image becomes unstable.

Several robotized assistance systems have been developed to deal with these new problems [10] [11] [12]. All different approaches presented in the literature use image processing techniques to track the instrument so that it is always centered in the displayed image. Some works [13] use colour information, but because of the high variability of colours in different scenes, the instruments must be coloured with an artificial colour. Others [12] use instruments that present artificial marks.

¹ This research has been partially supported by The Science and Technology Interministerial Commission CICYT TAP99-1163-C02-01.

We present in this paper a system that tracks instruments that have neither been marked nor coloured. No specific marks are needed; the system works with the usual surgical instruments.

The system is based on a sequence of different image processing techniques. The objective of this paper is to show, in detail, the way that these techniques have been used, and the results obtained.

2. System description

A system overview is shown in figure 1.

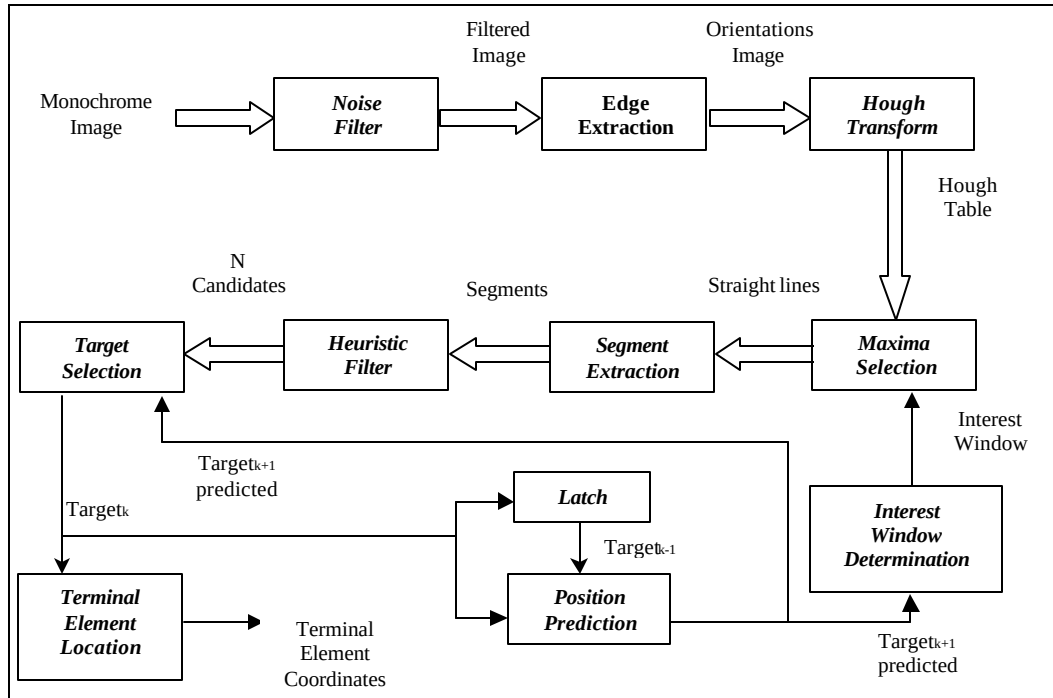


Figure 1. System Overview

The original image is filtered in order to reduce the influence of noise. To extract edge orientations, a classical edge extractor is applied afterwards. The straight lines present in the image are detected using the Hough transform. The most prominent lines are selected. All straight lines that satisfy some heuristic criteria are considered as possible targets.

According to the last position of the instrument in the scene, the best candidate between straight lines selected is chosen, and its ending point marks the position on the terminal element.

Once the actual position of the instrument is determined, its position in the next frame will be predicted. This position determines the location of the processing window of the Hough table for the next frame.

A description of all these stages is detailed in the following subsections.

2.1. Filtering Stage

The effects of noise on the results of the transformation are a matter of concern when dealing with real image data. Such noise may be due to random corruption of the signal during the data acquisition process or it

may be a normally distributed random error in the localization of image points due to the effects of digitizing continuous data. The characterization and prediction of the effects of noise have been studied extensively in [3] and [4].

In order to reduce the effects of noise on the edge orientation determination, a Gaussian filter is applied to the original image in this first stage:

$$h(x, y) = \frac{1}{2\pi s^2} e^{-\frac{(x^2+y^2)}{2s^2}}$$

As a smoothing mask, it has optimal properties in a particular sense: it removes small-scale texture and noise as effectively as possible for a given spatial extent in the image.

Another interesting property of the Gaussian is that it is rotationally symmetric. This means that in all directions smoothing will be the same; there will not be distortion in any direction. Since we are dealing with orientations, this isotropic property is mandatory.

Finally, the Gaussian filter is separable:

$$h(x, y) = \frac{1}{2\pi s^2} e^{-\frac{(x^2+y^2)}{2s^2}} = \frac{1}{\sqrt{2\pi} s} e^{-\frac{x^2}{2s^2}} \cdot \frac{1}{\sqrt{2\pi} s} e^{-\frac{y^2}{2s^2}} = h_{1D}(x) \cdot h_{1D}(y)$$

A 2D Gaussian convolution can be implemented using two orthogonal 1D Gaussian convolutions. Thus, the computational cost of the filtering stage is linear instead of quadratic.

Figure 2.b shows the image output after the Gaussian filtering with a standard deviation $\sigma=1.5$ and a kernel size of 7.

2.2. Edge orientation extraction

Edge orientations are needed to compute the Hough transform. The prior filtering stage is mandatory since the precision in the orientation of gradient operators is very sensitive to noise.

For extracting edge orientation a simple gradient operator is used. Given the filtered image $f(x,y)$, an approximation of the gradient direction $\theta(x,y)$ is computed as:

$$q(x,y) = \text{atan} \frac{\Delta y}{\Delta x}$$

where $\Delta y = f(x, y-1) - f(x, y+1)$ and $\Delta x = f(x-1, y) - f(x+1, y)$

The Sobel operator is not used in this stage. The justification is quite simple. The Sobel operator performs local smoothing. In our application there is no need of new smoothings since the image has been filtered previously. So, the computational load may be reduced by using the central difference masks instead.

The computational load is also reduced by considering only pixels whose gradient magnitude is above a certain threshold, Th . Figure 2.c shows the orientation image. Orientations have been quantified in 256 levels.

2.3. Hough transform computation

Paul Hough introduced the Hough transform in 1962 [1]. It is known that it gives good results in the detection of straight lines and other shapes even in the presence of noise and occlusion.

Our vision system detects the surgical instruments using the Hough transform. Since the instruments show a structured shape, mainly straight lines, the Hough transform is a powerful tool to detect them. There can be found in the literature other tracking applications that also use the Hough transform; see, for example [2]. Other works within the medical imaging discipline that make use of the Hough transform include [6][7][8].

At this stage, the normal parameterization of the Hough transform is used to extract the most significant straight lines in the scene.

$$x \cos q + y \sin q = r$$

where r and q are the length and orientation of the normal vector to the line from the image origin. Each straight line is uniquely defined by r and q , and for every point in the original image (x, y) it is possible to create a mapping from feature to the parametric space.

If we divide the parameter space into a number of discrete accumulator cells, we can collect 'votes' in the (r, q) space from each data point in the (x, y) space. Peaks in (r, q) space will mark the equations of lines of co-linear points in the (x, y) space.

Interested readers can find a good survey about Hough transform in [5]. A book which makes easily assimilated theory and advice available to the nonspecialist concerning state of the art Hough transform techniques is [9].

For every pixel in the image, the gradient direction has been determined in the last stage. Thus, the computation of distance r , becomes a single operation. Edge direction information made available at the edge detection stage is the most commonly used constraint on the range of parameters to be calculated [14].

The (r, q) space has been implemented using a 256x256 array of accumulators. All pixels, except those whose gradient magnitude is below the threshold Th , are mapped to one point in the (r, q) space. The corresponding cells in the accumulator are incremented every time a new pixel is mapped into it. Figure 2.d shows a 3D representation of the Hough table in the (r, q) space.

Peaks in the (r, q) space correspond to the presence of straight lines in the scene. The maximum peak is selected as the longest straight line in the image.



(a)



(b)

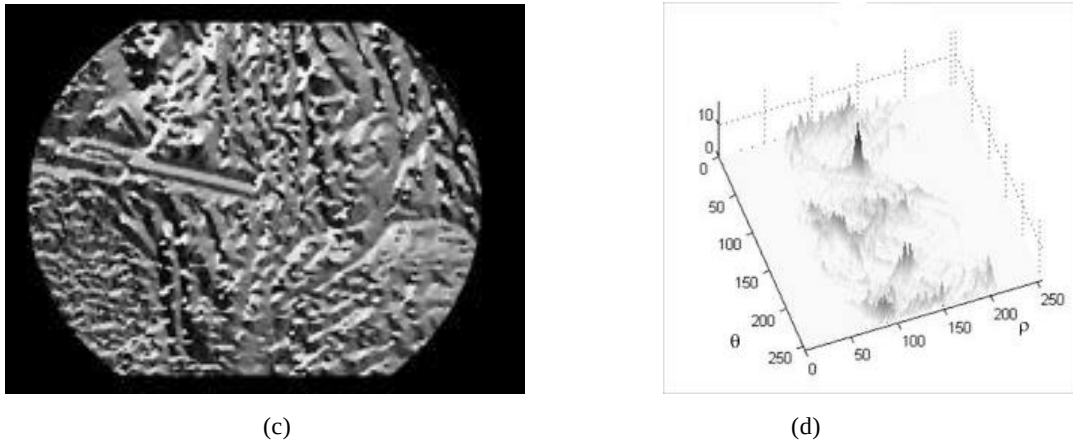


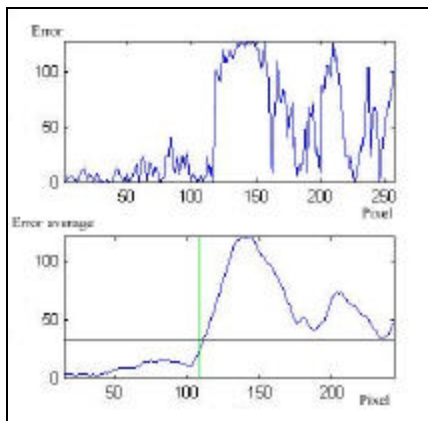
Figure 2. (a). Original image. (b). Filtered image. (c). Gradient orientations. (d). 3D representation of Hough table

2.4. Segment Extraction

One useful property of the Hough transform is that the pixels that lie on the line do not need to be contiguous. On the other hand, it can also give misleading results when objects happen to be aligned by chance. Some pixels are not a part of surgeon's tool but noise pixels or some other edge with the same line components.

The tool ending is determined by the loss of continuity in gradient direction along the straight line. Pixels along the straight line are traced until their orientations present a significant change with respect to the line orientation. For every pixel on the line we compute the difference between its gradient direction and line direction. This difference is considered as a local error. Then this error is averaged along the line. Figure 3.a shows the error in the orientations of all pixels along the main straight line. Part of this line belongs to the surgeon's tool. It can be seen that the continuity in orientations is lost when the tool ends.

Figure 3.b shows, in white, pixels selected as belonging to the tool, and, in black, those belonging to the background. The tool ending can then be located once the pixels not belonging to the correct segment have been removed.



(a)

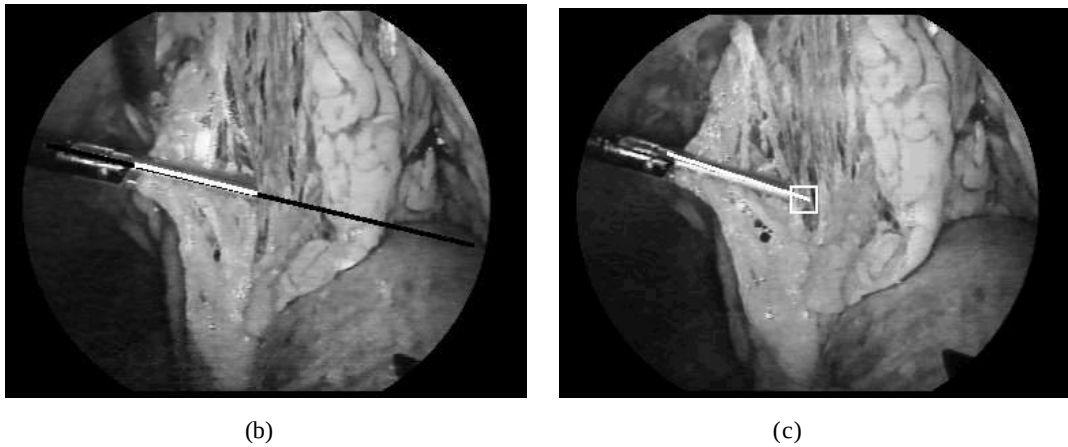


Figure 3. (a). Error line profile. (b). Segment extraction. (c). Tool ending location.

2.5. Heuristic Filter

The longest line in the scene does not always belong to the instrument. Thus, some extra information is needed to decide which one among the possible candidate lines is the most suitable. Figure 4 shows the ten longest straight lines in the image. A heuristic filter and motion information must be used to reject false candidates.

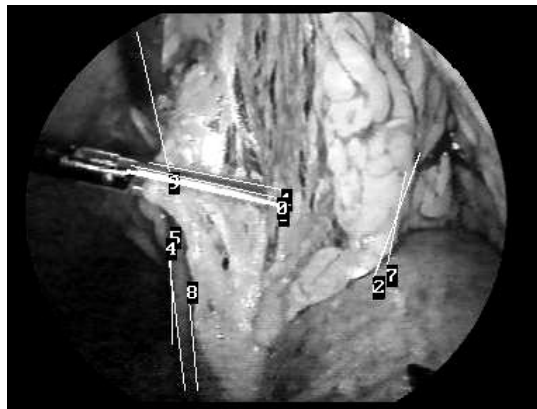


Figure 4. 10 candidates, the longest straight lines in the image.

Some heuristic information is used in order to reject false candidates.

The length of the straight line must be greater than a minimum fixed value.

Instruments always come into scene from out of the field of view of the camera. Thus, straight lines must finish in a boundary of the scene.

Since the tracking system focuses the area of interest in the centre of the image, only radial lines are candidates to be selected as surgeon's tools.

All candidates that do not satisfy these conditions are automatically rejected.

2.6. Position prediction

Given the actual and last frame positions of the instrument, the next position is predicted. A first order model is used. First, the speed of the instrument v_k is computed from two consecutive positions x_{k-1} and x_k , with being $x_k = (r, q)$ the position in Hough table at the k-th iteration:

$$v_k = x_k - x_{k-1}$$

Afterwards, the next position x_{k+1} is estimated:

$$\hat{x}_{k+1} = x_k + v_k T$$

where T is the sample period between two processed frames. We process a frame in 125 ms.

There must be certain continuity between the location of the peak in the (r, q) space of the current frame and the location of the peak in the last frame. Thus, not all the Hough table must be computed at every iteration, we only process those pixels whose (r, q) pair is within a window of acceptable values. Once the next position $\hat{x}_{k+1}$ has been estimated, the processing window is centered at this new position for the next frame. The objective of this processing window is, of course, to increase the processing speed. The size of the processing window is programmable. In our application it has been fixed to ± 20 degrees.

The estimated position is also used to select the best target from the list of candidates.

2.7. Target selection

The position $\hat{x}_{k+1}$ estimated in the last stage, will determine which one of the remaining candidates is the most probable. The error between the positions of all candidates, and the estimated position is then computed:

$$\Delta q_i = q_i - \hat{q}_T, \quad \Delta r_i = r_i - \hat{r}_T$$

Once these errors are computed for all candidates, the one closest to the estimated position and with the highest value in the Hough transform accumulator is selected. The function used is:

$$d_i = \frac{\text{valHough}_i}{\max(\text{valHough})} \cdot \left(1 - \frac{|\Delta q_i|}{\max|\Delta q_i|} \right) \cdot \left(1 - \frac{|\Delta r_i|}{\max|\Delta r_i|} \right)$$

where $\max(\text{valHough})$ is the absolute maximum of the Hough table values.

3. Results

Our vision system has been implemented on a PC system with a 1.7Ghz. Pentium III processor and a commercial image acquisition board. Processing time for the complete process is 125 ms., which is suitable for a real-time tracking application. The size of the processed images is 768x576 pixels.

Some parameters are configurable and must be tuned by the user. We show in table 1 the values assigned in our application:

Parameter	Description	Value
Th	Minimum gradient module	10
$\max \Delta q_i $	Maximum variation of θ between consecutive frames	20
$\max \Delta r_i $	Maximum variation of ρ between consecutive frames	35

Table 1. Parameter values

Table 2 shows the results of an experiment designed to show the percentage of cases in which the correct tool corresponds to the straight lines detected. A set of 128 images extracted from a real operation video have been used for the experiment. The experiment has been divided in two stages: the first one is a static test, and uses only the information provided by the image itself. The second one is a dynamic test, and it takes into account the information obtained from the previous images of the video sequence. The static test uses only the information given by the Hough transform. The dynamic test uses the static information plus the position prediction detailed in section 2.6.

For each image, the ten top values of the Hough table are selected and sorted by decreasing order. The coordinates (r, q) of these maxima in the Hough table correspond to ten different straight lines in the image. The objective of the experiment is to show when the correct surgeon's tool straight line corresponds with the lines detected in the Hough transform stage. The first column of table 2 are the longest straight lines in the Hough transform table, the second are the probabilities that these lines correspond with the correct tool.

Longest lines in scene	% of correct identifications
1 st	77 %
2 nd	11 %
3 rd	4%
4 th	4%
5 th	2%
6 th	1%
7 th	1%
8 th	0 %
9 th	0 %
10 th	0 %

Table 2. Static test results

The dynamic test takes into account the information obtained from the previous frames in the sequence, a position prediction (described in section 2.6) and a target selection (described in section 2.7) are performed in order to detect the tool. Using dynamic information the rate of correct detections goes up to 99%.

Finally, figures 5 and 6 show the results obtained with other different images.

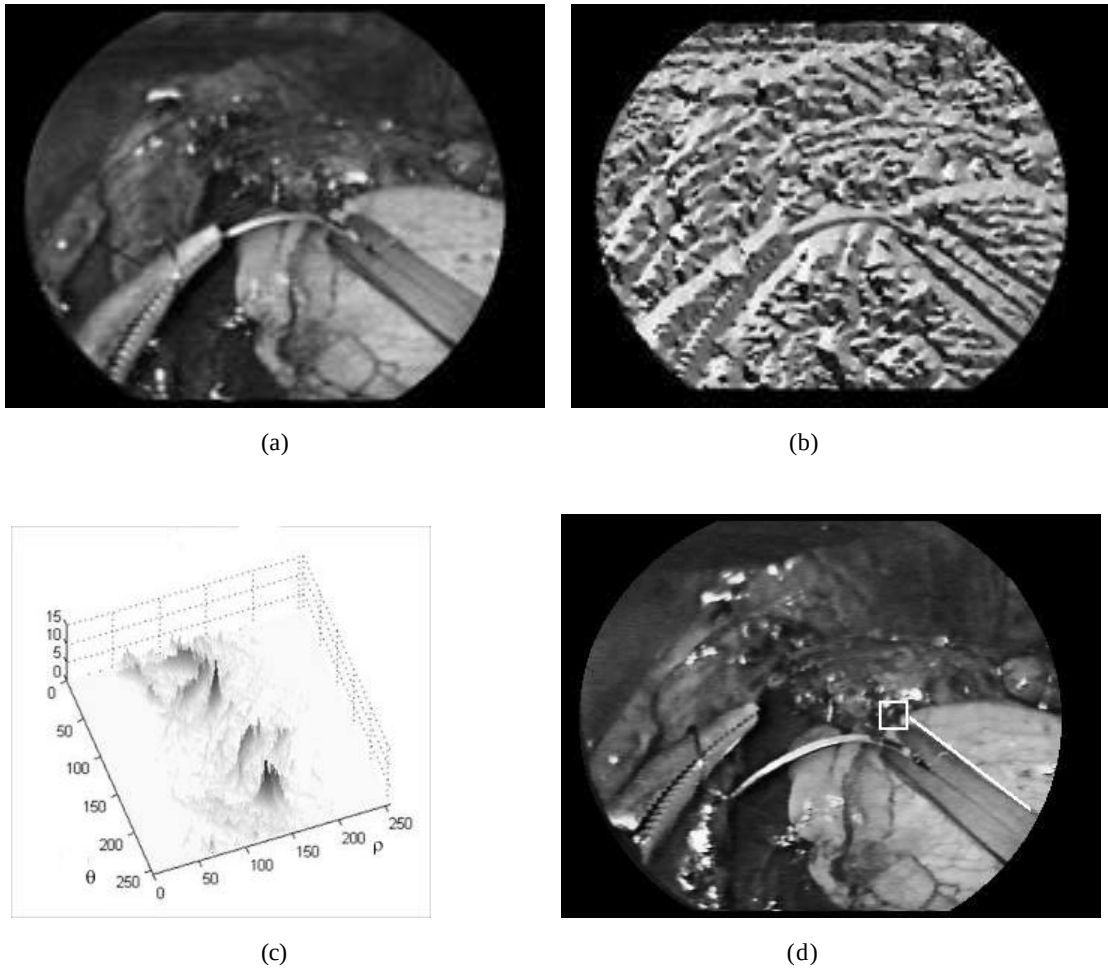
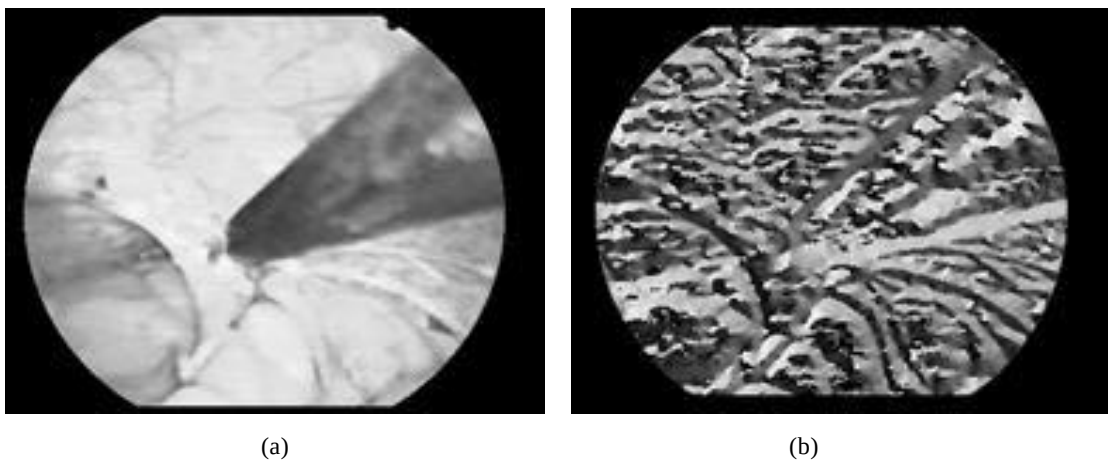
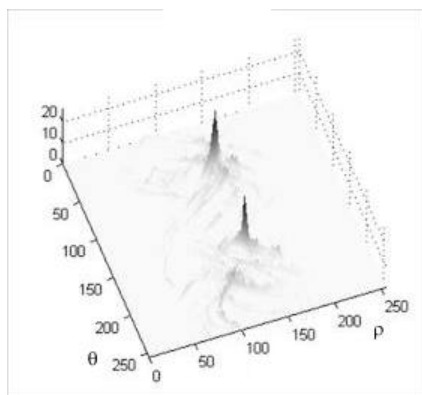
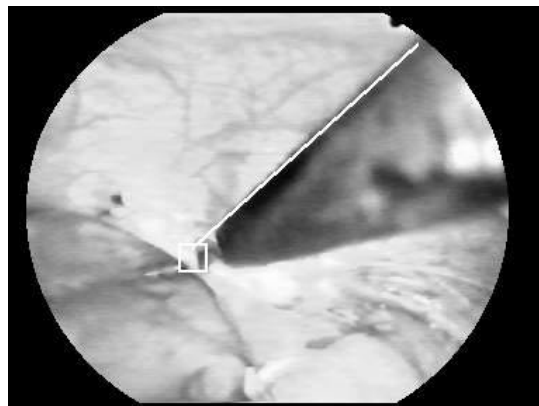


Fig 5. (a). Original image. (b). Gradient orientations. (c). 3D representation of Hough table (d). Tool ending location





(c)



(d)

Fig. 6 (a). Original image. (b). Gradient orientations. (c). 3D representation of Hough table (d). Tool ending location

4. Conclusion

We have presented a detailed description of all stages of a vision system. This system performs a real-time tracking of the surgical tools in laparoscopic operations. The method presented uses the Hough transform to detect the presence of structured objects in the scene. This technique has permitted the system to work without colour restrictions or special marks on the instruments.

The system can track tools whose orientations are within a 20 degrees interval between two consecutive frames. Since eight frames are processed per second, this means that angular speed of the tracked tool must be below 160 degrees per second.

We get some false detections in some conflictive cases, due to:

- Tool goes out of the field of view. We get a very short straight line when the tool progressively goes out from the visible area. This problem will be solved when the robotic arm closes the loop and surgeon tools are always into the visible area.
- Low contrast between the tool and the background. This problem is present in areas with bad quality illumination. These areas are not suitable to perform any surgical task.
- Sudden movements of the tool. The constraints shown in table 1 must be respected. Anyway, it is not recommended that the robotic arm make sudden movements in a live surgical operation. The tracking system must be inhibited when such movements occur.

Acknowledgement

The authors thank Dr. Silvina Thevenet, from the *Universidad Nacional del Centro in Tandil*, for her very efficient and pleasant cooperation.

References

- [1] P.V.C. Hough, *Method and Means for Recognising Complex Patterns*, U.S. Patent No. 3069654, 1962.

- [2] Y.Ermolin and C.Ljustin, *The use of image processing in tracking*, Nucl. Instrum. Methods Phys. Res. Sect. A Vol A 289(3), pp.592-596, 1990.
- [3] W.E.L. Grimson and D.P.Huttenlochner, *On the sensitivity of the Hough transform for object recognition*, IEEE Pattern Anal. Mach. Intell. 10(3), pp.255-274, 1990.
- [4] D.J.Hunt, L.W.Nolte, A.R.Reibman, and W.H.Ruedger. *Hough transform and signal detection theory performance for images with additive noise*, Comput. Vision Graphics Image Process. 52(3), pp. 386-401, 1990.
- [5] V.F.Leavers, *Which Hough transform?*, CVGIP: Image Unders., 58(2), pp.250-264, 1993.
- [6] M.S.Nixon, T.K.Hames, P.Martin, S.Powell, and S.de Pavia, *3-D arterial modelling using feature extraction in ultrasound images*, Int. Conf. on Image Processing and its Applications, Maastricht, Netherlands, pp. 373-376, April 1992.
- [7] M.S.Nixon and T.K.Hames, *New technique for 3D artery modelling by noninvasive ultrasound*, IEE Proc. I Commun. Speech and Vision 140(1), pp. 86-94, 1993.
- [8] A.D.H. Thomas, T.Davies and A.R.Luxmoore, *The Hough transform for locating cell nuclei*, IEE Colloquium on Applications of Image Processing in Mass Health Screening, London, pp. 81-84. March 1992.
- [9] V.F.Leavers, *Shape Detection in Computer Vision Using the Hough Transform*, Springer-Verlag, New York/Berlin, 1992.
- [10] R.Hurteau, S.DeSantios, E.Begin, M.Gagner, *Laparoscopic Surgery assisted by a robotic cameraman: Concept and experimental results*, Proc.IEEE Int. Conf. Robotics and Automation, pp.2286-2289, San Diego, May 8-13, 1994.
- [11] R.H.Taylor, J.Funda, B.Eldridge, S.Gomory, K.Gruben, *A telerobotic assistant for laparoscopic surgery*, IEEE Engineering in Medicine and Biology, 14 (3), pp.279-288, 1995.
- [12] A.Casals, J.Amat, D.Prats, E.Laporte, *Vision guided robotic system for laparoscopic surgery*, Proc.Int. Conf. Advanced Robots, pp.33-36, Catalonia, 1995.
- [13] G.Wei, K.Arbter, G.Hirzinger, *Real-time visual servoing for laparoscopic surgery*, IEEE Engineering in Medicine and Biology, 16 (1), pp.40-45, 1997.
- [14] D.H.Ballard, *Generalizing the Hough transform to detect arbitrary shapes*. Pattern Recognit. 13(2), pp. 111-122, 1981.