

2D/3D Real-Time Tracking of Surgical Instruments Based on Endoscopic Image Processing

Anthony Agustinos¹(✉) and Sandrine Voros²

¹ UJF-Grenoble 1, CNRS, TIMC-IMAG UMR 5525, Grenoble 38401, France
Anthony.Agustinos@imag.fr

² UJF-Grenoble 1, INSERM, TIMC-IMAG UMR 5525, Grenoble 38401, France
Sandrine.Voros@imag.fr

Abstract. This paper describes a simple and robust algorithm which permits to track surgical instruments without artificial markers in endoscopic images. Based on image processing, this algorithm can estimate the 2D/3D pose of all the instruments visible in the image, in real-time (30 Hz). The originality of the approach is based on the use of a Frangi filter for detecting edges and the tip of instruments. The accuracy of the instruments' location in the image is evaluated using an extensive dataset (1500 images, 3 laparoscopic surgeries). Pose estimation of instruments in space is quantitatively evaluated on a test bench through comparison with the ground truth positioning provided by a calibrated robotic instrument holder. This method opens perspectives in the real-time control of surgical robots and the intra-operative recognition of surgical gestures.

Keywords: Laparoscopy · Image processing · Surgical instruments · Real-time tracking

1 Introduction

Laparoscopic surgery is a minimally invasive procedure. This technique reproduces the principles of conventional surgery with minimal physical trauma. Compared to open surgery, this approach is more beneficial to the patient but significantly increases the complexity of the surgical gestures. The constraints for surgeons are mostly ergonomic with the manipulation of surgical instruments (reduction of instrument mobility due to fixed insertion points on the abdominal cavity, loss of tactile sense) and the visualization of the surgical scene (limited field of view, indirect view of the surgical scene, endoscope manipulation). The realization of a laparoscopy requires a large adaptability from surgeons and requires a long learning curve.

Electronic supplementary material The online version of this chapter (doi:[10.1007/978-3-319-29965-5_9](https://doi.org/10.1007/978-3-319-29965-5_9)) contains supplementary material, which is available to authorized users.

Automatic localization of instruments can be helpful to respond to several limitations of laparoscopy and to assist surgeons during an intervention. For instance, [1] propose to localize instruments in space in a surgical trainer, based on a projective model and gradient image processing. In [2], a similar approach is proposed (also in a surgical trainer), with the addition of an extended Kalman filter to extract the edges of instruments.

In [3], the authors use the instrument insertion point as a constraint and a probabilistic algorithm to find instruments with the aim of controlling a robotic endoscope holder to assist surgeons during surgery.

All these methods use a gradient approach to extract instrument edges in the image. However, such approaches are sensitive to noise, illumination and shadows that can lead to insufficient segmentation for robust localization of instruments in the image [4]. To overcome this problem, we propose to use a 2D Frangi filter [5] to obtain a robust instruments edge detection. We present an algorithm to localize and track surgical instruments in endoscopic images in real-time. Our algorithm also permits to estimate the 3D position and orientation of the instruments using 2D information in the images, knowing the camera and instrument models.

2 Instrument Localization and Tracking Framework

The principle of our instrument detection algorithm consists in:

- roughly identifying all regions corresponding to the location of an instrument in each laparoscopic image Sect. 2.1,
- refining the instruments detection within the identified regions Sect. 2.2,
- estimating the 3D pose of the instrument Sect. 2.3.

After an initial detection, the segmentation is constrained by the localization in the previous images to track the instrument.

2.1 Rough Extraction of Instruments Regions

First, the laparoscopic color image (Fig. 2a) is converted from the RGB color space to the CIE Lab color space. The L channel, corresponding to the luminance is removed to free ourselves from variations of light inherent to laparoscopic surgery. We thus obtain a grayscale image composed of the a and b channels (Fig. 2b) corresponding to the chromaticity $C_{ab} = \sqrt{a^2 + b^2}$. Using this color space is more robust for challenging images than color spaces commonly used such as HSV [7] or RGB, see Fig. 1. We then binarize this grayscale image using an automatic Otsu thresholding approach [8]. Since the laparoscopic instruments have a color very distinct from the background (laparoscopic tools are usually black, metallic, or blue/green), instrument pixels will appear as white pixels whereas background pixels will appear as black (Fig. 2c). Of course, this pre-processing step is noisy, with background pixels appearing as white and tool pixels appearing as black (Fig. 2c). We disconnect the regions by skeletonizing

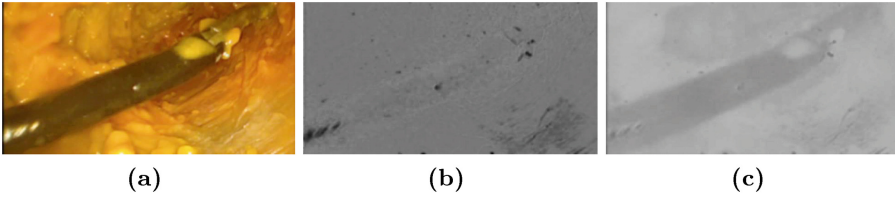


Fig. 1. Typical images obtained by color to grayscale conversion. (a) Original image (b) Saturation modified channels in HSV space [7] (c) Chromaticity C_{ab} of CIELab space (Color figure online)

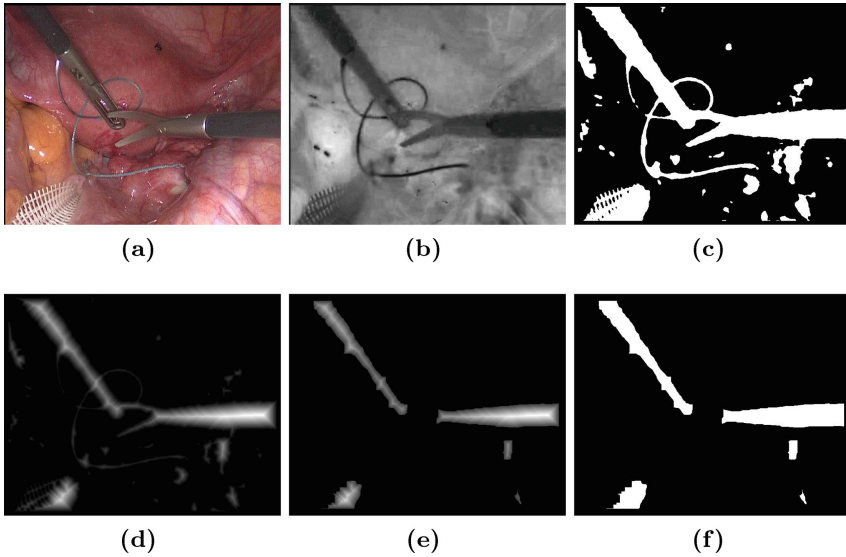


Fig. 2. Segmentation of a surgical instrument in 2D images. (a) Original image (b) Chromaticity C_{ab} of CIELab space (c) Segmentation using Otsu's thresholding (d) Conversion of the binary image using the distance transform (e) Disconnection of regions in the binary image using distance transform (f) Binarization of the distance transform image (Color figure online)

the image using a simple distance transform [9] and refine the separation by performing a simple erosion step on a cross-shaped kernel (Fig. 2d). Finally, we use a contour detection algorithm [10] to extract the extreme outer contour of each region as an oriented bounding box (see Fig. 3b). Based on the observation that laparoscopic instruments have a long and thin cylindrical shape, we eliminate bounding boxes with a width/length ratio inferior to 2 (red boxes in Fig. 3c).

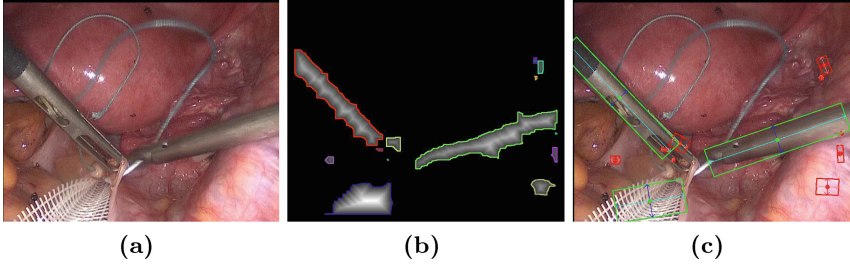


Fig. 3. Edge detection of a surgical instrument in 2D images. (a) Original image (b) Edge detection (c) Potential instrument bounding boxes obtained from image b (green) and incompatible bounding boxes (red) (Color figure online)

2.2 Fine Extraction of Instrument Edges

Now that we have potential bounding boxes for the instruments, we search for instrument edges within each bounding box. To do so, we use a Frangi filter [5], which is the major contribution of this paper. We compared the Frangi filter to the classical Canny filter [6] to search instrument edges (see Fig. 4). The Canny filter is the most classical gradient approach based on the Sobel filter. This filter uses a hysteresis thresholding that requires to find two optimal thresholds for accurate extraction of the edges of an instrument. However, as shown in Fig. 4, the conditions of the surgical scene evolves during an intervention, thresholds initially determined may no longer be optimal and cause of false detections. The advantage of the approach based on the Frangi filter is that it can be applied to different surgery conditions without adjusting the filter parameters. This filter is classically used in vessel detection in medical images. It is based on the computation of the eigenvalues of the image's Hessian matrix λ_1, λ_2 such that $|\lambda_1| \leq |\lambda_2|$. The Hessian matrix is obtained by convolving the image with derivatives of a Gaussian kernel with standard deviation σ .

The Frangi filter function can be defined as:

$$\begin{cases} 0 & \text{if } \lambda_2 > 0, \\ V_0 = \exp\left(-\frac{R_B^2}{2\beta^2}\right)\left(1 - \exp\left(-\frac{s^2}{2c^2}\right)\right) & \end{cases} \quad (1)$$

where, $R_B = \frac{\lambda_1}{\lambda_2}$ is the blobness measure, $s = \sqrt{\lambda_1^2 + \lambda_2^2}$ is the structureness measure and c, β are parameters to adjust the filter sensitivity. After applying the Frangi filter, each pixel value V_0 of the image indicates the pixel's probability of belonging to a tubular structure. Here, we do not use the Frangi filter to extract the whole cylindrical shape of the instrument. Indeed, the instrument's diameter in the image varies depending on its relative orientation with respect to the endoscope (i.e. we cannot fix the standard deviation σ). Instead, we apply the filter with a very low σ , in order to highlight the instrument edges (Fig. 5b). Finally, we identify the two borders of an instrument: the bounding box is extended and separated into two areas to search the top and bottom borders of the instrument separately using Hough transform [11] with a very low

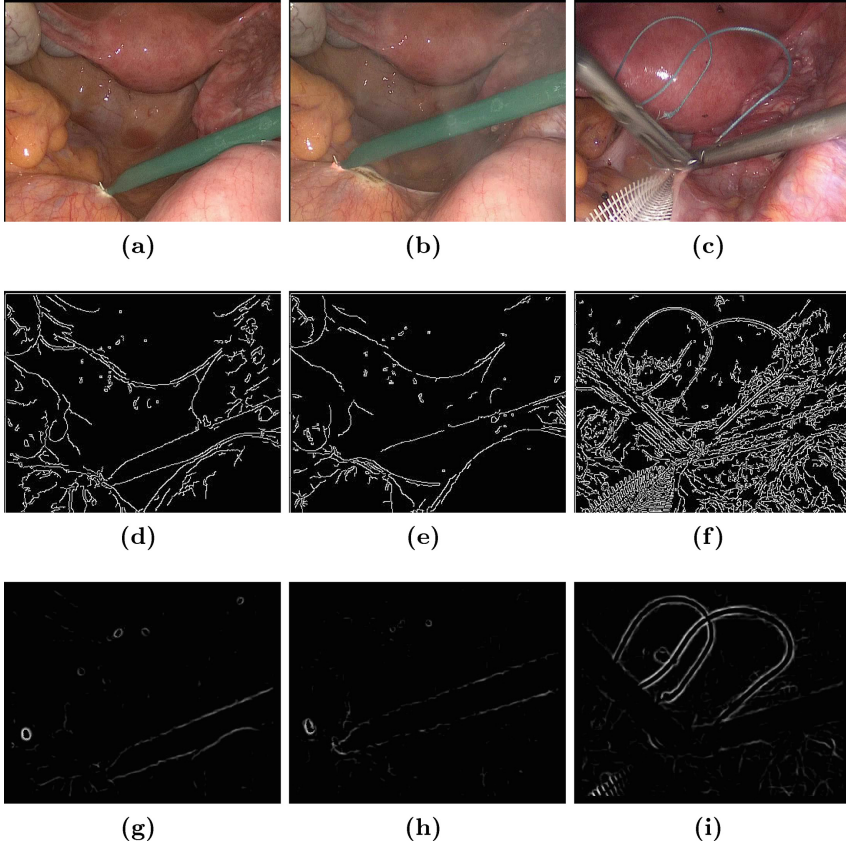


Fig. 4. Extraction of instrument edges. (a) and (b) Two images extracted from the same surgery at different time intervals (c) Image extracted from another surgery (d), (e) and (f) Edge detection in the images (a), (b) and (c) by the Canny filter with the thresholds $T_L = 30$ and $T_H = 90$ (h), (i) and (j) Edge detection by the Frangi filter in the images (a), (b) and (c) with the parameters $\sigma = 2$, $\beta = 0.5$ and $c = 0.5\max(S)$

threshold, as illustrated in Fig. 5b. At this step, we can eliminate lines that are incompatible with a surgical instrument based on the relative orientation and position of the detected lines (as illustrated by Fig. 5c).

2.3 Estimation of 3D Pose of the Instruments

The two borders of an instrument define two tangent planes \sum_i of normal \mathbf{n}_i passing through the optical center of the camera \mathbf{C} in space (see Fig. 5g). The camera calibration can be obtained with a classical chessboard calibration procedure such as [12]. The intersection of these two planes is a line $\mathbf{D} : (\mathbf{C}, \mathbf{e}_1)$ parallel to the central axis of the instrument passing through the optical center of the camera with a direction vector \mathbf{e}_1 . This line defines the instrument's central axis

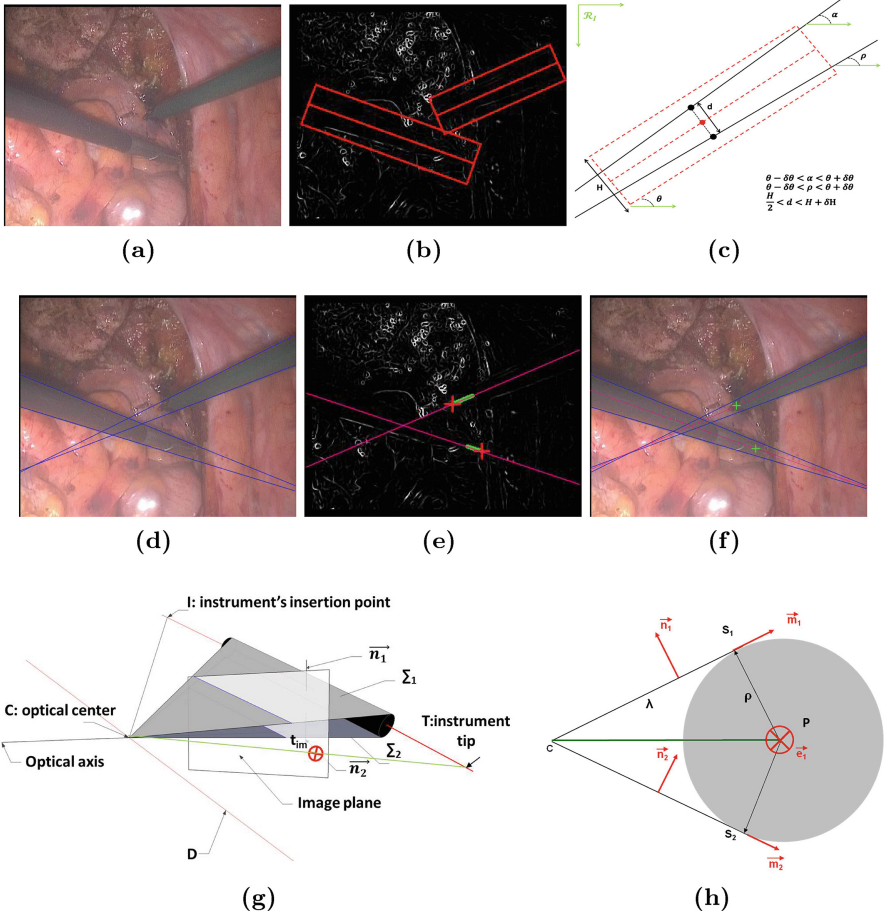


Fig. 5. Estimation of instruments poses in the image and in space. (a) Original image (b) Expansion and separation of a compatible bounding box in image filtered by the Frangi filter (c) Instruments' borders refinement process (d) Detection of the instruments borders (e) Instruments tips detection in the Frangi image (f) Instruments' pose in the image (g) Geometric representation of an instrument in space (h) Illustration of the compute instrument's position in space

direction in space. In order to fully describe the tool's orientation in space, we need to find a point \mathbf{P} on the instrument's axis. To do so, we follow the approach proposed in [3]: the instrument is modeled as a finite cylinder of radius ρ (see Fig. 5g). Such a point \mathbf{P} can be easily computed on the plane perpendicular to the instrument's axis (Fig. 5h). Indeed, \mathbf{P} must respect the condition:

$$\lambda \mathbf{m}_1 - \rho \mathbf{n}_1 = \lambda \mathbf{m}_2 + \rho \mathbf{n}_2 \quad (2)$$

where $\mathbf{m}_i = \mathbf{e}_1 \otimes \mathbf{n}_i$, λ is the distance from the optical center to tangent points \mathbf{S}_i and \mathbf{n}_i the normal to the plane \mathbf{i} . Using Eq. 2, we can compute λ and obtain:

$$\overrightarrow{\mathbf{CP}} = \lambda \mathbf{m}_1 - \rho \mathbf{n}_1 = \rho \frac{\|\mathbf{n}_1 + \mathbf{n}_2\|^2}{(\mathbf{m}_1 - \mathbf{m}_2) \cdot (\mathbf{n}_1 + \mathbf{n}_2)} \mathbf{m}_1 - \rho \mathbf{n}_1 \quad (3)$$

Then, we search the position of the instrument's tip \mathbf{t}_{im} , in the Frangi image along the projection of the instrument's axis $(\mathbf{P}, \mathbf{e}_1)$ in the image (see Fig. 5e). The pixel along the line with maximum grey level in the Frangi image is considered as the tip. Finally, we find the 3D position of the instrument's tip \mathbf{T} as the intersection of $(\mathbf{P}, \mathbf{e}_1)$ and the projection line of the tool's tip $(\mathbf{C}, \mathbf{t}_{\text{im}})$.

2.4 Tracking of Surgical Instruments

For our instrument tracking algorithm, we assume that between two successive images, an instrument does not undergo large displacements. In the initial step (first image), we find the instrument as described in Sects. 2.1 and 2.2. In the following images, we find the candidate bounding boxes, but we refine the instrument search only inside the bounding box best compatible with the position/orientation of the instrument in the previous image. If the instrument is not found in several images, we re-initialize the algorithm. In the case of several instruments, it is possible to track all the visible instruments or a particular one. Since only one instrument can be inserted at once through an insertion point \mathbf{I} on the abdominal wall, we can identify an instrument thanks to its insertion point, which can be easily computed using a pivot algorithm on $(\mathbf{P}, \mathbf{e}_1)$.

3 Experiments and Results

Our algorithm is implemented in C++ using OpenCV and OpenMP libraries. For the computations, we used an Intel Xeon PC 2.67 GHz, 3.48 GB RAM. The 2D evaluation was performed on real laparoscopic images (720×556). The 3D evaluation was performed on a laparoscopy test bench using an OLYMPUS OTV600 CCD and an IC Imaging Source grabber (720×480 , 25 fps). To achieve a fast processing time the image resolution is divided by 2 for the region extraction and by 4 for the Frangi filter. We evaluated 2D tracking of our algorithm on three in-vivo video sequences of laparoscopic rectopexies obtained through the Digestive Departement of Grenoble Hospital with challenging situations (see Fig. 6).

In these images, the tip position and orientation of the instrument were compared to manual annotation. The results obtained for each sequence are presented in Table 1 with a mean error of 16.10 pixels (std. dev. of 28.98) for the tip position, a mean error of 0.90° (std. dev. 0.88°) for the 2D orientation and a frequency of 30 Hz. Videos of this evaluation are included in supplementary material.

To evaluate the accuracy of the 3D pose estimation, we performed experiments on a testbench (see Fig. 7) consisting of a surgery trainer box on which a

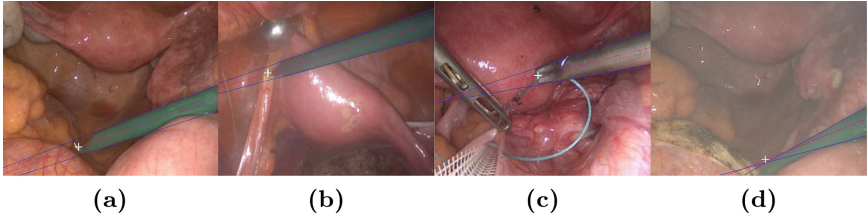


Fig. 6. Results of the tracking on the laparoscopic image sequences. (a) Sequence 1: Monopolar hook instrument (b) Sequence 2: Monopolar hook instrument (c) Sequence 3: Needle holder instrument (d) Example of a bad tip detection

Table 1. Laparoscopic images statistics

Instruments	# Images	Fps	Position error (pixels) Mean (Std. Dev.)	Orientation error (°) Mean (Std. Dev.)
Sequence 1	500	31.47	10.02 (11.2)	0.71 (0.55)
Sequence 2	550	29.76	15.46 (9.72)	0.75 (0.63)
Sequence 3	525	28.87	22.83 (47.14)	1.25 (1.19)
All	1575	30.03	16.10 (28.98)	0.90 (0.88)

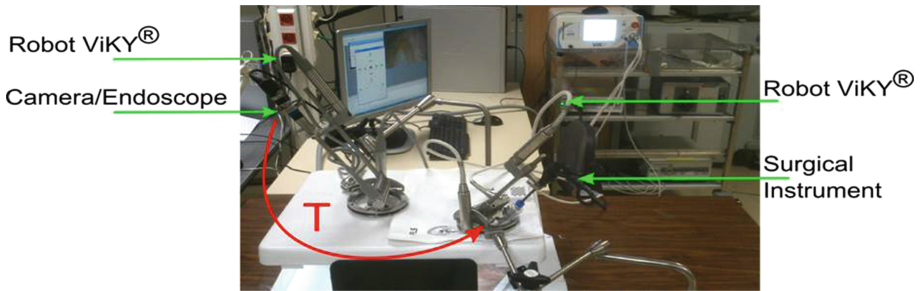


Fig. 7. Experimental test bench to evaluate the 3D pose estimation accuracy with a printout of a surgical scene as background.

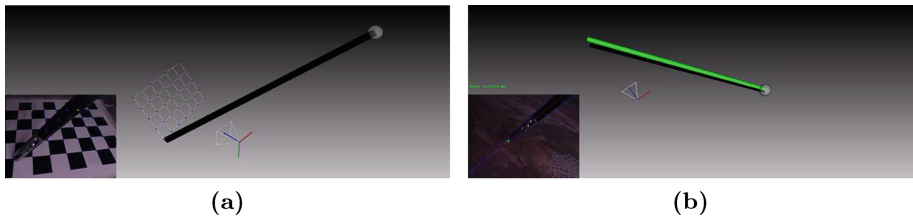


Fig. 8. Estimation of the instrument's pose in space. (a) Calibration step to find the rigid transformation \mathbf{T} (b) Evaluation of the 3D pose estimation accuracy with in black, the reference pose obtained with the robot, in green, the pose computed with our method (Color figure online)

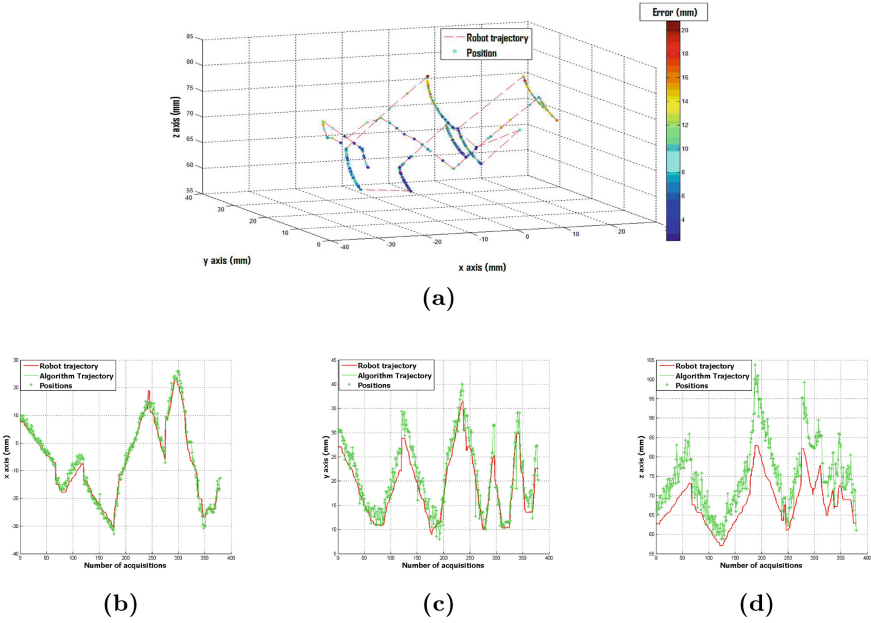


Fig. 9. Robot trajectory (red) and our tracking method trajectory (green) for 380 instrument positions. (a) 3D trajectories (b), (c) and (d) X, Y, Z trajectories with respect to the camera frame (of normal Z) (Color figure online).

commercial robotic instrument holder is directly positioned, and a printout of a surgical scene as background. We compared the 3D tip position of the instrument found by our algorithm to the 3D tip position given by the robot expressed in the camera referential. This required calibrating the system to find the rigid transformation \mathbf{T} between the robot and camera frame such that: $\mathbf{p}_{cam}^{frangi} = \mathbf{T}\mathbf{p}_{cam}^{robot}$. \mathbf{T} is obtained by pointing 12 points of a chessboard, for 6 chessboard positions, with the instrument carried by the robot (see Fig. 8). These 12 points can be expressed in the camera frame thanks to a standard extrinsic camera calibration procedure [12] and are also measured in the robot frame. We resolve a classical least squares system to find the rigid transformation between the two sets of 3D points coupled with a RANSAC to eliminate outliers. We obtain a camera calibration Root Mean Square (RMS) error of 0.25 pixels and \mathbf{T} with a RMS error of 1.2 mm. Figure 9 shows an example of the robot trajectory and of our tracking method for a series of instrument movements. The results for 380 measurements are presented in Table 2. In all results presented, we fixed the Frangi filter parameters as $\sigma = 2$, $\beta = 0.5$ and $c = 0.5\max(s)$, according to recommendations from the literature.

Table 2. Error of the 3D pose estimation with our method compared to the position obtained with a robotic instrument holder

Axis	Mean 3D position error (mm)	Std. Dev. (mm)
x	1.79	1.45
y	2.42	1.60
z	7.24	3.51

4 Conclusion

We presented a surgical instrument tracking algorithm based on image processing. It permits to estimate the 2D/3D instruments pose in real-time without artificial fiducials. An extensive 2D evaluation on real surgical videos shows that our 2D pose estimation is accurate and robust on wide range of realistic cases. In difficult situations as a suture gesture, we can lose accuracy in the instrument's tip position but the orientation is still correct. A machine learning approach as [13], applied in the neighbourhood of our estimated tip position could increase the accuracy of the tip detection. Our approach for 3D pose estimation was validated on a test-bench using a printout of a surgery background. Although this might lack realism we estimated that the robustness of the proposed method on realistic images was already shown extensively on the 2D case. This 3D evaluation provides us with the precision range we can expect when the 2D detection works well. The greatest errors are found in the depth estimation along the z axis. This error could be reduced by using a stereoscopic endoscope.

Our 2D localization approach is robust and accurate enough to control a robotic endoscope holder. Even if the Frangi filter might not be the most obvious approach for edge detection, we showed that it works better than classical approaches. Other more sophisticated edge detection approaches could easily be compared on our image database. The 3D pose estimation could be useful for surgical gesture recognition or for co-manipulation, if we are able to increase the depth precision. Another application could be the online calibration of no rigidly-linked robotic endoscope and instrument holders, which could lead to less bulky surgical systems. Our next step will be to evaluate the 3D pose estimation more extensively in conditions closer to the clinical reality (cadaver experiments).

References

1. 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)
2. Zhou, J., Payandeh, S.: Visual tracking of laparoscopic instruments. *J. Autom. Control Eng.* **2**(3), 234–241 (2014)

3. Wolf, R., Duchateau, J., Cinquin, P., Voros, S.: 3D tracking of laparoscopic instruments using statistical and geometric modeling. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 203–210. Springer, Heidelberg (2011)
4. Allan, M., Ourselin, S., Thompson, S., Hawkes, D.J., Kelly, J., Stoyanov, D.: Toward detection and localization of instruments in minimally invasive surgery. *IEEE Trans. Biomed. Eng.* **60**(4), 1050–1058 (2013)
5. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
6. Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 679–698 (1986)
7. 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), 429–442 (2005)
8. Otsu, N.: A threshold selection method from gray-level histograms. *Automatica* **11**(285–296), 23–27 (1975)
9. Felzenszwalb, P., Huttenlocher, D.: Distance transforms of sampled functions. Cornell University. (2004)
10. Suzuki, S.: Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.* **30**(1), 32–46 (1985)
11. Hough, V., Paul, C.: U.S. Patent No. 3,069,654. U.S. Patent and Trademark Office, Washington (1962)
12. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000)
13. Sznitman, R., Becker, C., Fua, P.: Fast part-based classification for instrument detection in minimally invasive surgery. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 692–699. Springer, Heidelberg (2014)