STAG: Smart Tools & Apps for Graphics (2015)
Silvia Biasotti and Marco Tarini and Andrea Giachetti (Editors)

Robust 3D Pose Estimation of a Laparoscopic Instrument with three Landmarks

M. Carletti, D. Zerbato, D. Dall'Alba, A. Calanca and P. Fiorini, *IEEE Fellow*

Department of Computer Science, University of Verona, 37134 Verona, Italy

Abstract

Knowing surgeon movements during laparoscopic training may provide useful data to speed up the learning process by means of instantaneous error correction and performance evaluation. The first step toward this knowledge is the identification of laparoscopic tool pose in the training environment. In this paper we propose a method to estimate in real time the 3D pose of laparoscopic instruments using a standard camera and three non invasive colored markers applied on the tool stem.

The proposed method takes advantage of closed form solution for the problem which speeds up the computation and improves the precision and accuracy of the results. In addition the method handles occlusions even without any marker tracking algorithm thanks to the automatic identification of the insertion point. The method is evaluated in terms of precision, accuracy and real time execution. Results show that it can be effectively used in common training scenarios.

1. Introduction

Minimally invasive surgery (MIS) is becoming more and more popular as it is characterized by faster recovery and reduced pain and scarring. The principal drawback of MIS is that it requires the surgeon to undergo long and specific training to acquire the skills needed to safely and effectively operate the surgical instruments. One of the most widespread MIS techniques is laparoscopy, which operates in the abdominal area with long instruments which access the intervention area through trocars. To become proficient in laparoscopy, surgeons currently start training on very simple tasks (pick and place, bimanual coordination, ...) in synthetic environments and then move toward more complex tasks (dissection, suturing, ...). The training may involve animals or cadavers and eventually real patients. All the phases of the training require the supervision of an expert surgeon who evaluates and corrects the trainee. Automating the evaluation and correction of trainee performance can lead to important advantages in terms of objectiveness and repeatability of the evaluation and in terms of cost reduction as it will allows the trainee to train without the presence of the expert surgeon.

One of the first steps toward the automation of training is the identification of the pose of the laparoscopic tool. The bare pose of the tool provides a lot of information on surgeon's dexterity as it allows to evaluate many different parameters, e.g. the total distance covered by the tool, the smoothness of the trajectory and so on. The knowledge of the relative pose of the tools adds further information about the bimanual proficiency of the trainee or the correct execution of specific tasks.

For these reasons, in recent decades, the estimation of laparoscopic instruments pose and their insertion point locations is one of the challenges that has produced many different solutions. Some works - [HWKH13], [KGD*03] - rely on data obtained from specific sensors added to laparoscopic instruments or introduces as additional probes while others - [DNdM07], [DNdM06] - focus their efforts on image analysis with the advantages of reduced invasiveness but at the cost of complex algorithms.

This paper presents an analytic solution for the estimation of the pose of the instruments by recovering the relative orientation and position of three collinear points such as colored markers applied on the instruments. The main contribution is a closed form algorithm to estimate the instrument poses even under occlusions of one reference point. The presented solution works in real-time and is based on a simple setup consisting of a monocular camera and three colored markers applied on each instrument.

The paper is organized as follows: in the next section we summarize the related works about visual servoing and la-

DOI: 10.2312/stag.20151287

paroscopic instrument identification referring to both real and synthetic environments. Section 3 describes our method along with a short explanation about computer vision theory, then we analyze each component of the pipeline in depth. The results of our tests are presented in Section 4 and some conclusions and future work are inferred in Section 5.

2. State of the art

One of the problems in visual servoing for laparoscopic abdominal surgery is the estimation of the pose (position and orientation) of laparoscopic instruments. The solution to this problem permits to deploy and improve assistive technologies, for example, to measure the surgeon ability and to improve his/her skills during training sessions.

To calculate the position of a laparoscopic instrument, many techniques were developed. We identify three classes of solutions based on the analysis of the image frames. The first class collects all the methods that exploit a particular hardware setup to improve instruments recognition and tracking: this approach usually leads to accurate results and improves the capability of the system but it is often more expensive and invasive than other solutions due to the added probes. The second class is based on probabilistic methods that model the workspace to identify the area where the instruments are. The last class is the largest and it includes the methods that use only image analysis techniques to identify the instruments without special hardware. These methods are often based on edge detection, color segmentation and marker-based approaches to identify the tool tips.

Hardware based methods usually lead to accurate systems: [HWKH13] uses an endoscope with a Time-of-Flight sensor and benefits from both color and range information to achieve robust results but it suffers of a slow framerate (10 Hz). [MRA*13] works with a transrectal ultrasound probe that requires a 2 minutes long manual or automatic calibration. [KGD*03] proposes a sophisticated method based on a laser emitter with three blinking markers on the side. This method works in real-time (50 Hz) thanks to a marker identification algorithm based on a 5×5 high-pass filter that works on interlaced images. However, hardware based approaches often present encumbrance problems, require the sterilization of the instruments and are usually more expensive than simpler setups.

Probabilistic approaches work on the image acquired by the endoscope: [CHS13] proposes a neural network that estimates the type of the instruments and its position simultaneously; it takes advantage of a Kalman filter (for tracking purposes) but only reaches an update frequency of 8 Hz. In addition, this method works only for 2D tracking and it is not tested for real-time applications. [WDCV11] proposes a so called *Condensation algorithm* that works on a mathematical model of the abdomen and tracks the instruments providing an automatic estimation of the insertion points positions. This solution has an important mean error (27.8 pixels on 768x576 images) of the identification of the 2D tool tip while automatic 3D is still imperfect. Moreover, despite the *Condensation algorithm* may work in real-time, the addition of some image segmentation procedures slows down the performance below 16 Hz.

The last class includes the methods based only on image analysis techniques. This class can be subdivided in two groups: the first is based on the identification of some structured markers, while the second includes those solutions that do not use any kind of extra marker or instrument.

[GJMSMCMJ14] and [KTC*13] are generic marker based methods that calculate in real-time the pose of a quadrangular marker. Using more than a single marker, these solutions manage partial or complete occlusions of the markers, despite they usually require a lot of space. [NZDdM06] implements a 20 Hz tracker that follows a 4-points pattern (similarly to [AC95] and [CA90]) combined to an edge detection algorithm that improves the system performance. The main problem of this solution consists of the delay that the markers detection involves (up to 300 ms) when the tracker loses all markers.

Markerless solutions mainly aim at identifying the edges of the instruments using only the information from the images captured by the endoscope. They are usually robust to occlusions despite they often involve heavy computations that do not allow real-time applications. At the best of our knowledge, current solutions use a combination of Hough transform, simple operators like erosion, dilation, thresholding and color segmentation: once the edge of the instruments are found, the instrument axes are computed and the tool tips are searched along them.

[AKN*11] calculates the pose of the instruments by singular value decomposition (SVD), but it is tested only in standard FLS (*Fundamentals of Laparoscopic Surgery*) training boxes assuming the insertion point is known. [CGL*08] proposes a solution that is robust to tool tip occlusions, but it suffers from the presence of blurred tool edges caused by its movement and it does not work in real-time. [VLC06] proposes a solution that assumes the knowledge of the insertion point: then, using Hough transform, it searches the axis of the instruments and searches the tool tip along it. This method is very slow (10 Hz) even on small images (200x100 pixels) which has good contrast and lightening. A promising solution is presented in [CLGG06]: the study of the vanishing point of the axis of the instruments permits to compute a geometric solution for the tool tip. This solution is not proved to work in real-time and assumes a performant edge detection algorithm. In [CA96] a 4×4 edge extractor filter is used before a 8×8 filter for straight line identification. This approach is slow (10 Hz) and slows down even more if it is used to estimate the tool tip position (5 Hz). To improve the performance, it uses a colored marker to identify a particular point of the instrument: comparing the apparent

size of the projected instrument and its real size (the tool diameter), it is possible to estimate the depth of the tool at that point as the height of a isosceles triangle.

Color segmentation is a simple method to identify points of interest in images. [WAH97] convert the color space of every frame from RGB to HSV to reduce color distortion: a 7×7 low pass filter, followed by a thresholding process, provides noise removal and a precise identification of a colored marker. One flaw of this approach is that it needs a stereo endoscope and only reaches 17 Hz. [DNdM07] and [DNdM06] propose markerless solutions that analyze images in HSV space, perform a real-time color segmentation of the laparoscopic instruments and work in a *in-vivo* environment. However, it is unclear what the overall performance is because results related to the framerate and image resolutions were not presented.

To overcome hardware based solutions and to reduce computation to meet real-time requirements, our solution proposes a closed-form algorithm to calculate the 5 DOF pose of the tool tip, in camera-space, exploiting the projection of only three non invasive collinear markers applied on the instruments: we estimate the 3D position of the tool tip and its distance from the insertion point but we cannot evaluate the roll of the instruments. The main benefit of a marker-based method lies in its simplicity to precisely compute the position of the reference points. Moreover, applying markers on surgical instruments drastically reduces the computation and overcomes the complexity that are typical of markerless algorithms. Also, using the markers is an inexpensive and safe solution. The closed-form solution allows to speed up the computation, improves the accuracy and precision of the results and works in real-time. Estimating the insertion point of the instruments allows our method to be robust even under partial occlusions without losing accuracy. Our algorithm does not depend in principle on the environment, however we tested it only on synthetic environments.

3. Method

We assume a pinhole camera model to describe our system, so, given an unknown point $\mathbf{q} = [X \quad Y \quad Z]^T$ in camera space, its projection $\mathbf{p} = [u \quad v]^T$ on the image plane is not enough to calculate the 3D position of $\mathbf{q}$: perspective projection implies the loss of the z-coordinate, so the coordinates of $\mathbf{q}$ are related to its camera coordinates $\mathbf{p}$ and the focal value f by:

$$u = X\frac{f}{f-Z} \qquad v = Y\frac{f}{f-Z}. \tag{1}$$

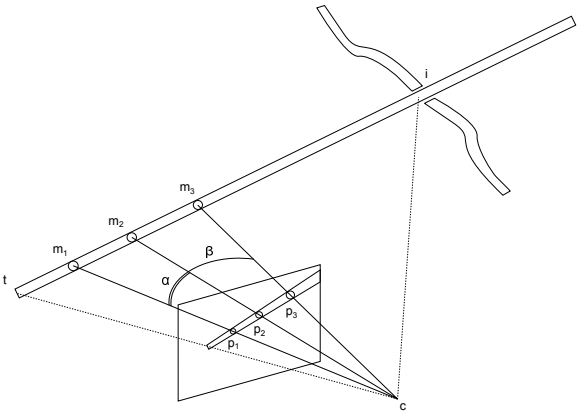
That is why we need at least three collinear markers and constrain the problem to a plane.

The only hardware modification required by our approach consists of three colored markers on each instrument, so that the distances between each marker and the tool tip are known: as shown in Fig. 1, the difference between the marker $\mathbf{m}_i$ and the tool tip is indicated with h_i, with $i \in \{1,2,3\}$. Each marker position is expressed in camera space, as:

$$\mathbf{m}_i = \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} \tag{2}$$

and its projection to the image plane is expressed in homogeneous coordinates:

$$\mathbf{p}_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \tag{3}$$



For all figures please keep in mind that you **must not** use images with transparent background!

Figure 1: *Graphic representation of the values used for the identification of the tool tip when three markers are visible.*

The preliminary step of our method, the calibration, is done once offline and calculates intrinsic and distortion parameters of the camera using the algorithm proposed by Zhang [Zha00] and the camera calibration toolbox for Matlab by Bouguet [Bou13]. From now on, we call K the 3×3 intrinsic parameters matrix and $\mathbf{d}$ the 5×1 distortion parameter vector.

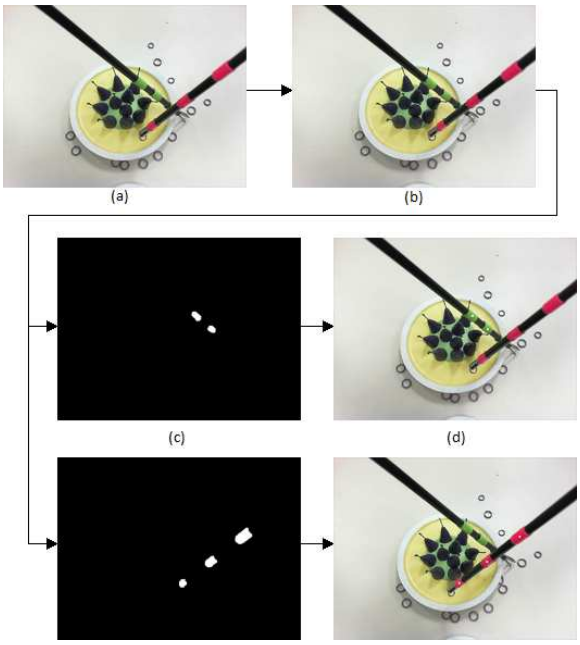
Our method consists of 2 phases: the first phase is the estimation of the insertion point position $\mathbf{i}$ in camera space: we calculate the orientation of the tool in different frames, then we estimate via SVD the intersection point of all these directions. This phase is performed at the beginning of the intervention/training and is required to make the method robust to occlusions. The second phase is described by the pipeline in Fig. 2: to identify markers, we convert every single frame to HSV color-space and apply a threshold to the color. Once we have found the centroid projection to the image plane, we compute the 3D position and orientation of the tool constraining the problem to the plane passing through the origin of the camera $\mathbf{c} = [0 \quad 0 \quad 0]^T$ and the normalized coordinates of marker projections. In this case, *normalized* means

that the coordinates do not depend on the camera matrix. In case of occlusions of one marker, we exploit the estimated position of the insertion point to calculate the tool pose. Point **i** is not usually visible on the image plane.

The initial step provides an estimation of **i**. In case of marker occlusions, we estimate the pose of the tool tip exploiting the estimated location of **i** in camera space, and comparing the result with the last valid pose. To do this we calculate a solution for each possible pair of markers, then we estimate the tool tip pose with a modified version of the algorithm described in subsection 3.1. At the end of this procedure, we select the solution whereby the difference with the last valid solution is minimal. The two phases are done separately for each instrument but while the first is done once at the beginning of the training session, the second is done automatically for each new frame.

3.1. Marker detection

In our setup each instrument has three markers whose color is unique. We thus apply a color segmentation procedure for every single instrument. We also categorize the tools by their insertion point, i.e. right or left of the camera.



For all figures please keep in mind that you **must not** use images with transparent background!

Figure 2: *Color segmentation pipeline for markers centroid identification. (a) frame captured by the camera in RGB color space, then (b) it is converted in HSV color space. For each instrument, (c) the HSV frame is binarized by thresholding all the channels of the HSV space, then (d) the centroid of every white area is calculated.*

Due to camera color distortion, we convert every frame from RGB to HSV color-space and create a binary image thanks to color threshold boundaries of the markers. The binary image is eroded and dilated with a 5×5 cross kernel to remove noise. Our experiments show that a 5×5 filter size is a good tradeoff to remove impulsive noise and not to slow down the performance but, as a side effect. Smaller filters are more sensible to find markers but they are more sensitive to noise; bigger filters are not recommended because they remove too much information. If two or three colored areas are found, we assume we have found the markers of the instruments so their respective centroids are computed; otherwise the current frame is discarded. If three markers are found, we estimate the pose with the method explained in subsection *B*. If two markers are found, we exploit **i** to estimate the instrument pose as presented in subsection *D*. All these markers are sorted according to the current instrument insertion point: we consider as the first marker the nearest to the tool tip and as the last one the nearest to the insertion point projection. This operation simplifies our algorithm and introduces a geometric constraint: we assume that a particular 3-markers-pattern belongs to a specific instrument of which we know the main direction respect the camera position.

3.2. 3D pose estimation

The estimation of the instrument position requires the projection of the markers to the image plane. Our algorithm takes as input the 2D coordinates of these markers, the matrix K and the real distances between the markers applied to the instrument, and gives a 5 DOF pose as output. Due to camera distortion, we use **d** to correct the coordinates of the given points.

For every instrument, starting from the markers projections $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$ in homogeneous coordinates, we calculate two angles as shown in Fig. 1. Respectively, we define the angles $\alpha = \widehat{\mathbf{p}_1\mathbf{c}\mathbf{p}_2}$ and $\beta = \widehat{\mathbf{p}_2\mathbf{c}\mathbf{p}_3}$:

$$\alpha = \arccos(\mathbf{n}_1 \cdot \mathbf{n}_2) \tag{4}$$

$$\beta = \arccos(\mathbf{n}_2 \cdot \mathbf{n}_3). \tag{5}$$

where

$$\mathbf{n}_i = \frac{K^{-1}\mathbf{p}_i}{\|K^{-1}\mathbf{p}_i\|} = \frac{\mathbf{m}_i}{\|\mathbf{m}_i\|} \tag{6}$$

Let h_{12} and h_{23} be the real distance between $\mathbf{m}_1$ and $\mathbf{m}_2$, and $\mathbf{m}_2$ and $\mathbf{m}_3$, respectively. h_{01} is the distance between the tool tip and the first marker. From these values, we can constrain our problem to a plane that contains the origin of the camera **c** and the instrument.

Let β' be the angle $\widehat{\mathbf{c}\mathbf{m}_3\mathbf{m}_1}$:

$$\beta' = \arctan\left(\frac{h_{12}\sin\beta\sin(\alpha+\beta)}{h_{23}\sin\alpha - h_{12}\sin\beta\cos(\alpha+\beta)}\right). \tag{7}$$

Thanks to (7), we can compute the length of the position vectors of the markers:

$$l_2 = h_{23}\frac{\sin\beta'}{\sin\beta} \tag{8}$$

$$l_3 = l_2\cos\beta + h_{23}\cos\beta' \tag{9}$$

$$l_1 = l_3\frac{\sin\beta'}{\sin(\alpha+\beta+\beta')} \tag{10}$$

than define the 3D position of each marker as:

$$\mathbf{m}_i = l_i\mathbf{n}_i \tag{11}$$

Due to inevitable errors during marker detection and numerical approximation, we compute the tool tip position (orientation) as the mean of the positions (orientations) estimated using every pair of markers. Let

$$\mathbf{m}_{ij} = \mathbf{m}_i - \mathbf{m}_j \tag{12}$$

be the vector that goes to m_i from m_j, and let

$$\mathbf{dir} = \frac{1}{3}\left(\frac{\mathbf{m}_{12}}{\|\mathbf{m}_{12}\|} + \frac{\mathbf{m}_{13}}{\|\mathbf{m}_{13}\|} + \frac{\mathbf{m}_{23}}{\|\mathbf{m}_{23}\|}\right) \tag{13}$$

be the averaged sum of the the vectors that connect the markers positions. So it is possible to estimate the orientation $\mathbf{o}$ of the tool as the versor:

$$\mathbf{o} = \frac{\mathbf{dir}}{\|\mathbf{dir}\|} \tag{14}$$

and to estimate the position $\mathbf{t}$ of the tool tip as:

$$\mathbf{t} = h_{01}\mathbf{o} + \mathbf{m}_1. \tag{15}$$

The accuracy and precision of this method will be discussed in Section 4.

Our method is similar to [DNMK08], where an elegant matrix-based algorithm is shown. This method is a linear algorithm that works with $n \geq 3$ collinear points but the matrix notation is ill-conditioned and introduces some significant bias in the results, in presence of noise and close points in the object pattern. Furthermore, it does not manage reference points occlusions. Our solution proposes an analytic approach that resolves these problems and works also for $n = 2$, as Section 3.4 shows.

3.3. Insertion point estimation

Once we have at least two frames which permit us to estimate two different tool poses, we have the possibility to estimate the intersection point $\mathbf{i}$ as the intersection between the estimated 3D lines that approximate the tool. The greater is the number of the estimated poses, the more robust will be the estimation of $\mathbf{i}$. We tested our algorithm with 100 different frames in which there are no marker occlusions: a lower number of frames risks to be too noisy due to orientation estimation. If the location of the real insertion point is too far from the estimated one, our algorithm generates results with larger errors. We compute the insertion point offline, thanks to Eikenes' algorithm [Eik12] that computes an approximated solution that minimizes the quadratic error between the estimated insertion point and the calculated orientations.

3.4. 3D pose estimation under occlusions

During laparoscopic task execution, occlusions may occur both in synthetic and real environments. Under these circumstances the described method fails because of the lack of enough information. To handle these events, our method takes advantage of the knowledge of the position of the insertion point measured at the beginning of the task.

Since we have no information about which marker is occluded, we must estimate three different tool poses, one for each pair of markers. The algorithm starts with two markers, $\mathbf{m}_i$ and $\mathbf{m}_j$, where $\mathbf{m}_j$ is the nearest marker to $\mathbf{i}$ and consider the pivot point $\mathbf{i}$ as the third marker. However, unlike in subsection 3.2, we do not know the actual distance between the markers and the insertion point but we know $\|\mathbf{i}\|$. We exploit this knowledge to obtain an analytic solution for the tool pose.

To estimate the tool orientation, we consider the coplanar lines identified by the segments $\overline{\mathbf{cm}_i}$, $\overline{\mathbf{cm}_j}$, $\overline{\mathbf{ci}}$ and $\overline{\mathbf{im}_i}$, then we define a reference system with the origin in $\mathbf{i}$, where the x axis along $\overline{\mathbf{ci}}$ pointing toward $\mathbf{c}$ and the y axis pointing toward the half-plane where the instrument is.

Defining the angles α and β like in (4) and (5) we can calculate the gradient of the lines passing through the markers as:

$$L_i = -\tan(\alpha+\beta) \qquad L_j = -\tan\beta. \tag{16}$$

So we define the fourth degree polynomial and identify its roots:

$$\begin{aligned}(i^2a^2 - k^2)x^4 +\\ 2bk^2x^3 +\\ (i^2a^2 - k^2(b^2+2c))x^2 +\\ 2bck^2x -\\ i^2c^2\end{aligned} \tag{17}$$

where k is the distance between the markers, $a = L_i - L_j$, $b = L_i + L_j$ and $c = L_iL_j$. Due to the geometrical structure of the problem, we discard imaginary solutions and the solution that does not converge with the optical ray in front of the camera. So, we consider only the real and positive solution L_o of (17) and if more than one valid solution is found, the current frame is discarded.

Let ω be the angle between the tool direction and the vec-

tor **i**, then:

$$\omega = \arccos\sqrt{\frac{1}{L_o^2+1}}. \tag{18}$$

Thanks to (17) and (18), the distance between the insertion point **i** and the marker $\mathbf{m}_j$ can be defined as

$$\|\mathbf{m}_j - \mathbf{i}\| = \frac{\mathbf{i}\sin\beta}{\sin(\beta+\omega)}. \tag{19}$$

After this operations, we compute the positions of the markers like in (11) where:

$$l_i = \frac{k+\|\mathbf{m}_j - \mathbf{i}\|\sin\omega}{\sin(\alpha+\beta)} \tag{20}$$

$$l_j = \frac{\|\mathbf{m}_j - \mathbf{i}\|\sin\omega}{\sin\beta} \tag{21}$$

$$\mathbf{m}_i = l_i\mathbf{n}_i \qquad \mathbf{m}_j = l_j\mathbf{n}_j. \tag{22}$$

Finally, it is possible to estimate the pose of the tool tip **t** using (15) where $\mathbf{m}_3 = \mathbf{i}$. It is important to underline that we do not know the pair of markers whose included angle is α. So, unless we include a marker tracking algorithm, we have to calculate ω and **t** three times setting $\|\mathbf{m}_i - \mathbf{m}_j\|$ as h_1, h_2 and (h_1+h_2). At the end, we choose the solution that minimizes the Euclidean distance between the last valid pose and the current one.

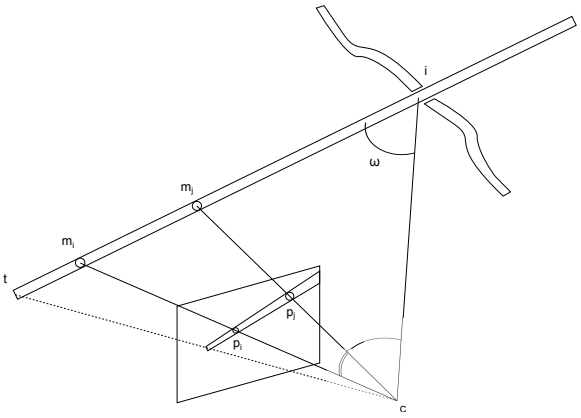


Figure 3: *Graphic representation of the values used for the identification of the tip of the instrument when only two markers are visible due to occlusions.*

4. Results

We use a support for a Microsoft LifeCam HD-6000 camera and two Olympus HiQ+ Bipolar Hand Instruments to work on a controlled environment. The proposed method was implemented in C++, using OpenCV 2.4.9, CvBlobsLib v83 and GLM 0.9.5.3 on a Q8200 (2.33 GHz) PC. For the results, we compare our algorithm with ArUco 1.2.5 [GJMSMCMJ14]. On each instrument three colored markers were applied whose color boundaries are computed by hand: we choose fluorescent colors like magenta and orange, for the right and left instrument respectively, to ensure robustness to illumination change. The color boundaries used in segmentation are computed to get the widest search window preserving robustness. For magenta, we choose [154, 184] Hue-boundaries (on 255). For orange, we choose [0, 24]. Both colors are characterized by Saturation and Value [128, 255] and [92, 255] respectively.

In our experiments, we assume to use a calibrated camera and to move the instruments no farther than 30 cm from the camera center. The markers have size of 1.0 cm and the distance h_{ij} between adjacent markers are [$h_{01} = 1.0, h_{12} = 2.0$, $h_{23} = 1.0$] cm for every instrument.

We show two studies performed on a single instrument at time: first one compares the insertion point estimated positions and the second one compares the tool tip position with and without occlusions. Every test has been reproduced for three different image resolutions (640x480, 800x600 and 1280x800).

Table 1: *Insertion point estimation [cm]*

	ArUco			our method		
	X	**Y**	**Z**	**X**	**Y**	**Z**
640x480						
Mean	9.48	-7.24	-0.09	9.47	-7.37	-0.26
STD	0.10	0.00	0.10	0.17	0.22	0.66
800x600						
Mean	9.61	-7.32	0.11	9.13	-7.09	-0.80
STD	0.20	0.14	0.28	0.00	0.10	0.89
1280x800						
Mean	9.31	-7.09	-0.04	9.80	-7.54	-0.56
STD	0.56	0.33	0.57	0.26	0.10	0.39

To estimate the insertion point coordinates, we applied two markers of ArUco to the instrument as in [KTC*13], then we compute the difference of their 3D positions to identify the axis of the instrument. Finally we estimate the insertion point as the point of intersection of all the estimated directions. Results are provided in table 1.

For the accuracy test, we draw four points on a line on the ground plane and touch them in sequence with the tool tip. The distance between two consecutive points is 4 cm. We then compare the relative positions of the actual points and the estimated ones. Table 2 shows the accuracy test results.

Position errors, which in some cases reach nearly a centimeter, are probably due to a rough camera calibration and to the manual positioning of the instrument.

For the precision test, we move the instrument along a

Table 2: *Accuracy and precision [cm]*

	w/o occlusions				occlusions			
	I	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
Ground	0.00	4.00	8.00	12.00	0.00	4.00	8.00	12.00
640x480								
Error	0.00	0.09	0.22	0.24	0.00	-0.01	0.20	0.55
STD	0.20	0.20	0.28	0.35	0.02	0.01	0.01	0.01
800x600								
Error	0.00	0.37	0.70	0.95	0.00	0.09	0.41	0.90
STD	0.26	0.26	0.17	0.10	0.02	0.01	0.01	0.01
1280x800								
Error	0.00	0.07	0.17	0.18	0.00	0.09	0.27	0.68
STD	0.24	0.14	0.14	0.14	0.01	0.01	0.01	0.00

Table 3: *Std.Dev. of the tool tip position during motion [cm]*

	w/o occlusions	**occlusions**
640x480	0.52	0.20
800x600	0.42	0.17
1280x800	0.44	0.14

line and calculate the average standard deviation of the measurements in respect of the main direction of the estimated poses.

Both table 1 and 2 show that higher resolution does not significantly affect the accuracy of the measurements despite a significant improvement in the precision, especially in case of occlusions. The higher precision in case of occlusions is due to the assumption that the insertion point is considered as a *fixed* marker. Precision results are shown in table 3. Fig.4 graphically shows the estimated poses along the tracked line.

Another important result concerns the performance: our solution works at an average framerate of 27 frames per second at a resolution of 640x480. Higher resolution results in lower framerates (about 10 FPS at 1280x800) due to poor segmentation algorithm performance. Occlusions do not influence the speed of the system.

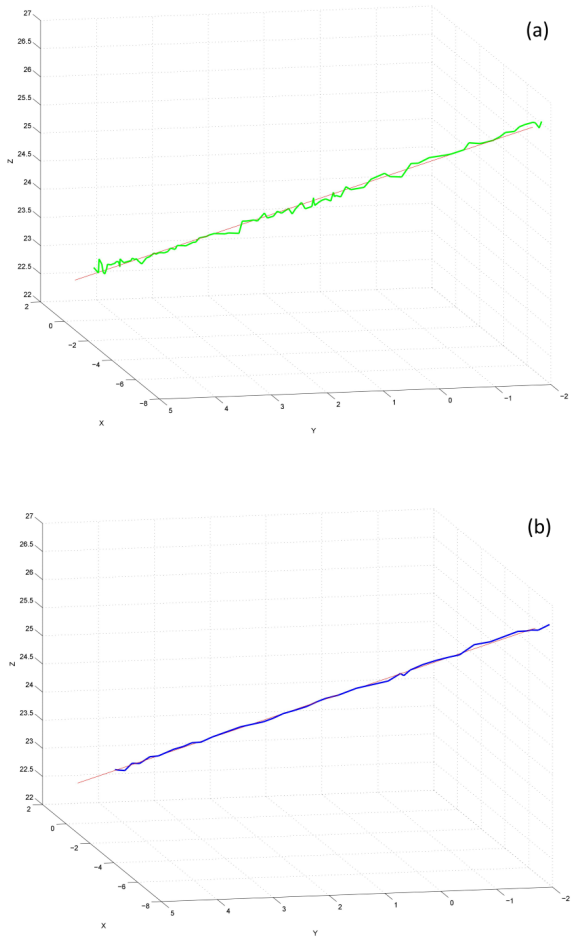


Figure 4: *Estimated tool tip position (green and blue) versus its regression line (red) at different resolutions, a) 640x480 and b) 1280x800.*

5. Conclusion and future work

We have developed a closed-form solution that works in real-time to estimate the 3D position of laparoscopic instruments by using a simple setup composed by a standard camera and three colored markers applied on the instruments. The proposed approach works even in presence of the occlusion of one of the reference points and guarantees robustness thanks to its analytic formulation.

Despite the speed of the system is slowed down by an inefficient segmentation algorithm, our solution works up to 29 FPS on a 640x480 image. Experimental results show that the resolution of the image does not significantly influence the accuracy of the tool tip pose estimation. This will permit to deploy our solution on a low-performance embedded system using a low resolution camera.

The proposed method may take advantage of a smart segmentation algorithm to calculate marker centroids and of a marker tracking system to understand which marker is occluded and to reduce computational time.

Future works include more realistic scenarios, like *in-vivo* environments, to test our algorithm. Although the closed-form solution proposed in our method does not depend on the algorithm that identifies the reference points, more experiments must be executed to stress the robustness of the colored markers under important environment changes, like limited illumination and lower distance between the instruments and the camera center.

Acknowledgment

The research leading to these results has been partly funded by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement n. 270396 (Intelligent Surgical Robotics, I-SUR).

References

[AC95] ABIDI M. A., CHANDRA T.: A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 17*, 5 (1995), 534–538. 2

[AKN*11] ALLEN B. F., KASPER F., NATANELI G., DUTSON E. P., FALOUTSOS P.: Visual tracking of laparoscopic instruments in standard training environments. In *MMVR* (2011), pp. 11–17. 2

[Bou13] BOUGUET J. Y.: Calibration toolbox for matlab, 2013. http://www.vision.caltech.edu/bouguetj/calib_doc/. 3

[CA90] CHANDRA T., ABIDI M. A.: A new all-geometric pose estimation algorithm using a single perspective view. In *1989 Advances in Intelligent Robotics Systems Conference* (1990), International Society for Optics and Photonics, pp. 318–331. 2

[CA96] CASALS A., AMAT J.: Automatic guidance of an assistant robot in1 laparoscopic surgery. 2

[CGL*08] CANO A. M., GAYÁ F., LAMATA P., SÁNCHEZ-GONZÁLEZ P., GÓMEZ E. J.: Laparoscopic tool tracking method for augmented reality surgical applications. In *Biomedical Simulation*. Springer, 2008, pp. 191–196. 2

[CHS13] CHEN C.-J., HUANG W. S.-W., SONG K.-T.: Image tracking of laparoscopic instrument using spiking neural networks. In *Control, Automation and Systems (ICCAS), 2013 13th International Conference on* (2013), IEEE, pp. 951–955. 2

[CLGG06] CANO A. M., LAMATA P., GAYÁ F., GÓMEZ E. J.: New methods for video-based tracking of laparoscopic tools. In *Biomedical Simulation*. Springer, 2006, pp. 142–149. 2

[DNdM06] 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 *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*. Springer, 2006, pp. 527–534. 1, 3

[DNdM07] DOIGNON C., NAGEOTTE F., DE MATHELIN M.: Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision. In *Dynamical Vision*. Springer, 2007, pp. 314–327. 1, 3

[DNMK08] DOIGNON C., NAGEOTTE F., MAURIN B., KRUPA A.: Pose estimation and feature tracking for robot assisted surgery with medical imaging. In *Unifying perspectives in computational and robot vision*. Springer, 2008, pp. 79–101. 5

[Eik12] EIKENES A.: Intersection point of lines in 3d space, 2012. http://www.mathworks.it/matlabcentral/fileexchange/37192-intersection-point-of-lines-in-3d-space. 5

[GJMSMCMJ14] GARRIDO-JURADO S., MUÑOZ-SALINAS R., MADRID-CUEVAS F. J., MARÍN-JIMÉNEZ M. J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition 47*, 6 (2014), 2280–2292. 2, 6

[HWKH13] HAASE S., WASZA J., KILGUS T., HORNEGGER J.: Laparoscopic instrument localization using a 3-d time-of-flight/rgb endoscope. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on* (2013), IEEE, pp. 449–454. 1, 2

[KGD*03] KRUPA A., GANGLOFF J., DOIGNON C., DE MATHELIN M. F., MOREL G., LEROY J., SOLER L., MARESCAUX J.: Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *Robotics and Automation, IEEE Transactions on 19*, 5 (2003), 842–853. 1, 2

[KTC*13] KE M.-C., TSENG Y.-H., CHEN C.-W., HO M.-C., LIAN F.-L., YEN J.-Y., LIN W.-L., CHEN Y.-Y.: Preliminary study of intracorporeal localization for endoscopy tracking. In *Automatic Control Conference (CACS), 2013 CACS International* (2013), IEEE, pp. 130–134. 2, 6

[MRA*13] MOHARERI O., RAMEZANI M., ADEBAR T. K., ABOLMAESUMI P., SALCUDEAN S. E.: Automatic localization of the da vinci surgical instrument tips in 3-d transrectal ultrasound. *Biomedical Engineering, IEEE Transactions on 60*, 9 (2013), 2663–2672. 2

[NZDdM06] NAGEOTTE F., ZANNE P., DOIGNON C., DE MATHELIN M.: Visual servoing-based endoscopic path following for robot-assisted laparoscopic surgery. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (2006), IEEE, pp. 2364–2369. 2

[VLC06] VOROS S., LONG J.-A., CINQUIN P.: Automatic localization of laparoscopic instruments for the visual servoing of an endoscopic camera holder. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*. Springer, 2006, pp. 535–542. 2

[WAH97] WEI G.-Q., ARBTER K., HIRZINGER G.: Real-time visual servoing for laparoscopic surgery. controlling robot motion with color image segmentation. *Engineering in Medicine and Biology Magazine, IEEE 16*, 1 (1997), 40–45. 3

[WDCV11] WOLF R., DUCHATEAU J., CINQUIN P., VOROS S.: 3d tracking of laparoscopic instruments using statistical and geometric modeling. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*. Springer, 2011, pp. 203–210. 2

[Zha00] ZHANG Z.: A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 22*, 11 (2000), 1330–1334. 3