

Real-time 3D visual tracking of laparoscopic instruments for robotized endoscope holder

Zijian Zhao *

School of Control Science and Engineering, Shandong University, Jinan 250061, Shandong China

Abstract. Robotized endoscope holder in laparoscopic surgeries allows surgeons to control the endoscope without the assistants' intervention. In this paper, a new method is proposed for the automatic 3D-tracking of laparoscopic instruments in real-time to provide more convenient interactions between surgeons and the robotized endoscope holder. The method is based on the 3D position measurements of the insertion points of the instruments and the strip markers, combined with the depth estimation of the instruments. The results of our experiments show that our method is fast and robust in the simulated laparoscopic surgeries.

Keywords: Laparoscopic surgery, tool tracking, localization of instrument, robotized camera holder

1. Introduction

In the recent years, minimally invasive surgeries (MIS), including laparoscopy, have been commonly adopted. In laparoscopic surgeries, small incisions are made in the human abdominal cavity to introduce surgical instruments and a laparoscope through trocars. This type of minimally invasive surgeries has several advantages over normal open surgeries including less blood loss, fewer transfusions, less consumption of analgesics and shorter hospitalization time. However, laparoscopies also have some drawbacks. For example, they usually lose depth information of the surgeries, the field of view of the surgeries is limited, and they are constrained by the abdominal wall. That is why the laparoscopic procedure is more complex for surgeons than the open surgeries are. To overcome the drawbacks, robotized laparoscopic systems (include the robotized camera holder) were developed to assist surgeons in performing laparoscopy. There are some outstanding robotized endoscopic holders commercialized in the market, for example, AESOP[®] (Computer Vision Inc.), EndoAssist[®] (Armstrong Healthcare) [1] and ViKY[®] (EndoControl) [2]. Each robotized system has its corresponding tool-tracking method, so that the endoscope can be controlled easily by the surgeon according to the positions of operation instruments during the surgical procedure. Our research focuses on realizing automatic tool-tracking by means of computer vision techniques under ViKY system.

* Address for correspondence: Zijian Zhao, School of Control Science and Engineering, Shandong University, 17923 Jingshi Road, Jinan 250061, China. Tel.:+8653188392591; Email:zhaozijian@sdu.edu.cn.



Fig. 1. The endoscope mounted on the ViKY robot.

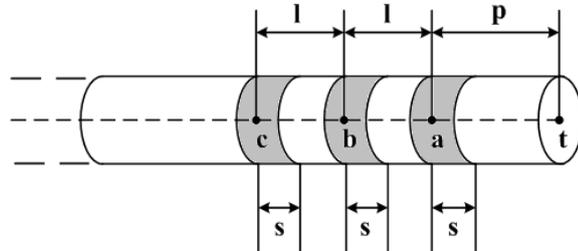


Fig. 2. The designed markers on the instruments.

It should be mentioned that there are several works proposed to track the operation instruments in 2D image space. Wei et.al [3] proposed a real-time detection algorithm based on the color-marked instruments, where the color regions are segmented by computing the color histogram on the H-S (Hue-Saturation) plane. Wang et.al [4] constructed a simple Bayesian classifier to segment the colored instruments. Doignon et.al [5] put forward a real-time color segmentation method to detect instruments by using a joint Hue-Saturation color feature. Climent et.al [6] proposed an approach to detect the line features in the images, by which the instruments can be located. Besides the image information of the instruments, some 3D based detection methods were proposed based on other feature information [7-11]. Zhang et.al [7] designed a kind of special markers for instruments detection, but they did not apply the 3D geometric properties of these markers. Voros et.al [8-10] utilized a 3D geometric tool model and insertion points and got some reasonable results, but their approaches are not capable of real-time 3D detection and tracking, and their precision of tip-positioning was not satisfactory. Allan et.al [11] applied a classifier to the tool-tracking, which was more complex and time-consuming.

In this paper, we introduced a new 3D locating and tracking approach for surgical instruments, which completely relies on the measurement information of insertion points and color markers of the instruments. In this approach, all the color markers are first segmented, and the 2D image positions of instruments are then determined. Having a calibrated camera mounted on the robotized camera holder [2], we can project all the insertion points of the instruments onto the image plane, and the image points of each insertion point can be used to constrain the image positions of instruments. Based on the 3D measurement of insertion points and the 2D image positions of color markers, the 3D positions of the instruments then can be induced rapidly and accurately. The experiment results show the validity, accuracy and robustness of our approach. With our approach, we can carry out the automatic localization for the visual servoing of robotized camera holders.

2. Material

Our approach is developed based on a system consists of an endoscope camera and a ViKY robot, as shown in Figure 1. We use Zhang's method [12] to calibrate the endoscope camera for acquiring the intrinsic matrix and the distortion parameters. It is also necessary to know the camera's displacement in the robot's coordinate system, which requires the "hand-eye" calibration. We choose the method in [13] to compute the hand-eye transformation between the robot end-effector and the camera.

We designed three round-strip markers, which are attached to the instrument (Figure 2). To segment the markers easily in the image, the color of all markers is set to green, also mentioned in [5]. As shown in Figure 2, points **a**, **b** and **c** are at the centers of each strip-marker, and point **t** is at the tip position of the instrument. The center-to-center distances of these strips are l ($l = 10mm$). The distance between the tip of instrument and its closest strip-center is p ($p = 10mm$). All the four points **t**, **a**, **b**, **c** satisfy the following cross-ratio equation:

$$(\mathbf{t}, \mathbf{a}; \mathbf{b}, \mathbf{c}) = \frac{2l + 2p}{2l + p}. \quad (1)$$

3. Method

The task is to track the motion of a laparoscopic instrument inside the abdominal cavity only using the information from the endoscopic images without any optical or magnetic localizers [14,15]. After calibration, we determine the insertion point of each instrument in 3D using a priori knowledge about its geometry. We then carry out the color segmentation to estimate the markers' image positions. The 3D positions of all markers and the instrument tips can be computed based on the image positions of the markers according to the computer vision algorithm. The Kalman filter is applied in tracking using the position data above to ensure the stable tracking results. In case of situations where markers are invisible, the visual servoing algorithm is also added in the ViKY system to keep the instrument in the field of view all the time.

The general framework of our method can be summarized into the following 5 steps:

- (1) Calibrate the 3D positions of the insertion points of the instruments.
- (2) Segment the color markers of the instruments.
- (3) Locate the 3D positions of all markers and the instrument tips.
- (4) Apply Kalman filter.
- (5) Visual servoing.

Every step will be explained in detail in the following section.

3.1. Calibration of the 3D positions of the insertion points

The 3D positions of the insertion points of the instruments are measured at the beginning of the intervention. First, we can compute the images of the insertion points on the endoscopic image manually by mouse-click [10]. Second, at every two different positions of the camera where the insertion points are visible, the 3D positions of the insertion points can be estimated in the camera's coordinate through stereo reconstruction algorithm. Since the robot moves under the surgeon's commands at any time during the operation procedure, the 3D positions of the insertion points should be transformed to the robot's base coordinate, according to the hand-eye parameters discussed in Section.2. The reason is that the insertion points are stable under the robot's based coordinate system.

3.2. Color segmentation of markers

In color endoscopic images, a pixel is a mixture of the three fundamental colors, Red, Green and Blue (RGB), but there exist many representations to encode this information. Because the markers we used are green, we only need to segment all the green pixels in the endoscopic image. To avoid the disturbing of

the luminance changing [11], we applied the color space of Hue, Saturation and Value (HSV) to represent a pixel. Any pixel satisfying the following conditions can be segmented as the pixel belonging to the markers.

$$\mathbf{V} = \text{Green}, \mathbf{S} \geq St,$$

where St is a set threshold.

After having all the green pixels of the markers, we utilized the fast k-means algorithm to identify three cluster regions of pixels in the endoscopic image. These regions are the images of three markers on the instrument. Then we can compute the median centers of the three regions. According to the distances between these center-points and the image of insertion points, we are able to recognize the three center-points as the images of points $(\mathbf{a}, \mathbf{b}, \mathbf{c})$, denoted by $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{c}})$. Suppose $\tilde{\mathbf{t}}$ is the image of the instrument tip \mathbf{t} , according to the property of cross-ratio, we can determine $\tilde{\mathbf{t}}$ based on the following equation

$$(\tilde{\mathbf{t}}, \tilde{\mathbf{a}}; \tilde{\mathbf{b}}, \tilde{\mathbf{c}}) = (\mathbf{t}, \mathbf{a}; \mathbf{b}, \mathbf{c}). \quad (2)$$

3.3. Locating 3D positions of the instruments

With the four image points $(\tilde{\mathbf{t}}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{c}})$, we can estimate their depths in the camera's coordinate system, with the intrinsic parameters' matrix \mathbf{K} . Suppose (z_t, z_a, z_b, z_c) are the depths of the image points $(\tilde{\mathbf{t}}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{c}})$. According to the pin-pole model [16], we have

$$\mathbf{t} = z_t \mathbf{K}^{-1} \tilde{\mathbf{t}}, \quad \mathbf{a} = z_a \mathbf{K}^{-1} \tilde{\mathbf{a}}, \quad \mathbf{b} = z_b \mathbf{K}^{-1} \tilde{\mathbf{b}}, \quad \mathbf{c} = z_c \mathbf{K}^{-1} \tilde{\mathbf{c}}. \quad (3)$$

The points $(\mathbf{t}, \mathbf{a}, \mathbf{b}, \mathbf{c})$ lie on the same line, so we have the following equations:

$$(\mathbf{t} - 2\mathbf{a} + \mathbf{b}) = 0, \quad (\mathbf{a} - 2\mathbf{b} + \mathbf{c}) = 0, \quad (\mathbf{t} - \mathbf{a} - \mathbf{b} + \mathbf{c}) = 0.$$

Substituting Eq. (3) into the above formulas, we have

$$\mathbf{M} [z_t, z_a, z_b, z_c]^T = \mathbf{0}_{9 \times 1}, \quad (4)$$

with

$$\mathbf{M} = \begin{bmatrix} \mathbf{K}^{-1} \tilde{\mathbf{t}} - 2\mathbf{K}^{-1} \tilde{\mathbf{a}} & \mathbf{K}^{-1} \tilde{\mathbf{b}} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}^{-1} \tilde{\mathbf{a}} - 2\mathbf{K}^{-1} \tilde{\mathbf{b}} & \mathbf{K}^{-1} \tilde{\mathbf{c}} \\ \mathbf{K}^{-1} \tilde{\mathbf{t}} & -\mathbf{K}^{-1} \tilde{\mathbf{a}} & -\mathbf{K}^{-1} \tilde{\mathbf{b}} & \mathbf{K}^{-1} \tilde{\mathbf{c}} \end{bmatrix}.$$

The right null-vector \mathbf{m} of \mathbf{M} is actually the depth vector $[z_t, z_a, z_b, z_c]^T$ up to a scale s_d . Knowing the measurements of the instrument, we can construct the following equations

$$s_d \| [\mathbf{K}^{-1} \tilde{\mathbf{t}} - \mathbf{K}^{-1} \tilde{\mathbf{a}} \ \mathbf{0} \ \mathbf{0}] \mathbf{m} \|_2 = p, \quad (5)$$

$$s_d \| [\mathbf{0} \ \mathbf{K}^{-1} \tilde{\mathbf{a}} - \mathbf{K}^{-1} \tilde{\mathbf{b}} \ \mathbf{0}] \mathbf{m} \|_2 = l, \quad (6)$$

$$s_d \| [\mathbf{0} \ \mathbf{0} \ \mathbf{K}^{-1} \tilde{\mathbf{b}} - \mathbf{K}^{-1} \tilde{\mathbf{c}}] \mathbf{m} \|_2 = l. \quad (7)$$

Based on Eqs. (5,6,7), the scale factor s_d can be determined. With the estimated depths $s_d \mathbf{m}$, the 3D position of the instruments can be determined according to Eq. (3). To make sure the instrument-tip lies on the same line with the insertion point \mathbf{I} , we can also get the rectified tip position $\hat{\mathbf{t}}$ by

$$\hat{\mathbf{t}} = \mathbf{I} + \frac{\|\mathbf{t} - \mathbf{I}\|_2}{\|\mathbf{a} - \mathbf{c}\|_2} \cdot (\mathbf{a} - \mathbf{c}). \tag{8}$$

3.4. Kalman filter

To acquire the stable tracking results, we applied a Kalman filter to process the position data from Eq. (8). The state vector x_k is defined as $[\mathbf{t}, d\mathbf{t}]^T$, the observation vector is $z_k = \hat{\mathbf{t}}$. The state-transition model and the observation model are respectively

$$\mathbf{F}_k = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}, \quad \mathbf{H}_k = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}.$$

According to the Kalman filter theory, we have the following equations:

$$z_k = \mathbf{H}_k x_k + v_k, \quad x_{k+1} = \mathbf{F}_k x_k + w_k,$$

where v_k and w_k are the noise interruptions. For the k th frame, the Kalman filter will update based on $z_k = \hat{\mathbf{t}}$ and get x_{k+1} as the final tracking position of instrument-tip.

3.5. Visual servoing of ViKY robot

In order to keep the instrument tip inside of the imaging field of laparoscopy, we need to control the ViKY robot's motions based on the feedback information extracted from the image point $\tilde{\mathbf{t}}$. Suppose $\tilde{\mathbf{t}} = [u_t, v_t, 1]^T$, the ViKY robot's three dimensional position parameters [10] are $[q_1, q_2, \rho]^T$, then the relative image plane position about $\tilde{\mathbf{t}}$ is related to the position velocity of ViKY robot by the following equation

$$\begin{bmatrix} u_t - u_0 \\ v_t - v_0 \end{bmatrix} = \begin{bmatrix} \frac{-f}{z_t} & 0 & \frac{u_t}{z_t} & \frac{u_t v_t}{f} & \frac{-(f^2 + u_t^2)}{f} & v_t \\ 0 & \frac{-f}{z_t} & \frac{v_t}{z_t} & \frac{f^2 + u_t^2}{f} & \frac{-u_t v_t}{f} & -u_t \end{bmatrix} \begin{bmatrix} 0 & -\rho & 0 \\ -\rho \sin(q_2) & 0 & 0 \\ 0 & 0 & 1 \\ -\sin(q_2) & 0 & 0 \\ 0 & 1 & 0 \\ \cos(q_2) & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{\rho} \end{bmatrix} \tag{9}$$

where f is the focal length of laparoscopy, $[u_0, v_0]^T$ is the principal point of the image plane, and the symbol "()" denotes the time derivative of coordinate (velocity). By using the Singular Value Decomposition and Tikhonov regularization, $[\dot{q}_1, \dot{q}_2, \dot{\rho}]^T$ can be resolved. Based on these position velocity parameters, we can perform the visual servoing of ViKY robot.

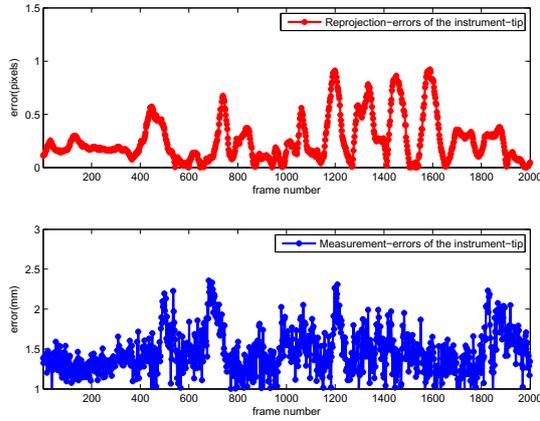


Fig. 3. The estimated errors on the instrument-tracking.

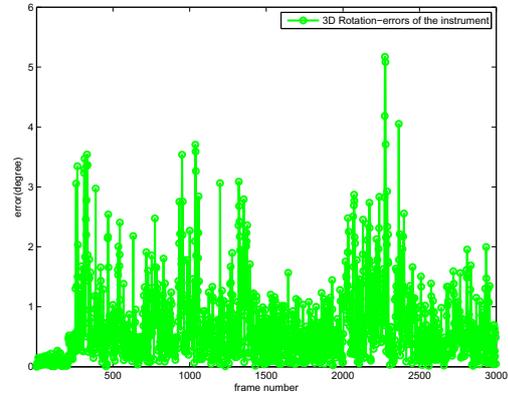


Fig. 4. The 3D direction errors of the instrument's gestures.

4. Results

In the experiment, a plastic phantom with a test-bench inside was used for the modeling of a laparoscopic surgery with one surgical instrument. The surgical instrument moved randomly in the view of the endoscope. Because there was a gap between the instrument and the trocar, the positions of the insertion point were always floating to some extent during the operation. To simulate the floating positions of the insertion point, we added a random noise to the calibrated insertion point in our program. Under such a tough condition, we performed our method to track and locate the surgical instrument in 2D and 3D spaces.

First, we measured the 2D image positions of the instrument-tip along time, and re-projected its 3D points onto the image plane. Comparing the two results, we were able to compute the re-projection error of the instrument-tip for every frame, as shown in Figure 3 (red). The statistical values of the 2D re-projection error are shown in Table 1, where the maximum error is less than 1 pixel. This is more accurate than the errors given in reference [8], which is 5 pixels.

Second, we measured the 3D positions of the instrument-tip and three markers along time. Then we computed the distance between the observation $\hat{\mathbf{t}}$ and the state \mathbf{t} for each frame. The metric errors for all image frames were estimated, as shown in Figure 3 (blue). According to the statistical values in Table 1, it can be seen that the metric mean-error of our method is $1.42mm$, which is very close to the error standard for human surgery ($1mm$).

After we had the 3D directions $(\hat{\mathbf{t}} - \mathbf{I})$ and $(\mathbf{t} - \mathbf{I})$ of the instrument at real-time, we could compute the 3D rotation error of the instrument (the angle between the above two directions) for each frame according to Eq. (8), which is shown in Figure 4. As shown in Table 1, the maximum 3D rotation error is less than 6° , and is much better than the mean 3D error in reference [9], which is 14.8° . We also re-projected the 3D model of the instrument back to the image plane. The results are shown in Figure 5. The pink lines in the images are the re-projected model of the instrument in the tracking process. From Figure 5, it can be seen that the tracking results of the proposed method are accurate and robust. The computation time of our method is $50ms$ per frame for 640×480 image (probed with the Windows timer API), for an Intel Core-i3 2.53GHz, 1.86Go RAM PC.

Table 1
The statistical error results of Figure 3 and Figure 4.

	2D Re-projection Error	3D Measurement Error	3D Rotation Error
Max	0.93	2.46	5.36
Mean	0.18	1.42	1.07



Fig. 5. Examples of tracking and locating of the instrument.

5. Discussion

Our approach allows surgeons to track the motion of their instruments during an intervention, in order to either automatically control the motions of the robot, or evaluate the quality of the gestures of surgical tools. As shown in the experimental results, our locating approach is precise, robust and quick enough for the anticipated uses in both the orientation and the position of the instrument under the image/camera coordinate. The color segmentation and the 3D depth evaluation of markers were still not perfect enough and should call for improvements in our algorithm.

In Figure 3, it can be seen that the re-projection error is a function of time, which is mainly because the color segmentation of markers is not stable. Unstable images of the markers will lead to unstable instrument-tip image point. In fact, the green color is very special in the abdominal cavity. Its segmentation is usually influenced by the luminance changing, with the Saturation of pixel affected the most. Therefore, it is not suitable to set a fixed threshold in the segmentation as mentioned in Section 3.2. We are planning to apply an adaptive threshold for the color segmentation, which could be more robust for the estimations of the image positions.

We only used three markers sticking to the instrument for tool-tracking in this paper. From Figure 3, it can be seen that we have a metric precision that is less than $2.5mm$, which could still be improved to a higher level. If we use more markers, we will have extra data to construct the coefficient matrix, which highly tolerates the random noise. Then we can estimate the depths of all points more accurately than before. The extra data may not obviously slow down the algorithm's speed, so the tracking and locating can still be performed in real-time.

The floating of the insertion points is inevitable in real surgeries, and it is necessary to find a way to compensate the errors due to floating. In our paper, we apply a random noise to the calibrated position of the insertion points during the program. It is not really a good idea, because we do not know exactly the model of the floating, and the random noise sometimes does worse. Therefore, our next task is to find the floating model.

After all the above improvements, we plan to test our method on anatomic specimens mimicking true surgical interventions, where we can understand how these improvements really work in surgeries.

6. Conclusion

We developed a novel method to locate the 3D positions of the instruments in laparoscopy. The method is automatic and can be realized in real-time. Our experimental tests showed encouraging results and allowed us to find some difficult obstacles that were still not solved. Some high level implementations based on this method could be implemented to supervise surgical tasks such as the 3D-positioning of special organs or focuses.

Acknowledgements

This work is supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (20130131120036), the Promotive Research Fund for Excellent Young and Middle-aged Scientists of Shandong Province (BS2013DX027), and the French National Research Agency (ANR) through TecSan program (project DEPORRA ANR-09-TECS-006).

References

- [1] P. B. Nebot, Y. Jain, K. R. Haylett and R. F. McCloy, Comparison of task performance of the camera-holder robots EndoAssist and Aesop, *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques* **13**(2003), 334-338.
- [2] P. J. Berkelman, P. Cinquin, J. Troccaz, J-M. Ayoubi, C. Létoublon, Development of a compact cable-driven laparoscopic endoscope manipulator, *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention* **2488**(2002), 17-24.
- [3] G. Wei, K. Arbter, G. Hirzinger, Real-time visual servoing for laparoscopic surgery, controlling robot motion with color image segmentation, *IEEE Engineering in Medicine and Biology* **16**(1997), 40-45.
- [4] Yuang Wang, D. R. Uecker and Yulun Wang, A new framework for vision enabled and robotically assisted minimally invasive surgery, *Computerized Medical Imaging and Graphics* **22**(1998), 429-437.
- [5] C. Doignon, F. Nageotte, and M. De Mathelin, Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature, *Real Time Imaging* **11**(2005), 429-442.
- [6] J. Climent, P. Marés, Automatic instrument localization in laparoscopic surgery, *Electronic Letters on Computer Vision and Image Analysis* **4**(2004), 21-31.
- [7] X. Zhang, S. Payandeh, Application of visual tracking for robotic-assisted laparoscopic surgery, *Journal of Robotics Systems* **19**(2002), 315-28.
- [8] S. Voros, J.A. Long, and P. Cinquin, Automatic localization of laparoscopic instruments for the visual servoing of an endoscopic camera holder, *Proceedings of the 9th international conference on Medical Image Computing and Computer-Assisted Intervention* **4190**(2006), 535-542.
- [9] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros, 3D tracking of laparoscopic instruments using statistical and geometric modeling, *Proceedings of the 14th international conference on Medical image computing and computer-assisted intervention* **6891**(2011), 203-210.
- [10] S. Voros, J.A. Long, and P. Cinquin, Automatic detection of instruments in laparoscopic images: a first step towards high-level command of robotic endoscopic holders, *International Journal of Robotics Research* **26**(2007), 1173-1190.
- [11] M. Allan, S. Ourselin, S. Thompson, D. Hawkes, J. Kelly and D. Stoyanov, Towards detection and localization of instruments in minimally invasive surgery, *IEEE Transactions on Biomedical Engineering* **60**(2013), 1050-1058.
- [12] Z. Zhang, A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(2000), 1330-1334.
- [13] Z. Zhao, Hand-eye calibration using convex optimization, *Proceedings of IEEE International Conference on Robotics and Automation* **1**(2011), 2947-2952.
- [14] O. Tonet, T.U. Ramesh, G. Megali, and P. Dario, Image analysis-based approach for localization of endoscopic tools, *Proceedings of Surgetica* **1**(2005), 221-228.
- [15] R. Aggarwal, T. Grantcharov, K. Moorthy, T. Milland, and A. Darzi, Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Annals of surgery* **247**(2008), 372-379.
- [16] R. Hartley, A. Zisserman, Computation of the camera matrix P, in: *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, 2004, pp.178-193.