32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Robotic-Surgical Instrument Wrist Pose Estimation

Stephan Fabel†, Kyungim Baek† and Peter Berkelman‡

Human-Robot Interaction Laboratory
†Department of Information and Computer Sciences
‡Department of Mechanical Engineering
University of Hawaii at Mānoa
2540 Dole St., Honolulu HI 96822
{sfabel,kyungim,peterb}@hawaii.edu

***Abstract*— The Compact Lightweight Surgery Robot from the University of Hawaii includes two teleoperated instruments and one endoscope manipulator which act in accord to perform assisted interventional medicine. The relative positions and orientations of the robotic instruments and endoscope must be known to the teleoperation system so that the directions of the instrument motions can be controlled to correspond closely to the directions of the motions of the master manipulators, as seen by the the endoscope and displayed to the surgeon. If the manipulator bases are mounted in known locations and all manipulator joint variables are known, then the necessary coordinate transformations between the master and slave manipulators can be easily computed. The versatility and ease of use of the system can be increased, however, by allowing the endoscope or instrument manipulator bases to be moved to arbitrary positions and orientations without reinitializing each manipulator or remeasuring their relative positions.**

The aim of this work is to find the pose of the instrument end effectors using the video image from the endoscope camera. The P3P pose estimation algorithm is used with a Levenberg-Marquardt optimization to ensure convergence. The correct transformations between the master and slave coordinate frames can then be calculated and updated when the bases of the endoscope or instrument manipulators are moved to new, unknown, positions at any time before or during surgical procedures.

I. Introduction

Teleoperated surgical robot systems typically can be classified as a common-base setup, in which all manipulators and sensors are attached to a large fixed base, or a modular-base setup, in which endoscope and instrument manipulators are independently positioned and attached to the operating table. The latter has several advantages in that this configuration supports smaller, simpler manipulators which can be made to be portable, modular, sterilizeable by autoclave, and more easily integrated into the surgical environment without extensive initialization or set-up time.

An example of a modular-base robot design is the *Compact Lightweight Surgical Robot* (CLSR) [1] system developed at the University of Hawaii, which evolved from the design of the ViKY by *EndoControl Medical* [2]. The most prominent example of a common-base surgical robot setup is the *daVinci* by *Intuitive Surgical* [3]. Common-base systems are typically restricted to a single configuration or surgical quadrant, and also require some surgical preplanning and training of the users to avoid interference between the manipulators and situations in which the correspondence between each master and instrument pair is ambiguous. Typically these situations are prevented by shutting down the actuation of the instrument arms, thus forcing a manual reset into a valid geometric configuration.

However, being able to arbitrarily reposition or exchange the insertion points of the endoscope and surgical instruments, as is frequently done in standard laparoscopic surgery, is crucial for new applications of robotic medical interventions such as multi-quadrant surgery. While a modular-base robotic system allows for greater flexibility in setup and usage, the problem of finding the poses of the manipulators with respect to each other becomes critical. In a teleoperation system, the motion directions of the master console handles must correspond to the motions of the slave manipulators as seen by the operator. In practice we have found that the 3D motion directions must correspond to within approximately 10 to 15 degrees in order to prevent disorientation of the operator. The relative positions and orientations between the instruments and the endoscope necessary to establish master-slave control direction correspondence to this degree of accuracy can be found from the existing endoscope video image using a pose estimation technique, requiring only that the dimensions of the instrument graspers are previously known and that the endoscope camera is intrinsically calibrated, which can be done well in advance of the actual procedure.

In Section II, the 3D instrument pose estimation problem from a 2D endoscope image is defined, and the basic coordinate frames and notations are described. In Section III, an extended version of the P3P algorithm is used to accurately estimate the pose relative to the endoscope point of view. Experimental results are presented in Section IV after which an outlook on future work is provided.

II. Problem Statement

Notation: We will denote points in 3D italic uppercase, vectors lowercase bold and matrices uppercase bold. Co-

This work is part of a project funded by NIH grant #R21EB006073.

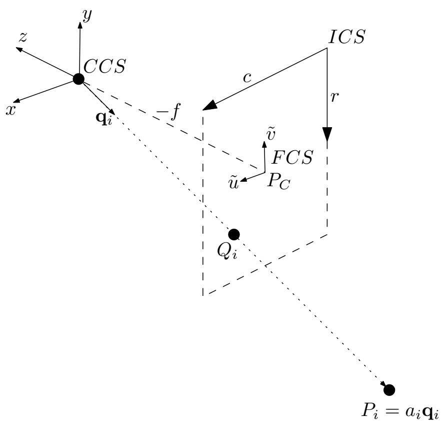

Fig. 1. Relevant coordinate systems. Includes the Camera Coordinate System (CCS), the Frame Coordinate System (FCS) and the Image Coordinate System (ICS). The 3D point P_i is a multiple of some constant a_i and the unit vector $\mathbf{q}_i$ pointing towards its image point Q_i.

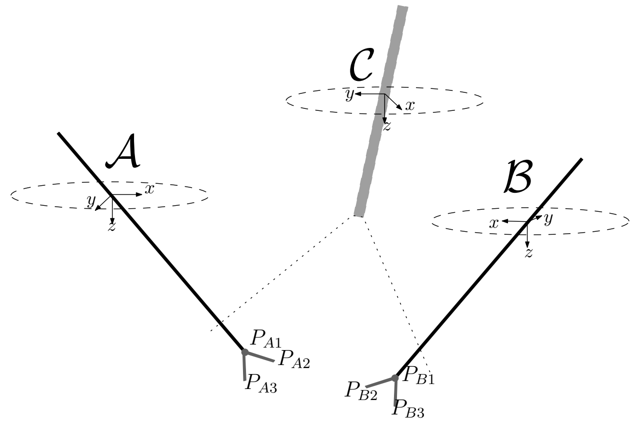

Fig. 2. Scene and configuration of the endoscope and instrument robots. The relationship between the different origins (which constitute the insertion points of the robot) are unknown.

ordinate frames are represented with 3 uppercase letters if related to computer vision, and one uppercase letter in physical 3D geometry. Estimated values are denoted with a hat.

A. Coordinate Frames

Each robot provides information about its own workspace, with the robot coordinate system based on the insertion point of the instrument, and is controlled by adjusting the inclination ϕ, azimuth θ, insertion depth ρ and the roll (instrument shaft rotation) α [4, Fig. 5].

Note that the endoscope coordinate system (C) shown in Fig. 2 is *not* the same as the *camera coordinate system* (CCS) with its origin at the focal point as shown in Fig. 1. The CCS is set to be the origin of the “world” coordinate system.

B. Method

In order to compute the pose of the instrument grasper with respect to the endoscope view, the endoscope is used to extract the pose of the MIS-instruments, in this case two graspers, using a modification of the well-known *P3P* method [5].

The P3P algorithm is a common method for 3D pose estimation with respect to the focal point. However, in order to increase the robustness of the pose estimation method, a non-linear minimization using the Levenberg-Marquardt algorithm (first introduced in [6]) is used to guarantee convergence. The three points used in the P3P algorithm are the tip of the instrument and the ends of the grasper end-effector.

In order to detect these points accurately within the image frame, various methods and algorithms exist, such as using colored markers, light sources or other fiducials to enable an automatic point detection. A typical image read from the frame grabber is presented in Fig. 3.

III. 3D Pose Estimation

A. Perspective 3 Point Problem

Fig. 1 shows the geometry of different coordinate systems involved, where a 3D point P_i is projected onto the projection (or image) plane. The camera coordinate system (CCS) is defined by having its z-axis (or principal axis) intersect the projection plane (here the *frame coordinate system*, or FCS) at the principal point P_C. The frame point ${}^{FCS}P_i = (\tilde{u}, \tilde{v})$, which is at a constant distance from the focal point (i.e. the focal length), is then being subjected to radial and tangential undistortion [7]. The projection is finally displayed in the rectified *image coordinate system* (ICS) as $Q_i = (c, r)$ where r is the pixel row and c is the pixel column. Fig. 3 shows a typical image as recorded by the frame grabber hardware used in the image processing machine.

In the *Perspective 3 Point Problem* (P3P), three 3D scene points are observed in the ICS as $Q_{i=1\ldots3}$. Let the distances d_{mn}, $m \neq n$, between points P_m and P_n be known. In addition, if the object is rigid, we can assume that these distances will not vary over time. If $\mathbf{q}_i$ denotes the *unit vector* from the origin of the CCS in the direction of Q_i, then there exist some scale parameters $a_{1,\ldots,3}$ such that

$$\hat{P}_i = a_i\,\mathbf{q}_i, \tag{1}$$

with $\hat{P}_i$ as the estimated points in 3D. We can iteratively find combinations of the scale factors with the unit vectors such that the following error functions are minimized:

$$e_{mn}(a_1, a_2, a_3) = \underbrace{||a_m\mathbf{q}_m - a_n\mathbf{q}_n||}_{\hat{d}_{mn}}{}^2 - d_{mn}^2 = 0. \tag{2}$$

By linearizing Eq. 2 using a Taylor expansion in the neighborhood of $A = (a_1, a_2, a_3)$ and dropping the higher order terms, the function becomes

$$e_{mn}(A) + \left[\frac{\partial e_{mn}}{\partial a_1}\;\frac{\partial e_{mn}}{\partial a_2}\;\frac{\partial e_{mn}}{\partial a_3}\right]\begin{bmatrix}\delta_1\\ \delta_2\\ \delta_3\end{bmatrix} = 0 \tag{3}$$

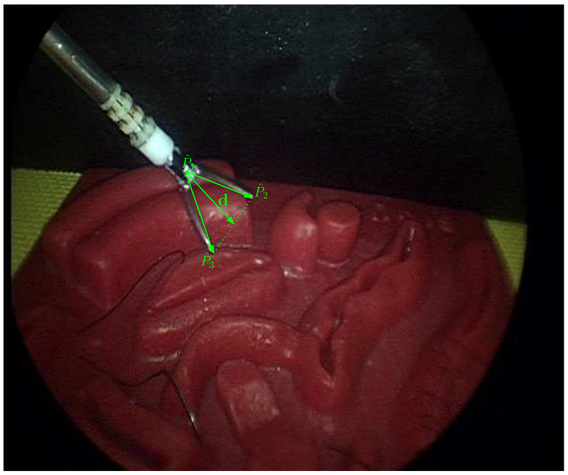

Fig. 3. Typical image captured using our endoscope camera. The image is distorted, and the gripper is clearly visible. The vectors are added to illustrate the directional vector **d** of the instrument grasper attached to the flexible wrist.

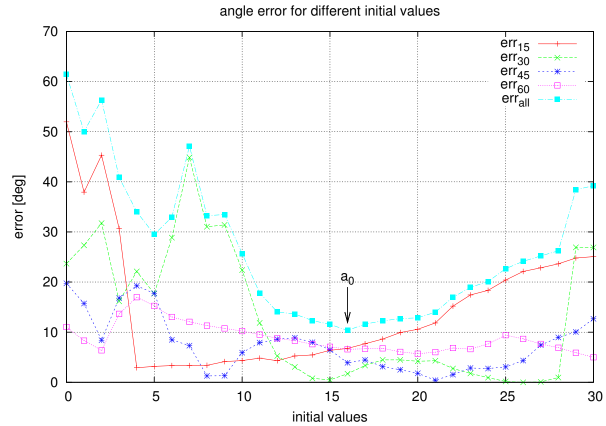

Fig. 4. Different initial values and the angular error of the normal vector as calculated in Sec. III-C. The lines show the angular error depending on different initial values of the pose estimation algorithm for each true angle as given in Table IV-B.

with $\delta_{1\ldots3}$ as the change in each iteration. Expanding this for three points and solving for $\delta_{1\ldots3}$ leads to

$$\begin{bmatrix}\delta_1\\ \delta_2\\ \delta_3\end{bmatrix} = -\mathbf{J}^{-1}(A)\,\mathbf{E}(A), \quad \text{with}\,\mathbf{E}(A) = \begin{bmatrix}e_{12}(A)\\ e_{23}(A)\\ e_{31}(A)\end{bmatrix} \tag{4}$$

where $\mathbf{J}$ is the Jacobian of the error function $\mathbf{E}(A)$ [5].

Beginning with sensible initial values and abort criteria (e.g. $\delta_{1\ldots3} < \varepsilon$), the algorithm converges towards a solution. However, with tests using this standard method of update calculation (Newton's method), there are cases in which the algorithm does not converge or a significant error remains. In order to address these issues, the Levenberg-Marquardt minimization technique was chosen for its robustness and level of sophistication.

B. Update Calculation

The Levenberg-Marquardt approach allows to defensively navigate around non-linearities in the error function, while providing reasonable performance in all other regions [8]. The update is then calculated as:

$$\begin{bmatrix}\delta_1\\ \delta_2\\ \delta_3\end{bmatrix} = -\left(\mathbf{J}^T\mathbf{J} + \lambda\mathbf{I}\right)^{-1}\mathbf{J}^T\,\mathbf{E}(A^k) \tag{5}$$

where λ is a damping coefficient that can be dynamically adjusted during the minimization algorithm—a small λ leads to larger steps, and vice versa. A sufficient error limit is chosen so that when the updates are too small (e.g. $\varepsilon < 10^{-15}$), the algorithm stops. Also, the damping adjustment strategy needs to be addressed. A common one can be found in [9]. The algorithm converges within relatively few iterations using this strategy. In addition to the damping strategy, the results obtained are somewhat sensible to the chosen initial values. We found a value between 12 and 20 suitable enough to provide best overall performance, which can be seen from Fig. 4.

C. Determining the Wrist Pose

We assume an instrument grasper at the end of the instrument wrist, which supports an opening angle of 90°. The direction of the grasper is defined as the vector **d** point from the central joint out in between both instrument grippers (see Fig. 3).

The directional vector **d** is found using the estimated object points. This vector gives the direction of the grasper opening.

$$\begin{aligned}\mathbf{a} &= \hat{P}_2 - \hat{P}_1, & \mathbf{c} &= \mathbf{b} - \mathbf{a}\\ \mathbf{b} &= \hat{P}_3 - \hat{P}_1, & \mathbf{d} &= \mathbf{a} + \frac{1}{2}\mathbf{c}\end{aligned} \tag{6}$$

Using the motorized instrument grasper, the opening of the grasper can be set to any angle within the mechanical boundaries. The estimated points can be used to calculate two unit vectors, whose cross-product determine the attitude of the plane in which the grasper points lie.

$$\mathbf{u}_z = \mathbf{u}_x \times \mathbf{u}_y = \frac{\mathbf{b}}{|\mathbf{b}|} \times \frac{\mathbf{a}}{|\mathbf{a}|} \tag{7}$$

IV. Experimental Results

The endoscope and the instrument robots are arranged in a set of known configurations. A Polaris Vicra position sensor was used to establish the geometric relations between the endoscope and the grasper points.

Fig. 6 shows a typical setup used in laparoscopic surgery, with the endoscope in the middle and both instruments entering the image from opposite sides.

A. Levenberg-Marquardt vs. Newton's Method

To compare the Levenberg-Marquardt approach with the Newton Minimization technique, the P3P algorithm was implemented to use both algorithms. Fig. 5 shows an example where all points were at equal distance from the camera frame origin. The results are similar for other configurations.

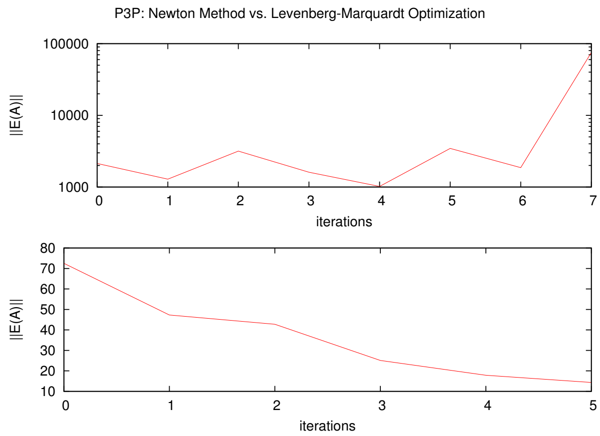

Fig. 5. Newton Minimization (top) vs. Levenberg-Marquardt (bottom) in P3P. Note the semi-log scale using the (non-converging) Newton Method. The line represents the error function minimized as described in eqs. 3 and 4.

TABLE I

EST. ANGLES OF NORMAL VECTOR WITH INIT. VAL. $a_0 = 17$

true angle	**est. angle**	**error**
15.0	21.758	6.759
30.0	31.707	1.707
45.0	48.898	3.898
60.0	66.595	6.595

B. Accuracy of estimated wrist orientation

In order to measure the accuracy of the wrist orientation, the vector normal to the plane containing the three points was calculated (see Sec. III-C) while the grasper was positioned at different angles with respect to the camera frame. The normal vector was compared to the vector resulting from a zero degree rotation.

Assuming that the error in the orientation matters only if it is larger than 10 to 15 degrees, the P3P gives results well within this boundary as can be seen in Table IV-B. Once this correspondence is established, an accurate mapping of the haptic master console to the wrist orientation in 3D is possible so that directions of master and slave match as seen through the endoscope camera.

V. CONCLUSION AND OUTLOOK

An effective method to estimate the pose of a robotic surgical instrument wrist with respect to the camera coordinate system was implemented. Using the estimated points of the P3P pose estimation algorithm, two unit vectors in the object coordinate system were found, providing a complete rotational registration of the wrist with respect to the point of view of the operator.

Solving the problem of 3D pose estimation of the instrument wrist is important especially in the modular-base robot system because it allows for a more realistic mapping between the viewpoint of the surgeon and the robotic actuator. Our method provides unambiguous orientation and perception in space, and lays the foundation for visual augmentation methods during surgical procedures. Using

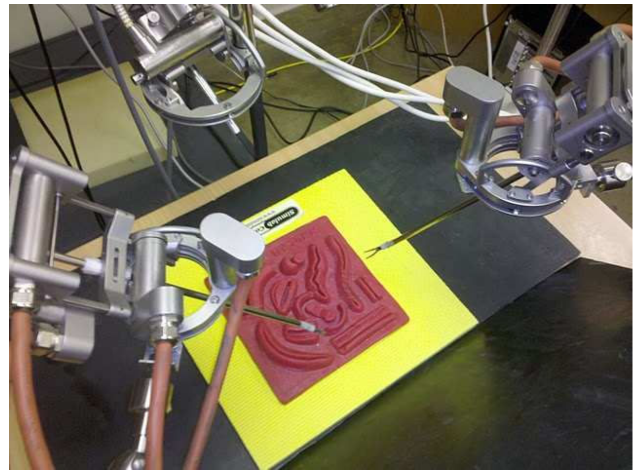

Fig. 6. CLSR with the ViKY with three independently positioned insertion points (middle of the base rings).

the Levenberg-Marquardt method enhances the accuracy and ensures convergence of the pose estimation.

Based on the assumption that the orientation needs to be determined only within 10 to 15 degrees, our algorithms shows promising first results. However, the next step for this pose estimation method will be to evaluate the performance of the algorithm in a real-world scenario.

Future applications of this algorithm include an automatic registration system which can be performed online during an exchange of instruments and the endoscope, or whenever necessary. Since the P3P estimation typically converges within 5 to 20 iterations, a continuous estimation of the grasper points is feasible to provide real-time updates of the wrist attitude with respect to the operator.

REFERENCES

[1] P. Berkelman and J. Ma, "A compact modular teleoperated robotic system for laparoscopic surgery," *Int'l J Robotics Research*, vol. 28, no. 9, pp. 1198–1215, Sept 2009.

[2] J.-A. Long, P. Cinquin, J. Troccaz, S. Voros, P. Berkelman, J.-L. Descotes, C. Letoublon, and J.-J. Rambeaud, "Development of miniaturized light endoscope- holder for laparoscopic surgery," *Journal of Endourology*, vol. 8, pp. 911–914, 2007.

[3] G. S. Guthart and J. K. Salisbury, "The Intuitive (TM) telesurgery system: Overview and application," in *International Conference on Robotics and Automation*. San Francisco: IEEE, April 2000, pp. 618–621.

[4] P. Berkelman and J. Ma, "A compact, modular, teleoperated robotic minimally invasive surgery system," in *Int. Conf. on Biomedial Robotics and Biomechatronics*. Pisa Italy: IEEE/RAS-EMBS, Feb 2006.

[5] L. Shapiro and G. Stockman, *Computer Vision*. Prentice Hall, NJ, 2001.

[6] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, June 1963.

[7] J.-Y. Bouguet, "Camera calibration toolbox for matlab," http://www.vision.caltech.edu/bouguetj/calib_doc/index.html, Dec 2008.

[8] A. W. Fitzgibbon, "Robust registration of 2d and 3d point sets," in *In Proc. British Machine Vision Conference, volume II*, 2001, pp. 411–420.

[9] M. I. A. Lourakis and A. A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm," Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), Tech. Rep., 2004.