2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Assisting Minimally Invasive Surgery through Exterior Orientation to Enhance Perception

Agustin A. Navarro and Joan Aranda

Abstract—**A diversity of action-perception applications relies on 2D information to perform 3D tasks. Minimally Invasive Surgery (MIS) is one of these applications. It requires a high degree of sensory-motor skills to overcome the disengagement between action and perception caused by the physical separation of the surgeon with the operative site. The integration of body movements with visual information serves to assist the surgeon providing a sense of position. Our purpose in this paper is to present the exterior orientation as a tool in assisted interventions locating the instruments with respect to the surgeon. An enhanced perception is obtained by augmenting the 2D information imposing position and orientation data through a human-machine interface. Applying motion analysis in a sequence of images and having knowledge of the 3D transformations implemented to the instrument, we show it is possible to estimate its orientation with only two different rotations and also its position in the case its length information is supplied.**

I. Introduction

THE knowledge of 3D scene information obtained from a video camera allows to perform a diversity of action-perception tasks in which 2D data are the only inputs. Exterior orientation techniques are aimed to calculate the position and orientation of the camera with respect to other objects in the scene and perceptually contribute to control action in this kind of applications. In Minimally Invasive Surgery (MIS) a 2D window on a 3D world is imposed. The surgeon is limited to work physically separated from the operation site and must rely heavily on indirect perception [1]. To effectively link action to perception it is necessary to integrate body movements with other senses such as vision. Therefore, locating the instruments with respect to the surgeon through computer vision techniques serves to enhance the cognitive mediation between action and perception and can be considered as an important assistance in MIS.

The 2D-3D pose estimation problem serves to map this relation estimating the transformation between coordinate frames of objects and the camera. There are several methods proposed to estimate the pose of a rigid object. The first step of their algorithms consists on the identification and location of some kind of features that represent an object in the image plane. Most of them rely on feature points and apply closed-form or numerical solutions depending on the number of correspondences between objects and image features.

Lines are the kind of features of interest in this paper. As higher-order geometric primitives, describe objects where part of its geometry is previously known. These kinds of features have been incorporated to take advantage of its inherent stability to solve pose estimation problems. Approaches as the one by Navab and Faugeras [2] proposed a solution using line correspondences representing them as Plücker lines, in [3] Homer discussed the problem for line-plane correspondences and Kumar and Hanson [4] and Ansar and Daniilidis [5] estimated the pose from a set of lines and study its sensitivity. Other approaches as the one proposed by Liu et al. [6] combine line and point features.

In the case of image sequences motion and structure parameters of the 3D scene can be determined. Correspondences between selected features in successive images must be established. This provides motion information and can be used, as is our case, to estimate the pose of a moving object. The uniqueness of the structure and motion was discussed by Holt and Netravali [7] for combinations of lines and point correspondences, and their result was that three views with a set of correspondent features, two lines a one point, or two points and one line give a unique solution.

It is of our interest the analysis of changes in the image plane determined through selected feature correspondences induced by 3D transformations applied to objects. Some properties of these motions are useful to estimate the pose of an object addressing questions as the number of movements or motion patterns required which give a unique solution. Straight lines are the features to be used in this paper and we show how knowing the transformations applied, it is possible to estimate their relative orientation with only two different rotations. Providing cognitive information represented by augmented sensory input to the surgeon.

The remainder of this paper discusses in the next section the mediation between action and perception in MIS. An analysis of motion from line correspondences is presented in Section III, as a synthesis of the particular problem. Then a mathematical description of the proposed pose estimation algorithm using only rotations is explained in Section IV. Following in the next sections with simulations and experimental results, to finalize with concluding remarks in Section VII.

Manuscript received September 15, 2006. This work was supported in part by the Spanish MCYT under Grant BES-2003-1659.

A. Navarro is with the Automatic Control Department-UPC, Center of Bioengineering of Catalonia, Barcelona, 08028 Spain (e-mail: agustin.navarro@upc.edu).

J. Aranda is with the Automatic Control and Computer Engineering Department, Technical University of Catalonia (UPC), Barcelona, 08028 Spain (e-mail: joan.aranda@upc.edu).

II. Mediating Action and Perception in MIS

The visual sense in the MIS environment is limited. It imposes a 2D window of the operative site. Thus, approaches focused to assist the surgeon are fundamentally based on image content recognition and presentation. Dutkiewicz et al. [8] reported an experimental verification of surgical tool tracking to be presented in the center of the image. Sun et al. [9] studied the distribution of markers to accurately track the instruments, and Payandeh et al. in [10] established models for the lens distortion. These are examples of emergent techniques to assist the surgeon. Our approach, however, is based on the integration of visual and motion information to perceptually locate the instruments with respect to the surgeon.

Healey in [1] describes the mediation between action and perception in the MIS environment. There he states that it is necessary to effectively link action to perception in egocentric coordinates to overcome the indirect cognitive mediation. We suggest that the estimation of the position and orientation of the surgery instruments with respect to the camera is capable to provide this egocentric information. Computer vision issues as the 2D-3D pose estimation and exterior orientation deal with this problem and can be applied to aid the surgeon in this kind of procedures.

It can be seen in Fig. 1 an application where exterior orientation is used and presented through enhanced visual information to assist the surgeon. This presentation is commonly performed by augmented reality. From early approaches as the one by Milgram et al. [11] in different kinds of applications, to more specialized in surgery as the works of Devernay [12], recognizing objects seen by the endoscope in cardiac MIS, or Pandya and Auner [13], designing a system for surgical guidance, this visual enhancement has served as a human-machine interface. In this work, the position and orientation of surgery instruments is the information to be imposed in the visual information. It serves to integrate egocentric information, as vision and limb movements, to provide a sense of presence and relate it with the external environment to give a sense of place. Nevertheless, the camera-tool calibration must be calculated. This problem can be tackled by computer vision techniques. We propose a pose estimation algorithm based in motion analysis of lines and serves as part of the process shown in Fig. 2.

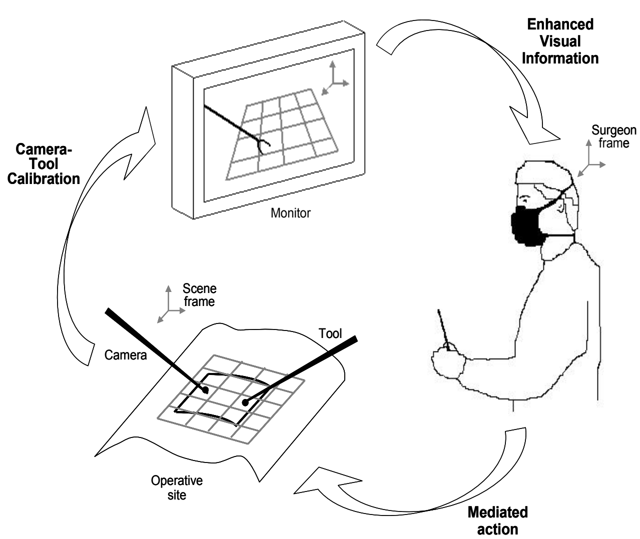

Fig. 1. Application of the exterior orientation as a tool to assist the surgeon in MIS through perception enhancement to control the action

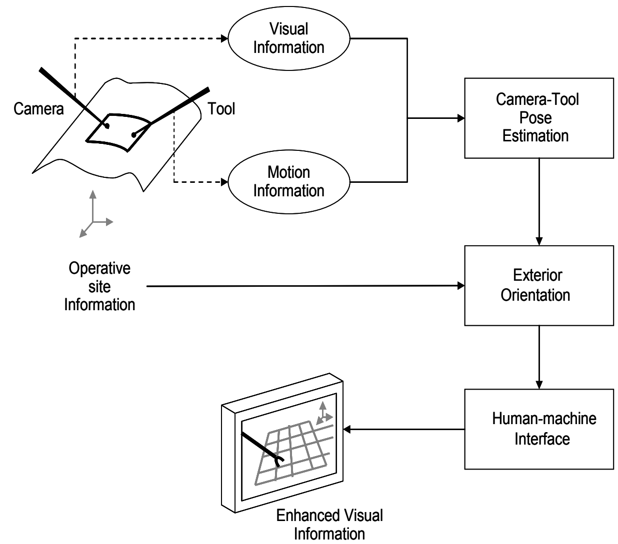

Fig. 2. Visual enhancement process to assist the surgeon. Motion and visual information is needed to calibrate the surgical instrument with respect to the camera.

III. Motion Analysis of Feature Lines

In this work straight lines were the features extracted from a sequence of images. They provide motion information through its correspondences. There are several approaches that use this kind of features to estimate motion parameters. Some early works solve a set of nonlinear equations, as the one by Yen and Huang [14], or use iterated extended Kalman filters, as showed by Faugeras et al. [15], through three perspective views. Weng et al. [16] discussed the estimation of motion and structure parameters studying the inherent stability of lines and explained why two views are not sufficient.

An important property in using lines, as was reported by Mitiche et al. [17], is their angular invariance between them. Then our purpose was to study this property to provide a robust method to solve pose estimation problems. It is possible to compute the position and orientation of an object through the analysis of angular variations in the image plane induced by its 3D rotations with respect to the camera. It can be seen as an exterior orientation problem where objects in the scene are moved to calculate their pose.

Some action-perception applications, as MIS, can be seen as a fixed camera visualizing objects to be manipulated. In our case, these objects were surgical instruments represented by lines. Three views after two different rotations generate three lines in the image plane, each of them define a 3D

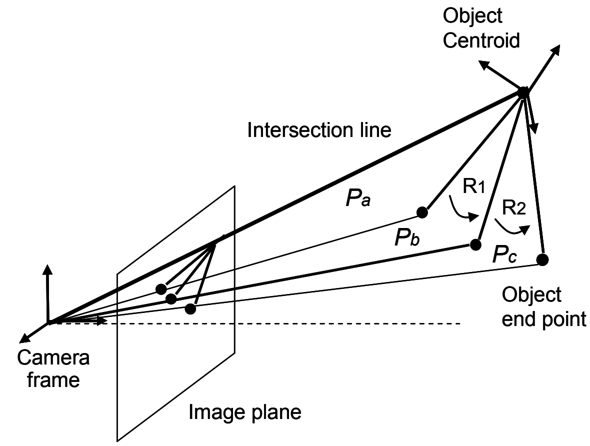

Fig. 3. A 3D line through the origins is the intersection of the projection planes after two rotations of the object.

plane called the projection plane of the line. These planes pass through the projection center and their respective lines. Their intersection is a line in 3D from the origin of the perspective camera frame to the centroid of the rotated object, as seen in Fig. 3.

The motion analysis of angular variations between lines permitted us to estimate the pose of a given object. Therefore, we propose a robust method to compute the orientation through rotations. These rotations must be known. They could be sensed or given and controlled as is the case of robotic applications. This information also gives rise to address the uniqueness issue and study motion patterns that permit to solve the problem in an easier way. Chen in [18] addressed questions that are useful for our analysis such as the number of movements needed, and the motion patterns required to obtain a unique solution for the relative orientation and consequently to the relative position. These questions were applied in hand-eye calibration problems which in the case the camera is not fixed to a robot gripper are similar to ours.

IV. Pose Estimation Algorithm

Motion analysis of feature lines was the base of our pose estimation algorithm. In this case known 3D rotations of a line and its subsequent projections in the image plane were related to compute its relative position and orientation with respect to a perspective camera. Vision problems as feature extraction and line correspondences are not discussed and we suppose the focal distance f as known. Our goal is, having this image and motion information, estimate the pose of an object represented by feature lines with the minimum number of movements and identify patterns that permits to compute a unique solution without defined initial conditions.

A. Mathematical Analysis

A 3D plane is the result of the projection of a line in the image plane. It is called the projection plane and passes through the projection center of the camera and the 3D line. This 3D line, in this case, is the representation of an object. With three views after two different rotations of the object, three lines are projected in the image plane. Thus three projection planes can be calculated. These planes are P_a, P_b and P_c, and their intersection is a 3D line that passes through the projection center and the centroid of the rotated object. Across this line a unit director vector $\boldsymbol{v}_d$ can be determined easily by knowing f and the intersection point of the projected lines in the image plane. Our intention is to use this 2D information to formulate angle relations with the 3D motion data.

Working in the 3D space permits to take advantage of the motion data. In this case where the object is represented by a 3D line, the problem could be seen as a unit vector across its direction that is rotated two times. In each position of the three views this unit vector lies in one of the projection planes as seen in Fig. 4. It is first located in P_a, then it rotates an angle α_1 to lie in P_b and ends in P_c after the second rotation by an angle α_2. To estimate the relative orientation of the object we first must obtain the location of three unit vectors, $\boldsymbol{v}_a$, $\boldsymbol{v}_b$ and $\boldsymbol{v}_c$, that coincide with the 3D motion data and lie on their respective planes. To do this we know that the scalar product of:

$$v_a . v_b = \cos\alpha_1 \tag{1}$$

$$v_b . v_c = \cos\alpha_2 \tag{2}$$

Calculating the angle γ between the planes formed by $\boldsymbol{v}_a\boldsymbol{v}_b$ and $\boldsymbol{v}_b\boldsymbol{v}_c$ from the motion information, we have

$$\left(v_a \times v_b\right)\left(v_b \times v_c\right) = \cos\gamma \tag{3}$$

And applying vector identities

$$v_a . v_c = \cos\alpha_1 \cos\alpha_2 - \cos\gamma \tag{4}$$

With the set of equations conformed by (1), (2) and (4) we have the three unit vectors to calculate. However, there is not a unique solution, thus constraints must be applied.

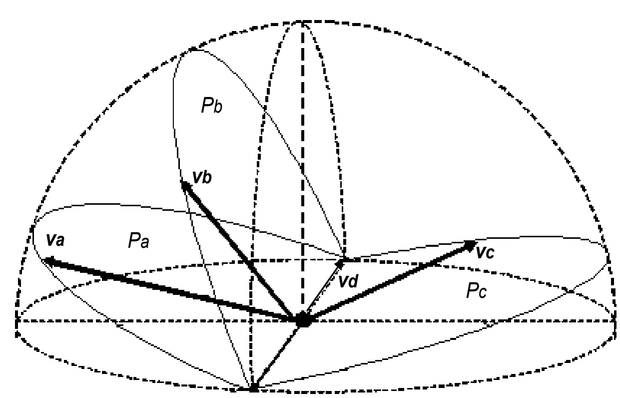

Fig. 4. Unit vectors *Va, Vb* and *Vc* are constrained to lie on planes *Pa*, *Pb* and *Pc* respectively. Their estimation can be seen as a semi sphere where their combination must satisfy the angle variations condition.

B. Projection Planes Constraint

There are many possible locations where the three unit vectors can satisfy the equations in the 3D space. To obtain a unique solution unit vectors $\boldsymbol{v}_a$, $\boldsymbol{v}_b$ and $\boldsymbol{v}_c$ must be constrained to lie in their respective planes. Unit vector $\boldsymbol{v}_a$ could be seen as any unit vector in the plane P_a rotated through an axis and an angle. Using unit quaternions to express $\boldsymbol{v}_a$ we have

$$v_a = q_a v q_a^* \tag{5}$$

where $\boldsymbol{q}_a$ is the unit quaternion applied to $\boldsymbol{v}_a$, $\boldsymbol{q}_a^*$ is its conjugate and $\boldsymbol{v}$ is any vector in the plane. For every rotation about an axis $\boldsymbol{n}$, of unit length, and angle Ω, a corresponding unit quaternion $\boldsymbol{q} = (\cos \Omega/2, \sin \Omega/2\ \boldsymbol{n})$ exists. Thus $\boldsymbol{v}_a$ is expressed as a rotation of $\boldsymbol{v}$, about an axis and an angle by unit quaternions multiplications. In this case $\boldsymbol{n}$ must be normal to the plane P_a if both unit vectors $\boldsymbol{v}_a$ and $\boldsymbol{v}$ are restricted to be in the plane.

Applying the plane constraints and expressing $\boldsymbol{v}_a$, $\boldsymbol{v}_b$ and $\boldsymbol{v}_c$ as mapped vectors through unit quaternions, equations (1), (2) and (4) can be expressed as a set of three nonlinear equations with three unknowns

$$q_a v_d q_a^* q_b v_d q_b^* = \cos\alpha_1 \tag{6}$$

$$q_b v_d q_b^* q_c v_d q_c^* = \cos\alpha_2 \tag{7}$$

$$q_a v_d q_a^* q_c v_d q_c^* = \cos\alpha_1 \cos\alpha_2 - \cos\gamma \tag{8}$$

The vector to be rotated is $\boldsymbol{v}_d$, which is common to the three planes, and their respective normal vectors are the axes of rotation. Extending the equations (6), (7) and (8), multiplying vectors and quaternions, permits to see that there are only three unknowns which are the angles of rotation Ω_a, Ω_b and Ω_c.

C. Relative Position and Orientation Estimation

Applying iterative numerical methods to solve the set of nonlinear equations, the location of $\boldsymbol{v}_a$, $\boldsymbol{v}_b$ and $\boldsymbol{v}_c$ with respect to the camera frame in the 3D space are calculated. Now we have a simple 3D orientation problem that can be solved easily by a variety of methods as least square based techniques. However, in the case where motions could be controlled and selected movements applied, this last step to estimate the relative orientation would be eliminated. Rotation information would be obtained directly from the numerical solution. If we assume one of the coordinate axes of the object frame coincide with the moving unit vector and apply selected motions, as one component rotations, a unique solution is provided faster and easier.

The estimation of the relative orientation serves now to compute the relative position. It is necessary to track the projection of the end point of the 3D line that represents the object. The relation between the 3D position of this end point, (X, Y, Z), and its projection in the image plane, (x, y), can be expressed as $x=fX/Z$ and $y=fY/Z$. In our case where we know the 3D transformations applied to the object and its image projections, the relative position can be easily

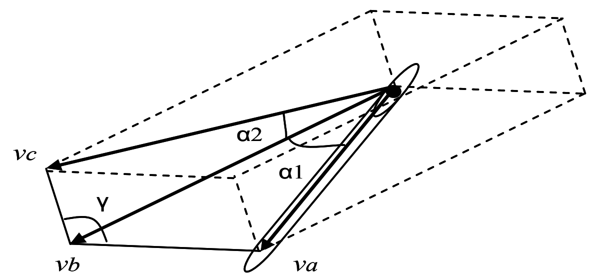

Fig. 5. Two selected rotations through normal axes simplify computations.

calculated rotating them to camera frame coordinates and by relating point position differences ΔX, ΔY and ΔZ and their projections Δx and Δy. Thus two views are sufficient to obtain the position of the centroid if the length to its end point is provided.

D. Uniqueness of Solution and Motion Patterns Analysis

The pose estimation algorithm explained above guarantee a unique solution. With two views, there would be many combinations of unit vectors constrained to their planes that satisfy the angular variations condition. The third view constrains the set of combinations, thus the solution is unique. It is estimated from two rotations and is considered as a robust orientation method due to the use of angle between lines. A pattern of motions analysis is useful in the case movements are controlled and can be selected to simplify calculations. This is the case of robotic surgery.

In this case where small rotations in different axes provide a unique solution and the object is represented by a 3D line, some selected rotations provide faster computations. If it is assumed that one of the axes of the coordinate frame of the object is the unit vector that describes the 3D line direction, with two different one component rotations, the relative orientation result comes directly from the numerical solution. With normal axes of rotations the angle between the planes formed by $\boldsymbol{v}_a\boldsymbol{v}_b$ and $\boldsymbol{v}_b\boldsymbol{v}_c$ is $\pi/2$, as seen in Fig. 5. It also serves to reduce calculations. This is relevant when applications require real time performance, providing fast results which in many cases should be updated frequently.

V. Simulations

Simulations were performed through a graphics interface application where 3D transformations could be handled. The object to be transformed was a line segment where one of its end points was the centroid. The position in the 3D space of the view point was known and represented the camera. Input parameters were the respective angles for each rotation of the object, α_1 and α_2, and the angle between the planes generated by these rotations, γ. Image changes were identified through straight line detection algorithms. The parameters to measure from the 2D data were the angular variation, β_1 and β_2, corresponding to the two 3D rotations,

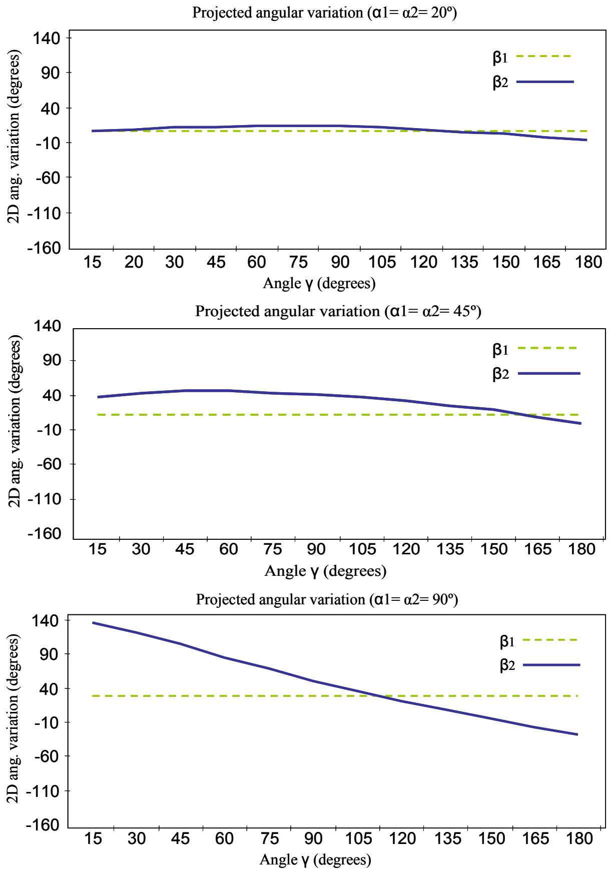

Fig. 6. 2D angular variations induced by 3D rotations. With γ variable, different values of α_1 and α_2 produce changes in the 2D angular variation.

and the unit vectors that describe the projection planes of each line to compute the relative orientation.

In Fig. 6 can be observed the angular variations projected in the image plane, β_1 and β_2, through different 3D rotations of α_1 and α_2. There is an increment of the angular variation with greater changes of motion. With only variations of γ and static rotation angles, the angular difference β_2 also varies. And there are values of γ where the angular variation is higher with small variations of α. Thus γ can be employed to determine the minimum number of movements to estimate the orientation easier and decrease errors. In the case of 3D transformations with high angular variations the 2D difference decreases constantly. This case is not useful; motions with such a magnitude are not always applicable. However, observing the diversity of variations in the 3D space projected in the plane give us an idea of the information acquired to solve the orientation problem.

VI. Experimental Results

Real world data was used to validate the algorithm. Experiments were carried out through a robotic test bed that was developed in order to get high repeatability. It consists on an articulated robotic arm with a calibrated tool frame equipped with a tool, easily described by a line, which is presented in different precisely known orientations to a camera. The camera field of view remains fixed during the image acquisition sequence. The tool center of rotation is programmed to be out of the field of view, as it would be presented in an action-perception application.

With this premises, a standard analog B/W camera equipped with known focal length optics has been used. This generates a wide field of view that is sampled at 768x576 pixels resolution. After image edge detection, Hough Transform is used in order to obtain the tool contour and the straight line in the image plane associated to it. Tool contour is supposed to have the longest number of aligned pixel edges in the image.

Having this setup, feature lines were identified and located in a sequence of images. This images captured selected positions of the rotated tool. Once the equation of the lines projected in the image plane were acquired, unit vectors normal to the constraint planes and $\boldsymbol{v}_d$ could be calculated. This unit vectors and the motion angles α_1 and α_2 served as the input to the proposed algorithm. The intersection of the lines was needed to calculate $\boldsymbol{v}_d$. This calculation is prone to errors due to be located out of the field of view. It means the intersection of a different number of lines is not usually the same point. The intersections converge to a single point when the angles between lines are higher.

The 3D transformation resultant from the algorithm was tested reprojecting 3D lines, derived by new tool rotations, in the image plane and comparing them with the line detected by the vision system. Tests for motion angles between 5 and 20 degrees were carried out. Fig. 7 shows the error between lines through different angles. There can be seen how the error is minimum at the position where the transformation was calculated, it means at its second motion or third image. This error varies depending on the position of the tool, it increases with higher angles, when the position of the tool separates from the minimum error position. Fig. 8 compares the algorithm performance for 5 and 20 degrees motion angles. There the error varies differently. In the case of 5 degrees the error increases greatly with each motion that separates the tool from the minimum error position. While for 20 degrees this error also increments, but remains stable.

This results validate the line-based algorithm and its simplicity demonstrates its real-time performance. The error

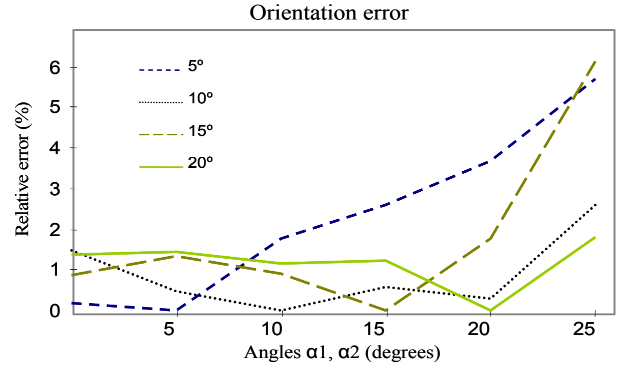

Fig 7. Relative error using the rotational motion analysis algorithm.

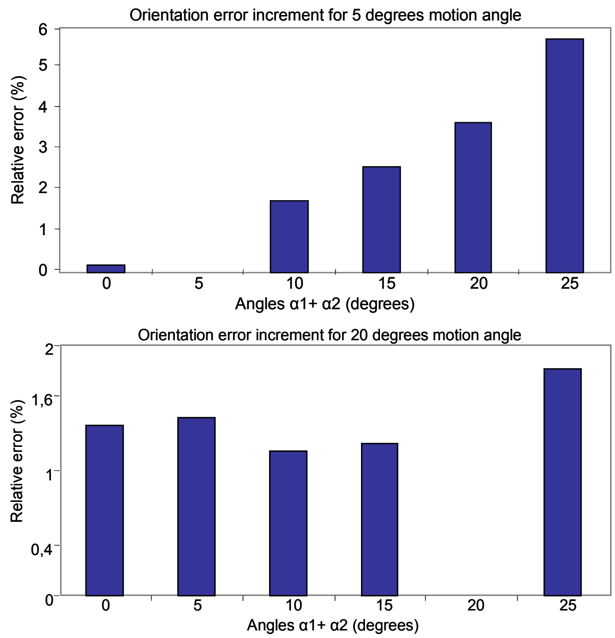

Fig 8. Algorithm performance comparison between 5 and 20 degrees, with a first rotation α1 followed by a second α2 of the same magnitude.

increment with large position separations is mainly product of the deviation at the intersection point. It can be seen that the calculation of $\boldsymbol{v}_d$ has a great impact in the result and future work should be focused in this issue.

VII. Conclusion

A robust method to estimate the relative pose of an object moved by a robot with respect to a camera has been presented. The object was assumed to be a surgical tool that could be represented by feature lines. 2D correspondences of a line due to known 3D transformations of the tool were the information used to calculate its pose. We have showed that with only two rotations the angular variation between lines provides sufficient information to estimate the relative orientation. The relative position can also be calculated by tracking the end point of the object after the orientation had been determined if its length from the centroid is given or other scale information is known. This motion analysis led to address questions as the uniqueness of solution for the minimum number of movements and possible motion patterns to solve it directly. In the case of controlled motions, one component rotations through normal axes simplify calculations to provide a robust technique to estimate the relative orientation without defined initial conditions.

This method can be used in assisted MIS to locate the instruments with respect to the surgeon. An enhanced perception can then be offered by augmenting the 2D information imposing position and orientation data through a human-machine interface.

References

[1] A. Healey, "Speculation on the neuropsychology of teleoperation: implications for presence research and minimally invasive surgery," *Presence: Teleoperators and Virtual Environments*, MIT Press, 2006.
[2] N. Navab and O. Faugeras, "Monocular pose determination from lines: Critical sets and maximum number solutions," *Proc. Computer Vision and Pattern Recognition*, New York, pp. 254-260, 1993.
[3] H.C. Homer, "Pose determination from line-to-plane correspondences: Existence condition and closed-form solutions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 530-541, 1991.
[4] R. Kumar and A.R. Hanson, "Robust methods for estimating pose and sensitivity analysis," *Computer Vision and Image understanding*, vol. 60, pp. 313-342, 1994.
[5] A. Ansar and K. Daniilidis, "Linear pose estimation from points or lines," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 4, 2003.
[6] Y. Liu, T.S. Huang and O. Faugeras, "Determination of camera location from 2-D to 3-D line and point correspondences," *IEEE Trans. Pattern Analysis and machine Intelligence*, vol. 12, pp. 28-37, 1990.
[7] J.R. Holt and A.N. Netravali, "Uniqueness of solution to structure and motion from combinations of point and line correspondences," *Journal of Visual Communication and Image Representation*, vol. 7, no. 2, pp. 126-136, 1996.
[8] P. Dutkiewicz, M. Kielczewski, M. Kowalski and W. Wroblewski, "Experimental verification of visual tracking of surgical tools," *Fifth International workshop on Robot Motion and Control*, 2005.
[9] J. Sun, M. Smith, L. Smith and L.P. Nolte, "Simulation of an optical-sensing technique for tracking surgical tools employed in computer-assisted interventions," *IEEE Sensors Journal*, vol. 5, no. 5, 2005.
[10] S. Payandeh, Z. Xiaoli and A. Li, "Application of imaging to the laparoscopic surgery," *IEEE Int. Symp. Computer Intelligence in Robotics and Automation*, pp. 432-437, 2001.
[11] P. Milgram, S. Zhai, D. Drascic and J. Grodski, "Applications of augmented reality for human-robot communication," *Proc. Int. Conf. on Intelligent Robotics and Systems*, pp. 1467-1472, 1993.
[12] F. Devernay, F. Mourgues and E. Coste-Maniere, "Towards endoscopy augmented reality for robotically assisted minimally invasive cardiac surgery," *IEEE Proc. Int. Workshop on Medical Imaging and Augmented Reality*, pp. 16-20, 2001.
[13] A. Pandya and G. Auner, "Simultaneous augmented and virtual reality for surgical navigation," *IEEE Annual Meeting of the North American Fuzzy Information Processing Society*, 2005.
[14] B.L. Yen and T.S. Huang, "Determining 3-D motion and structure of a rigid body using straight line correspondences," *Image sequence Processing and Dynamic Scene Analysis*, Heidelberg, Germany: Springer-Verlag, 1983.
[15] O. Faugeras, F. Lustman and G. Toscani, "Motion and structure from point and line matches," *Proc. First Int. Conf. Computer Vision*, 1987.
[16] J. Weng, T.S. Huang and N. Ahuja, "Motion and structure from line correspondences; closed-form solution, uniqueness, and optimization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 318-336, 1992.
[17] A. Mitiche, S. Seida and J.K. Aggarwal, "Interpretation of structure and motion using straight line correspondences," *Proc. Int. Conf. Pattern Recognition*, pp. 110-1112, 1986.
[18] H. Chen, "A screw-motion approach to uniqueness analysis of head-eye geometry," *IEEE Proc. Conf. Computer Vision and Pattern Recognition*, pp. 145-151, 1991.