Position and Motion Estimation for Visual Robot Control with Planar Targets

H.L. Yung, G. Chesi and Y.S. Hung
Department of Electrical and Electronic Engineering, The University of Hong Kong
Pokfulam Road, Hong Kong
E-mail: hlyung@eee.hku.hk, chesi@eee.hku.hk, yshung@eee.hku.hk

Abstract

This paper addresses two problems in visually-controlled robots. The first consists of positioning the end-effector of a robot manipulator on a plane of interest by using a monocular vision system. The problem amounts to estimating the transformation between the coordinates of an image point and its three-dimensional location supposing that only the camera intrinsic parameters are known. The second problem consists of positioning the robot end-effector with respect to an object of interest free to move on a plane, and amounts to estimating the camera displacement in a stereo vision system in the presence of motion constraints. For these problems, some solutions are proposed through dedicated optimizations based on decoupling the effects of rotation and translation and based on an a-priori imposition of the degrees of freedom of the system. These solutions are illustrated via simulations and experiments.

1 Introduction

Robot control based on artificial vision is an important area of robotics with useful applications. Indeed, artificial vision may allow robots to imitate human beings in performing simple operations such as grasping a cup of coffee, as well as difficult operations such as threading a needle. Technically speaking, artificial vision may be used as feedback information so that a robot can reach a desired location and/or touch a desired object with its end-effector. This information is provided by visual sensors such as cameras that acquire images of the scene around the robot. These images describe if and how the robot and its end-effector are moving toward the goal, and hence constitute a feedback information.

The applications of robotic systems with artificial vision are numerous and various. To name but a few, one can cite the industrial manufacture, where robotic arms are used for grasping and positioning tools and objects. Other applications can be in surveillance, where a mobile camera observes an area of interest such as an entrance, and in vehicles alignment, as in car parking and airplane landing. Also, robots equipped with vision find application in surgery, where an instrument has to be guided to an organ to operate, and in dangerous environments such as nuclear stations and spatial missions, where humans have to be replaced.

Depending on the number and position of the cameras, and on how the information provided by these cameras is exploited by the system, several configurations of robot control based on artificial vision can be obtained. For instance, the cameras can be mounted on the robot end-effector in the so called eye-in-hand configuration (also known as hand-eye), or can be positioned somewhere in the scene separately from the robot in the so called eye-to-hand configuration (also known as static-eye). Then, each camera may be used to inspect a different region of the scene (monocular vision), or all cameras may be used to observe a common set of objects (multi-camera vision, known as stereo vision in the case of two cameras). Lastly, the images acquired by the cameras can be used by the control system to define the goal only at the beginning of the task (open-loop control) or exploited during the robot motion in order to progressively update the goal (closed-loop control). See for instance [1]-[7] and [12]-[15] for details.

This paper presents some applications of robot control based on artificial vision, in particular considering the following two problems. First, the task of positioning the end-effector of a robot manipulator on a plane of interest by using the view of the scene provided by a camera is addressed. Specifically, the camera is supposed to observe the robot and its workspace, and one defines the target position to be reached by the robot end-effector in the view of the camera. The problem hence amounts to estimating the transformation relating the coordinates of a point chosen on the image and its three-dimensional location supposing that only the camera intrinsic parameters are known, and for this problem an optimization based on decoupling the effects of rotation and translation is proposed. The second problem consists of positioning this end-effector with respect to an object of interest which is free to move on a plane, and hence amounts to estimating the camera displacement in a stereo vision system in the presence of motion constraints. For this problem, a solution is proposed based on the estimation of the homography matrix and its decomposition in rotation and translation by taking into account the reduced degrees of

Acknowledgement: The work described in this paper was supported by the Research Grants Council of Hong Kong Special Administrative Region, China (Project Nos. HKU711208E and HKU712808E).

freedom. Simulations and experiments are reported to illustrate the proposed strategies.

The paper is organized as follows. Section 2 introduces the problem formulation. Section 3 presents the proposed strategies. Section 4 presents the simulations and experiments. Lastly, Section 5 provides some final comments.

2 Problem Formulation

We consider the problem of positioning a robot end-effector on a plane using a fixed camera. This problem is addressed in the following two situations.

2.1 Positioning with one view

Let F_R be the coordinate frame of the robot, and let F_C be the coordinate frame of a camera observing the robot end-effector. It is assumed that the origin and orientation of F_C coincide respectively with the center and axes of the camera, and that the motion of the robot end-effector is restricted on a plane Π as shown in Figure 1.

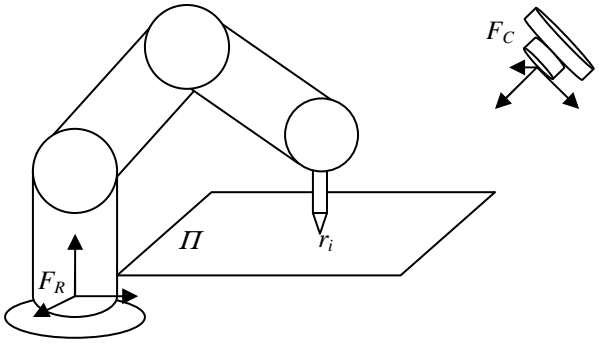

Figure 1: Problem formulation of position with one view

Let $r_i \in \mathbb{R}^3$ be a feature point on Π expressed with respect to F_R. This point satisfies the relation:

$$a^T r_i + b = 0 \tag{1}$$

where $a \in \mathbb{R}^3$ and $b \in \mathbb{R}$ are constants describing Π with respect to F_R.

Let $R_{RC} \in \mathbb{R}^{3\times3}$ and $t_{RC} \in \mathbb{R}^3$ be the rotation matrix and translation vector describing the motion between F_R and F_C. The point r_i can be hence expressed as:

$$r_i = R_{RC} q_i + t_{RC} \tag{2}$$

where $q_i \in \mathbb{R}^3$ is the point r_i expressed with respect to F_C.

Let $p_i \in \mathbb{R}^3$ denote the projection (in homogeneous coordinates) on the image plane of the camera of r_i. The standard pin-hole camera projection model provides the relation:

$$\lambda_i p_i = K q_i \tag{3}$$

where $\lambda_i \in \mathbb{R}$ is the scaling factor and $K \in \mathbb{R}^{3\times3}$ is the upper-triangular intrinsic camera calibration matrix.

The problem consists of estimating R_{RC}, t_{RC}, a and b satisfying (1)-(3) using a set of training pairs (r_i, p_i), i=1,…,k, that are assumed to be known initially. Subsequently, the estimated R_{RC}, t_{RC}, a and b allow one to estimate r_i from any user defined p_i, and hence to position the robot end-effector to a desired point lying on Π specified in the image.

2.2 Positioning with two views

Let us consider a planar object φ in two different locations with coordinate frames F_P and F_P^* observed by a fixed camera with coordinate frame F_C as shown in Figure 2. We assume that the object has undergone a planar motion from F_P to F_P^*, i.e. F_P^* is obtained from F_P via a translation on the object plane and a rotation about the normal to the object plane. Let $s_i \in \mathbb{R}^3$ denote the ith feature point on the object. Let $m_i, m_i^* \in \mathbb{R}^3$ denote the projections on the image plane of the camera of the point s_i with the object in the locations F_P and F_P^* respectively. These projections are given by:

$$m_i = \xi_i K(R_{PC} s_i + t_{PC}) \tag{4}$$

$$m_i^* = \xi_i^* K(R_{PC}^* s_i + t_{PC}^*) \tag{5}$$

where ξ_i , $\xi_i^* \in \mathbb{R}$ are scaling factors, and R_{PC}, $R_{PC}^* \in \mathbb{R}^{3\times3}$ and t_{PC} , $t_{PC}^* \in \mathbb{R}^3$ are the rotation matrices and translation vectors describing the motions between F_P and F_C and between F_P^* and F_C.

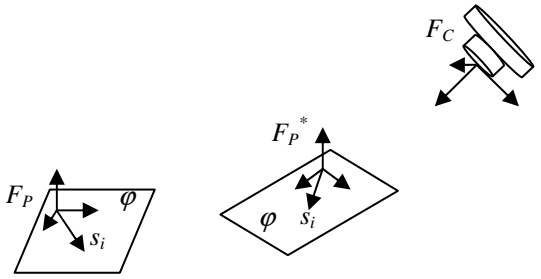

Figure 2: Problem formulation of position with two views

The problem consists of estimating the rotation matrix and translation vector describing the motion between F_P and F_P^*, denoted by $R \in \mathbb{R}^{3\times3}$ and $t \in \mathbb{R}^3$ respectively, supposing that an estimate of K and a set of image measurements (m_i, m_i^*), i=1,…,j, are available. In fact, R and t allow one to move the robot end-effector from the current location F_P to the desired (unknown) location F_P^*.

3 Proposed Solution

3.1 Positioning with one view

Assume p_i is expressed in homogenous coordinates, i.e.:

$$p_i = [u_i \quad v_i \quad 1]^T \tag{6}$$

for some $u_i, v_i \in \mathbb{R}$.

From (3), the following equations can be deduced:

$$u_i = \frac{e_1^T K q_i}{e_3^T K q_i} \quad , \quad v_i = \frac{e_2^T K q_i}{e_3^T K q_i} \tag{7}$$

where e_i is the ith column of the 3×3 identity matrix I_3.

The expression (7) can be rewritten as:

$$M_i q_i = 0 \tag{8}$$

where $$M_i = \begin{bmatrix} u_i e_3^T K - e_1^T K \\ v_i e_3^T K - e_2^T K \end{bmatrix} \in \mathbb{R}^{2\times 3} \tag{9}$$

is independent of q_i.

From (2), q_i can be expressed in terms of r_i as follows:

$$q_i = R_{RC}^T (r_i - t_{RC}) \tag{10}$$

By combining (8) and (10), the following relationship is obtained:

$$M_i R_{RC}^T r_i - M_i R_{RC}^T t_{RC} = 0 \tag{11}$$

With several samples of r_i and p_i, the problem is hence:

$$\min_{R_{RC}, t_{RC}} \sum_i \left\| M_i R_{RC}^T r_i - M_i R_{RC}^T t_{RC} \right\|^2 \tag{12}$$

Since R_{RC} is a rotation matrix, it can be expressed as a matrix exponential of a skew-symmetric matrix, i.e.:

$$R_{RC} = e^{[\theta]_\times} \tag{13}$$

where $\theta \in \mathbb{R}^3$ and $[\ ]_\times$ denotes the skew-symmetric matrix as follows:

$$[\theta]_\times = \begin{bmatrix} 0 & -\theta_3 & \theta_2 \\ \theta_3 & 0 & -\theta_1 \\ -\theta_2 & \theta_1 & 0 \end{bmatrix}, \qquad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \tag{14}$$

If θ is fixed, (11) reduces to a linear system in t_{RC}, in particular:

$$A_i t_{RC} = c_i \tag{15}$$

where $A_i = M_i R_{RC}^T$ and $c_i = M_i R_{RC}^T r_i$.

If k pairs of corresponding r_i and p_i are used to solve for t_{RC}, the equations are stacked up in the following form:

$$A t_{RC} = c \tag{16}$$

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix} \in \mathbb{R}^{2k\times 3}, \qquad c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} \in \mathbb{R}^{2k} \tag{17}$$

From (16) t_{RC} can be estimated via linear least-squares as follows:

$$\hat{t}_{RC} = (A^T A)^{-1} A^T c \tag{18}$$

The least-squares error is hence given by:

$$\varepsilon = \left\| A\hat{t}_{RC} - c \right\| \tag{19}$$

Since this error depends on θ, we can estimate θ by solving:

$$\min_\theta \varepsilon(\theta) \tag{20}$$

for instance via gradient-descent methods. This provides an estimate $\hat{\theta}$. From $\hat{\theta}$, we find $\hat{R}_{RC}$ via (13) and $\hat{t}_{RC}$ via (18). Lastly, a and b are estimated from (1) via simple linear least-squares, hence obtaining $\hat{a}$ and $\hat{b}$.

Let us observe now that, each pair of corresponding r_i and p_i, gives two scalar equations from (8) and one scalar equation from (1). Since in (11) there are six scalar variables (three in θ and three in t_{RC}) and in (1) there are four scalar variables, it follows that at least 4 pairs of corresponding r_i and p_i are required, i.e. $k \geq 4$.

Once $\hat{R}_{RC}$, $\hat{t}_{RC}$ $\hat{a}$ and $\hat{b}$ have been found, one can estimate the position of a new feature point $r \in \mathbb{R}^3$ on the plane Π from an estimate of its projection $p \in \mathbb{R}^3$ on the camera. Indeed, from (1) and (2), it follows that (replacing r_i and q_i with r and q respectively):

$$\hat{a}^T (\hat{R}_{RC} q + \hat{t}_{RC}) + \hat{b} = 0 \tag{21}$$

$$\bar{a}^T q + \bar{b} = 0 \tag{22}$$

where $\bar{a} = \hat{R}_{RC}^T \hat{a}$ and $\bar{b} = \hat{a}^T \hat{t}_{RC} + \hat{b}$. Then, by using (8) and (22), one has that:

$$\begin{bmatrix} M \\ \bar{a}^T \end{bmatrix} q = \begin{bmatrix} 0 \\ -\bar{b} \end{bmatrix} \tag{23}$$

where M is calculated as in (9) with p_i replaced by p. This provides q, and consequently r from (2).

3.2 Positioning with two views

Since the feature points are on the plane, m_i and m_i^* are related as follows:

$$m_i^* = G m_i \tag{24}$$

where $G \in \mathbb{R}^{3\times 3}$ is known as a collineation matrix which can be further decomposed as follows:

$$G = KHK^{-1} \tag{25}$$

where $H \in \mathbb{R}^{3\times 3}$ is known as a homography matrix, see for instance [11].

H can be estimated given an estimate of K and a set of estimates (m_i, m_i^*), i=1,…,j. From H, the rotation $R \in \mathbb{R}^{3\times 3}$ and translation $t \in \mathbb{R}^3$ between F_P and F_P^* can be obtained by using suitable decomposition methods (see, e.g., [8] and [10]).

However, these methods consider a general motion between F_P and F_P^* with six degrees of freedom, while in our case the motion between F_P and F_P^* is constrained to a planar class with three degrees of freedom. It can be expected that, by taking into account this constraint in the decomposition of H, more accurate results can be obtained compared with the case where this constraint is not considered. Therefore, our target is to derive a new procedure for decomposing H into R

and t where the motion constraint is taken into account a priori.

Let us start by observing that the problem of planar object in two positions can be reduced to stereo camera problem [9]. Using [8], the homography matrix H can be decomposed as:

$$H = R_{get} + \frac{t_{get} n^T}{d} \tag{26}$$

where: $$R_{get} = R_{PC} R^T {R_{PC}}^T \tag{27}$$

and $$t_{get} = (I_3 - R_{PC} R {R_{PC}}^T) t_{PC} - R_{PC} R^T t \tag{28}$$

Since the motion between F_P and F_P^* is restricted to be planar, R and t should have the following forms:

$$R = \begin{bmatrix} R_{\#} & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad t = \begin{bmatrix} t_{\#} \\ 0 \end{bmatrix} \tag{29}$$

where $R_{\#} \in \mathbb{R}^{2\times 2}$ is a rotation matrix and $t_{\#} \in \mathbb{R}^2$ is a translation vector.

If H is pre- and post-multiplied by ${R_{PC}}^T$ and R_{PC} respectively, we have:

$$\overline{H} = {R_{PC}}^T H R_{PC} = R^T + \frac{{R_{PC}}^T t_{get} (R_{PC}^T n)^T}{d} \tag{30}$$

Concerning ${R_{PC}}^T n$, since the motion between F_P and F_P^* is restricted to be planar, the normal vector is parallel to the z-axis of F_P or F_P^*, which means that if the normal vector is expressed with respect to F_C, the following relationship can be deduced:

$$n = w_3 \quad \text{where } R_{PC} = [w_1 \quad w_2 \quad w_3] \tag{31}$$

Therefore, $$ {R_{PC}}^T n = [0 \quad 0 \quad 1]^T \tag{32}$$

Concerning ${R_{PC}}^T t_{get}$, by using (29), it can be shown that:

$${R_{PC}}^T t_{get} = \begin{bmatrix} \tilde{t} \\ 0 \end{bmatrix} \tag{33}$$

Combining the results in (29), (30), (32) and (33), the following relationship can be found:

$$\overline{H} = \begin{bmatrix} R_{\#}^T & \tilde{t}/d \\ 0 & 1 \end{bmatrix} \tag{34}$$

As a result, the following relationship holds:

$$\overline{m}_i^* = \overline{H} \overline{m}_i \tag{35}$$

where $\overline{m}_i^* = {R_{PC}}^T K^{-1} m_i^*$ and $\overline{m}_i = {R_{PC}}^T K^{-1} m_i$.

In other words, the degree of freedom of the homography matrix is reduced in the case of planar motion and hence a more accurate result should be obtained.

4 Results

Experiments with simulation and real data are done in order to justify the proposed solution and compare with other existing methods.

4.1 Simulation results

Simulations are done with synthetic data in order to illustrate the performance of the proposed method.

4.1.1 Simulation results of positioning with one view

Eight points are selected as the position of the end-effector of the robot with their coordinates r_i being known such that all the points are coplanar. The camera, with its intrinsic parameters defined, is placed randomly with respect to the robot coordinate frame such that the rotation and translation between the robot and the camera are known. The points r_i are projected onto the image plane by (2) and (3) so that their corresponding pixel coordinates are calculated. Due to optimization error, the retrieved rotation and translation is not exactly the same as originally defined. The pixel coordinates are back-projected to the robot coordinate frame assuming that the equation of the plane is known. Simulation is done in order to examine the average rotation and translation errors and the average back-projection error.

Gaussian noise with standard deviation ranging from 0 to 4 units is added to the pixel coordinates. The algorithm is run 20 times for each noise level. The average rotation and translation errors are plotted in Figure 3 and Figure 4.

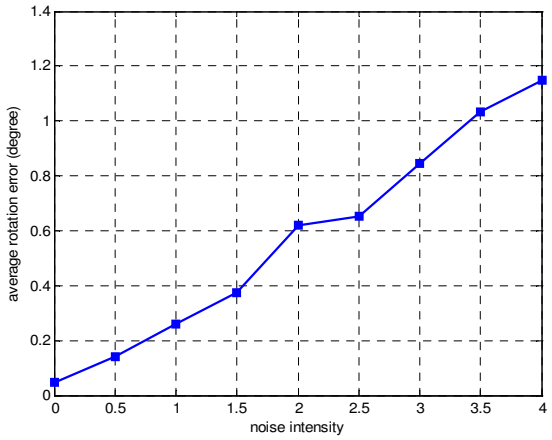

Figure 3: Average rotation error

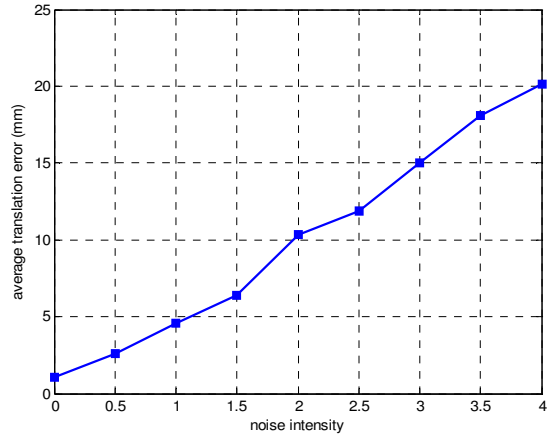

Figure 4: Average translation error

4.1.2 Simulation results of positioning with two views

Four feature points s_i are selected on a planar object. The camera is placed randomly with respect to the object provided that the rotation and translation between the camera and the object are known. The intrinsic camera calibration matrix is also defined. Using (4), the pixel coordinates can be calculated. Next the object is moved randomly provided that the rotation and translation of the object at two locations are known. The new pixel coordinates are calculated using (5).

The standard homography decomposition method (see, e.g., [8] and [10]) is applied to the same data for the purpose of comparison with the proposed method. Gaussian noise with standard deviation ranging from 0 to 4 pixels with 0.5 pixel interval is added to the pixel coordinates. The algorithms are run 100 times for each noise level and the average errors of rotation and translation of the object retrieved using two methods are plotted in Figure 5 and Figure 6.

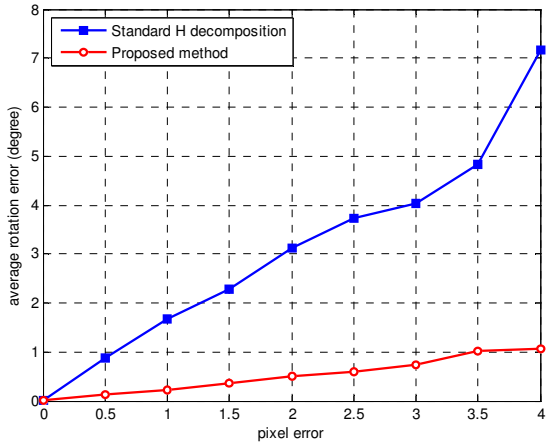

Figure 5: Average error in rotation

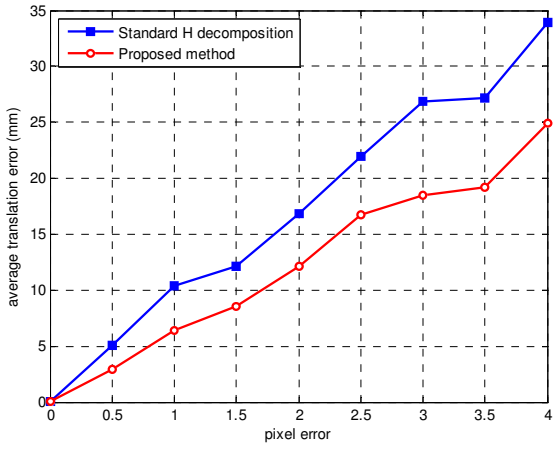

Figure 6: Average error in translation

It can be seen that the average errors on rotation and translation of the proposed method are significantly less than that of the standard homography decomposition method.

4.2 Experimental results

The algorithm is tested using a 6-DoF articulated robot arm and a calibrated camera mounted at a fixed position looking at the workspace of the robot where a piece of A4 paper with marks on it is used as the planar object as shown in Figure 7.

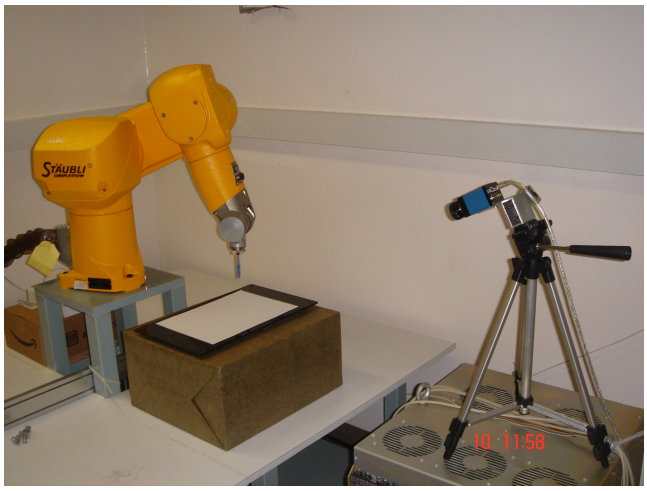

Figure 7: Testing environment

4.2.1 Experimental results of positioning with one view

The robot is manipulated to move to eight designated positions on a plane parallel to the x-y plane of the base coordinate frame so that their coordinates with respect to the base coordinate frame are known. The pixel coordinates of the end-effector of the robot are recorded corresponding to its 3D positions. The pairs of coordinates are used as training samples to calibrate the robot with respect to the camera using the method proposed in Section 3.1.

Thirteen points printed on a piece of paper are used to estimate the actual error of the algorithm. Those points visible by the camera are selected on the image so their pixel coordinates are known. Their back-projected coordinates are calculated where the robot is driven to. The errors are measured as the distance between the positions of the end-effector of the robot and the corresponding points.

We define x-direction being parallel to the long side of the paper while y-direction being parallel to the short side of the paper. The errors are measured and presented at the following table:

	max. error	min. error	avg. error
x-direction	3mm	0mm	1.31mm
y-direction	6mm	1mm	3.73mm

Table 1: Back-projection errors

That errors in the y-direction are larger than in the x-direction is expected because disparity in depth is less discernable than in the lateral direction of the camera, given that only one view of the scene is available.

4.2.2 Experimental results of positioning with two views

A sheet of A4 paper is used as the planar object with its four

corners and four additional markers on the paper as the feature points. The robot is used to verify the algorithm. First, the initial position of the paper, as shown in Figure 7, is identified by the camera and a line segment is drawn on the paper by a robot. Next, the paper is moved to an arbitrary position as shown in Figure 8 and the robot is to draw an extension of the line segment from its previous endpoint according to the rotation and translation calculated using both the standard and the proposed method of homography decomposition. If the situation is perfect, there should be no gap between the two line segments and their orientations should be consistent. This experiment can be used to estimate the rotation and translation errors of both methods in a real scenario.

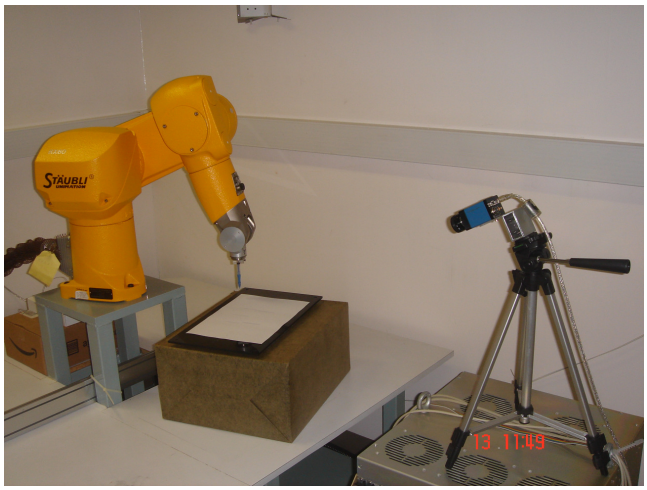

Figure 8: Paper in new position

The errors are measured and recorded. The orientations of the lines are basically consistent using both methods, which means that the rotation error is small. The translation error using proposed method is 7mm while that using standard homography decomposition method is 9mm, which means that the proposed method achieves about 22% improvement in translation.

5 Conclusion

We have considered two applications of visually- controlled robots, addressing the problems of positioning the end-effector of a robot manipulator on a plane of interest by using monocular and two-view vision. The problems amount to estimating the existing transformations between the coordinates of some available image points and the three-dimensional location of the robot end-effector corresponding to these points. Solutions have been proposed through dedicated optimizations based on a-priori imposition of the degrees of freedom of the system. Future work will be devoted to improving the proposed solutions in order to achieve higher positioning accuracies.

References

[1] S. Hutchinson, G. Hager, and P. Corke, “A tutorial on visual servo control,” IEEE Trans. on Robotics and Automation, vol. 12, no. 5, pp. 651–670, 1996.

[2] K. Hashimoto, “A review on vision-based control of robot manipulators,” Advanced Robotics, vol. 17, no. 10, pp. 969–991, 2003.

[3] G. Chesi and K. Hashimoto, “Effects of camera calibration errors on static-eye and hand-eye visual servoing,” Advanced Robotics, vol. 17, no. 10, pp. 1023–1040, 2003.

[4] F. Chaumette and S. Hutchinson, “Visual servo control, part I: Basic approaches,” IEEE Robotics and Automation Magazine, vol. 13, no. 4, pp. 82–90, 2006.

[5] F. Chaumette and S. Hutchinson, “Visual servo control, part II: Advanced approaches,” IEEE Robotics and Automation Magazine, vol. 14, no. 1, pp. 109–118, 2007.

[6] G. Chesi and Y. S. Hung, “Global path-planning for constrained and optimal visual servoing,” IEEE Trans. on Robotics, vol. 23, no. 5, pp. 1050-1060, 2007

[7] G. Chesi, “Visual servoing path-planning via homogeneous forms and LMI optimizations,” IEEE Trans. on Robotics, vol. 25, no. 2, pp. 281-291, 2009

[8] O. Faugeras and F. Lustman, “Motion and structure from motion in a piecewise planar environment,” Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 2, no. 3, pp. 485–508, 1988.

[9] J. Chen, A. Behal, D. Dawson, and Y. Fang, “2.5D Visual Servoing with a Fixed Camera”, *Proceedings of the American Control Conference*, Denver, Colorado, June 4-6, 2003.

[10] Z. Zhang and A.R. Hanson, “Scaled Euclidean 3D Reconstruction Based on Externally Uncalibrated Cameras”, IEEE Symp. on Computer Vision, 1995.

[11] E Malis, G Chesi, R Cipolla, “2 1/2 D visual servoing with respect to planar contours having complex and unknown shapes”, Int. Journal of Robotic Research, 22:10, 841-853, 2003.

[12] G. Chesi, “Camera displacement via constrained minimization of the algebraic error,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 370–375, 2009.

[13] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura, “Manipulator control with image-based visual servo,” in Proc. IEEE Int. Conf. on Robotics and Automation, 1991, pp. 2267–2272.

[14] C. Taylor and J. Ostrowski, “Robust vision-based pose control,” in Proc. IEEE Int. Conf. on Robotics and Automation, San Francisco, California, 2000, pp. 2734–2740.

[15] G. Chesi and Y. S. Hung, “Image noise induced errors in camera positioning,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pp. 1476-1480, 2007