Controlling robot by image-based visual servoing with stereo cameras

Fan Junmin, Sangchul Won
GIFT, POSTECH, Korea,fanjm@postech.ac.kr

Abstract - In this paper, an image-based "approach-align -grasp" visual servo control design is proposed for the problem of object grasping, which is based on the binocular stand-alone system. The basic idea consists of considering a vision system as a specific sensor dedicated a task and included in a control servo loop, and we perform automatic grasping follows the classical approach of splitting the task into preparation and execution stages. During the execution stage, once the image-based control modeling is established, the control task can be performed automatically. The proposed visual servoing control scheme ensures the convergence of the image-features to desired trajectories by using the Jacobian matrix, which is proved by the Lyapunov stability theory. And we also stress the importance of projective invariant object/gripper alignment.

The alignment between two solids in 3-D projective space can be represented with view-invariant, more precisely; it can be easily mapped into an image set-point without any knowledge about the camera parameters. The main feature of this method is that the accuracy associated with the task to be performed is not affected by discrepancies between the Euclidean setups at preparation and at task execution stages.

Then according to the projective alignment, the set point can be computed. The robot gripper will move to the desired position with the image-based control law. In this paper we adopt a constant Jacobian online.

Such method describe herein integrate vision system, robotics and automatic control to achieve its goal, it overcomes disadvantages of discrepancies between the different Euclidean setups and proposes control law in binocular-stand vision case. The experimental simulation shows that such image-based approach is effective in performing the precise alignment between the robot end-effector and the object.

Keywords: Image-based servoing, projective invariance, constanst Jacobian, real-time robot control

1 Introduction

Robotic visual servoing and manipulation has received significant attention nowadays for it can mimic the human sense of vision and allow non-contact measurement of the environment. Still most of the existing systems rely on one visual servoing control strategy or one sensory modality. This commonly limits the flexibility and accuracy of the system. Then it has been pointed out that one of the key research areas in the field of visual sevoing is the integration of existing techniques, regarding both the estimation and control [1].

In this paper, we apply visual-feedback loop to increase the overall accuracy of the system—a principal concern in most applications. Also the binocular system is used to provide a wide view and avoid the singularity points of object. The image-based method herein we adopted has been developed, in parallel, by a number of researchers [2], [3].

The main feature of this paper is that the desired object to gripper alignment will be represented in three-dimensional (3-D) projective space rather than in 3-D metric space.

Such a nonmetric representation can be obtained with an uncalibrated pair of camera, or a stereo rig. The main advantage of this method is that accuracy associated with the task to be performed is not affected by discrepancies between the Euclidean setups corresponding to different stages.

The remainder of this paper is structured as follows. Section II introduces the image-based visual servoing with stereo vision. In section III, we show how to represent an alignment between two objects in 3D projective space. The alignment condition thus derived is projective invariant in the sense that it can be used in conjunction with two different setup pairs to compute a goal position for visual servoing. In section IV and section V, we give the simulation results about this paper and conclude our topic respectively.

2 Image-based visual servoing

In image-based method, the error to be controlled is defined directly in terms of image feature parameters (in contrast to position -based methods that define error signal in the task space coordinates). So it can be known that, given the proper knowledge on an observed 3-D point set $\{S_i\}_{i=1\dots m}$ and its(fixed) position within the reference frame of a robot end-effector, it is possible to control this robot so to align the projection s of the point set $\{S_i\}$ in an image with a predefined goal position in this image (generally denoted s*) under condition that s* is the projection of an attainable 3-Dposition S*. This is represented in Figure 1 for the monocular case.

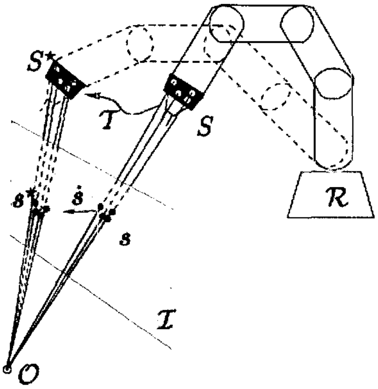

Figure1. Fixed monocular visual servoing

In that case, the observed image speed $\dot{s}$ of the considered point set is related to the effector speed, animated with the kinematics screw τ, through the image Jacobian J by the following equation:

$$\dot{s} = J\tau \tag{1}$$

By imposing that the observed point set s moves towards the required goal position s* (i.e. $\dot{s}$ =g(s*-s)), we can compute the kinematic screw τ to be sent to the robot end-effector within a control loop:

$$\tau = gJ^{+}(s^{*}-s) \tag{2}$$

where g is a scalar gain factor, J^{+} the $6 \times 2m$ pseudo-inverse of the image Jacobian and (s*-s) the observed 2m error vector in the image.

In stereo visual servoing, two cameras in a stereo arrangement are used to provide complete 3D information about scene, which is adopted as stand-alone configuration in this paper. A schematic overview of task using binocular system is shown in figure 3.3. The error function is defined for the left, $e_l = f_l^c - f_l^*$, and the right image, $e_r = f_r^c - f_r^*$. To drive this error to zero the image Jacobian is estimated by stacking of a two monocular image Jacobian defined for each of the camera. According to the image-based servoing control law, suppose we observe the movement of a set of 3-D points $\{s_i\}_{i=1\cdots m}$ in two different cameras c^l and c^r. We suppose also that these points are rigidly fixed to the robot end-effector, and that the latter is animated by a kinematic screw τ. Then, because of the rigidity constraint, the following equations hold:

$$\begin{aligned} \dot{s}^{l} &= J_{l}\tau \\ \dot{s}^{r} &= J_{r}\tau \end{aligned} \tag{3}$$

or concatenating the two matrices,

$$\begin{pmatrix} \dot{s}^{l} \\ \dot{s}^{r} \end{pmatrix} = \begin{pmatrix} J_{l} \\ J_{r} \end{pmatrix} \tau \tag{4}$$

then

$$\dot{s}^{l-r} = J_{l-r}\tau \tag{5}$$

which is exactly the same equation as (1), and gives rise to the same solution as the one presented in equation (3).The 3-D velocity induced by the command

$$\tau = g\left[\left(J_{l-r}^{T} J_{l-r}\right)^{-1} J_{l-r}^{T}\right]\left(s^{l-r^{*}} - s^{l-r}\right)$$
$$\tau = gJ_{l-r}^{+}\left(s^{l-r^{*}} - s^{l-r}\right) \tag{6}$$

now tends to obtain a simultaneous convergence in two images rather than in just one single view, as was the case in equation (2) . This situation has the above advantages just mentioned.

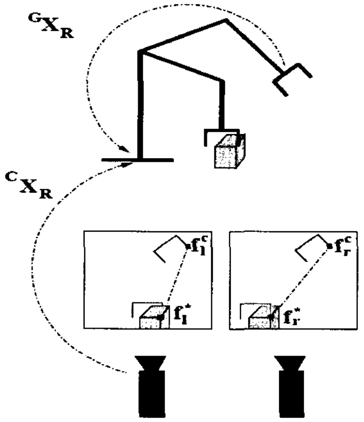

Figure 2. A schematic overview of the binocular camera system

3 Projective alingment

The visual servoing method described in the previous section requires knowledge of the set-point s* which is a set of image points. In this section we show how to compute the set-point s* such that it is "view-invariant," i.e., it is independent of both intrinsic and extrinsic camera parameters.

We consider a pair of uncalibrated cameras which observe the gripper aligned with the object. We denoted by P^x and P'^x the two projection matrices. Let m^x and m'^x be the projections of a 3D point M onto the left and right images associated with the two cameras. The equations

$$m^x = P^x M^x \,,\; m'^x = P'^x M^x \tag{7}$$

allow to compute the 3D projective coordinates M^x of the 3D point M in a projective basis x attached to the camera pair. Since the geometry of the camera pair may change over time, the camera pair is not a rigid object. However it is possible to compute a projective transformation mapping the sensor centered projective reconstruction x into the object centered projective reconstruction. For the sensor and object projective coordinates of a point A_i, we have

$$A_i^o = H^{xo} A_i^x \tag{8}$$

where A_i^x is obtained by applying (3) to an object point being observed with the camera pair, and H^{xo} is a 4×4 transformation.

At runtime, another stereo pair observes the object to be grasped. However, the gripper is at some distance from the object and the task is to move the gripper form its initial position to a virtual position. The latter gripper position corresponding to the gripper-to-object alignment defined during the off-line stage.

Let P^y and P'^y be the matrices associated with the runtime camera pair y and therefore we have

$$m^y = P^y M^y, m'^y = P'^y M^y \tag{9}$$

again, the sensor centered 3-D projective coordinates of an object can be mapped in a object centered description

$$A_i^o = H^{yo} A_i^y \tag{10}$$

By combining (9) and (10) we obtained a relationship between the projective coordinates of an object point expressed in the two projective bases x and y

$$A_i^y = (H^{yo})^{-1} H^{xo} A_i^x = H^{xy} A_i^x \tag{11}$$

Equation (11) allows computing a 4×4 homogeneous matrix H^{xy} from point matches between two setups, x and y, $(a^x, a'^x) \leftrightarrow (a^y, a'^y)$. With five point matches one obtains an exact solution. However, if a larger number of point matches are available, a least square solution can be computed [4]. To summarize, the following procedure transfers gripper points from the learning setup to the runtime setup:

1) For each gripper point $B_j, j = 1 \ldots n$
2) Reconstruct the projective coordinates of a gripper point from its images associated with the setup x
$$b_j^x = P^x B_j^x, b_j'^x = P'^x B_j^x$$
3) Map these point coordinates from one projective basis to the oyher projective basis
$$B_j^y = H^{xy} B_j^x$$
4) Project the gripper point onto the images associated with the runtime setup
$$b_j^y = P^y B_j^y, b_j'^y = P'^y B_j^y$$

Then set-point s* is simply derived by transforming the 2-D homogeneous coordinates of an image point onto its image coordinates

$$s^* = \begin{pmatrix} \tilde{b}_1^x \\ \vdots \\ \tilde{b}_n^x \end{pmatrix} \quad \text{with } b_j^x = \lambda \begin{pmatrix} \tilde{b}_1^x \\ 1 \end{pmatrix} \tag{12}$$

4. Simulation

In this section, we make a simulation about the visual servoing algorithm just mentioned with MATLAB. As already mentioned, the image Jacobian in visual servoing is key which decides the robot convergence to the desired position. In this paper, we follow the classical way to use a measured value of Jacobian at equilibrium configuration—the robot lies in the desired position, i.e., constant Jacobian matrix.

First, we establish a 3-D camera model where two cameras observe the object as figure 3 shown. Here we suppose that the object is stable and apply the control law in cameras. If the object image in both cameras coincides simultaneously, we can prove that this image-based is effective.

Then by the analysis of simulation, we have known that this method has a very good convergence to the desired position as we expected, as shown in figure 4.

We also give the error decrease diagram between the initial point and final point in figure 5.

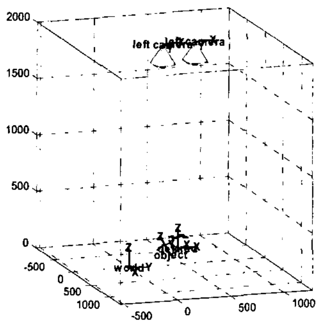

Figure 3. The model of stereo vision servoing

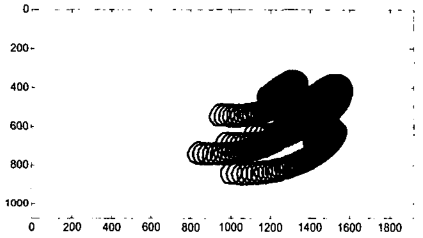

Figure 4. Convergence diagram

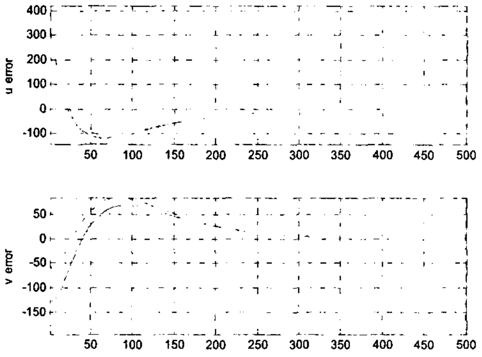

Figure 5. The error analysis diagram

So from the above simulation we can conclude that image-based servoing is a very robust way to control robot. Also the adoption of projective alignment and stereo system can compute the final points easily and give a wider view for control field. The method described here gives a deeper understanding for image-based visual servoing.

References

[1] Work shop on visual servoing, *IEEE International Conference on Intelligent Robots and Systems,IROS2002,* Lausanne, Switzerland,2002.

[2] Bernard Espiau, Francois, and Patrick Rives, "A New Approach to Visual Servoing in Robotics", *IEEE Transaction on Robotics and Automation,* vol.8, no.3, pp.313-326, 1992.

[3] K. Hashimoto, T.Kimoto, and H. Kimura. Manipulator control with image-based visual sero. In *proceedings of the 1991 IEEE International Conference on Robtics and Automation,* vol.3, pp.2267-2272, California, 1991.

[4] R. Horaud and G. Csurka, "Self-calibration and Euclidean reconstruction using motions of a stereo rig," in *Proc. 6th Int. Conf. Comput. Vision,* Bombay, India, Jan. 1998, pp.96-103

[5] Emanuele Trucco, Alessandro Verri, "Introductory Techniques for 3D Computer Vision", *Prentice-Hall, Inc,* 1998.

[6] David A.Forsyth, Jean Ponce, "Computer Vision", *Prentice-Hall, Inc,* 2003.