Visual Servoing using Image Motion Information

V. Sundareswaran,* F. Chaumette, and P. Bouthemy

IRISA/INRIA-Rennes

Campus de Beaulieu

35042 Rennes France

Abstract

Visual servoing is a framework for achieving the tight coupling of camera movements and information from images. We consider a typical visual servoing approach that uses geometric information about image features for controlling the position and attitude of a camera. We claim that image motion information can be used as well. We substantiate this claim by presenting two different approaches to visual tasks that use motion information. The first one uses the focus of expansion. The second one incorporates the parameters of the 2D affine motion model in the control equations. We illustrate both these approaches by means of a task to align the optical axis of the camera with the unknown direction of the translational motion of the system on which it is mounted, and present results of experiments done with a six DOF robot. The contribution of this work is in showing that a tight coupling between the camera behavior and image motion is possible.

1 Introduction

Recently, many new approaches have been taken to perform tasks based on responses to visual information. Visual inputs are processed, and based on this information, camera behavior is controlled, usually with the goal of attaining a configuration that simplifies further action. The approaches span a broad spectrum ranging from methods that perform very specific tasks, to methods that provide general framework. An extensive survey of all the methods is outside the scope of this paper. We mention just a few in each of the broad categories.

There are many examples for approaches that perform specific tasks. Nelson and Aloimonos proposed

a scheme to detect obstacles by using flow field divergence [13]; Santos-Victor et. al. [17], and Coombs and Roberts [6] present methods to steer a camera between two walls, and to veer around obstacles, both methods being based on a simple analysis of the computed optic flow fields. Performing saccades in real time to moving regions of interest has been demonstrated in [11].

In the middle of the spectrum, there are methods that perform more general tasks such as gaze control and fixation; these are more general because fixation or gaze control can be used as a means to accomplish other tasks. In [2], the uses of gaze control, and fixation in particular, in obtaining an object-centered reference-frame and in figure-ground discrimination have been outlined. A system performing several active visual tasks, including closed-loop gaze-control (based on a differential analysis) for fixating on an object has been presented in [9]. A fixation method running in real-time on a head-eye system can be found in [15]. Target-tracking methods by cameras mounted on robots have been demonstrated [1, 16].

On the farther side of the spectrum, there are methods which provide more general solutions. Visual servoing methods [7, 8, 19] present control-theoretic approaches for controlling the position and attitude of the camera by using visual information. These methods are general, and can be used to perform a variety of tasks with specific goals. Our present work is in extending the power of this class of approaches, by showing how motion information can be incorporated.

We start from one of the visual servoing formalisms [7]. The procedure outlined in [7] for determining the control equations is suited for the use of geometric information about features such as points, lines and circles. We show how this formalism can make use of motion-based information. In particular, the methods described in this paper use dynamic image parameters such as the focus of expansion or the coefficients in the 2D affine motion model. More details can be found in [18]. Such a use of the dynamic image parameters for the purpose of visual servoing is

new, and it provides a starting point for other interesting closed-loop methods using these parameters.

2 Visual Servoing

In this section, we review the basic principles of visual servoing. Detailed descriptions can be found in [7, 8, 19].

The principle of visual servoing is to use visual information as observation in closed-loop control when the desired configuration can be described as a particular visual observation. The control is effected on the camera position and orientation, or on an external object such as a robot arm. The only condition is that the instantaneous change in the visual information (in the sense of temporal derivative) as a function of the controllable parameters be known analytically. Intuitively, if the effect of the control parameters on the observation is known, we could provide the appropriate control in order to result in obtaining the desired observation. The visual servoing theory provides a framework for determining the *control law* which is simply a set of equations to calculate the control parameters.

More precisely, for a given vision-based task, we have to choose a set $\mathbf{s}$ of visual features (for example, the coordinates of an image point, the parameters of a selected line, etc.) suited for achieving the task. In order to perform a control law based on $\mathbf{s}$, we need to know the equations for the temporal variation of $\mathbf{s}$ with respect to camera translational and rotational motion (T, Ω). In other words, we have to determine the matrix L described by the following equation:

$$\dot{\mathbf{s}} = L \left(\begin{array}{c} T \\ \Omega \end{array} \right) \qquad (1)$$

A *task function* $\mathbf{e}$ can be defined as

$$\mathbf{e} = M \, (\mathbf{s} - \mathbf{s}^*), \qquad (2)$$

where $\mathbf{s}$ is the measured visual features currently observed by the camera, $\mathbf{s}^*$ is the desired final configuration for $\mathbf{s}$ in the image, and M is a constant matrix which allows, for robustness issues, to take into account more visual features than necessary.

The control problem thus appears as the regulation of the task function $\mathbf{e}$ to zero or, equivalently, as the minimization of $\|\mathbf{e}\|$ in the image by appropriate camera motion. We would like the task function to decay exponentially towards zero. For such a requirement, $\dot{\mathbf{e}} = -\lambda \mathbf{e}$, where $\lambda(> 0)$ is the exponent that controls the speed of the decay. Noting that M and $\mathbf{s}^*$ are

constant and assuming for simplicity that the scene is static, we obtain from Eqns. (1) and (2),

$$\left(\begin{array}{c} T \\ \Omega \end{array} \right) = -\lambda L^+ M^+ \mathbf{e}, \qquad (3)$$

where L^+ and M^+ are the pseudo-inverses of L and M.

Thus the principle of visual servoing is to use visual information to perform a closed-loop control to reduce an "error" in the visual information. Various tasks have been performed within the framework of the visual servoing outlined here and described in detail in [7].

3 Image motion information

We claim that the visual servoing approach can be applied to features other than geometric image features. We confirm this claim by using parameters obtained from image motion. First, we show that we can use the focus of expansion. This is a only a modest step forward because the focus of expansion (FOE) is a geometric property in some sense; this is so because the FOE is a location on the image plane, and can be treated as such. Nevertheless, this viewpoint brings out the notion of using motion information available as a simple geometric feature.

Secondly, we show how to use the parameters of the affine model (henceforth, *2D affine motion parameters*) of the optic flow. Here, we truly deviate from the tradition of using geometric information for visual servoing. The affine motion parameters have been well-studied theoretically as well as empirically, and have been used in many applications. But they have not been exploited so far as entities useful in directly controlling camera behavior.

We will illustrate these two approaches by their application in a task to align the optical axis of the camera with its unknown direction of translation. For the purposes of this paper, we restrict ourselves to the pure translation situation, expecting that this will provide an initiative to solve the more general cases.

4 Control using the FOE

The focus of expansion (x_f, y_f) is the projection of the 3D translational vector on the image plane: $x_f = \frac{U}{W}, \quad y_f = \frac{V}{W}$. The location (x_f, y_f) may be treated as a geometric feature and used to obtain control laws. For instance, we could apply control to bring

this *feature* to the center of the image: this control corresponds to the task of aligning the camera optical axis with the direction of translation.

For the alignment task, it is easy to see that two of the rotational parameters are sufficient to provide the necessary control. We only control camera pan A and tilt B, we know, from the optic flow equations [10],

$$\left(\begin{array}{c} \dot{x}_f \\ \dot{y}_f \end{array} \right) = L \left(\begin{array}{c} A \\ B \end{array} \right), \quad L = \left(\begin{array}{cc} -(1 + x_f^2) & x_f y_f \\ -x_f y_f & 1 + y_f^2 \end{array} \right).$$

In that case, the matrix L can be estimated on-line since it only depends on the position (x_f, y_f) of the FOE measured in the image. In Eqn 3, we can set M as the identity matrix and we finally obtain:

$$\left(\begin{array}{c} A \\ B \end{array} \right) = -\lambda L^{-1} \left(\begin{array}{c} x_f - x_f^* \\ y_f - y_f^* \end{array} \right), \qquad (4)$$

For the alignment task, $x_f^* = y_f^* = 0$. One could also position the FOE at any desired location (x_f^*, y_f^*) using exactly the same method.

5 Control using the affine parameters

The 2D affine motion model is often useful. It is possible to derive expressions for the affine motion parameters assuming that an analytical surface is imaged (i.e., it is possible to describe the depth by a Taylor series expansion).

Let the first-order approximation to the optic flow be

$$\begin{array}{rcl} u(x, y) & = & a_1 + a_2 x + a_3 y, \\ v(x, y) & = & a_4 + a_5 x + a_6 y. \end{array}$$

Using the first-order model for the imaged surface, $Z = Z_0 + \gamma_1 X + \gamma_2 Y$, we get [3, 12] expressions of the form

$$a_1 = -\frac{U}{Z_0} - B, \quad a_4 = -\frac{V}{Z_0} + A.$$

Consider the two "parameters" $U_z = \frac{U}{Z_0}$ and $V_z = \frac{V}{Z_0}$. If we apply control in such a way to result in zero values for these variables, we will achieve the goal of setting the components U and V of the translational velocity to zero (the tacit assumption is that infinite depth does not occur).

The components of T, which are assumed to remain constant in a global (world) coordinate system during the time taken for the alignment, change however in the camera coordinate system as $\dot{T} = -\Omega \times T$.

Figure 1: A typical image in the sequence processed.

Starting from the time derivatives of U_z and V_z, it is possible to obtain the following control law [18]:

$$\left(\begin{array}{c} A \\ B \end{array} \right) = \frac{1}{2} \left(\begin{array}{cc} 0 & -\tau_c \\ \tau_c & 0 \end{array} \right) \left(\lambda \left(\begin{array}{c} U_z \\ V_z \end{array} \right) - \left(\begin{array}{c} \frac{a_1}{\tau_c} \\ \frac{a_4}{\tau_c} \end{array} \right) \right),$$

$$(5)$$

where τ_c can be computed from the affine parameters, and the observations U_z and V_z are given by

$$U_z = -a_1 - B, \quad \text{and} \quad V_z = -a_4 + A,$$

where we use the previous measured values for A and B (under normal conditions, these are simply the control rotational velocity components applied at the preceding instant).

6 Experiments

We used a camera with a field of view of about 35 degrees mounted on a six degrees-of-freedom cartesian robot (AFMA). The size of the images processed was 128×182 pixels. All the image processing and control velocity computations are carried out on the host (Sparc IPX) and the computed control is transmitted to the robot controller.

The experiments were conducted indoors; a sample image can be seen in Fig. 1. The translational motion was towards the floor with cluttered objects; the floor was not fronto-planar, but with an average angle of inclination in the range 45-70 degrees between the floor surface and the optical axis.

6.1 The FOE-based method

The FOE-based method described in Section 4 has been implemented. The FOE is calculated as in [4].
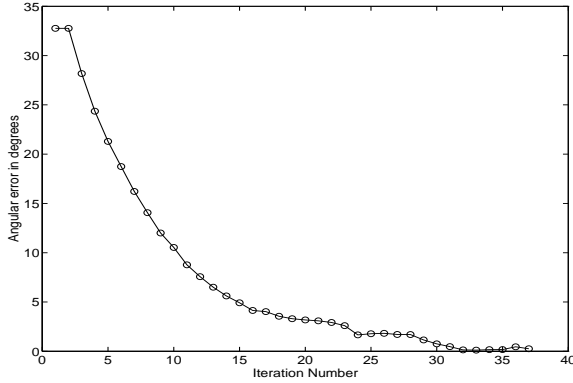
104

Figure 2: The angular error plot for the FOE method.

Figure 3: The error plot for the affine parameter method.

The normal flow field is computed, and the fraction h_p of positively oriented (i.e., towards the left) horizontal components and the fraction v_p of positively oriented (upwards) vertical components are determined by counting. The approximate FOE is computed as $(-(h_p - 0.5)*W_I/2, -(v_p - 0.5)*H_I/2)$, where W_I and H_I are the width and the height of the image in focal length units. As explained in [4], this gives a location on the image if the FOE is within the image, or a location on the border of the image if the FOE is on the border or is outside the image; in the last situation where the FOE is outside the image, the computed position is in the direction of the FOE.

The camera is attached to the end-effector of the robot undergoing an arbitrary translational motion. The control loop utilizing Eqn. 4 is repeated to obtain alignment. Note that the control is applied for only a finite duration during each iteration. This is because the qualitative method used for the FOE computation works only for pure translation. The total time spent in one iteration is about two seconds.

The result of a typical experiment is shown in Fig. 2. The plot shows the variation of the angle between the direction of translation T and the optical axis Z with respect to time. As expected, the angle decreases and converges to zero. The final error plotted in Fig. 2 is 0.25 degrees.

6.2 The Affine Parameters method

The affine parameters method described in Section 5 has also been implemented. The spatiotemporal derivatives of the (smoothed) intensity function are calculated using a simple procedure [10]. The affine parameters are computed using an over-constrained set of equations by considering points from all over the image, thresholded by gradient magnitude to suppress contribution from relatively uniform regions where the estimates are noisy.
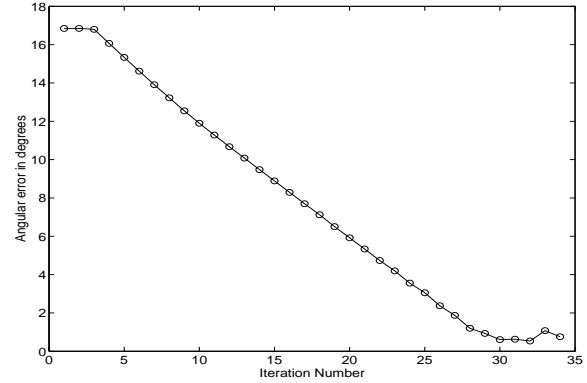
The error plot from an experiment using the control law in Eqn. 5 is shown in Fig. 3 (final error = 0.75 degrees). The linear (as opposed to an exponential) decrease in the error is due to experimental constraints.

7 Conclusions

We restricted ourselves to the pure translation situation. It would be interesting to examine the general case where there is also rotation. We know that if there is a small amount of external rotation, it could be still accommodated in the "pure translation" situation, with residual errors remaining as *tracking errors*, which can be in principle compensated for by estimating their effects and representing them using an additional term in the control law [5]. With large rotational velocities, it remains to be seen if workable methods can be developed.

We have also not examined the possibility of predicting the motion. Standard methods such as Kalman filtering could be employed to predict the focus of expansion or the affine motion parameters; this would increase the stability of the methods.

In the affine parameter method, we assume that the affine approximation to the optical flow field is valid. This is supported by several useful methods based on the affine approximation [3, 12]. Nevertheless, this approximation can fail for the entire image when there are objects located at very different depth in the scene, or moving objects of significant size. However, we have recently developed a robust estimation method able to correctly compute the 2D dominant affine motion in the presence of outliers [14].

105

An application for the task described here can be found in [4]. It consists in detecting moving objects in a qualitative way as explained in [3], once alignement is achieved.

In summary, we have proposed the use of motion information in the visual servoing framework where only geometric information has been used so far. Two control schemes, one using the focus of expansion, and the other using 2D affine motion parameters, were presented. Experimental results from a camera mounted on a robot serve to validate our proposal. We believe that tightly coupled visual behavior formalized in the fashion described here, and the kind of coupling between qualitative and active visual methods like the one we have proposed [4] have significant roles to play

Future work will include processing on the image-processing board to improve speed and performance, and to investigate other forms of tight coupling between camera behavior and motion information.

References

[1] P.K. Allen, A. Timcenko, B. Yoshimi, and P. Michelman. Automated tracking and grasping of a moving object with a robotic hand-eye system. *IEEE Transactions on Robotics and Automation*, 9(2):152–165, 1993.

[2] D.H. Ballard and C.M. Brown. Principles of animate vision. *CVGIP: Image Understanding*, 56(1):3–21, July 1992.

[3] P. Bouthemy and E. François. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *International Journal of Computer Vision*, 10(2):157–182, 1993.

[4] P. Bouthemy and V. Sundareswaran. Qualitative motion detection with a mobile and active camera. In A. G. Constantinides et. al., editor, *Procs of the Intl. Conf. on Digital Signal Processing*, pages 444–449, Cyprus, July 1993.

[5] F. Chaumette and A. Santos. Tracking a moving object by visual servoing. In *Proc. 12th world congress IFAC, Vol. 4*, pages 409–414, Sydney, July 1993.

[6] D. Coombs and K. Roberts. Centering behavior using peripheral vision. In CVPR, pages 440–445, New York City, June 1993.

[7] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics.

IEEE Transactions on Robotics and Automation, 8(3):313–326, 1992.

[8] J.T. Feddema and O.R. Mitchell. Vision-guided servoing with feature-based trajectory generation. *IEEE Journal of Robotics and Automation*, 5(5):691–700, October 1989.

[9] E. Grosso and D. Ballard. Head-centered orientation strategies in animate vision. In *Proc. of the 4th ICCV*, pages 395–402, Berlin, May 1993.

[10] B.K.P Horn. *Robot Vision*. The MIT Press, 1987.

[11] D.W. Murray, P.F. McLauchlan, I.D. Reid, and P.M. Sharkey. Reactions to peripheral image motion using a head/eye platform. In *Proc. of the 4th ICCV*, pages 403–411, Berlin, May 1993.

[12] S. Negahdaripour and S. Lee. Motion recovery from image sequences using only first order optical flow information. *IJCV*, 9(3):163–184, December 1992.

[13] C. Nelson and J. Aloimonos. Obstacle avoidance using flow field divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1102–1106, 1989.

[14] J.-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models in complex image sequences. In *Proc. 7th Conf. Eusipco*, Edinburgh, Scotland, September 1994.

[15] K. Pahlavan, T. Uhlin, and J-O. Eklundh. Dynamic fixation. In *Proc. of the 4th ICCV*, pages 412–419, Berlin, May 1993.

[16] N. Papanikolopoulos, P. Khosla, and T. Kanade. Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision. *IEEE Transactions on Robotics and Automation*, 9(1):14–35, 1993.

[17] J. Santos-Victor, G. Sandini, F. Curotto, and S. Garibaldi. Divergent stereo for robot navigation: learning from bees. In CVPR, pages 434–439, New York City, June 1993.

[18] V. Sundareswaran, P. Bouthemy, and F. Chaumette. *Visual servoing using dynamic image parameters*. Research Report, IRISA, Rennes, 1994. To appear.

[19] L.E. Weiss, A.C. Sanderson, and C.P. Neuman. Dynamic sensor-based control of robots with visual feedback. *IEEE Journal of Robotics and Automation*, RA-3(5):404–417, October 1987.