Active Vision II

Active Camera Self-orientation using Dynamic Image Parameters

V. Sundareswaran, P. Bouthemy and F. Chaumette

IRISA / INRIA Rennes, Campus de Beaulieu,
35042 Rennes Cedex, France

Abstract. We are interested in the realization of active visual tasks, and we propose an innovative visual servoing method using parameters obtained from visual motion processing. In particular, we consider the task of dynamically aligning the optical axis of a translating camera with its unknown direction of translation, by controlling the orientation of the camera. Parameters of the 2D affine motion model are used. We show experimental results of this method implemented on a six d.o.f. robot arm carrying a camera on its end-effector.

1 Introduction

Combining active manipulation of the camera along with processing of visual motion could allow us to perform certain tasks in a more robust fashion. In general, the importance and advantages of active vision have been well-understood; in [6], it is shown that ill-posedness of certain problems can be eliminated by active perception. 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 [7]. Real-time responses that include saccades to moving regions of interest, using a transputer-based system has been demonstrated in [9]. A fixation method running in real-time on a head-eye system has been presented in [12]. Santos-Victor et. al. [8], and Coombs and Roberts [4] 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.

The work presented here is within the visual servoing formalism [5]. Visual servoing provides a framework to determine the control equations for camera motion to perform useful tasks. A strong motivation for the research described here is to explore the possibility of using visual information that is not simply a geometric feature, as has been done in visual servoing applications in the past [5]. The method described in this paper uses dynamic image parameters such as the coefficients in 2D affine motion model. Such a use of the dynamic image parameters is innovative, and it provides a starting point for other interesting closed-loop methods using these parameters. The use of the focus of expansion has already been investigated in [2] and [13].

The method presented here examines the possibility of controlling the orientation of a camera that is under motion due to an external agent such as the vehicle on which the camera is mounted. We consider a method to apply

a control in order to align the optical axis of the camera with the direction of its unknown translation. We would simply refer to this process as the task of camera *alignment*. Such a reorientation would facilitate tasks to avoid obstacles and to perform other visual processing useful in navigation. For the purposes of this paper, we restrict ourselves to the pure translation situation, and hope that this will provide an initiative to solve the more general cases. In the section on experiments, we point out the errors that arise if external rotation is present and propose a way to successfully compensate for the errors.

2   Visual servoing using 2D affine motion parameters

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. More precisely, for a given vision-based task, we have to choose a set s of visual features suited for achieving the task (for example, the coordinates of an image point, the parameters of a selected line, etc). In order to obtain a control law based on s, we need to know the equations for the variation of 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:

$$\dot{\mathbf{s}} = L \begin{pmatrix} T \\ \Omega \end{pmatrix} \tag{1}$$

A *task function* $\mathbf{e} = M \, (\mathbf{s} - \mathbf{s}^*)$ can thus be defined where s is the measured visual features currently observed by the camera, $\mathbf{s}^*$ is the desired final configuration for s in the image, and M is a constant matrix which allows, for robustness issues, to take into account more visual features than necessary. Let us note that M can simply be chosen as the identity matrix when the number of the selected visual features is equal to the number of the camera d.o.f. controlled by the task.

The control problem thus appears as the regulation of the task function e to zero or, equivalently, as the minimization of $\|\mathbf{e}\|$ in the image by appropriate camera motion. If we would like the task function to decay exponentially towards zero, we get the control law

$$\begin{pmatrix} T \\ \Omega \end{pmatrix} = -\lambda \hat{L}^+ M^+ \mathbf{e}, \tag{2}$$

where $\lambda (> 0)$ is the exponent that controls the speed of the decay, $\hat{L}^+$ and M^+ are the pseudo-inverses of $\hat{L}$ and M, $\hat{L}$ being a model of L. The convergence of the control law will be ensured under the sufficient condition [5]:

$$M L \hat{L}^+ M^+ > 0. \tag{3}$$

Usually, $\hat{L}$ is chosen as L^*, the value of L at convergence. Indeed, in that case, the positivity condition is valid around the desired configuration.

We begin by describing the parameters of the 2D affine motion model, and then derive a control law to achieve an alignment task using these parameters. We

note here that two other recently proposed methods [8, 4] use motion information for the task of steering a camera between two walls.

The affine motion model is often useful. It is possible to derive expressions for the first-order parameters (affine parameters) assuming that an analytical surface is imaged (i.e., it is possible to describe the depth by a Taylor series expansion). It has been shown [10] that the affine parameters can be reliably estimated. Multiresolution methods for the estimation of the affine parameters [11] have proved to yield accurate values.

The optical flow equations, with translational velocity T (components U, V, W), and rotational velocity Ω (A, B, C), are

$$
\begin{aligned}
u(x, y) &= \tfrac{1}{Z(x,y)} \left[-U + xW\right] + A\left[xy\right] - B\left[1 + x^2\right] + Cy, \\
v(x, y) &= \tfrac{1}{Z(x,y)} \left[-V + yW\right] + A\left[1 + y^2\right] - B\left[xy\right] - Cx,
\end{aligned}
\tag{4}
$$

assuming unit focal length. Let the first-order approximation be

$$
u(x, y) = a_1 + a_2 x + a_3 y, \quad v(x, y) = a_4 + a_5 x + a_6 y.
\tag{5}
$$

If we assume that the imaged surface has $Z = Z_0 + \gamma_1 X + \gamma_2 Y$ as the first-order approximation, from Eqns. (4), and (5), we get [1, 10]

$$
\begin{aligned}
a_1 &= -\tfrac{U}{Z_0} - B, & a_4 &= -\tfrac{V}{Z_0} + A, \\
a_2 &= \tfrac{1}{Z_0}(\gamma_1 U + W), & a_5 &= \tfrac{1}{Z_0}(\gamma_1 V) - C, \\
a_3 &= \tfrac{1}{Z_0}(\gamma_2 U) + C, & a_6 &= \tfrac{1}{Z_0}(\gamma_2 V + W).
\end{aligned}
\tag{6}
$$

Our problem situation is when there is an unknown translational velocity T of the camera, for example due to the motion of the host vehicle. The task is to align the optical axis Z with the translation T by utilizing the free rotational parameters. Indeed, it is easy to see that the control of two rotational parameters (say, tilt A and pan B) is sufficient to align the optical axis with the translational direction We assume that acceleration, if present, manifests from time to time but not constantly. We further assume that the only rotational velocity arises due to the control action that is responsible for achieving the alignment.

Consider the two "parameters" $U_z = \tfrac{U}{Z_0}$ and $V_z = \tfrac{V}{Z_0}$. 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 derivatives of these parameters are given by

$$
\begin{pmatrix} \dot{U}_z \\ \dot{V}_z \end{pmatrix} = \begin{pmatrix} \dfrac{\dot{U}}{Z_0} - \dfrac{U}{Z_0}\dfrac{\dot{Z_0}}{Z_0} \\ \dfrac{\dot{V}}{Z_0} - \dfrac{V}{Z_0}\dfrac{\dot{Z_0}}{Z_0} \end{pmatrix}.
\tag{7}
$$

The components of T, which remain constant in a global (world) coordinate system (during the time taken for the alignment), change however in the camera coordinate system because of the rotation of the camera axes. Since the rotational velocity is $\Omega = (A, B, C)$ about the three axes, the variation of T is simply the

cross product $\dot{T} = -\Omega \times T$. Using this observation to substitute for $\dot{U}$ and $\dot{V}$ in Eqn. 7, and noting that $U_z = -a_1 - B$, $V_z = -a_4 + A$, we get

$$\begin{pmatrix} \dot{U}_z \\ \dot{V}_z \end{pmatrix} = \begin{pmatrix} 0 & -\frac{W}{Z_0} + \frac{\dot{Z}_0}{Z_0} \\ \frac{W}{Z_0} - \frac{\dot{Z}_0}{Z_0} & 0 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} + \begin{pmatrix} C\frac{V}{Z_0} + a_1\frac{\dot{Z}_0}{Z_0} \\ -C\frac{U}{Z_0} + a_4\frac{\dot{Z}_0}{Z_0} \end{pmatrix}.$$

When the planar approximation to the viewed surface does not have a large angle of inclination with respect to the camera, we have

$$\frac{1}{\tau_c} = \frac{W}{Z_0} \approx -\frac{\dot{Z}_0}{Z_0} \approx \frac{a_2 + a_6}{2}, \tag{8}$$

where τ_c is used to denote the instantaneous *time-to-collision*. Furthermore, if we assume for simplicity that $C = 0$, we obtain, for an exponential decay of the task function, the following control law:

$$\begin{pmatrix} A \\ B \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & -\tau_c \\ \tau_c & 0 \end{pmatrix} \left(\lambda \begin{pmatrix} U_z \\ V_z \end{pmatrix} - \begin{pmatrix} \frac{a_1}{\tau_c} \\ \frac{a_4}{\tau_c} \end{pmatrix} \right). \tag{9}$$

Using the approximations (8), the positivity condition (3) is ensured when the task is realized. Convergence of the control law will thus be obtained if the initial configuration is *not too far* from convergence. Our experiments, wherein divergence was never observed, confirm that such approximations do not disturb the task behavior.

Let us finally note that the observations U_z and V_z are given by

$$U_z = -a_1 - B, \quad \text{and} \quad V_z = -a_4 + A, \tag{10}$$

where we use the previous measured values for A and B (under normal conditions, it is nothing but the control rotational velocity applied at the preceding instant). Other control strategies to perform the same task are described in [13] (using other combinations of the affine parameters or using the focus of expansion).

3 Alignment: experiments

The method based on the use of the parameters U_z and V_z has been implemented in a real system. Simulation experiments were also carried out on the other methods [13], but they are not detailed here for lack of space.

The camera with a field of view of about 35 degrees has been mounted on a six degrees-of-freedom cartesian robot (AFMA). The size of the images processed is 128 × 182 pixels. All the image processing and control velocity computations are carried out on the host (a Sun 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 in the range 45-70 degrees between the floor surface and the optical axis.

For the image processing, the spatiotemporal derivatives of the (smoothed) intensity function are calculated using a simple procedure. The affine motion parameters are computed using a 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 robot is commanded to move the camera with a certain translational velocity. The control loop consists of the following steps which are repeated: obtain two successive images, compute the affine parameters of the flow field, compute the rotational velocity control required using the control law in Eqn. 9, and apply the control rotational velocity. Here, two different programs, one in which the control is applied for a finite duration and another in which the control is applied in a continuous manner, have been implemented. Each iteration took around three seconds.

The error plots from experiments using the two different implementations are shown in Fig. 1. The final error is 0.5 degrees for the implementation where the control is applied for a finite duration (discrete control), and for continuous control, the final error is 0.75 degrees. For the discrete control, while the two successive images are acquired, the control is withdrawn.

We restricted ourselves to the pure translation situation. It would be interesting to examine the general case where there is rotation also. We know that if there is external rotation, it could still be accommodated in the "pure translation" situation, with residual errors remaining as *lag*, which can be compensated for by estimating their effects and representing them using an additional term in the control law [3]. Furthermore, we assume that the affine approximation to the optical flow field is valid or at least sufficient for the task at hand. This is supported by several useful methods developed for scene motion recovery based on the affine approximation [1, 10]. 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. Motion-based segmentation of the image into regions [1] could be one possible solution, but far too complex to be implemented in such a closed-loop procedure. However, we have recently designed a multi-resolution robust estimation method which could cope with these situations [11].

The use of this task in aiding a qualitative method for motion detection is described in [2].

4 Conclusions

In this report, we have proposed a new active vision approach and showed how it can be performed by visual servoing methods using parameters derived from the motion information contained in a sequence of images. We presented the task of aligning the optical axis with the translational direction, and derived control equations for this task. The implementations on a real robot-camera configuration validate the methods and prove that it is possible to do motion information-based servoing at a reasonable frame-rate. We would like to stress

in concluding that such innovative use of the dynamic image parameters can be expected to be fruitful in the approaches for various active vision tasks.

References

1. Bouthemy, P., François, E.: Motion segmentation and qualitative dynamic scene analysis from an image sequence. IJCV. **10-2** (1993) 157-182
2. Bouthemy P., Sundareswaran, V.: Qualitative motion detection with a mobile and active camera. Int. Conf. on Digital Signal Processing, Cyprus. (1993) 444-449
3. Chaumette, F., Santos, A.: Tracking a moving object by visual servoing. 12th World Congress IFAC, Sidney. **4** (1993) 409-414
4. Coombs, D., Roberts, K.: Centering behavior using peripheral vision. CVPR 93, New York. (1993) 440-445
5. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. IEEE Trans. on Robotics and Automation. **8-3** (1992) 313-326
6. Fermüller, C., Aloimonos, Y.: The role of fixation in visual motion. IJCV. **11-2** (1993) 165-186
7. Grosso, E., Ballard, D.: Head-centered orientation strategies in animate vision. 4th ICCV, Berlin. (1993) 395-402
8. Santos-Victor, J., Sandini, G., Curotto, F., Garibaldi, S.: Divergent stereo for robot navigation: learning from bees. CVPR 93, New York. (1993) 434-439
9. Murray, D., McLauchlan, P., Reid, I., Sharkey, P.: Reactions to peripheral image motion using a head/eye platform. 4th ICCV, Berlin. (1993) 403-411
10. Negahdaripour, S., Lee, S.: Motion recovery from image sequences using only first order optical flow information. IJCV. **9-3** (1992) 163-184
11. Odobez, J.-M., Bouthemy, P.: Robust multiresolution estimation of parametric motion models applied to complex scenes. IRISA (1994) Report PI 788
12. Pahlavan, K., Uhlin, T., Eklundh, J.-O.: Dynamic fixation. 4th ICCV, Berlin, (1993) 412-419
13. Sundareswaran, V., Bouthemy, P., Chaumette, F.: Visual servoing using dynamic image parameters. IRISA (to appear 1994)
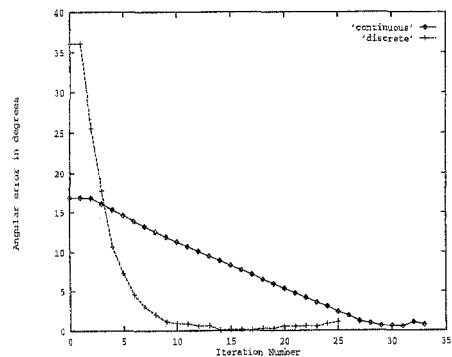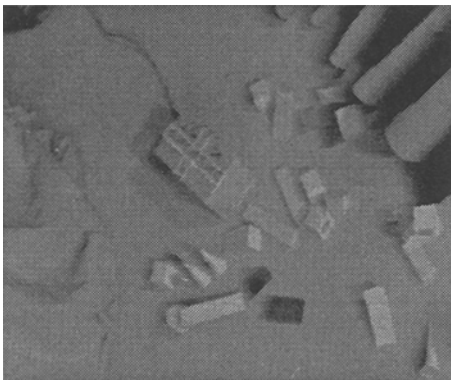
Fig. 1. On the left is a sample image from the sequence; on the right are the angular error plots for the experiment.