Real-time Visual Tracking for Surveillance and Path Planning

R. Curwen, A. Blake, and A. Zisserman

Department of Engineering Science
University of Oxford
Oxford OX1 3PJ, U.K.
Tel. +44-865-273154, Fax. +44-865-273908, email rupert@uk.ac.ox.robots.

Abstract. In this paper we report progress towards a flexible, visually driven, object manipulation system. The aim is that a robot arm with a camera and gripper mounted on its tip should be able to transport objects across an obstacle-strewn environment. Our system is based on the analysis of moving image contours, which can provide direct estimates of the shape of curved surfaces. Recently we have elaborated on this basis in two respects. First we have developed real-time visual tracking methods using "dynamic contours" with Lagrangian Dynamics allowing direct generation of approximations to geodesic paths around obstacles. Secondly we have built a $2\frac{1}{2}$D system for incremental, active exploration of free-space.

1 Introduction

Over the last few years, significant advances have been made in estimation of surface shape from visual motion, that is from image sequences obtained from a moving camera. By combining differential geometry [6] with spatio-temporal analysis of visual motion [9, 10, 7] it has been shown that local surface curvature can be computed from moving images [8, 3, 4, 11, 1]. The computation is robust with respect to surface shape, configuration of the surface relative to the camera and the nature of the camera motion. For example the ability to discriminate qualitatively between rigid features and silhouettes on smooth surfaces has been demonstrated [3]. Particularly important for collision-free motion, the "sidedness" of silhouettes is computed, that is, which side is solid surface and which is free space.

This paper reports on progress in building a robot with active vision that can manipulate objects in the presence of obstacles. Our Adept robot has a camera and gripper on board and is able to make exploratory "dithering" movements around its workspace. As it moves, it monitors the image motion and deformation of contours in real-time using parallel dynamic contours. Real-time performance depends on appropriate internal dynamical modelling with adaptive control of scale. As in earlier versions of our system [3, 4], contour motion is used to interpret occlusion and surface curvature. More recently we have built on further features: incremental building of a free-space model, incremental planning of robot motion and search strategies for navigation.

2 Visual Tracking

The curvature analysis described above relies on the ability to track moving curved image-contours. We have implemented a series of deformable cubic B-spline "dynamic contours" that run at video rates using a Transputer network. Points along the dynamic contour are programmed to have an affinity for image features such as brightness or high intensity-gradient. The contour can now be used to make curvature estimates in a few seconds. Substantial improvements in tracking performance are realised when Lagrangian dynamics formalisms are used to model mass distributed along a dynamic contour which moves in a viscous medium. We have found that use of large Gaussian blur is unnecessary, a crucial factor in achieving real-time performance.

2.1 Coupled B-spline model

Defining a dynamic contour with inertia [5], implies a model in which the image velocity of features is assumed uniform. However, stronger assumptions about the feature can be incorporated when appropriate, to considerable effect. This is done by making the further assumption that the feature shape will change only slowly. As an illustration, with no shape assumption, the contour flows *around* corners (Fig. 1a). When the shape assumption is incorporated it moves almost rigidly with the corner feature (Fig. 1b).

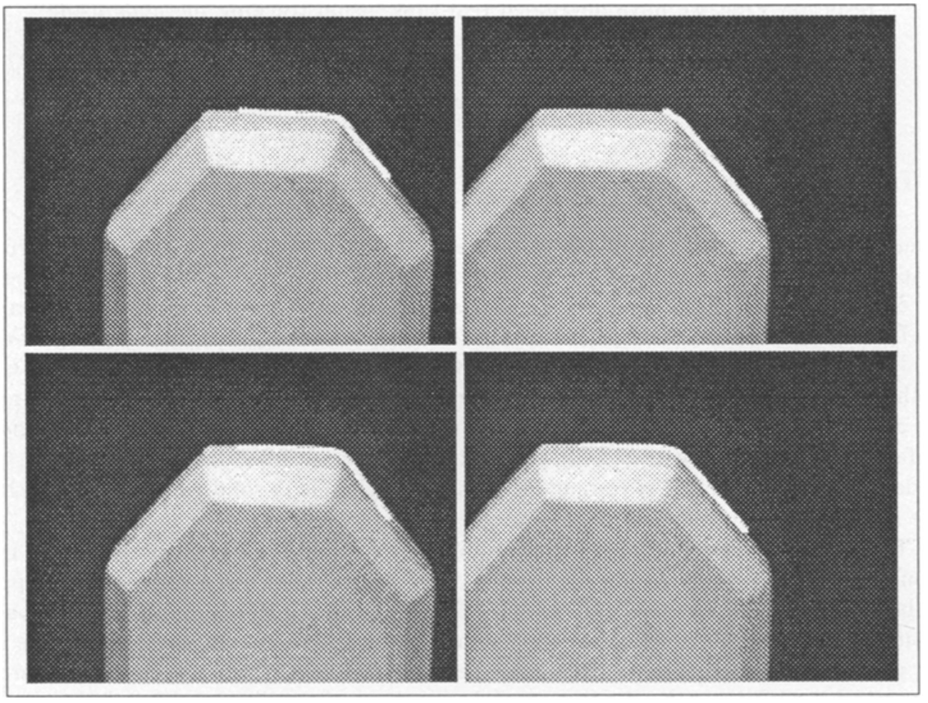

Fig. 1. Dynamic contours tracking a corner. In (a) the dynamic contour is flexible, and slides round the corner as the object moves to the left, whereas the "coupled" dynamic contour in (b) follows the true motion of the corner.

The shape assumption is imposed by using a pair of coupled B-splines. The first can be trained by taking the original (uncoupled) dynamic contour and allowing it to relax onto a feature. Its shape is then frozen and becomes the "template" shape (see also Yuille et. al. [12]) in the new model. A second B-spline curve, initially an exact copy of the first, is then spawned and coupled to the template B-spline. The coupling is defined, naturally, by a new elastic energy term. In the case of first derivative coupling energy, the dynamic contour behaves like a one-dimensional membrane or string. Second derivative coupling produces a contour which acts like a one-dimensional thin plate or rod. Suppose the template contour has control point vector $\mathbf{Q}_s$, then the equations of motion, derived by an analysis similar to the uncoupled model [5]), are:

$$\ddot{\mathbf{Q}} = \omega_0(\mathbf{U} - \mathbf{Q}) - 2\beta_0\dot{\mathbf{Q}} + \mathbf{H}_0^{-1}\mathbf{H}_1\left[\omega_1(\mathbf{Q}_\mathbf{S} - \mathbf{Q}) - 2\beta_1\dot{\mathbf{Q}}\right] , \tag{1}$$

where $\mathbf{H}_0$ and $\mathbf{H}_1$ are constant matrices, simply compositions of B-spline coefficients, and $\mathbf{U}$ is the least squares B-spline approximation to the feature vector. The constants ω_0, β_0 govern elastic attraction towards the feature and velocity damping respectively. The constants ω_1 and β_1 govern elastic restoring forces and internal damping respectively.

2.2 Parallel implementation

The equations of motion are integrated using an implicit Euler scheme for speed on a network of transputers. The dynamic contours are allocated to worker transputers a span at a time. The individual spans of a contour communicate their contribution to $\mathbf{V}$ in (1) to the rest of the contour at each Euler step. With six worker transputers, Euler steps can be performed at frame rate (25Hz). To overcome this problem, three separate frame grabbers are used, sampling the sample video input. Figure 2 shows a sample of frames from a multiple contour tracking sequence using coupled dynamic contours.

The dynamic contour can successfully track features whose velocity is such that the lag caused by viscous drag does not exceed the radius of the tracking window. With a tracking window radius of approximately 35 mrad (in a field of view of 0.3 rad) maximum tracking velocity is about 1.4 rad/sec, for our system. Note that varying the tracking window radius is our mechanism for control of scale. For example, a large window is used during feature capture, getting smaller as the contour locks on.

3 Active Exploration of Free Space

Following the first version of our manipulation system [2], we have investigated a more complex version of the manipulation problem in which the workspace is cluttered with *several* obstacles. In addition to visual and spatial geometry, we now need to add Artificial Intelligence search techniques (A^* search). Path-planning works in an incremental fashion, repeating a cycle of exploratory motion, clearing a triangular chunk of freespace (figure 3), and viewpoint prediction. After several of these cycles, over which the robot "jostles" for the best view of the goal (box lid), it may find itself jammed between obstacles and the edge of its workspace. In that case it backtracks to an earlier point in its search of freespace and investigates a new path. So far we have demonstrated these techniques with up to three different, unmodelled obstacles in the workspace.

4 Conclusions

We have discussed the principles and practice of visual manipulation of objects in clutter from several points of view: tracking, dynamics, spatial geometry and geometry of grasp. We are currently working to extend this in a variety of ways, including:

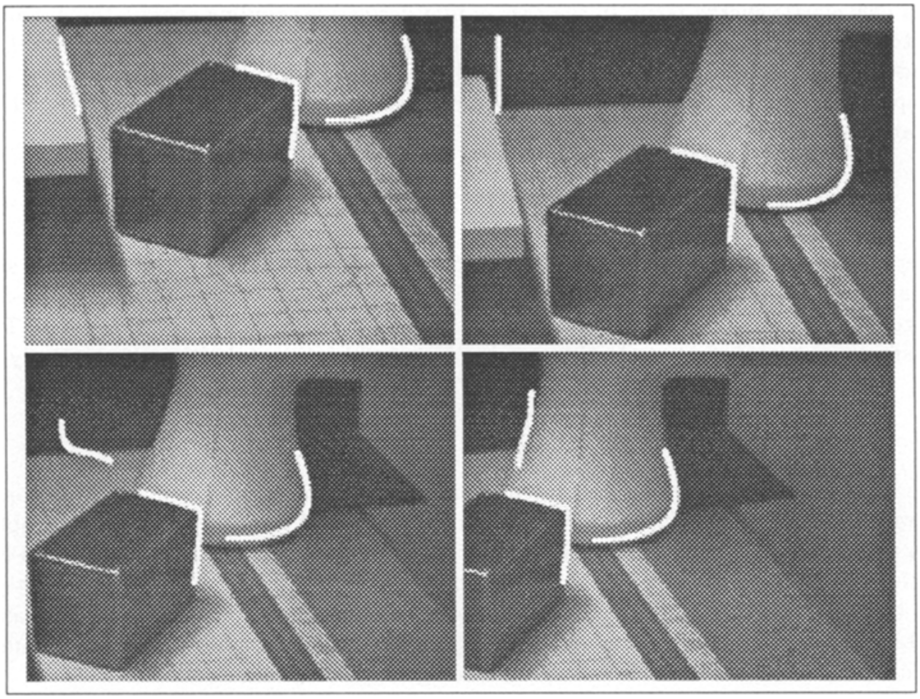

Fig. 2. Dynamic contours tracking objects in a scene (raster order). All dynamic contours are "coupled". Note the top left hand contour falling off its feature as the object moves over the image boundary.

- Developing more powerful internal models for tracking to improve ability to ignore clutter. This would enable the robot to perform efficiently with obstacles in the background as well as in the foreground.
- Employ more sophisticated geometric modelling to allow fine-motion planning that takes account of the shape of the gripped object, and the fact that the workspace is 3D not $2\frac{1}{2}$D.

Acknowledgments

We acknowledge the financial support of the SERC and of the EEC Esprit programme. We have benefitted greatly from the help and advice of members of the Robotics Research Group, University of Oxford, especially Professor Michael Brady, Roberto Cipolla and Zhiyan Xie.

References

1. E. Arbogast and R. Mohr. 3D structure inference from images sequences. In H.S. Baird, editor, *Proceedings of the Syntactical and Structural Pattern Recognition Workshop,*, pages 21–37, Murray-Hill, NJ, 1990.
2. A. Blake, J.M. Brady, R. Cipolla, Z. Xie, and A. Zisserman. Visual navigation around curved obstacles. In *Proc. IEEE Int. Conf. Robotics and Automation*, volume 3, pages 2490–2499, 1991.

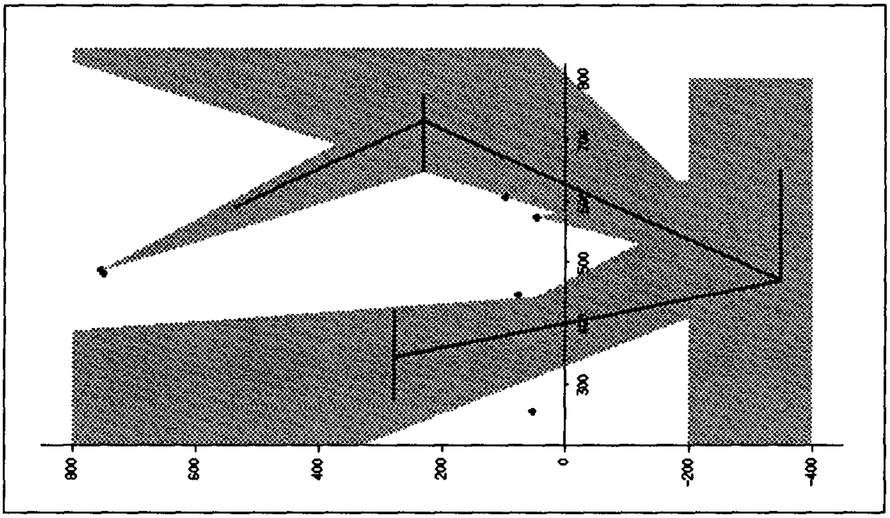

Fig. 3. Our visual object manipulation system clears triangular chunks of freespace visually, in an incremental fashion (freespace shown in grey). Accumulated freespace is represented in a 2D plane as shown, and is actually a projection of 3D obstacles onto a horizontal plane. The robot plans paths (black lines), restricted to space currently known to be free. Horizontal paths in the figure are exploratory "dithering" motions performed deliberately to facilitate structure from motion computations. The state of freespace is shown after navigation around three obstacles, towards the goal object at the left of the figure. The robot tried the path on the bottom, reached a fixed search depth-bound, backtracked, and successfully navigated on the top.

3. A. Blake and R.C. Cipolla. Robust estimation of surface curvature from deformation of apparent contours. In O. Faugeras, editor, *Proc. 1st European Conference on Computer Vision*, pages 465–474. Springer–Verlag, 1990.
4. R. Cipolla and A. Blake. The dynamic analysis of apparent contours. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 616–625, 1990.
5. R.M. Curwen, A. Blake, and R. Cipolla. Parallel implementation of lagrangian dynamics for real-time snakes. Submitted to 2nd British Machine Vision Conference, 1991.
6. M.P. DoCarmo. *Differential Geometry of Curves and Surfaces.* Prentice-Hall, 1976.
7. O.D. Faugeras. On the motion of 3-d curves and its relationship to optical flow. In *Proceedings of 1st European Conference on Computer Vision*, 1990.
8. P. Giblin and R. Weiss. Reconstruction of surfaces from profiles. In *Proc. 1st Int. Conf. on Computer Vision*, pages 136–144, London, 1987.
9. H.C. Longuet-Higgins and K. Pradzny. The interpretation of a moving retinal image. *Proc. R. Soc. Lond.*, B208:385–397, 1980.
10. S.J. Maybank. The angular velocity associated with the optical flow field arising from motion through a rigid environment. *Proc. Royal Society, London*, A401:317–326, 1985.
11. R. Vaillant. Using occluding contours for 3d object modelling. In O. Faugeras, editor, *Proc. 1st European Conference on Computer Vision*, pages 454–464. Springer–Verlag, 1990.
12. A.L. Yuille, D.S. Cohen, and P.W. Hallinan. Feature extraction from faces using deformable templates. *Proc. CVPR*, pages 104–109, 1989.

This article was processed using the LaTeX macro package with ECCV92 style