Model-Based Visual Feature Tracking for Assembly*

Srinivas Ravela Richard S. Weiss
Department of Computer Science
University of Massachusetts, Amherst, MA 01003-4610

Abstract

Visual tracking is applicable to many tasks in robotic assembly and inspection, including peg-in-hole insertion, alignment of parts, and gaze control. A fast, robust, hybrid approach to feature tracking is presented. Rotation-compensated normalized cross-correlation is used to handle large rotations. Adaptive templates allow for some changes in lighting and background. The tracking system runs at 8Hz. and is demonstrated in a peg-in-hole insertion task.

1. Introduction

Visual feature tracking is the dynamic registration of object features over a time varying sequence of images. Given initial correspondences between features of objects in the world and their image-space occurances, visual tracking attempts to preserve these model-image correspondences over time. Thus, visual feature tracking can be used to close a feedback loop around the pose of an object. This enables the construction of controllers that servo a manipulator or mobile robot to a desired pose or to trace a desired trajectory[3, 5].

In this paper we describe a new tracking technique and it's application to visually servo a manipulator with poor kinematics in a peg-in-hole insertion task described below. The kinematics are poor in the sense that the accuracy of spatial resolution provided by the kinematics is coarser than the tolerance requirements for insertion.

2. Task Specifications and Initialization

The insertion task consists of a stationary hole and a camera pair in a fixed convergent stereo configuration as the visual observer. The peg is assumed to be rigidly mounted on the manipulator tool frame and the arm kinematics are used as the nominal observer. Thus for any manipulator configuration the peg-to-base transformation is approximately known. The geometric model of the peg (vertex-edge representation) expressed in the manipulator base frame is matched with the images of the peg observed in each camera using the matching technique described in [1]. This matching is used to construct correspondences between the image-object features in either camera. Using at least four non-coplanar model points (such as corners) and their correspondences, a weak perspective affine invertible transform [6] (of dimensions 4×4) that is the base-to-camera transform is constructed. The hole-to-camera transform is computed similarly. By expressing the hole and peg frames in the manipulator base frame using these transforms , a pose error in the base frame can be computed. This pose error is used to derive manipulator commands that robustly traces a pre-planned contact space trajectory [2].

The process of line extraction and model matching is slow compared with filtering and correlation, and it is only performed once to initialize the appropriate number of templates. These templates are tracked over time, by first using the peg-to-camera (peg-to-base followed by base-to camera) transform to hypothesize their locations and then by localizing them in the next image. The image features used are image patches centered around dominant oriented edge points. Thus, conventional methods to extract lines for example can be replaced by instantiating two templates on the representative image edge and localizing them over the object motion. In addition features such as corners can be inferred by computing the intersections of the appropriate edges. Rotation-compensated normalized correlation is used to localize the templates in appropriate search windows around their hypothesized locations and is discussed below.

3. Template Localization

Normalized correlation [3] is a well known match measure and a search for the maximum correlation score of a template over a search window is typically use to localize an image patch in the window. However, this scheme is sensitive to rotations. In order to compensate for rotations, edges and their orientations are

*The authors received support from DARPA and TACOM under contract DAAE07-91-C-R035 and NSF under grants IRI-9208920 and IRI-9116297.

Figure 1: Tracking sequence showing peg-in-hole insertion.

detected in the search window using first derivative steerable filters [4]. Each edge location (gradient local maximum) is a candidate for a match with the template and correlation is performed at these locations. This reduces the search space. The orientation of the edge in the template and at the candidate location determines a sampling order of the template that cancels any relative angle. The re-sampled template is correlated with the image at the candidate edge location. This is performed over all edge candidates in the search window and the best correlation is picked as the matched location. This technique is fast and tracking with about 6 templates of sizes up to 11×11 and search windows of up to 15×15 has yielded an 8Hz tracking system on a conventional Sparc2 processor. This hybrid approach (combining image and low level features) is robust with respect to 2D rotations. After a match, the templates are updated by cutting out patches at matched locations in the new image and is therefore adaptive. The dominant centered edge assumption can be used to ensure that the templates do not drift away from the contour.

Figure 1 shows two images from the left camera of a servoing sequence. The lines indicate hypothesized peg locations. The large squares indicate the search windows and the the smaller, filled squares indicate the points of maximum correlation. The right image of the figure indicates the attainment of the point of first contact in the assembly plan.

Geometric models can be used to achieve robust tracking in real-time assembly-type operations. The use of steerable filters reduces sensitivity to rotations. Normalized cross-correlation with adaptive templates (captured from the image) reduces sensitivity to lighting changes. This approach is also being extended using aspect graphs to determine which features are most likely to remain visible in the next view. The hard problems that still remain concern occlusion and dis-occlusion.

REFERENCES

[1] J. Ross Beveridge, R. Weiss and E. M. Riseman, "Combinatorial optimization method applied to variable scale 2D model matching", *Proc. Int. Conf. Patt. Recognition*, pp. 18-23, June, 1990.

[2] G. Dakin and R.J. Popplestone, "Contact space analysis for narrow-clearance assemblies," *IEEE Symp. Intell. Control*, pp. 542-7.

[3] C. L. Fennema, "Interweaving Reason, Action and Perception", *COINS TR91-56*, Dept. of Computer Science, Univ. of Massachusetts, Amherst, 1991.

[4] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters", *IEEE Trans. Patt. Anal. and Mach. Intell.*, 13(9):891-906, Sept., 1991.

[5] K. Hashimoto (ed.), "Visual Servoing", *World Scientific*, 1994.

[6] N. J. Hollinghurst and R. Cipolla, "Uncalibrated stereo hand-eye coordination", *CUED/F-INFENG/TR 126*, Dept. of Engineering, Univ. of Cambridge, Cambridge, England, May, 1993.