

An Integrated Motion Analysis System Guided by Feedback Information *

Yong C. Kim and Keith Price

Institute for Robotics and Intelligent Systems
Departments of Computer Science and Electrical Engineering
University of Southern California
Los Angeles, CA 90089-0273

Abstract

In the feature-based motion analysis of an image sequence, consistent feature extraction and reliable matching are crucial factors for the motion estimation. In this paper, we address the issues of using errorful data in motion estimation and of using feedback to improve feature extraction and matching in incremental analysis. Initial noisy correspondence data are continuously refined by removing those parts that do not fit the estimated 3-D motion parameters. The extraction of a tracked region is guided by its expected properties obtained from the corresponding objects in the previous frames. Finally, the surfaces of objects are reconstructed from a sparse depth map.

1 Introduction

Conventional feature based motion analysis techniques use a sequential framework - feature extraction, establishment of correspondence, estimation of motion parameters and recovery of 3-D structure. Each of these have been studied extensively as a separate research topic. Little work has been done on the use of feedback of 3-D motion estimation to feature extraction and matching.

In this paper, we present a feedback approach in a feature-based motion analysis system that extracts the 3-D structure of a scene and estimates the 3-D egomotion of the observer, given a monocular image sequence. Feature extraction, matching and motion analysis are performed in a cooperative manner, exchanging information among the separate subsystems.

Our motion analysis system is based on very common subsystems. We believe that the idea of using feedback to improve the extraction and matching of features can be applied to other motion analysis system.

*This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F49620-90-C-0078. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

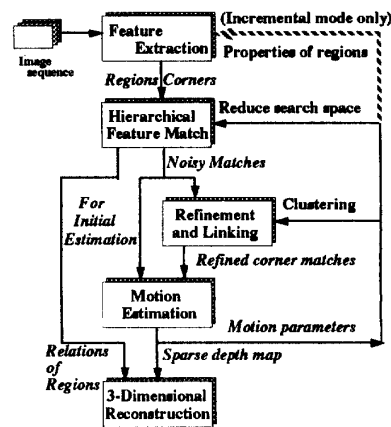


Figure 1: Block diagram and flow of data

2 System Configuration

Figure 1 shows the major components of the motion analysis system and the use of feedback information. The first few frames are analyzed all at once in a batch mode and the additional frames are processed in an incremental mode. In a batch mode, regions and corners are extracted and then matched between adjacent frames. 3-D motion and structure are estimated from the correspondence data. The initial noisy correspondence data are refined by removing those parts that do not fit well with the estimated 3-D motion. The 3-D motion estimation from this refined correspondence data provides the initial guide for incremental processing. Incremental analysis performs the same operations as done in batch mode on each incoming frame. This time, feature extraction and matching as well as refinement of matches are guided by feedback from 3-D motion. Finally, the 3-D surface of the environment is reconstructed from a sparse depth map of corners, utilizing the fact that the regions underlie the corners.

In order to accomplish consistent feature extrac-

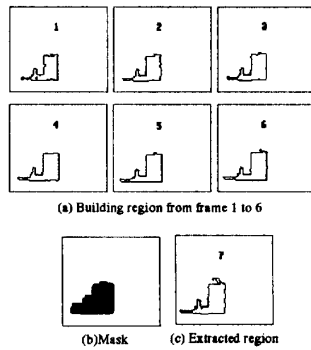


Figure 2: Guided segmentation in Rocket sequence

tion and reliable correspondence, the processing of features is guided by feedback of information in several ways, including guide in global segmentation, guide in feature matching, refinement of noisy correspondences and guide in local segmentation.

In local segmentation, the *reference regions* which are the most representative of tracked regions, focus the segmentation process on both the position and shape of the corresponding object. The mask that covers the object of interest is predicted from the motion and the shapes of the reference regions. The peak selection process in the intensity histogram of the masked area of the image is guided by the histograms of the reference regions.

The multiplicity of solutions from incremental analysis provides a means of selecting a good solution by measuring the confidence factor of the 3-D motion interpretation associated with each solution. The confidence factor of a solution is computed from the consistency of the 3-D structure over the sequence of the frames. We believe that a solution which represents the actual environment is more likely to have consistent solutions in future frames than a solution which is far from the real situation but happens to fit the given 2-D positions of the correspondence data.

3 Results

The motion analysis system has been tested for standard sets of real image sequences. We present the results for two image sequences provided by UMASS. The first one is the Rocket field sequence [1] and the second one is the Cone sequence [3]. More detailed results can be found in [2]. Figure 2 (a) shows a region tracked from frame 1 to frame 6 of the Rocket field sequence. In global segmentation, the building region is extracted from frame 1 to 6 but it is not available in frame 7. A mask around the predicted position obtained from the reference regions is shown in figure 2 (b). The region from the guided segmentation over the mask in the image is shown in figure 2 (c).

Figure 3 shows the image and intensity histogram used in global and local segmentation. While the image area corresponding to each peak continues to be

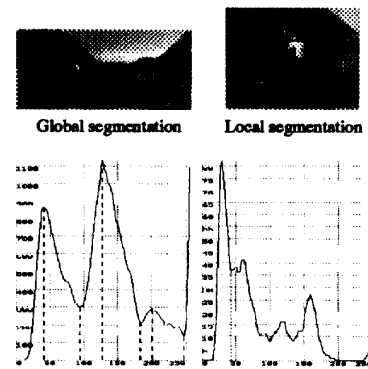


Figure 3: Histograms of intensity in segmentation

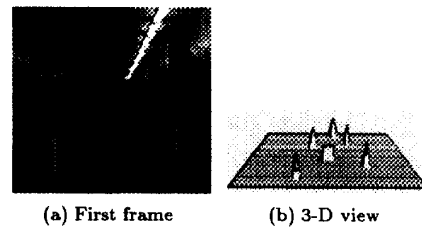


Figure 4: 3-D reconstruction of Cones and cans

segmented into smaller regions recursively, the area corresponding to the building is not extracted into a region in global segmentation. In local segmentation, the area surrounded by the white solid line is intensively examined. The building region is represented by the dominant peak in the intensity histogram and thus extracted into a region as shown in figure 2 (c).

Figure 4 (a) and (b) show the the first frame of the Cone sequence and the reconstructed 3-D structure of the cones and the trash box in the scene, seen from a different viewing angle.

References

- [1] R. Dutta, R. Manmatha, L. Williams, and E. Riseman. A data set for quantitative motion analysis. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 159–164, 1989.
- [2] Y. Kim and K. Price. A feature-based monocular motion analysis system guided by feedback information. In *Proceedings of the DARPA Image Understanding Workshop*, 1993.
- [3] H. Sawhney and A. Hanson. Tracking, detection and 3D representation of potential obstacles using affine constraints. In *Proceedings of the DARPA Image Understanding Workshop*, pages 1009–1017, 1992.