The Keystone Scavenger Team

Jacky Baltes and John Anderson
Autonomous Agents Laboratory
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba, R3T 2N2, Canada
`jacky,andersj@cs.umanitoba.ca`

Abstract

Stereo vision for small mobile robots is a challenging problem, particularly when employing embedded systems with limited processing power. However, it holds the promise of greatly increasing the localization, mapping, and navigation ability of mobile robots. To help in scene understanding, objects in the field of vision must be extracted and represented in a fashion useful to the system. At the same time, methods must be in place for dealing with the large volume of data that stereo vision produces, in order that a practical frame rate may be obtained. We have been working on stereo vision as the sole form of perception for Urban Search and Rescue (USAR) domains over the last three years. Recently, we have extended our work to include domains with more complex human robot interactions. Our entry in the 2006 AAAI Robotics competition embodies these ideas.

Introduction

This paper describes the 2006 Keystone Scavengers, a team participating in the AAAI 2006 Mobile Robotics competition. This team is based on the Keystone Rescue robotic Urban Search and Rescue (USAR) team, which has taken part in previous AAAI and RoboCup competitions.

The main goal of our research is to develop and implement practical algorithms for computer vision, scene understanding, and robot navigation in complex and unstructured environments. We focus on the development of fully autonomous robots that are able to use stereo vision to understand the world around them and navigate through this world. We have extended this work to apply to domains with more complex human-robot interactions and added face recognition, speech recognition, and speech generation to our approach.

Robot Hardware

We plan to use two robots in the competition: (a) Jet, a platform we have used previously, and (b) Slug, a new robot based on the Tarantula toy (Figure 1). Both robots employ a 533 MHz C3 Eden VIA Mini-ITX board, running a Linux kernel with realtime extensions that allows us to do time-sensitive tasks without the need for additional hardware. For

Figure 1: Jet (Left) and Slug (Right)

example, Jet's pan servo is controlled directly from the parallel port on the Mini-ITX board and does not jitter even under the heavy load of the operating system. We developed our own miniaturized version of the Debian Linux distribution for use on systems with reduced hard drive space.

Pragmatic Stereo Vision

The aim of our overall approach to vision is to identify useful regions and match these between stereo images, with limited computational resources and under difficult conditions. The matched regions are used as input for our ego-motion based visual SLAM algorithm (as employed originally in (Baltes *et al.* 2004; Baltes & Anderson 2002) using a single camera). We divide the process of performing region matching in stereo vision into six stages: color correction, image blur, edge detection, region extraction, region simplification, and stereo matching. The remainder of this section briefly describes the region extraction, simplification, and stereo matching components of our system. For more details, please refer to (McKinnon, Baltes, & Anderson 2005).

Region Extraction

Our approach to region extraction involves growing regions from individual pixels using a stack-based approach. At any point, the pixels on the stack represent those from which future expansion of the region will take place. In contrast to other approaches, we allow a pixel to be part of more than one region, since most images can reasonably be segmented into more than one segmentation set. Our algorithm selects the best of the possible segmentations by comparing the quality of the resulting stereo match.

Figure 2: Segmented regions (top), convex hulls and signature lines plotted (middle), matched regions (bottom).

Region Simplification and Stereo Matching

The process of matching regions between stereo images is normally very time consuming, since it may require checking for a match between hundreds of pixels for each potential match. We have sped up this process dramatically, while maintaining similar accuracy, by representing a region using the signature of a convex hull approximating the region.

We use Graham's Scan (Cormen *et al.* 2001) to compute each region's convex hull, and then derive a signature by recording the distance from the center point of the region to the convex hull at angle increments of 5 degrees. This results in 72 values representing lines that approximate the hull (Figure 2, middle). A comparison of the similarity of two regions can then be easily performed by summing the squares of the differences of the values in the 72 corresponding positions in the two arrays (implicitly superimposing the centroids). Beyond greatly decreasing the number of individual points to match, this representation makes the time required to make a comparison *independent of region size*. There is no particular threshold to a match: each region is matched to its strongest partner in the corresponding stereo image. Figure 2 (bottom) illustrates the matches in a sample stereo pair.

Face Recognition

We employ face recognition to detect the presence of humans, using an approach based on Harr features (Lienhart, Kuranov, & Pisarevsky 2002) using summed area tables. This approach provides fast object recognition and performs well in a variety of poses and lighting conditions. The system is trained using positive and negative examples of the target objects, and recognizes variations in faces including scale differences in the same image. Figure 3 shows the out-
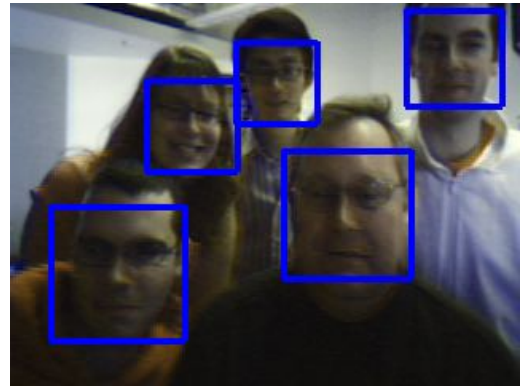

Figure 3: The face detection system wraps bounding boxes around detected faces.

put of our face detection system with an image of some of the team members.

By combining facial detection with the stereo vision system, the distance to a detected face can be determined, and the robot can then approach to a desired distance. Future work will focus on more general detection of human forms, including hands and feet.

Conclusion

The subsystems necessary for a successful scavenger robot (navigation through ego-motion detection, mapping, object recognition) are extremely challenging to support on mobile processors. A significant part of our work is the development of pragmatic techniques to make such processors feasible. With the region simplification methods described here, for example, we observe very good improvements in processing performance with similar matching ability. Our robotic hardware, running only its control software, can achieve a base frame capture rate of 2.5–2.9 fps. Using region matching without simplifications results in a frame rate of 1.5–1.8 fps, while employing the convex hull techniques described here allows a frame rate of 2.3–2.5 fps.

References

Baltes, J., and Anderson, J. 2002. A pragmatic approach to robot rescue: The keystone fire brigade. In *Proceedings of the AAAI Workshop on Rescue Robots*.

Baltes, J.; Anderson, J.; Schaerer, S.; and Wegner, R. 2004. The keystone fire brigade 2004. In *Proceedings of RoboCup-2004*.

Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; and Stein, C. 2001. *Introduction to Algorithms*. Cambridge, MA: MIT Press, 2nd edition.

Lienhart, R.; Kuranov, A.; and Pisarevsky, V. 2002. An extended set of haar-like features for rapid object detection. In *Proc. IEEE Int. Conf. on Image Processing*, 900–903.

McKinnon, B.; Baltes, J.; and Anderson, J. 2005. A region-based approach to stereo matching for USAR. In *Proceedings of RoboCup-2005*.