Learning Image Segmentations from Experience

Michael Ross & Leslie Pack Kaelbling

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: The image segmentation problem originated from the Gestalt school of psychological research, which focused on the organizational and grouping principles of human vision. Segmentation algorithms divide images into regions based on statistical or aesthetic qualities. Common goals are to encode the image efficiently, to separate foreground figures from background, or to locate the boundaries of objects in the image. Previous segmentation algorithms include the snakes, or contour, method by Kass, et al. [3], minimum description length regions by LeClerc [4], and normalized graph-cutting by Shi and Malik [6].

Motivation: We believe that new progress can be made on the segmentation problem by making the definition of "regions" less dependent on the immediate visual input. Traditional approaches to segmentation have two fundamental problems. First, it is difficult to quantitatively compare the outputs of two segmentation algorithms. There is no consensus in the literature about the optimal output of a segmentation algorithm. The most widespread notion is that regions should conform to the objects in a scene, but the definition of "object" is equally elusive. Unfortunately, this makes most comparisons of segmentation algorithms purely qualitative.

The second problem is that making the statistics of the raw image data the foundation of image segmentation is not well supported by our understanding of human visual development. Spelke, et al., have demonstrated that infants are strongly attuned to motion and depth cues when separating figures from backgrounds, but that they are insensitive to the traditional image segmentation cues - color, shape, and texture [7]. It seems likely that infants learn spatial segmentation based on their early experience with motion and depth segmentations, and we might derive better segmentation algorithms by attempting to do the same.

Approach: To address these two problems, we are pursuing two parallel research tracks. First, we are examining ways to optimize segmentations by their utility for performing quantifiable tasks. In an image feature space (composed of edges, shapes, and textures), segmentation is a measure of the distance between spatial features. Image elements that are adjacent in this space belong to the same image region, or segment. We can adjust this segmentation measure to improve performance on a task that requires segmentation information. An example of a segmentation-dependent task is motion prediction. In order to compute a dense field of motion vectors from a series of video frames, it is necessary to group image elements together [2]. When we use this flow to predict the location of the tracked object in future frames we get error information and can use this to derive a new segmentation metric that will produce better prediction results. Other, more sophisticated tasks, such as object manipulation, also require segmentation information and can provide useful, quantitative evaluations of segmentation quality.

We are also researching how to base spatial image segmentation on the primary cues of motion. The photorecptive neurons in the visual system adapt to the light and color levels of an input and become less sensitive to its presence [5]. When the stimulus changes, their activity levels increase strongly, producing sensitivity to motion in the world, a primary cause for change in visual input. Moving objects are highlighted by our visual system and our attention is naturally drawn towards them. The computational equivalent to this process is background subtraction, where we use probabilistic models to distinguish moving objects from static background pixels. Using background subtraction algorithms developed by Grimson, et al. [1], we can extract pixels that belong to moving objects, as shown in Figure . These pixels serve as sample figure-background segmentations that we hope to use to learn how to segment similar objects from novel, static scenes.

Impact: We hope that this work will produce segmentation results that are quantifiably justified by their external utility and their basis in reasonable psychophysical assumptions. Our ultimate goal is a system that can learn useful segmentations in real-world environments from self-labeled data. Learning segmentations from motion

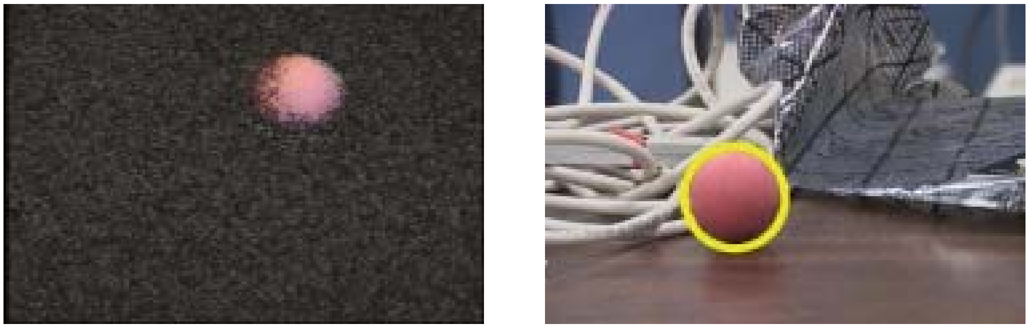

Figure 1: A moving object [left] produces a natural figure-ground segmentation due to its difference from the background. Learning the spatial properties of the object may make it possible to extract it from novel still scenes, as in the synthetic example on the right.

cues should provide more robust object separation than segmentations derived from pre-programmed statistical or aesthetic rules. Linking the segmentation task to higher-level goals will lead to segmentation algorithms that are principled communicators between low-level sensors and high-level artificial intelligence systems.

Research Support: This research is supported in part by a graduate student grant from the National Science Foundation, and in part by a research grant from the Singapore-MIT Alliance.

References:

[1] W.E.L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *Computer Vision and Pattern Recognition*, 1998.

[2] B.K.P. Horn. *Robot Vision*. The MIT Press, 1986.

[3] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *International Conference on Computer Vision*, 1987.

[4] Y.G. LeClerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[5] Stephen E. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, 1999.

[6] J. Shi and J. Malik. Normalized cuts and image segmentation. In *Computer Vision and Pattern Recognition*, June 1997.

[7] Elizabeth S. Spelke, Peter Vishton, and Claes Von Hofsten. Object perception, object-directed action, and physical knowledge in infancy. In *The Cognitive Neurosciences*, pages 165–179. MIT Press, 1994.