Object Segmentation with Pixel Stacks

Kinh Tieu & Eric Grimson

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: There is wide agreement that segmenting an image into objects is a useful pre-processing step in computer vision. Although the definition of "object" depends on the context, and currently we may only be able to achieve an over-segmentation of the image into parts of objects, this intermediate representation is nevertheless much more compact and less arbitrary than the raw pixel-based one.

Motivation: Higher-level recognition systems may use configurations of segments for tasks such as recognition. By reducing the dimensionality of the input, segmentation also makes estimating probabilistic models of objects more tractable. Our conjecture is that starting with some sort of segmentation defines a set of low-level "features" for higher-level tasks such as grouping based on similarity and further feature induction.

Previous Work: Much of previous and current work on segmentation has focused on efficient algorithms for optimal segmentation where the segmentation criterion is selected a priori based on heuristics such as homogeneous color. However, some objects such as pedestrians (with varied colored clothing) as shown in Figure 2 are not well represented with a fixed color model. Here we focus on developing a suitable criterion for grouping pixels into segments. Our work is related to that of [2, 3].

Approach: Our approach segments an object into parts using a set of spatially (approximately) aligned images of instances of the object. Form a cube $I(i, j, k)$, where for $1 \leq k \leq n$, i, j index the pixel locations in the images, by arranging the images in a stack. A "pixel stack" $p(i, j)$ is defined as the array of n pixel values (one from each image) in location i, j. The idea is to cluster the pixel locations into segments using the values of the corresponding pixel stacks. For example for a set of pedestrian images, one may want to group the pixel locations belonging to the shirt region. So although the value of the pixel location may be different from one image to another (e.g., different shirt colors), all pixel locations corresponding to the shirt will change similarly. Figure 1 shows a hierarchical segmentation of pedestrians using this method.

Impact: The pixel stack representation suggests a reasonable criterion for model-based object segmentation where traditional heuristics such as homogeneous color fail. It is naturally invariant to absolute values and instead considers the relative similarity of pixels. In fact the color representation is a special case of pixel stacks because a color image may be thought of as a stack of three perfectly aligned images of the same scene. Note that many existing methods which perform pattern classification on the raw pixels implicitly assume some sort of alignment.

Future Work: We plan to extend the work by integrating our segmentation approach with a registration algorithm to deal with image sets that are not spatially aligned. The goal is to incrementally align the images to maximize the likelihood of the data under the segmentation. In addition, we also plan to investigate probabilistically segmenting new images using a learned pixel stack representation. The "soft" segmentation boundaries may prove useful for deformable shape models of object parts.

Research Support: This research is supported in part by DARPA: DAAD19-00-1-0466 and ITRI.

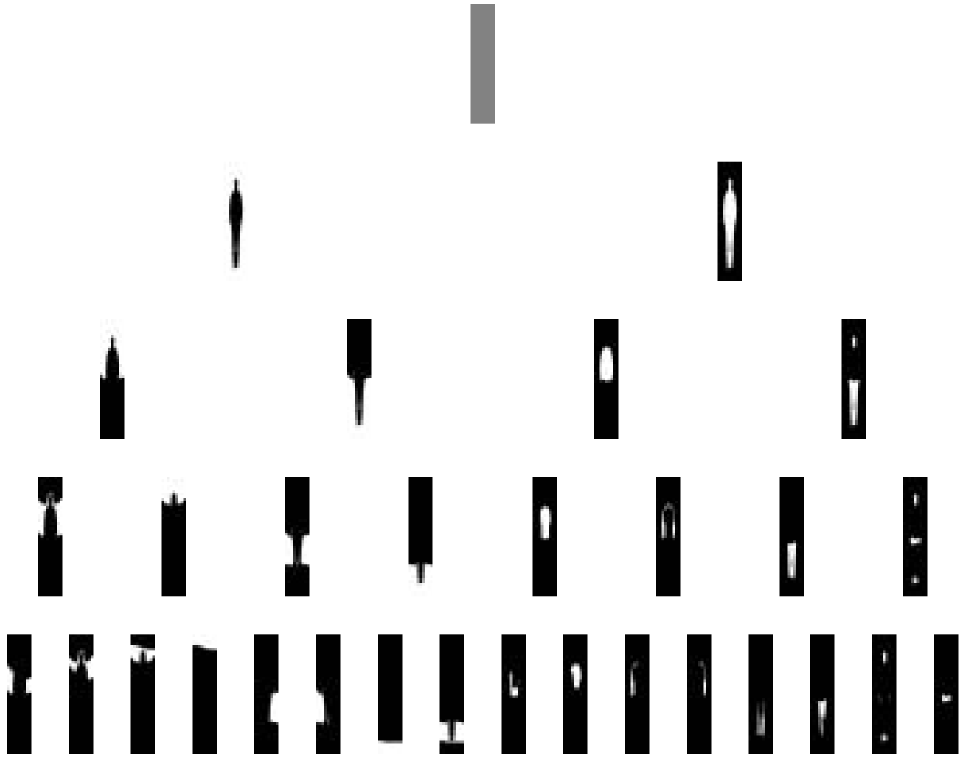

Figure 1: Hierarchical segmentation of pedestrians using the pixel stack representation.

Figure 2: Example images of pedestrians [1] with different colors for parts such as shirt and pants regions.

References:

[1] MIT CBCL. Cbcl pedestrian database #1, 2000.

[2] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[3] Chris Stauffer. Similarity templates for detection and recognition. In *IEEE Comp. Vis. Patt. Recog. Conf.*, 2001.