Category-Specific Object Recognition and Segmentation Using a Skeletal Shape Model

Nhon H. Trinh
ntrinh@lems.brown.edu
Benjamin B. Kimia
kimia@lems.brown.edu

Division of Engineering
Brown University
Providence, RI
USA

While appearance-based methods have achieved great success in image classification and object detection, their use can be limited because stable key feature points are not always abundantly available and intra-category variations of features can be as large as inter-category variations. This paper investigates the complementary use of shape cues for object recognition and segmentation. Two interconnected issues emerge: *(i)* the repeatability/distinctiveness tradeoff on the choice of shape features and *(ii)* how to model the spatial relationship among the features to fully and efficiently capture their variations within a category, which is especially important when the objects articulate or when the features are simple. We propose using the **skeletal/shock graph model**, a distributed pictorial-structure-type model, as the shape representation and using the "visual fragments," pairs of contour fragments corresponding to a skeleton segment, as the basic features for recognition. These two proposals are integrated in an object recognition/segmentation system which consists of three components: *(1)* a generative shape model which can accurately describe a wide range of free-form shapes, *(2)* an efficient search method, and *(3)* an objective function to rank-order the hypotheses.

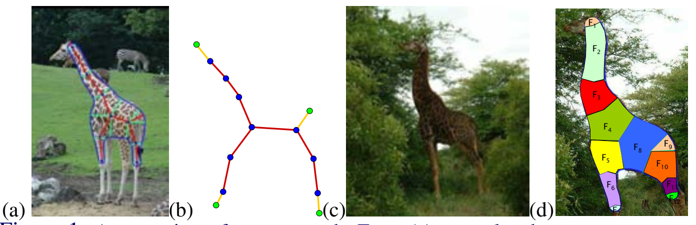

Figure 1: An overview of our approach: From (a) exemplar shapes, we construct (b) prototypical shock graph topologies and a range of metric variations associated with them. Given (c) a new image, we generate shapes from the constructed prototype and search for those with adequate image support. (d) The result of this search allows us to detect and delineate category instances in the image.

1. Fragment-Based Generative Shape Model Our generative model for shape is an extension of Trinh and Kimia's model (ICCV'07), which synthesize shapes sharing a given shock graph topology from a fairly low number of parameters. Like their model, ours describe a shape as a composition of **fragments**, our basic features for recognition, Figure 1a. However, through an extrinsic formation the fragments in our model are independent from one another since each fragment is fully described from the local first-order geometric properties at adjacent nodes.

Since the specification of a shock graph topology can still generate shapes well beyond one category and since each object category typically maps to a few shock graph topologies, we also model a shape prior capturing the allowed variation of the geometric attributes of the shock graphs. We developed a simple procedure to learn these prototypes from the training exemplars.

2. Single-Pass Multiple Solution Dynamic Programming Given a query image and an objective function which is a sum over image support of the fragments, the Dynamic Programming (DP) approach allows us to search among those generated by our model in polynomial time. However, the traditional DP fails to return additional instances of a category when present in the image because it only gives the globally optimal solution. We devise the **Differential Exclusion Principle** to find multiple solutions without having to run DP multiple times: the solution space is initially enlarged to one optimal DP solution per root node state, and then narrowed down by discarding solutions that are spatially not local minimum of the objective function. Lastly, we apply thresholding to obtain the final solution set.

3. Rank-Ordering Hypotheses Using Contour Chamfer Matching (CCM) The popular Oriented Chamfer Matching (OCM) cost (Shotton *et al.*, PAMI'08) may over-count or under-count image support by matching an edge to multiple contour points or by not matching the best edges to the contour points, Figure 2a. It may also reward accidental alignment between edges and the contour, Figure 2b. To avoid these pitfalls, we restrict the selection of matching edges to a thin window orthogonal to the contour point, Figure 2c, and match each contour point to the edge that minimizes the OCM cost. We also penalize the orientation difference between the model contour and the connected best edge correspondences, which "zig-zags" around the model curve if the alignment is accidental.

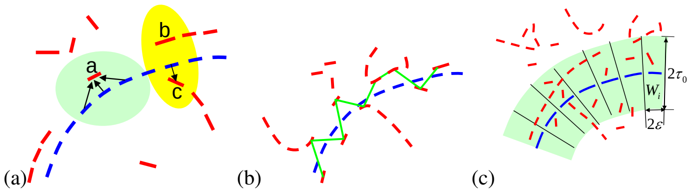

Figure 2: (a) OCM can overcount edge labeled "a", or may undercount by not using the best edge, edge "c" instead of "b"; (b) Accidental alignment between a contour and the image edges often forms a zig-zagging image contour. (c) In CCM, the best matching edge is restricted to a window around each contour point.

Experiments The 255-image ETHZ Shape Classes dataset (Ferrari *et al.*, ECCV'06) is used to measure the performance of our algorithm. This is a challenging data set in which objects from 5 categories vary in size and pose and often blend with cluttered backgrounds. For the object detection task, we show significant improvements at least in three categories (bottles, giraffes, and swans), same performance for the mugs, and worse performance for the applelogos when compared to results of Zhu *et al.* (ECCV'08) and Ferrari *et al.* (INRIA Tech Report'08). The worse performance in apple logo was probably because not enough prototype scales were used. In a pilot experiment where we used more prototype scales for larger images, our performance on the applelogos category exceeded that of both Zhu *et al.* and Ferrari *et al.* As for the object segmentation task, our algorithm has a stable performance in both coverage ($\sim$90%) and accuracy ($\sim$80%), and improves on Ferrari *et al.* in three categories, but not in the categories of applelogos and mugs.

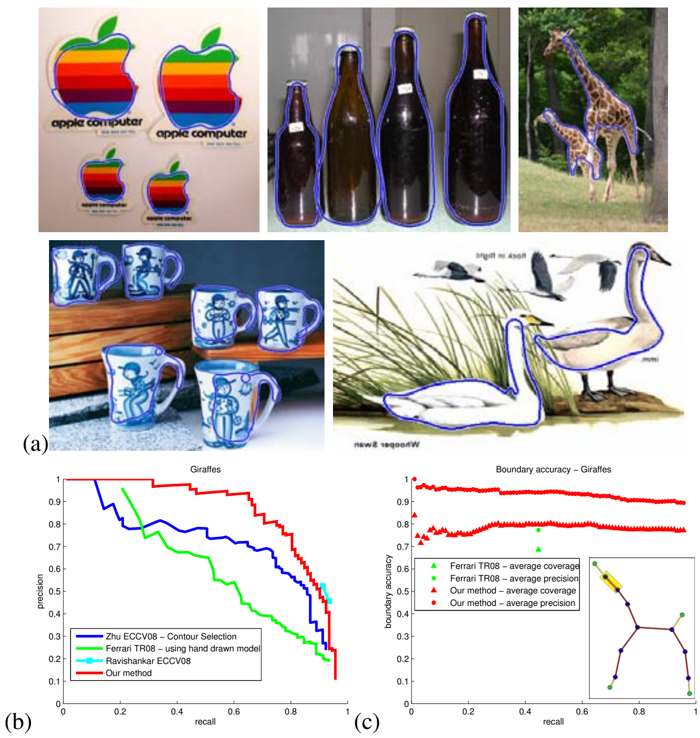

Figure 3: (a) Segmentation examples of the ETHZ dataset. (b-c) Comparisons of precision-recall and average boundary accuracy for the giraffe category in the ETHZ dataset.