Contour Segment Matching by Integrating Intra and Inter Shape Cues of Objects

Ishani Chakraborty
http://research.rutgers.edu/~ishanic
Ahmed Elgammal
http://www.cs.rutgers.edu/~elgammal

Dept. of Computer Science,
Rutgers, The State University of New Jersey
Piscataway, NJ
USA

Figure 1: Overview of our approach. Left: Input image. (b) Left Center: Line segments (in white) extracted to form an *edge image*. (c)Right Center: Contour segments detected by *inter-shape* matching only (d)Right: Contour segments detected by combining *inter-shape* correspondences and and *intra-shape* contextual constraints. This is the output of our framework.

Shape-based methods are a natural choice for color and texture invariant object detection. In recent years, a large body of research has focused on contour based techniques for shape representation. Most of the methods can be broadly classified as point-based approaches eg., [1] or boundary-curve based approaches eg., [4] and [2]. Our method follows the latter approach in which contours of an object shape are approximated as a set of line segments. We perform object detection by framing it as a correspondence problem between contours segments in an input image and an object model. The model is a line drawing that consists of a small number of line strokes defining the boundary contour of an object. In the input image, we identify an instance of the object category in a cluttered environment by searching for contour segments in a similar topology as that of the model.

Our framework focuses on a mid-level, local representation of shape. We adopt the *Contour Segment Network* (CSN) [2] formulation in which k contour segments are grouped based on spatial connectivity. An ordering of the segments is then enforced and a numerical descriptor encodes the structure as a vector. We define a *structure* as a group of connected line segments encoded by a descriptor. *Structural support* is then, the number of line segments in a structure. A single support is insufficient for encoding the complexities of various object parts efficiently. To overcome this problem, we generate and describe multi-support structures at each segment.

To match contours between image and model, we define two metrics: (a) distance *between* the model and image contours (termed as *inter-shape correspondence*) to find likely candidates for a match. and (b) spatial connectivities among contours *within* each image (termed as *intra-shape contextual constraints*) to find a coherent whole that matches the model.

Inter-shape correspondences are derived from inter-structure distances. An inter-structure distance is a distance between *two sets* of k *segments*. By slight twist of notation, we convert this distance to a segment-centric form, distance between k *sets* of *two segments*. I.e., each inter-structure distance $d^k_{S_D,S_M}$ is attributed to k ordered, inter-segment distances.

Each segment can belong to several structures and hence generate multiple inter-structure distances. To select a single representative inter-segment distance, we apply a minimum filter across all inter-structure distances and all k to compute the distance that is induced by the best matching structure. Let $a,b\cdots$ and $\alpha,\beta,\cdots$ be segments and $S^k_D = \cup_{1:k}\{a\}$ and $S^k_M = \cup_{1:k}\{\alpha\}$ be structures in the input (D) and model (M) images. Then, inter-segment distance is

$$d_{a\alpha} = min_k min_{i,j}(d(S^k_{i,D}, S^k_{j,M} | a \in_p S^k_{i,D}, \alpha \in_p S^k_{j,M})) \qquad (1)$$

Pairwise inter-segment distances can be used to solve for one-to-one correspondences between contours in input and model. Distances $d_{a\alpha}$ are converted into probabilities and and expressed in the matrix form Q. We implement a maximum weighted graph matching using the Hungarian algorithm.

Our part-based representation models local shape, ignoring the spatial context of those parts. As a result, background parts often hallucinate as objects and create wrong matches. To mitigate erroneous detections, we include contextual constraints and search for a connected, coherent whole within the input.

Intra-shape constraints are expressed in an image by the connectivities that underlie contour grouping in the formation of structures. We define the *intra-shape* adjacency matrices in which two nodes in a graph are connected by weight 1 if their representative contour segments are members of a common structure.

By including intra-shape constraints, we softly bias the matches towards a single, connected set of contour segments that match strongly with the model. This can be achieved by integrating the inter-shape matches with intra-shape grouping in an iterative framework.

Our iterative approach is based on the relaxation labeling framework [3] for contextual graph matching. We break the problem into a two-step iterative approach. In the first step, we find an assignment M_t based on the probability distances Q_t between the contours of input and model images. In the second step, we recalculate the probability matrix Q_{t+1} based on a *support function*. This is calculated as a function of matches M_t in the first step and the contextual cues expressed as intra-shape adjacency matrices. The iteration continues till a stable local point and the corresponding M is the optimal assignment that identifies the object contours in the input image. The relaxation rule is then:

$$Q^{t+1}_{a\alpha} = \frac{Q^t_{a\alpha} S^t_{a\alpha}}{\sum_b \sum_\beta Q^t_{b\beta} S^t_{b\beta}} \qquad (2)$$

where Q^t is the probability matrix and the support function S^t weighs the probabilities according to the intra-shape contextual constraints.

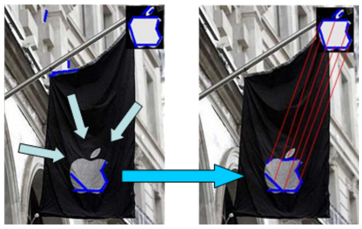

Figure 2: The first image shows the contour matches found using inter-shape correspondences only. These matches are enhanced by including contextual constraints that drive the matches towards a connected set of contour segments that match strongly with the model in the second image.

We tested our algorithm on the ETHZ shape dataset. The results obtained by this method outperforms previous work in contour-based object detection and can be also be used to localize the actual object contours in an image. Our method differs from its predecessors in (1) We use a *multi-level* shape description to capture complexities of different object parts (2) We match groups of line segments across images for discriminative correspondences and use a novel mechanism to induce inter-segment distances and (3) We follow a two-step intuitive approach for object detection: inter-shape correpondences are used to perform a dense search followed by a sparse, local search for the best matching candidate. We *integrate* inter and inter shape cues for contour segment matching in an iterative framework. Our proposed approach is less vulnerable to background clutter, more adaptive to different complexities in shape and yields much higher detection rate than its predecessors.

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 24(4):509–522, April 2002.

[2] V. Ferrari, T. Tuytelaars, and L.J. Van Gool. Object detection by contour segment networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages III: 14–28, 2006.

[3] A. Kostin, J.V. Kittler, and W.J. Christmas. Object recognition by symmetrised graph matching using relaxation labelling with an inhibitory mechanism. *Pattern Recognition Letters (PRL)*, 26(3):381–393, February 2005.

[4] S. Ravishankar, A. Jain, and A. Mittal. Multi-stage contour based detection of deformable objects. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages I: 483–496, 2008.