Perceiving Shapes through Region and Boundary Interaction [*]

Stella X. Yu[†‡] and Jianbo Shi[†]

Robotics Institute[†]

Carnegie Mellon University

Center for the Neural Basis of Cognition[‡]

5000 Forbes Ave, Pittsburgh, PA 15213-3890

{stella.yu, jshi} @cs.cmu.edu

Abstract

We present a computational algorithm for shape perception by studying the interaction between region and boundary cues. We formulate this problem in a graph partitioning framework, where region cues defined on a pixel graph and boundary cues defined on its dual edgel graph are coupled through edge-node incidence relationships. The consistency of simultaneous partitioning on such graphs can thus be guaranteed. We generalize normalized cuts criteria and algorithms to this model for globally optimal solutions. We demonstrate that by incorporating boundary smoothness, objects with heterogeneous region properties can stand out as one group and objects with weak contours can be segmented more readily without the suppression from objects of high contrast. This model can also encode higher-order shape information and preliminary results on shape selection are given.

1. Introduction

The perception of shapes arises from interactions between perceived boundaries and regions in the image. Boundaries usually coexist with regions as the locations where region properties break down, but they can also be quite independent [7], as in line drawings. Regions properties can sometimes distinguish one area from another simply by their within-area homogeneity and between-area contrast. However, when image features do not obey these homogeneity constraints as often the case in natural images, geometric regularities of the boundary can be utilized to bind image elements into one shape. In addition, our perception of region properties of a shape can be strongly influenced by bounding contours [2].

The exploitation of both region and boundary information in seeing a shape can be appreciated in Fig. 1, where a faint contour boarding a coherent region is perceive more strongly in segregating shapes (Fig. 1a), whereas regions with heterogeneous properties are seen as one shape due to their continuous enclosing boundaries (Fig. 1b).

a) Faint edges.　　b) Heterogeneous subregions.

Figure 1: Interaction of region and boundary for seeing a shape. a) The faint contour between two birds is enhanced due to region coherence. b) Despite of its various markings, the pumpkin is perceived as a whole due to its continuous enclosing boundary.

At a lower level, region grouping and contour integration are intimately coupled in determining image segmentation. At a higher level, since a shape is the 2D version of a 3D object, the spatial arrangement of features on both the silhouette and the interior work together to define shapes. This point is articulated by Koffka [13] through a physical analogy of an oil droplet forming its shape in water: *"We saw that the oil, immersed in a liquid with which it will not mix, will separate itself from it by the forces within and between the surfaces of the two media, and that the same surface forces will also give shape to the oil, under specially simple conditions, the shape of a sphere."* Such interaction becomes especially apparent in illusory contour formation and

---

amodal completion [8, 18, 9].

In computer vision, the two aspects of a shape, region and boundary, have both been exploited in image segmentation. Boundary-based methods include edge detection and active contour techniques. Edge detection [5] faces the difficult task of converting edge candidates to boundaries of regions of interest. Snakes [12] and balloons [6] evolve the boundary of some initial segmentation to a locally optimal energy state which demands both contour smoothness and high intensity gradient along the contours. Faint contours (Fig. 1a) can completely elude detection or attraction in such models. Region-based grouping methods include region growing and merging techniques [1]. These algorithms build up complicated regions by merging smaller regions that pass certain similarity test. They lack a global cost function and resulting regions may have noisy boundaries.

There have been numerous attempts to combine the boundary and region information to segment objects as well. Either the two processes are implemented sequentially [15], or they are formulated in one global energy functions [10, 4, 3, 20, 11]. The formulations in Markov random fields (MRF) have the flavor of considering two grouping problems simultaneously, however putting all the constraints on both regions and contours is difficult. Furthermore, energy minimization approaches through gradient descent often suffer from finding only locally optimal solutions. The computational difficulty of MRF forces one to seek an alternative solution.

In this paper, we formulate shape perception in a graph partitioning framework, where a global decision on the goodness of shapes is made based on local cues, evaluated on both boundaries and regions. We construct two dual graphs with image pixels and edgels as nodes respectively. Weights in the pixel graph encode region cues, whereas weights in the edgel graph encode contour cues. We generalize normalized cuts criteria [17] to a partitioning on this graph pair. A consistent partitioning on the two graphs is guaranteed by the coupling of their partitioning membership vectors through edge-node incidence matrices. Such a link through linear operators allows us to derive analogous eigendecomposition solutions to the global optimization criterion.

We demonstrate that by incorporating boundary smoothness, objects with weak contours and multiple structures can be segmented more readily from the background. This model can be extended to more general situations where the dual edgel graph becomes a hyperedge graph representing higher order boundary constraints. We present some preliminary results on shape selection as well.

The rest of the paper is organized as follows. Section 2 expands our grouping method in detail. Section 3 illustrates our ideas and results on synthetic images. Section 4

concludes the paper.

2. Method

In graph approaches for image segmentation, an image is described by a weighted graph $G = (V, E)$, in which each pixel is associated with a vertex $v \in V$ and an edge $e \in E$ between vertex j to k is associated with some weight W_{jk}, which measures the affinity between the nodes.

2.1. Dual graph representation

In order to model both region grouping and boundary grouping, we adopt two graphs for an image. One is the above pixel graph, $G = (V, E)$; the other is an edgel graph, $\partial G = (\partial V, \partial E)$. The edgel graph is the dual graph of the pixel graph, whose nodes in ∂V correspond to edges in E. Let E be associated with weight matrix A. It measures the pairwise fitness of grouping based on region cues, such as intensity, texture and motion similarities of two pixels. Let ∂E be associated with weight matrix B. It measures the pairwise fitness of grouping based on boundary cues, such as collinearity and cocircularity of two edgels.

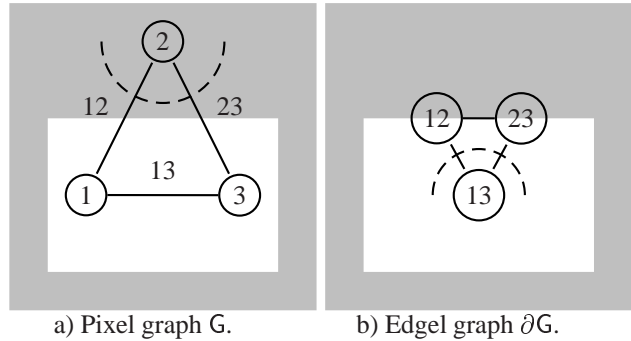

a) Pixel graph G.   b) Edgel graph ∂G.

Figure 2: Dual graph representation of region and contour cues. A white rectangular shape is displayed in a gray background. Graph representations are superimposed on the image, where circles are nodes and lines are edges between nodes. Dashed lines indicate graph cuts. a) Pixel graph G has pixels as its nodes. Bipartitioning of the pixels into figure and ground also induces a bipartitioning of edge set E into a cut set (edges 12, 23) vs. a non-cut set 23). The edges between nodes measure region cues such as intensity similarity. b) Edgel graph ∂G has edgels as its nodes, which are edges of G. The induced partitioning on E is thus translated into a vertex partitioning on ∂G. The edges in ∂G are associated with weights measuring contour continuity. The goal of partitioning is to achieve high within-group affinity both for nodes in G (surface property similarity) and for nodes in ∂G (boundary regularity).

In general, a node in the edgel graph can be defined as a hyperedge in the pixel graph characterizing relationships among many nodes. The simplest case is an ordinary edge defined by two nodes. Fig. 3 shows the pixel and edgel graphs first adopted in MRF formulations for image restoration [10]. We emphasize that an edgel graph can be defined in various ways other than having these horizontal and vertical line sites.

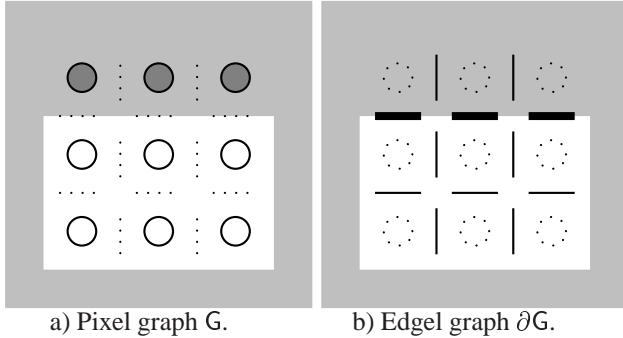a) Pixel graph G.   b) Edgel graph $\partial\mathsf{G}$.

Figure 3: An example of pixel and edgel graphs. A rectangular shape is displayed on a gray background. a) This divides the nodes in the pixel graphs into figure and ground sets (empty and filled circles). b) It also induces a partitioning on line sites, defined by two neighboring pixels, into figure-ground border and interior sets (thick and thin lines). To demonstrate the duality, we superimpose the dual graph in dotted version on each graph.

The connections between the two dual graphs is characterized by an edge-node matrix H which links the nodes in the edgel graph to the nodes in the pixel graph. Each row of H indicates the incidence relationships of a node in $\partial\mathsf{V}$ (i.e. an edge in E) with the nodes in V. With a particular choice of H, we can turn the evaluation on $\partial\mathsf{V}$ into that on V.

The goal we are seeking in segmenting objects is to find a partitioning on the pixel graph so that not only region coherence of a shape is maximized, but the induced partitioning on the edgel graph also maximizes contour coherence of the shape. This leads to a simultaneous partitioning on coupled graphs. Here we adopt normalized cuts criteria [17] for graph partitioning. Similar derivation can be made for other graph partitioning criteria. We will first briefly overview normalized cuts on a single graph. Then we extend the criterion to simultaneous partitioning on multiple graphs and derive its computational solutions. Shape perception through pixel and edgel graphs can then be cast as an application of the general approach.

2.2. Vertex partitioning on one graph

A weighted graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ consists of vertex set V and edge set E. An edge between node j and k is associated with weight W_{jk}. The weight matrix is assumed symmetrical: $W = W^T$, where ^{T} is the transpose operator. Let $W = W_+ + W_-$, where W_+ and W_- contain respectively all nonnegative and nonpositive entries of W. For two arbitrary vertex sets P and Q, let $\mathcal{D}_W(P)$ denote the degree of connections of P, and $\mathcal{C}_W(P, Q)$ denote the total W connections from P to Q, $\mathcal{R}_W(P, Q)$ denote connection ratio from set P to Q:

$$
\begin{aligned}
\mathcal{D}_W(P) &= \sum_{j \in P, k \in \mathsf{V}} |W(j, k)|, \\
\mathcal{C}_W(P, Q) &= \sum_{j \in P, k \in Q} W(j, k), \\
\mathcal{R}_W(P, Q) &= \frac{\mathcal{C}_W(P, Q) + \mathcal{D}_{W_-}(P)}{\mathcal{D}_W(P)}.
\end{aligned}
$$

Let $\Gamma_G^k = \{\mathsf{V}_l, l = 1, \dots, k\}$ denote a k-way vertex partitioning on graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$, where $\mathsf{V} = \cup_{l=1}^k \mathsf{V}_l$ and $\mathsf{V}_i \cap \mathsf{V}_j = \varnothing, \forall i \neq j$. It can be shown [19] that with these definitions, normalized association and cuts criteria [17] can be extended to any real matrices:

$$
\begin{aligned}
\epsilon_a(\Gamma_G^k) &= \sum_{l=1}^k \mathcal{R}_W(\mathsf{V}_l, \mathsf{V}_l), \\
\epsilon_c(\Gamma_G^k) &= \sum_{l=1}^k \mathcal{R}_W(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l).
\end{aligned}
$$

Since $\epsilon_a(\Gamma_G^k) + \epsilon_c(\Gamma_G^k) = k$, they form dual formulations: maximizing $\epsilon_a(\Gamma_G^k)$ is equivalent to minimizing $\epsilon_c(\Gamma_G^k)$. Therefore, a good partitioning only cuts off connections which are globally relatively weak weights.

Let partition assignment matrix be $X = [X_1, \dots, X_k]$, where X_l be the membership indicator vector for $\mathsf{V}_l \in \Gamma_G^k$. Let the degree matrix D_W of weight matrix W be a diagonal matrix, where $D_W(i, i) = \sum_j |W(i, j)|$. It can be verified that with $Y = X(X^T D_W X)^{-\frac{1}{2}}$,

$$
\begin{aligned}
\epsilon_a(\Gamma_G^k) &= \mathrm{trace}(Y^T(D_{W_-} + W)Y), \\
\epsilon_c(\Gamma_G^k) &= \mathrm{trace}(Y^T(D_{W_+} - W)Y), \\
\text{s. t.} \quad & Y^T D_W Y = I.
\end{aligned}
$$

By relaxing discreteness constraints, the maximizer of ϵ_a is obtained from the first k largest generalized eigenvectors of $(D_{W_-} + W, D_W)$, or equivalently the last k smallest generalized eigenvectors of $(D_{W_+} - W, D_W)$.

The generalization of graph partitioning from a nonnegative weight matrix to an arbitrary matrix is important for our model. As we will see, even if both region affinity and boundary affinity are nonnegative, our formulation leads to equivalent weight matrices that necessarily involve negative weights.

2.3. Simultaneous partitioning on multiple graphs

Suppose we are given n graphs $\mathsf{G}_m = (\mathsf{V}_m, \mathsf{E}_m)$, with weight matrix $W_m, m = 1, \dots, n$. We would like to find a

simultaneous partitioning on these graphs optimizing a joint criterion. If these graphs are unrelated, the optimization of the criterion can be done independently. However, if these graphs are coupled together, such that vertex partitioning in a master graph induces corresponding vertex partitioning in other graphs, we need to find one optimal partitioning that takes all weight matrices into consideration.

The simplest case of such coupled graphs is that they share the same topology but different weight matrices: $\mathsf{V}_i = \mathsf{V}_j$, $\mathsf{E}_i = \mathsf{E}_j$, $\forall i \neq j$. Let the graph be $\mathsf{G} = (\mathsf{V}, \mathsf{E})$. In this case, we choose to optimize the linearly weighted association and cut criteria (we can show that the criteria have a probabilistic view of finding low conductivity sets in a stochastic jump-diffusion process):

$$
\begin{aligned}
\epsilon_a(\Gamma_G^k) &= \sum_{l=1}^{k} \frac{\sum_{m=1}^{n} \mathcal{R}_{W_m}(\mathsf{V}_l, \mathsf{V}_l) \mathcal{D}_{W_m}(\mathsf{V}_l)}{\sum_{m=1}^{n} \mathcal{D}_{W_m}(\mathsf{V}_l)}, \\
\epsilon_c(\Gamma_G^k) &= \sum_{l=1}^{k} \frac{\sum_{m=1}^{n} \mathcal{R}_{W_m}(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l) \mathcal{D}_{W_m}(\mathsf{V}_l)}{\sum_{m=1}^{n} \mathcal{D}_{W_m}(\mathsf{V}_l)}.
\end{aligned}
$$

The duality between the two criteria is maintained through the convex combination of connection ratios. It can be shown that the optimal partitioning on graph G with $\{W_m : m = 0, \ldots, n\}$ is given by the first k largest generalized eigenvectors of

$$
(\sum_{m=1}^{n} D_{W_{m-}} + \sum_{m=1}^{n} W_m , \sum_{m=1}^{n} D_{W_m}).
$$

Note that it is *not* equivalent to the normalized cuts on G with weight matrix $\sum_{m=1}^{n} W_m$, unless each weight matrix W_m is nonnegative, $m = 1, \ldots, n$.

If graphs G_m's do not share the same topology, one way to connect their partitioning is to find a mapping between them. We assume that the master graph is $\mathsf{G} = (\mathsf{V}, \mathsf{E})$. A vertex partitioning Γ_G^k induces a corresponding vertex partitioning on G_m, $m = 1, \ldots, n$. Since our criteria respect individual grouping on each of the graphs while coupling them through a weighted addition procedure, it allows certain independence of individual graph partitioning. If we can turn the scoring of partitioning on G_m to that on G, we essentially transform the partitioning on heterogeneous graphs to that on homogenous graphs, the case of which has been studied in the above. Such transformation relies intimately on incidence matrices that link one graph topology to another.

We now come back to the pixel and edgel graphs for shape perception and use this as a concrete example to illustrate how such transformations can be done.

2.4. Coupling between dual graphs

An edge-node incidence matrix H in graph G, relates edges in E to the nodes in V. Let $|P|$ denote the cardinality of set P. The edge-node incidence matrix H is an $|\mathsf{E}| \times$

$|\mathsf{V}|$ matrix. An example is given in Fig. 4. The choice of a directed edge-node incidence matrix over an undirected edge-node incidence matrix (in which the elements have no signs) will become clear shortly.

a) Nodes in a pixel graph and its edgel graph.

$$
H = \begin{bmatrix}
n_1 & n_2 & n_3 & n_4 & n_5 & n_6 & n_7 & n_8 \\
\hline
-1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{matrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{matrix}
$$

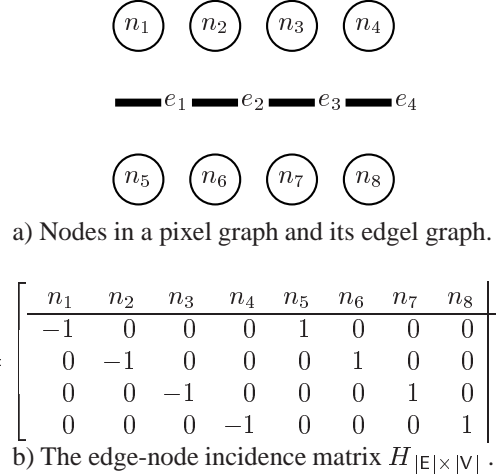b) The edge-node incidence matrix $H_{|\mathsf{E}| \times |\mathsf{V}|}$.

Figure 4: An example of directed edge-node incidence matrices. Horizontal line site e_j is defined as the edge pointing from the north pixel neighbor n_j to the south pixel neighbor n_{j+4}, $j = 1, \ldots, 4$. The direction is reflected in the 4×8 egde-node incidence matrix H, where $H_{j,j} = -1$ for the source and $H_{j,j+4} = 1$ for the destination.

In general, a node in $\partial \mathsf{V}$ can be related to more than two pixel nodes and thus can be defined as a hyperedge on a set of nodes in V. In this sense, we can interpret every 2D convolution kernel O in the image pixel plane as an edge-node incidence matrix $H(O)$, where each row of $H(O)$ is a translated 1D version of O. For example, for edgels of four orientations at each pixel location, we can define four types of dual graphs with H's generated by the following oriented difference operators,

$$
\begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}_{O_1} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}_{O_2} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}_{O_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}_{O_4}
$$

where each edgel is incident on two pixels along a particular orientation. Likewise, Prewitt difference operators can also be used and they involve 6 pixel nodes instead. Fig. 5 shows $H(O_3)$ on a 9×9 image.

The goodness of boundary grouping for an inducing partitioning Γ_G^k is evaluated on the pairwise boundary elements using contour affinity matrix B. The key observation is that this evaluation can be recast into that on associated region grouping using an equivalent region affinity matrix $H^T B H$.

Let $\partial X = HX$, where X is the $|\mathsf{V}| \times k$ partition assignment matrix for the pixel graph, i.e, $X_{i,r} = 1$ if pixel i belongs to region r and 0 otherwise. Since H is essentially a difference operator, HX picks up changes in pixel node valuation given by a partitioning Γ_G^k. This results in
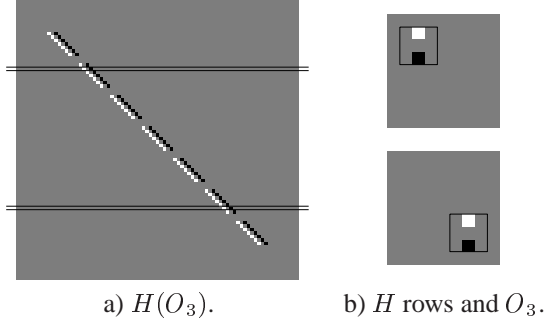
a) $H(O_3)$.   b) H rows and O_3.

Figure 5: Edge-node incidence matrix H can be generated by a 2D convolution kernel O in the image plane. a) H for a 9×9 image is generated by O_3 detecting horizontal edges. All-zero rows correspond to edgels at image boundary where the kernel cannot be applied. b) Each row of H is a translated 1D version of the 2D kernel O_3. Shown here are the results of reshaping two rows outlined in H into 9×9 images, with kernel O_3 outlined.

edgel node valuation for the inducing partitioning Γ_G^k. Note that ∂X is an $|\mathsf{E}| \times k$ matrix. Suppose edgel e is incident on pixel i (source) and j (destination). For region index $r = 1, \dots, k$,

$$\partial X_{e,r} = \begin{cases} 0, & X_{i,r} = 0, X_{j,r} = 0 : & e \text{ outside region } r \\ 0, & X_{i,r} = 1, X_{j,r} = 1 : & e \text{ within region } r \\ 1, & X_{i,r} = 0, X_{j,r} = 1 : & e \text{ into region } r \\ -1, & X_{i,r} = 1, X_{j,r} = 0 : & e \text{ out of region } r \end{cases}$$

In other words, $\partial X_{e,r}$ only fires when e is a crossing edgel of region r and its sign indicates the direction of the crossing with reference to e's own direction, which is irrelevant to any partitioning.

With X labeling region belongingness of pixel nodes, $X^T A X$ evaluates region grouping; whereas with ∂X signaling only the boundary edgels of the inducing pixel partitioning, $(\partial X)^T B (\partial X)$ evaluates boundary grouping. The signs of ∂X tell us whether two edgels are crossing the boundary in the same direction. This information encodes border ownerships [14] and we impose contour continuity on edgels with the same partitioning polarity.

Note that this partitioning induction relationship is asymmetrical: every pixel graph partitioning has an edgel graph partitioning, but not every edgel graph partitioning has a valid pixel graph partitioning. To reflect this dependence and partitioning consistency, for inducing partitioning Γ_G^k, we denote the induced partitioning on $\partial \mathsf{G}$ as $\partial \Gamma_G^k$.

To help understand the process we described so far, we analyze the example in Fig. 4. Suppose that we have a

bipartitioning on its pixel graph, given by:

$$\begin{aligned} \Gamma_G^2 &= \{\mathsf{V}_1, \mathsf{V}_2\}, \\ \mathsf{V}_1 &= \{n_1, n_2, n_3, n_5, n_8\}, \\ \mathsf{V}_2 &= \{n_4, n_6, n_7\}. \end{aligned}$$

This induces a 3-way partitioning on the edgel graph, which includes interior edgels and crossing edgels originating from V_1 and V_2 respectively:

$$\begin{aligned} \partial \Gamma_G^2 &= \{\partial \mathsf{V}_0, \partial \mathsf{V}_1, \partial \mathsf{V}_2\}, \\ \partial \mathsf{V}_0 &= \{e_1\}, \\ \partial \mathsf{V}_1 &= \{e_2, e_3\}, \\ \partial \mathsf{V}_2 &= \{e_4\}. \end{aligned}$$

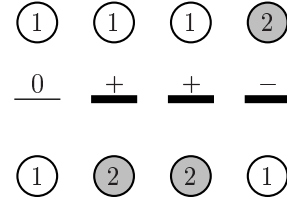This dual partitioning is illustrated in Fig. 6.

Figure 6: Dual partitioning on pixel and edgel graphs. The pixel labels are given by the numbers in circles. The induced edgel labels are marked on top of each line, with 0 for interior edgels, $+$ and $-$ for crossing edgels originating from pixel partition 1 and 2 respectively.

With these notations, we can compute:

$$\begin{aligned} X^T &= \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}, \\ (\partial X)^T &= \begin{bmatrix} 0 & -1 & -1 & 1 \\ 0 & 1 & 1 & -1 \end{bmatrix}, \\ X^T A X &= \begin{bmatrix} \mathcal{C}_A(\mathsf{V}_1, \mathsf{V}_1) & \mathcal{C}_A(\mathsf{V}_1, \mathsf{V}_2) \\ \mathcal{C}_A(\mathsf{V}_2, \mathsf{V}_1) & \mathcal{C}_A(\mathsf{V}_2, \mathsf{V}_2) \end{bmatrix}, \\ \partial X^T B \partial X &= \sum_{i,j>0} (-1)^{i-j} \begin{bmatrix} \mathcal{C}_B(\partial \mathsf{V}_i, \partial \mathsf{V}_j) & -\mathcal{C}_B(\partial \mathsf{V}_i, \partial \mathsf{V}_j) \\ -\mathcal{C}_B(\partial \mathsf{V}_i, \partial \mathsf{V}_j) & \mathcal{C}_B(\partial \mathsf{V}_i, \partial \mathsf{V}_j) \end{bmatrix}, \end{aligned}$$

where $\mathcal{C}_W(P, Q)$ (defined earlier) is the total W connection from node set P to Q.

As can be seen, the resulted $k \times k$ matrix $X^T A X$ has within-region affinity as diagonal elements and between-region affinity as off-diagonal elements. From the formula for $(\partial X)^T B (\partial X)$, we see that only crossing edgels contribute to boundary grouping. Those edgels of the same polarity ($i = j$) support the inducing partitioning by adding positively to within-region affinity and negatively to

between-group affinity; whereas those of the opposite polarities ($i \neq j$) undermine the score of the inducing partitioning by contributing in an opposite way so as to decrease within-region affinity and increase between-region affinity.

This example illustrates that $(\partial X)^T B (\partial X)$ provides an equivalent scoring on the inducing partitioning Γ_G^k for the edgel graph partitioning $\partial \Gamma_G^k$. In fact, through the edge-node incidence matrix H, we can transform the partitioning on graph ∂G to that on G with an equivalent weight matrix $H^T B H$:

$$(\partial X)^T B (\partial X) = X^T (H^T B H) X.$$

Since H is a directed edge-node incidence matrix, the equivalent weight matrix necessarily contains negative weights. This is also evidenced by the above example, where for the H in Fig. 4, we have

$$H^T B H = \left[\begin{array}{cc} B & -B \\ -B & B \end{array} \right].$$

Therefore, the contribution from an edgel to the regions of its incident pixels has always two folds: whenever it increases(decreases) within-region affinity, it also decreases(increases) between-region affinity by the same amount. These two aspects are consistent in supporting (undermining) the inducing pixel partitioning.

We would like to point out that not every edge-node incidence matrix H generated by an arbitrary 2D kernel O (e.g. a Gaussian kernel) leads to a reasonable equivalent weight matrix $H^T B H$. We essentially need to find weights on pixels so that their difference represents the weight on an edge. This is in a close analogy (theoretically and methodologically) to the concepts in physics about electrical currents and potentials. In a larger framework, it resembles the approach taken by the Stokes theorem on differentiable functions (not every function!), where an integral of a vector field over a closed region boundary can be turned into an integral over the region. This conversion also requires the region and the boundary to be oriented, just as we use a directed edge-node incidence matrix instead of the undirected version.

2.5. Algorithm

After the edgel graph partitioning is converted into an equivalent pixel graph partitioning, the image segmentation with both region affinity A and boundary affinity B becomes a simultaneous partitioning on the pixel graph with weight matrices of A and $\{H(O_m)^T B H(O_m), m = 1..4\}$. The complete algorithm is summarized in Table 1.

This algorithm on a pair of A and B can be easily generalized to deal with multiple region affinity matrices $\{A_m : m = 1, \ldots, n_a\}$ and multiple boundary affinity matrices $\{B_m : m = 1, \ldots, n_b\}$.

Step 1: Compute A - pairwise pixel feature similarity.
Step 2: Compute B - pairwise edgel continuity.
Step 3: Compute $H(O_m)$, where O_m is a difference operator defining edgels from pixels, $m = 1..4$.
Step 4: Compute $W_m = H(O_m)^T B H(O_m)$, $m = 1..4$. Let $W_5 = A$.
Step 5: Normalized cuts on G with weight matrices $\{W_m : m = 1..5\}$ by solving the largest eigenvector of
$$\left(\sum_{m=1}^{5} D_{W_{m-}} + \sum_{m=1}^{5} W_m , \sum_{m=1}^{5} D_{W_m} \right).$$

Table 1: Algorithm for object segmentation with region affinity matrix A and boundary affinity matrix B.

3. Results

We illustrate our ideas and method on synthetic images. We calculate the affinity between two $d-$dimensional features using a Gaussian function of their difference: $g(X; \Sigma)$, where

$$g(X; \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp^{-\frac{1}{2}(X)^T \Sigma^{-1} X},$$

and Σ's are $d \times d$ covariance matrices.

For an image f, the pairwise pixel affinity is measured by proximity and intensity similarity, whereas pairwise edgel affinity is measured by proximity and curve collinearity:

$$A_{ij} = g(X_a; \Sigma_a), \ X_a = \left[\begin{array}{c} l_{ij} \\ f(i) - f(j) \end{array} \right];$$

$$B_{ij} = g(X_b; \Sigma_b), \ X_b = \left[\begin{array}{c} l_{ij} \\ 2 - \cos(2\beta_i) - \cos(2\beta_j) \end{array} \right];$$

where l_{ij} is the distance between pixel i and j, β_i is the relative edgel angle with respect to the line connecting pixel i and j [16]. We confine this affinity evaluation to a local neighborhood of a pixel. For all our examples, we use a neighborhood radius of 3, $\Sigma_a = \left[\begin{array}{cc} 36 & 0 \\ 0 & 0.01 \end{array} \right]$ and $\Sigma_b = \left[\begin{array}{cc} 100 & 0 \\ 0 & 1 \end{array} \right]$.

We use the four 2D kernels O_m defined earlier to generate the edge-node incidence matrix H. We scale A so that the maximum affinity in A is equal to that in $\{H(O_m)^T B H(O_m), m = 1..4\}$. After this balancing operation, we study how partitioning changes with the relative weighting between the two using a homotopy parameter t:

$$\{(1 - t) \cdot A, \ t \cdot H(O_m)^T B H(O_m), m = 1, \ldots, 4\}.$$

As t varies from 0 to 1, we shift from solely region grouping to solely boundary grouping.

Fig. 7 shows an image of a high contrast shoe-like shape and a weak contrast rectangle. If only region grouping based on intensity similarity is considered, the shoe stands out since it has much higher contrast against the background. If we incorporate boundary grouping based on contour collinearity, the rectangle can compete with the shoe despite of its weak contrast. In fact, when the contour information is weighted heavily against the region information, the shoe can lose the battle completely.
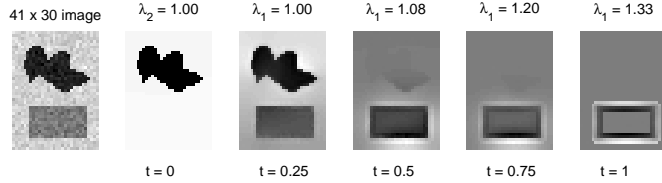

Figure 7: Contour regularity helps the segmentation of an object with weak contrast. The image has a shoe shape of intensity 0.2 and a rectangle of intensity 0.4 against a background of intensity 0.6. Gaussian noise with standard deviation 0.03 is added to the image. Next to the image are normalized cuts results from solely region cues to solely boundary cues. Their eigenvalues are given at the top and homotopy parameter t at the bottom. When $t = 0$, there is essentially no contour information, the irregular shoe shape of high contrast pops out from the rest (in the second eigenvector). When $t = 1$, there is no contrast information, the boundary of the weak-contrast rectangle pops out. When t varies from 0 to 1, the shoe gradually fades into the background while the rectangle starts to dominate in the partitioning.

Fig. 8 demonstrates that a figure made of two objects with opposite contrast polarities can be grouped together if their enclosing boundary forms good grouping. Note that their internal boundary is zigzagged, so that it does not compete with the external boundary grouping.

We note that the above framework can take more general boundary constraints into consideration so that a particular shape can be selected. Since we only consider edgel collinearity, we cannot encode higher-order relationships between edgels at this point. However, with this limited feature input, we can make a simple shape selection. Fig. 9 is an illustration along this direction, where two identical bars can be differentiated depending on their orientations: the weak vertical bar stands out if vertical boundary is favored in boundary grouping.

These examples serve as a proof of concepts and we will apply the method to real images such as those shown in Fig. 1 in the future.
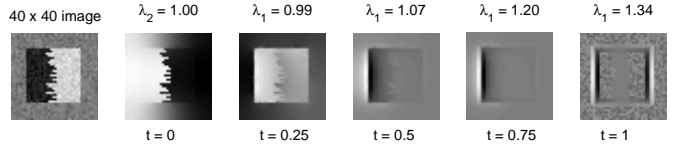

Figure 8: Contour regularity helps binding heterogeneous structures into one shape. The average intensity values of figures and ground are 0.4, 0.8 and 0.6. The results are organized as in Fig. 7. Without boundary constraints, the two substructures of the figure have more contrast against each other than each with the ground, so they cannot be grouped together. However, when the boundary grouping is incorporated, they come out as one shape due to the regularity formed by their external boundaries.
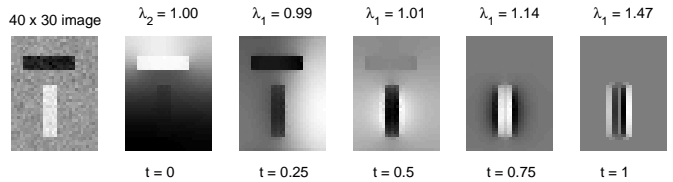

Figure 9: Biased boundary grouping can lead to shape selection. The image has two identical bars. One is placed horizontally and the other vertically. The horizontal bar has a higher intensity (0.3) contrast against ground (0.6), while the vertical bar is 0.8. The results are organized as in Fig. 7. The horizontal bar stands out when region cues are considered. Since only O_1 is used for boundary grouping, objects with vertical boundaries are favored. When t increases, the vertical bar pops out.

4. Conclusions

We formulate object segmentation through region and boundary interaction in the computational framework of graph partitioning. Representing region and boundary explicitly by dually coupled pixel graph and edgel graph, we show that boundary grouping can be turned into an equivalent region grouping. This allows the two grouping processes to be combined as one partitioning problem and globally optimal solutions can be obtained by simple eigendecomposition techniques.

We generalize normalized cuts criteria on a single graph to simultaneous partitioning on multiple graphs. This provides a general framework for integrating multiple cues from multiple levels. With the power of topology conversion through incidence matrices, we can encode higher-order boundary constraints that describe a shape and integrate them with region constraints.

We show on synthetic images that objects with regular geometrical properties can be segmented out more readily despite of their weak contours or internal variation of region properties. With bias in boundary grouping, shapes of interest can be selected first in segmentation.

5. Acknowledgements

References

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–7, 1994.

[2] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, editor, *The cognitive neurosciences*, pages 339–51. MIT Press, Cambridge, MA, 1999.

[3] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection and robust statistics with applications in early vision. *International Journal of Computer Vision*, 1995.

[4] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.

[5] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–98, 1986.

[6] L. D. Cohen. On active contour models and ballons. *Computer Vision, Graphics and Image Processing*, 53(2):211–8, 1991.

[7] J. H. Elder and S. W. Zucker. Evidence for boundary-specific grouping. *Vision Research*, 38(1):143–52, 1998.

[8] D. Geiger and K. Kumaran. Visual organization of illusory surfaces. In *European Conference on Computer Vision*, Cambridge, England, April 1996.

[9] D. Geiger, K. Kumaran, H. kuo Pao, and N. Rubin. The shape of illusory figures. In *International Conference on Image Processing*, volume 2, pages 6–10. 1999.

[10] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–41, 1984.

[11] I. H. Jermyn and H. Ishikawa. Globally optimal regions and boundaries. In *International Conference on Computer Vision*, 1999.

[12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, pages 321–331, 1988.

[13] K. Koffka. *Principles of Gestalt Psychology*. A Harbinger Book, Harcourt Brace & World Inc., 1935.

[14] K. Nakayama, S. Shimojo, and G. H. Silverman. Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18:55–68, 1989.

[15] T. Pavlidis and Y.-T. Liow. Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):225–33, 1990.

[16] P. Perona and W. Freeman. A factorization approach to grouping. In *European Conference on Computer Vision*, pages 655–70, 1998.

[17] J. Shi and J. Malik. Normalized cuts and image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 731–7, June 1997.

[18] L. R. Williams and D. W. Jacobs. Local parallel computation of stochastic completion fields. *Neural computation*, 9(4):859–81, 1997.

[19] S. X. Yu and J. Shi. Segmentation with pairwise attraction and repulsion. In *International Conference on Computer Vision*, 2001.

[20] S. C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.