Understanding Popout:
Pre-attentive Segmentation through Nondirectional Repulsion [*]

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

The goal of pre-attentive segmentation is to mark conspicuous image locations such as region boundaries, smooth contours and popout targets against backgrounds. This salience detection relies on not only feature similarity analysis but also local feature contrast. We identify these two measures with attraction and nondirectional repulsion, and unify the dual processes of association by attraction and segregation by repulsion in one grouping framework. We generalize normalized cuts to multi-way partitioning with these dual measures and show that the criterion can be viewed as a stochastic jump-diffusion process, where the probability of jump is determined by the relative strengths of attraction and repulsion. We demonstrate that this extended model can deal with salience detection under various situations as well as the asymmetry in visual search. Through these results, we provide a clear understanding of the role of negative weights in the graph partitioning framework. This opens up the possibilities of encoding negative correlations in constraint satisfaction problems, where solutions by simple and robust eigendecomposition become possible.

1. Introduction

Visual processing starts by extracting local features such as oriented edges. As a prerequisite for higher-level tasks such as objection recognition, these features detected at an early stage must be grouped into meaningful global entities such as regions, boundaries and surfaces. The goal of pre-attentive visual segmentation [15] is to mark conspicuous image locations and make them more salient for perceptual popout. These locations not only include boundaries between regions, but also smooth contours and pop-out targets against backgrounds (Fig. 1).

It has long been assumed that regions are first characterized by features which are homogeneous within the areas.
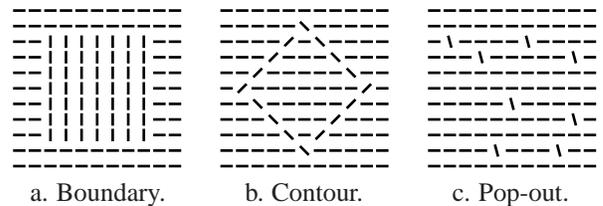

Figure 1: Pre-attentive segmentation is to mark conspicuous image locations, which could be caused by a) region boundaries, b) smooth contours and c) pop-out targets. In these examples, the similarity of features within figure and ground compounds with the dissimilarity between figure and ground. Figure and ground are well segregated in feature maps tuned to different orientations.

These feature measures are then compared at neighboring locations to locate boundaries between regions [15]. This view of feature discrimination for grouping is supported by evidence in neurophysiology on elaborate feature detectors in visual cortex [6], in psychophysics on visual search [26] and in modeling on texture segmentation [12, 4, 18]. Some other approaches of texture segmentation go beyond the analysis of features obtained from image filters and model the interactions between filters [30]. These Markov Random Field models [8] capture context dependences and other statistical characteristics of texture features [15].

However, it has been shown [3, 13, 23, 19] that when feature similarity within an area and feature differences between areas are teased apart, the two aspects of perceptual organization, grouping and segregation, can contribute independently to perception. In particular, when features are varied continuously within areas, it is the local feature contrast, rather than the feature properties themselves, that is more important for the perceived grouping. Fig. 2 demonstrates that local feature contrast plays an active role in binding (even dissimilar) elements together [19].

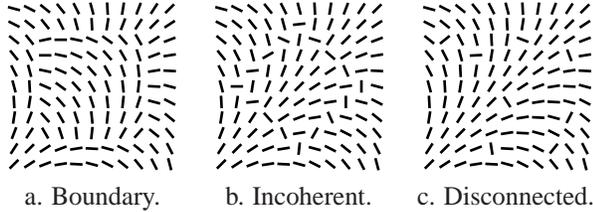| a. Boundary. | b. Incoherent. | c. Disconnected. |

Figure 2: Local feature contrast alone is sufficient to perceptually link dissimilar elements together. a) Boundary by local orientation contrast. b) Figure without curvilinearity. c) Spatially disconnected figure without element similarity.

These results motivate models of preattentive vision where region boundaries are directly localized through lateral interactions between edge detectors [20, 15]. The breakdown of spatial homogeneity in features causes changes in contextual influences, resulting in higher responses at the border than at surrounding locations.

This contextual feature analysis for grouping can be best understood in a relational graph framework, where each location is denoted by a node and feature compatibility between locations is captured by a weight associated with the edge between nodes. A good segmentation is a graph partitioning which cuts the graph into components of relatively large interior weights and relatively small exterior weights. Gestalt grouping factors, such as proximity, similarity, continuity and symmetry, can be encoded and combined in pairwise similarity measures [28, 25, 22, 7, 24]. Complex grouping phenomena can emerge from simple computation on these local cues [10, 17].

While Gestalt laws [27] have always stressed the aspects of similarity of elements in grouping, the effect of saliency, or feature contrast, or local dissimilarity, cannot be described in a framework that models similarity grouping. It has already been pointed out [21] the asymmetry between figure and ground might have to be described by an unbalanced criterion which favors figure (but not ground) being coherent. We see in Fig. 2c that, completely dissimilar elements that are spatially disconnected can be grouped together in a figure simply due to the fact they are locally dissimilar to a common ground. This suggests that it may not be the problem with a balanced criterion, but rather it is the transcription problem from an image to a relational graph where dissimilarity measures are lacking.

In this paper, we present a grouping method which integrates pairwise attraction and repulsion information. Whereas the attraction measures the degree of association by feature similarity, the repulsion measures the segregation by feature dissimilarity. We generalize normalized cuts criteria [25] to a multi-way partitioning on these dual measures in one framework. We derive the necessary and sufficient conditions on the graph weights for objects to be segmented in a variety of settings. We show that grouping by the dual procedures of association and segregation can be viewed as a jump-diffusion process [9], where the probability of jump between two diffusion processes is determined by the relative strengths of attraction and repulsion. We demonstrate that both salience detection and the asymmetry in visual search [14] can be accounted for by our method.

The rest of the paper is organized as follows. Section 2 expands our grouping method in detail. Section 3 provides a probabilistic view of our criterion. Section 4 presents experimental results. Section 5 concludes the paper.

2. Method

In graph approaches, an image is described by an undirected weighted graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$, in which each pixel is associated with a vertex $v \in \mathsf{V}$ and an edge $e \in \mathsf{E}$ between vertex j and k is associated with weight W_{jk}. Suppose we are given two *nonnegative* measures, A_{Ga} and A_{Gr}, for attraction and nondirectional repulsion (as opposed to directional repulsion that describes ordinal relationships such as relative depth cues [29]). Both of them are assumed symmetrical: $A_{Ga} = A_{Ga}^T$ and $A_{Gr} = A_{Gr}^T$.

2.1. Criteria

For two vertex sets P and Q, let $\mathcal{A}(P, Q)$ and $\mathcal{R}(P, Q)$ denote the total attraction and repulsion from P to Q, i.e.

$$
\begin{array}{rcl}
\mathcal{A}(P, Q) & = & \sum_{j \in P, k \in Q} A_{Ga}(j, k), \\
\mathcal{R}(P, Q) & = & \sum_{j \in P, k \in Q} A_{Gr}(j, k).
\end{array}
$$

In particular, $\mathcal{A}(P, \mathsf{V})$ and $\mathcal{R}(P, \mathsf{V})$ measure total attraction and repulsion of set P respectively. For attraction, we ask the association by within-group attraction to be as large as possible; whereas for repulsion, we ask the segregation by between-group repulsion to be as large as possible (Fig. 3).

We extend the bipartitioning in normalized cuts [25] to a k-way partitioning based on both attraction and repulsion. A k-way vertex partitioning $\{\mathsf{V}_l, l = 1, \ldots, k\}$ on graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ has $\mathsf{V} = \cup_{l=1}^k \mathsf{V}_l$ and $\mathsf{V}_i \cap \mathsf{V}_j = \varnothing, \forall i \neq j$. We unify the association by attraction and segregation by repulsion in one criterion by a linear combination of normalized ratios weighted with their relative strengths. Our normalized association and normalized cuts criteria are defined as:

$$
\epsilon_a = \sum_{l=1}^k \left[\frac{\mathcal{A}(\mathsf{V}_l, \mathsf{V}_l)}{\mathcal{A}(\mathsf{V}_l, \mathsf{V})} \cdot \frac{\mathcal{A}(\mathsf{V}_l, \mathsf{V})}{\mathcal{A}(\mathsf{V}_l, \mathsf{V}) + \mathcal{R}(\mathsf{V}_l, \mathsf{V})} + \frac{\mathcal{R}(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l)}{\mathcal{R}(\mathsf{V}_l, \mathsf{V})} \cdot \frac{\mathcal{R}(\mathsf{V}_l, \mathsf{V})}{\mathcal{A}(\mathsf{V}_l, \mathsf{V}) + \mathcal{R}(\mathsf{V}_l, \mathsf{V})} \right],
$$

$$
\epsilon_c = \sum_{l=1}^k \left[\frac{\mathcal{A}(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l)}{\mathcal{A}(\mathsf{V}_l, \mathsf{V})} \cdot \frac{\mathcal{A}(\mathsf{V}_l, \mathsf{V})}{\mathcal{A}(\mathsf{V}_l, \mathsf{V}) + \mathcal{R}(\mathsf{V}_l, \mathsf{V})} + \frac{\mathcal{R}(\mathsf{V}_l, \mathsf{V}_l)}{\mathcal{R}(\mathsf{V}_l, \mathsf{V})} \cdot \frac{\mathcal{R}(\mathsf{V}_l, \mathsf{V})}{\mathcal{A}(\mathsf{V}_l, \mathsf{V}) + \mathcal{R}(\mathsf{V}_l, \mathsf{V})} \right].
$$

In ϵ_a, for each of the k partitions, we measure the goodness of association by its within-group attraction ratio, the goodness of segregation by its between-group repulsion ratio, resulting in the goodness of grouping as the linear summation
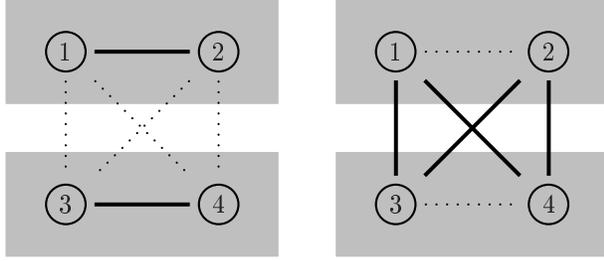
a) Association by attraction.   b) Segregation by repulsion.

Figure 3: Grouping criterion. Illustrated here is $\Gamma_G^2 = \{\{1,2\},\{3,4\}\}$. We want to maximize normalized sums of thick-lined edge weights in ϵ_a, while minimizing those of dotted-lined weights in ϵ_c. a) For attraction, we aim at maximizing within-group association. b) For repulsion, we aim at maximizing between-group segregation. These two goals are combined into one grouping criterion ϵ_a weighted by their relative strengths of total connections.

of the two ratios, weighted by their relative strengths. A good partitioning is the one which maximizes the sum of such k ratios. Because $\epsilon_a + \epsilon_c = k$, ϵ_c is a dual formulation of ϵ_a. Therefore, maximizing ϵ_a is equivalent to minimizing ϵ_c. Intuitively, we aim at having tight attraction within clusters and loose attraction between clusters at the same time, strong repulsion between clusters and weak repulsion within clusters at the same time. The duality between ϵ_a and ϵ_c relieves us of such two goals to either one of them. Whenever one is maximized, the other is automatically minimized. Such a balance in the criterion underlies its success in real image processing.

2.2. Computational solution

It is readily seen from the criterion that repulsion can be regarded as negative attraction:

$$
\begin{aligned}
\epsilon_a(\Gamma_G^k) &= \sum_{l=1}^{k}\left[\frac{\mathcal{A}(\mathsf{V}_l,\mathsf{V}_l)+\mathcal{R}(\mathsf{V}_l,\mathsf{V}\setminus\mathsf{V}_l)}{\mathcal{A}(\mathsf{V}_l,\mathsf{V})+\mathcal{R}(\mathsf{V}_l,\mathsf{V})}\right] \\
&= \sum_{l=1}^{k}\left[\frac{\mathcal{A}(\mathsf{V}_l,\mathsf{V}_l)-\mathcal{R}(\mathsf{V}_l,\mathsf{V}_l)+\mathcal{R}(\mathsf{V}_l,\mathsf{V})}{\mathcal{A}(\mathsf{V}_l,\mathsf{V})+\mathcal{R}(\mathsf{V}_l,\mathsf{V})}\right].
\end{aligned}
$$

We introduce a few symbols to turn the above criterion into a computable form. Let X_l be a membership indicator vector for class l, $l = 1, \ldots, k$. Let the degree matrices D_{Ga} and D_{Gr} be diagonal matrices, where $D_{Ga}(j,j) = \sum_k A_{Ga}(j,k)$ and $D_{Gr}(j,j) = \sum_k A_{Gr}(j,k), \forall j$. Let

$$
\begin{aligned}
W &= A_{Ga} - A_{Gr} + D_{Gr}, \\
D &= D_{Ga} + D_{Gr}.
\end{aligned}
$$

It can be verified that [31]

$$
\begin{aligned}
\epsilon_a(\Gamma_G^k) &= \sum_{l=1}^{k}\frac{X_l^T W X_l}{X_l^T D X_l^T} = \mathrm{trace}(Y^T W Y), \\
\text{s. t.} \quad & Y^T D Y = I,
\end{aligned}
$$

where $Y = X(X^T D X)^{-\frac{1}{2}}$ and $X = [X_1, \ldots, X_k]$. Therefore, if we relax the discreteness constraints on Y, by using the method of Lagrange multipliers, the above quadratic formulation leads to the standard generalized eigenvalue problem, i.e., the maximizer of ϵ_a is given by the k largest generalized eigenvectors of (W, D). Based on Gershgorin's theorem, we can estimate $|\lambda_l(W, D)| \leq 2, \forall l$, where $\lambda_l(W, D)$ is the l-th largest eigenvalue of (W, D).

Let's examine two extreme cases. If there is no repulsion, i.e. $A_{Gr} = 0$, we have $W = A_{Ga}$ and $D = D_{Ga}$, which reduces to the conventional normalized cuts [25], where 1 is the eigenvector of (W, D) with the largest eigenvalue of 1. If there is no attraction, i.e. $A_{Ga} = 0$, we have $W = D_{Gr} - A_{Gr}$ and $D = D_{Gr}$, where 1 is the eigenvector of (W, D) with the smallest eigenvalue of 0. Between these two extremes is the case where we have both attraction and repulsion, usually 1 is no longer an eigenvector of (W, D). Indeed, attraction tends to bind elements together, while repulsion tends to break elements apart. The optimal partitioning results from the balance of such two forces.

2.3. Bipartitioning case

The formulation in the continuous domain can become even stronger if we exploit the constraints among indicator vectors. It seems nontrivial to derive the constraints for general k-way partitioning, but we can get a good handle on $k = 2$. Let $M = 1^T D 1$, $\alpha = \frac{X_1^T D X_1}{M}$ and $y = (1-\alpha)X_1 - \alpha X_2$. By using $X_1 + X_2 = 1$ and $X_1^T X_2 = 0$, we can prove that:

$$
\epsilon_a(\Gamma_G^2) = \frac{y^T W y}{y^T D y} + C, \quad C = \frac{1^T W 1}{1^T D 1}.
$$

Notice that C is a constant and it measures the attraction to total degree ratio, i.e. $0 \leq C = \frac{1^T A_{Ga} 1}{1^T D 1} \leq 1$: $C = 1$ when and only when $A_{Gr} = 0$, $C = 0$ when and only when $A_{Ga} = 0$. Ideally, a good grouping seeks the solution of the following optimization problem,

$$
\begin{aligned}
y_{opt} &= \arg\max_y \frac{y^T W y}{y^T D y} \\
\text{s.t.} \quad & y^T D 1 = 0, \quad \forall j, y_j \in \{(1-\alpha), -\alpha\}.
\end{aligned}
$$

With attraction and repulsion working together, the constraint $y^T D 1 = 0$ is not automatically satisfied by the solution of the generalized eigensystem. Instead, according to the Courant minimax theorem, we have

$$
\lambda_2(W, D) = \min_{y_1}\max_{\{y:\, y^T D y_1 = 0\}}\frac{y^T W y}{y^T D y} \leq \max_{\{y:\, y^T D 1 = 0\}}\frac{y^T W y}{y^T D y} \leq \lambda_1(W, D).
$$

3

Let the set of eigenvectors for eigenvalue $\lambda(W, D)$ be $\Upsilon(W, D, \lambda)$. In other words, $y_1 \in \Upsilon(W, D, \lambda_1)$ and $y_2 \in \Upsilon(W, D, \lambda_2)$ yield an upper and lower bound for the optimal ϵ_a. Stated in Theorem 1 are the necessary and sufficient conditions when the second largest eigenvector does give the optimal solution in the continuous domain.

Theorem 1 *For $W = A_{Ga} - A_{Gr} + D_{Gr}$, where both A_{Ga} and A_{Gr} are nonnegative and symmetric, $y_{opt} = \arg\max_{\{y:\, y^T D1 = 0\}} \frac{y^T W y}{y^T D y} \in \Upsilon(W, D, \lambda_2)$ iff $\exists \mu \geq 0$, s.t. $D_{Gr} = \mu D_{Ga}$ and $\lambda_1(W, D) \leq \frac{1}{1+\mu}$.*

Theorem 1 states two requirements. One is the balance between attraction and repulsion among all vertices. The ratio of total repulsion to total attraction for every node should be the same across the network. The other requirement can be interpreted as the dominance of attraction, since when μ is small, the second condition is more likely to be satisfied. Due to the continuity, the closer these conditions are satisfied, the closer the solution to the original formulation.

2.4. Partitioning with a symmetrical matrix

In the framework of attraction and nondirectional repulsion, we can formulate the normalized cuts on an arbitrary symmetrical weight matrix W. Let

$$W = W_+ - (-W_-) = A_{Ga} - A_{Gr},$$

where W_+ and W_- contain respectively all nonnegative and negative entries of W. We regard W_+ as attraction A_{Ga} and $-W_-$ as repulsion A_{Gr}, and interpret the normalized cuts on this pair of A_{Ga} and A_{Gr} as that on W.

For vertex sets P and Q, let $\mathcal{D}_W(P)$ denote the degree of connections of P, $\mathcal{C}_W(P, Q)$ denote the total W connections from P to Q, $\mathcal{S}_W(P, Q)$ denote the connection ratio from set P to Q, based on which we can define the normalized association (ϵ_a) and cuts (ϵ_c) criteria:

$$
\begin{aligned}
\mathcal{D}_W(P) &= \sum_{j \in P, k \in \mathsf{V}} |W(j, k)|, \\
\mathcal{C}_W(P, Q) &= \sum_{j \in P, k \in Q} W(j, k) + \mathcal{D}_{W_-}(P), \\
\mathcal{S}_W(P, Q) &= \frac{\mathcal{C}_W(P,Q)}{\mathcal{D}_W(P)}; \\
\epsilon_a(\Gamma_G^k) &= \sum_{l=1}^k \mathcal{S}_W(\mathsf{V}_l, \mathsf{V}_l), \\
\epsilon_c(\Gamma_G^k) &= \sum_{l=1}^k \mathcal{S}_W(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l).
\end{aligned}
$$

Since $\mathcal{S}_W(\mathsf{V}_l, \mathsf{V}_l) + \mathcal{S}_W(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l) = 1$, $\epsilon_a + \epsilon_c = k$. The maximizer of ϵ_a is given by the k largest eigenvectors of an equivalent eigensystem (W_{eq}, D_{eq}):

$$W_{eq} = W + D_{W_-}, \ \ D_{eq} = D_W,$$

where D_W is the diagonal degree matrix for any matrix W: $D_W(i, i) = \sum_j |W(i, j)|$.

With the introduction of nondirectional repulsion, we remove the nonnegative constraints on weight matrices and

thus graph partitioning algorithms can directly apply to any symmetrical matrices. This opens up new possibilities of encoding negative correlations among constraints in optimization problems, e.g. those formulated in an energy function on Markov Random Fields [11, 5], whereby global optimal solutions through simple and robust eigendecompositions become available.

2.5. Regularization

The above decomposition of W into an attraction field and a repulsion field is not unique. In fact,

$$W = (W_+ + \Delta) - (\Delta - W_-) = A_{Ga} - A_{Gr},$$

where Δ could be any nonnegative matrix. The previous case corresponds to $\Delta = 0$. If we interpret W using $A_{Ga} = (W_+ + \Delta)$ and $A_{Gr} = (\Delta - W_-)$, the partitioning is then given by the eigensystem

$$(W_{eq} + D_\Delta, \ D_{eq} + 2D_\Delta).$$

We see that the only effect of Δ on the solution is the matrix D_Δ. As D_Δ increases its magnitude, the first largest eigenvalue approaches 0.5.

This extra degree of freedom provides us with solution regularization. Practical experiences have indicated that when D_{eq} have near-zero connection degrees for some nodes, the segmentation computed by (W_{eq}, D_{eq}) becomes highly unstable. This situation occurs when a coherent figure is embedded in a random ground. In the attraction case, this problem can be remedied by the addition of a small constant baseline pairwise connections. However, such a technique lacks any theoretical justification and the resulted solutions are no longer the same. In our current framework, we can in fact introduce any constant baseline pairwise connection to the attraction matrix W_+, and its effect can be cancelled by the same addition to the repulsion matrix $-W_-$. In other words, we choose $D_\Delta = \delta I$, where δ is a scalar and I is the identity matrix. This does not change the information contained in the variation of the original graph weights, but as it increases the degree of total connections, it regularizes the solutions in eigendecomposition.

2.6 Conditions for popout

To help understanding repulsion measures and regularization process in this computational framework, we study a simple case of 4-node graph (Fig. 3). Let

$$
A_G = \begin{bmatrix} 0 & x & y & y \\ x & 0 & y & y \\ y & y & 0 & z \\ y & y & z & 0 \end{bmatrix}, V = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},
$$

4

where x, y, z denote figure-to-figure, figure-to-ground, ground-to-ground connections [1, 2] respectively; each column of V gives a labeling of the graph. Due to the weight symmetry, we only need to consider the four cases in V, all others leading to one of the four ϵ_a values. We analyze what are the conditions on x, y and z so that figure-ground as $\Gamma_G^2 = \{\{1, 2\}, \{3, 4\}\}$ can be guaranteed.

Since scaling on A_G does not change grouping, we assume z of unit affinity. The feasible set of x and y can be found (after a lengthy derivation!) by requiring ϵ_a for the first column of V having the largest value over the other three cases (Table 1a) . The change of feasible sets with regularization is illustrated in Fig. 4. The closed-form feasible solutions are also given for the limit case where regularization parameter δ approaches infinity (Table 1b).

	z	y	x
a)	1	$(-\infty,\ 0)$	$(1 - y - \sqrt{1 - 2y + 9y^2},\ +\infty)$
	1	$[0,\ 1]$	$(\frac{2y^2}{1+y},\ +\infty)$
	1	$(1,\ +\infty)$	$(-y + 2y^2,\ +\infty)$
	-1	$(-\infty,\ -1)$	$(\frac{-2y^2}{1-y},\ \frac{-1+2y+8y^2}{2})$
	-1	$[-1,\ -\frac{1}{2}]$	$(-y - 2y^2,\ \frac{-1+2y+8y^2}{2})$

	z	y	x
b)	1	$(-\infty,\ 0)$	$(-1 + 2y,\ +\infty)$
	1	$[0,\ 1]$	$(\max(0, \frac{-3+24y}{21}),\ +\infty)$
	1	$(1,\ +\infty)$	$(0,\ +\infty)$
	-1	$(-\infty,\ -1)$	$(\frac{3+24y}{21},\ \infty)$
	-1	$[-1,\ -\frac{7}{8}]$	$(\frac{-3+5y+2y^2}{2},\ 0)$
	-1	$[0,\ \infty)$	$(1 + 2y,\ \infty)$

Table 1: Feasible sets of parameters for Fig. 3. a) No regularization: $\delta = 0$. b) Regularization at a limit: $\delta = \infty$.

The effects of both repulsion measures and regularization are evident in Fig. 4, which can be summarized in the Table 2. We see that 1) with negative figure-ground connections such as those in figures defined by local feature contrast, repulsion measures allow weak figure-figure connections for an object to be segmented, while 2) with negative ground-ground connections such as those in fragmented or incoherent background cases, regularization allows moderately coherent foreground to stand out.

The problem of fragmented background has led [21] to adopt an unbalanced criterion which favors figure (but not ground) being coherent. However, an unbalanced criterion tends to pick out small local clusters and thus miss global structures. Here we show that the same goal can be achieved with a balance criterion in the attraction-repulsion framework. We demonstrate these effects in the results section.
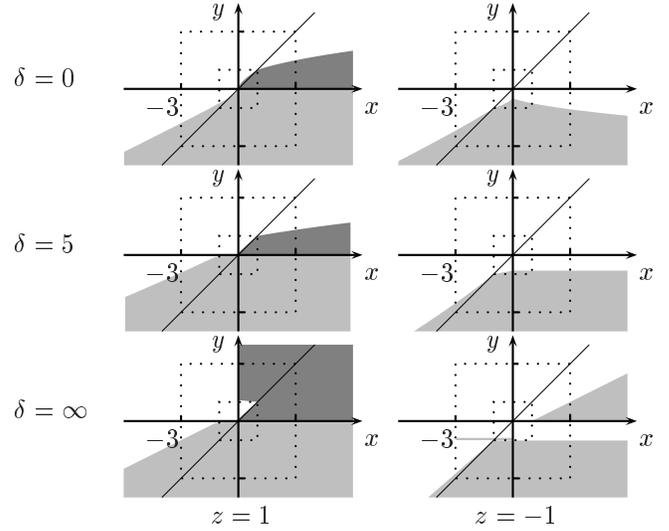

Figure 4: Repulsion and regularization help figural popout. Here x, y and z are figure-figure, figure-ground and ground-ground affinity. The shaded areas indicate feasible sets for figural popout. The darker areas are attraction alone cases. When $z = 1$, the ground is made of similar elements. When y gains its strength as attraction, x has to increase rapidly (quadratic). However, if y is repulsion, x can be even more repulsive than y. Therefore, with attraction, only coherent figures can popout, while with repulsion, even incoherent figures can popout. When $z = -1$, the ground is incoherent. If y is attraction, any coherent figure ($x > 0$) will not popout. If y is repulsion, then it will popout even if $x < y$. With regularization, measured by δ, the solution space in general expands. In particular, a sufficiently coherent figure (with linear $x - y$ relationship) can popout from random ground, which would be otherwise impossible.

Figure \ ground	coherent	incoherent
coherent	Attraction	Regularization
incoherent	Repulsion	No figure-ground

Table 2: Popout through normalized cuts criteria on a weight matrix with negative weights illustrates distinct major contributions of attraction, repulsion and regularization to various figure-ground combinations. Attraction is most effective at detecting a coherent figure against a coherent ground. With repulsion, dissimilar figural elements against a common ground popout. With regularization, a coherent figure can be segmented out from a random ground.

5

3. Jump-diffusion process view

With attraction alone, the normalized cuts criterion can be viewed as finding low conductivity sets in Markov random walks [16]. When we have attraction and repulsion, the goal of partitioning is to find low conductivity sets in a stochastic jump-diffusion process, where the walk jumps between two Markov chains with a probability determined by the relative strengths of total attraction and total repulsion.

Let's define two probability transition matrices, $P_a = D_{Ga}^{-1}A_{Ga}$ and $P_r = D_{Gr}^{-1}A_{Gr}$. $P_a(i,j)$ describes the one-step likelihood of a random walker A walking from i *toward* j based on attraction measures, while $P_r(i,j)$ describes the one-step likelihood of a random walker R walking from i *away from* j based on repulsion measures. The two random walkers live in an identical state space. If both Markov chains are ergodic, which happens under mild conditions and means that the random walker can go between any two states, then there exist unique stationary distributions $\pi_a = \frac{1^T D_{Ga}}{1^T D_{Ga} 1}$ and $\pi_r = \frac{1^T D_{Gr}}{1^T D_{Gr} 1}$. Here we have followed the convention of representing stationary distribution π in row vectors such that $\pi P = \pi$ for transition probability matrix P. Since A_{Ga} and A_{Gr} are symmetric, the two random walks are reversible, i.e.,

$$\pi_a(i)P_a(i,j) = \pi_a(j)P_a(j,i) = \frac{A_{Ga}(i,j)}{1^T D_{Ga} 1},$$

$$\pi_r(i)P_r(i,j) = \pi_r(j)P_r(j,i) = \frac{A_{Gr}(i,j)}{1^T D_{Gr} 1}.$$

Let $\omega(t)$ denote the state of a mind-jumping random walker at time t, where the change of the mind is governed by a two-choice walker identity variable, denoted by $H(t)$. $H(t)$ has a homogeneous distribution over attraction walker A and repulsion walker R, i.e.,

$$Pr(H(t) = A) = \frac{1^T D_{Ga} 1}{1^T D 1} = \frac{\mathcal{A}(\mathsf{V},\mathsf{V})}{\mathcal{A}(\mathsf{V},\mathsf{V}) + \mathcal{R}(\mathsf{V},\mathsf{V})},$$

$$Pr(H(t) = R) = \frac{1^T D_{Gr} 1}{1^T D 1} = \frac{\mathcal{R}(\mathsf{V},\mathsf{V})}{\mathcal{A}(\mathsf{V},\mathsf{V}) + \mathcal{R}(\mathsf{V},\mathsf{V})}.$$

We can interpret our criterion in terms of a stochastic jump-diffusion process [9]. When $H(t) = A$, it acts as random walker A in its equilibrium π_a; When $H(t) = R$, it acts as random walker R in its equilibrium π_r. There is diffusion within each random walk space:

$$Pr(\omega(t+1) \in \mathsf{V}_l | H(t) = A, \omega(t) \in \mathsf{V}_l) = \frac{\mathcal{A}(\mathsf{V}_l,\mathsf{V}_l)}{\mathcal{A}(\mathsf{V}_l,\mathsf{V})},$$

$$Pr(\omega(t+1) \in \mathsf{V} \setminus \mathsf{V}_l | H(t) = R, \omega(t) \in \mathsf{V}_l) = \frac{\mathcal{R}(\mathsf{V}_l,\mathsf{V}\setminus\mathsf{V}_l)}{\mathcal{R}(\mathsf{V}_l,\mathsf{V})}$$

The probability of the jump-walker staying in one set of states is its average staying time with respect to H:

$$Pr(\omega(t) \in \mathsf{V}_l) = E_{H(t)}(\omega(t) \in \mathsf{V}_l) = \frac{\mathcal{A}(\mathsf{V}_l,\mathsf{V}) + \mathcal{R}(\mathsf{V}_l,\mathsf{V})}{\mathcal{A}(\mathsf{V},\mathsf{V}) + \mathcal{R}(\mathsf{V},\mathsf{V})}.$$

Putting these equations together, we can parse the normalized association criterion below:

$$\begin{aligned} \epsilon_a &= \sum_{l=1}^{k} Pr(\omega(t+1) \in \mathsf{V}_l, H(t) = A | \omega(t) \in \mathsf{V}_l) \\ &+ \sum_{l=1}^{k} Pr(\omega(t+1) \in \mathsf{V} \setminus \mathsf{V}_l, H(t) = R | \omega(t) \in \mathsf{V}_l). \end{aligned}$$

The solution ω lies in a mixture space which consists of two identical subspaces $\mathsf{V} \cup \mathsf{V}$. If the normalized association ϵ_a is large for a k-way vertex partitioning $\{\mathsf{V}_l, l = 1, \ldots, k\}$ on graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$, then it means that once the walk is in V_l, the average of the probabilities of staying in it as the attraction walker and those of departing from it as the repulsion walker are large. Intuitively, we look for partitions that once the walk enters one of the parts, it tends to remain in it, due to strong attraction and weak repulsion from the inside. This probabilistic view provides a framework for learning parameters in feature integration.

4. Results

To calculate the affinity between two $d-$dimensional features, we use a Mexican hat function of their difference. It is implemented as the difference of two Gaussian functions:

$$\begin{aligned} h(X; \Sigma_1, \Sigma_2) &= g(X; 0, \Sigma_1) - g(X; 0, \Sigma_2), \\ g(X; \mu, \Sigma) &= \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)}, \end{aligned}$$

where Σ's are $d \times d$ covariance matrices. The evaluation signals pairwise attraction if positive, repulsion if negative and neutral if zero. Assuming $\Sigma_2 = \beta^2 \Sigma_1$, we can calculate two critical radii, r_0, where affinity changes from attraction to repulsion and r_-, where affinity is maximum repulsion:

$$\begin{aligned} r_0(\beta, d) &= \sqrt{\frac{2d \ln(\beta)}{1 - \beta^{-2}}}, \\ r_-(\beta, d) &= \sqrt{2 + d} \cdot r_0(\beta, d). \end{aligned}$$

The case of $d = 1$ is illustrated in Fig. 5. With this simple change from Gaussian functions [25, 17, 21] measuring attraction to Mexican hat functions measuring both attraction and repulsion, we will show that negative weights play a very unique and effective role in graph partitioning.

We compute pairwise affinity in a local neighborhood of any pixel. We denote the neighborhood radius by r. The larger the radius, the more computation is needed and larger structures can be discovered by grouping.

Fig. 6 shows that repulsion has the advantages of computational efficiency and binding objects against one common ground. For an image of two rectangles with equal intensity against ground. With $r = 1$, incorporating repulsion measures by $\beta = 5$, which leads to only 3% negative weights among evaluated affinity, can effectively segment both rectangles out of the ground. The uniform grouping valuation of the repulsion result on both figure and ground is in sharp contrast with attraction results. For attraction, since
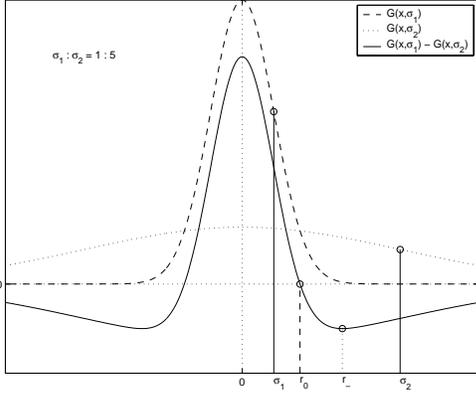
Figure 5: Calculate pairwise affinity using Mexican hat functions based on difference of Gaussians. When two features are identical, it has maximum attraction; when feature difference is r_0, it is neutral; when feature difference is r_-, it has maximum repulsion.
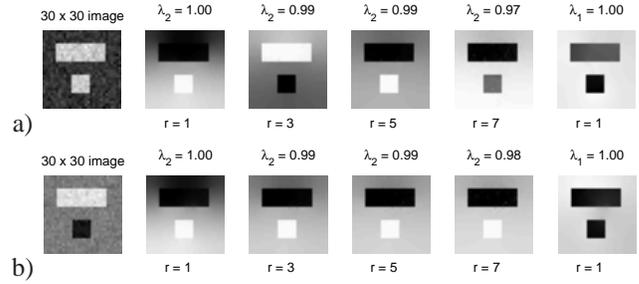

Figure 6: Repulsion has the advantages of computational efficiency and binding objects against one common ground. We choose $\sigma_1 = 0.1$ for pairwise intensity difference so that there is high figure-figure connections but weak figure-ground connections, which become negative with $\beta = 5$. The normalized cuts results are shown next to the images, with corresponding eigenvalues on top and neighborhood radii on bottom. The first four are the results with attraction measured by $g(X; 0, \sigma_1)$ confined to progressively larger r. The rightmost is the result with repulsion measured by $h(X; \sigma_1, 5\sigma_1)$. a) The image has two rectangles of equal average intensity 0.8 against background of 0.5, added by Gaussian noise with standard deviation 0.03. A much larger neighborhood size is needed for attraction to achieve a comparable result with repulsion. b) The smaller object now has an average intensity of 0.2. The results show that attraction encodes feature similarity and repulsion encodes feature contrast and they play very different roles in grouping.
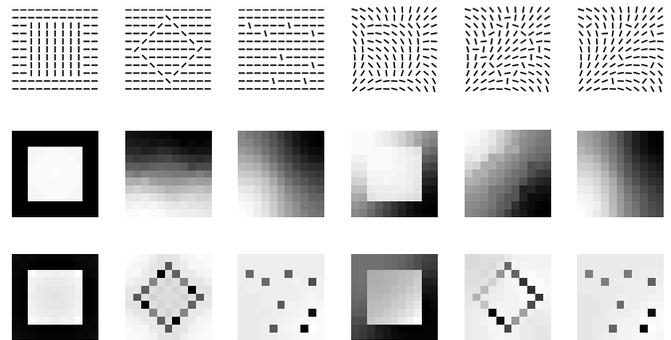
zero affinity could mean either evaluated high dissimilarity or out of evaluation neighborhood, the result with $r = 1$ has graded valuation over the ground and unclear boundary over the larger rectangle. With larger r, the meaning of zero connection becomes disambiguated and both objects come out as different groups until $r = 7$. If we make the rectangles have opposite intensity polarities, attraction cannot possibly group the two because they are more different in their intensity values than each from the ground, whereas repulsion captures local feature contrast and thus they are readily grouped together and in turn lead to sharper contrast between figure and ground.

Fig. 7 are results on bar configurations used in the introduction (Fig. 1 and 2). As can be expected, attraction is good at detecting groups of interior coherence, but poor at salience detection, especially when figure-figure connections are weak. When ground-ground connections are attraction (Fig. 4), repulsion between figure-ground can greatly reduces the pressures on figural coherence, thus dissimilar elements can be grouped together easily.

Fig. 8 shows the case of coherent figures against a random ground. Since ground-ground connections are weak, any coherent figures cannot popout without regularization. Adding a small baseline connection to each weight results in figural popout, whereas too much regularization overwhelms useful information and the results get worse.

Finally, Fig. 9 shows that the asymmetry in visual search can be accounted by the asymmetry in ground-ground connections resulted from contextual influence (collinearity). The figure-ground connections are comparable, but ground-ground connections are weaker for vertical bars among $45°$ bars, which leads to smaller figure-ground contrast.

Figure 7: Pre-attentive segmentation on line segments (first row). Row 2 and 3 are results by attraction and repulsion respectively, $\sigma_1 = 30°$ for orientation, $\sigma_2 = 10$ for distance, $r = 2$ and $\beta = 2$.
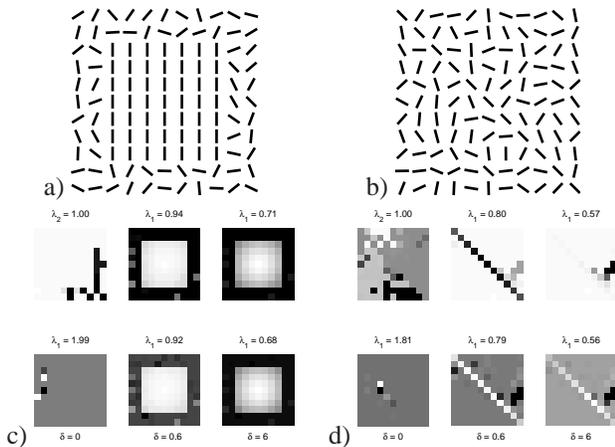
Figure 8: Regularization helps coherent figures to popout from a random ground. a) Region. b) Contour. Their results in c) and d) are organized row-wise for attraction and repulsion, with $\sigma_1 = 5°$ for orientation, $\sigma_2 = 5$ for distance, $r = 2$ and $\beta = 2$. Across the columns varies regularization constant δ, the last of which shows that regularization becomes saturated as evidenced by eigenvalues approaching 0.5. Small amount of regularization helps popout.
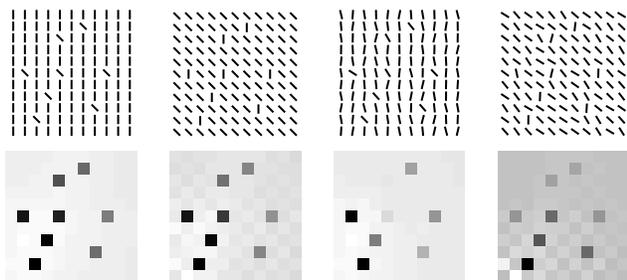

Figure 9: Asymmetry in visual search. The images contains vertical and $45°$ bars. The first two are noiseless conditions, the last two are added with the same $\pm 15°$ noise field. The asymmetry is reflected in different figure-ground contrast.

5. Conclusions

We develop a computational grouping method with dual procedures of association by attraction and segregation by repulsion. Within this framework, we provide a theoretical ground for solution regularization in normalized cuts algorithms. We show that our criterion can be viewed as finding low conductivity sets in a jump-diffusion process.

With attraction measuring feature coherence, with repulsion measuring local feature contrast, with regularization improving signal-to-noise ratios, we show that all popout phenomena can be modeled with a balanced criterion. The conditions on element affinity for popout are derived.

We expand graph partitioning to weight matrices with negative weights, which are shown to lead to computational efficiency. This approach provides a representation for negative correlations in constraint satisfaction problems and simple solutions to such formulations can thus be possible.

Acknowledgements

References

[1] A. Amir and M. Lindenbaum. Ground from figure discrimination. In *ICCV*, pages 521–7, 1998.

[2] A. Amir and M. Lindenbaum. Grouping-based nonadditive verification. *PAMI*, 20(2):186–92, 1998.

[3] J. Beck. Textural segmentation. In J. Beck, editor, *Organization and representation in perception*, pages 285–317. Hillsdale, NJ: Erlbaum, 1982.

[4] J. R. Bergen and E. H. Adelson. Early vision and texture perception. *Nature*, 333:363–4, 1988.

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *ICCV*, 1999.

[6] D. C. V. Essen. Functional organization of primate visual cortex. In *Cerebral Cortex*, pages 259–329. 1985.

[7] Y. Gdalyahu, D. Weinshall, and M. Werman. A randomized algorithm for pairwise clustering. In *Neural Information Processing Systems*, pages 424–30, 1998.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI*, 6(6):721–41, 1984.

[9] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistics Society, Series B*, 56(4):549–603, 1994.

[10] G. Guy and G. Medioni. Inferring global perceptual contours from local features. *International Journal of Computer Vision*, pages 113–133, 1996.

[11] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *CVPR*, 1998.

[12] B. Julesz. A brief outline of the texton theory of human vision. 7:41–5, 1984.

[13] B. Julesz. Texton gradients: the texton theory revisited. *Biological Cybernetics*, 54:245–51, 1986.

[14] Z. Li. Contextual influences in v1 as a basis for pop out and asymmetry in visual search. *Proceedings of National Academy of Science*, 96:10530–5, 1999.

[15] Z. Li. Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, 13(1):25–50, 2000.

[16] M. Maila and J. Shi. Learning segmentation with random walk. In *Neural Information Processing Systems*, 2001.

[17] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 2001.

[18] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of Optical Society of America*, 7:923–32, 1990.

[19] H.-C. Nothdurft. The role of features in preattentive vision: comparison of orientation, motion and color cues. *Vision Research*, 33(14):1937–58, 1993.

[20] H.-C. Nothdurft. Different approaches to the coding of visual segmentation. In L. Harris and M. Kjenkius, editors, *Computational and psychophysical mechanisms of visual coding*. Cambridge University Press, New York, 1997.

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

[22] J. Puzicha, T. Hofmann, and J. Buhmann. Unsupervised texture segmentation in a deterministic annealing framework. *PAMI*, 20(8):803–18, 1998.

[23] D. Sagi and B. Julesz. Short-range limitations on detection of feature differences. *Spatial Vision*, 2:39–49, 1987.

[24] E. Sharon, A. Brandt, and R. Basri. Fast multiscale image segmentation. In *CVPR*, pages 70–7, 2000.

[25] J. Shi and J. Malik. Normalized cuts and image segmentation. In *CVPR*, pages 731–7, June 1997.

[26] A. Treisman. Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, 31:156–77, 1985.

[27] M. Wertheimer. Laws of organization in perceptual forms (partial translation). In W. B. Ellis, editor, *A sourcebook of Gestalt Psychology*, pages 71–88. Harcourt Brace and company, 1938.

[28] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *PAMI*, 11:1101–13, 1993.

[29] S. X. Yu and J. Shi. Segmentation with pairwise attraction and repulsion. In *ICCV*, 2001.

[30] S. C. Zhu, Y. N. Wu, and D. Mumford. Filters, random field and maximum entropy: — towards a unified theory for texture modeling. *International Journal of Computer Vision*, 27(2):1–20, 1998.

[31] J. Y. Zien, M. D. F. Schlag, and P. K. Chan. Multilevel spectral hypergraph partitioning with arbitrary vertex sizes. *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, 18(9):1389–99, 1999.