Understanding Popout through 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
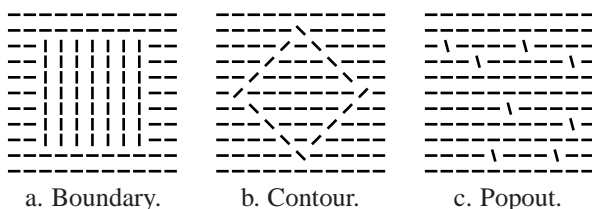
{stella.yu, jshi}@cs.cmu.edu

Abstract

Perceptual popout is defined by both feature similarity and local feature contrast. We identify these two measures with attraction and repulsion, and unify the dual processes of association by attraction and segregation by repulsion in a single grouping framework. We generalize normalized cuts to multi-way partitioning with these dual measures. We expand graph partitioning approaches to weight matrices with negative entries, and provide a theoretical basis for solution regularization in such algorithms. We show that attraction, repulsion and regularization each contributes in a unique way to popout. Their roles are demonstrated in various salience detection and visual search scenarios. This work 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 object 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 [12] 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 popout targets against backgrounds (Fig. 1).
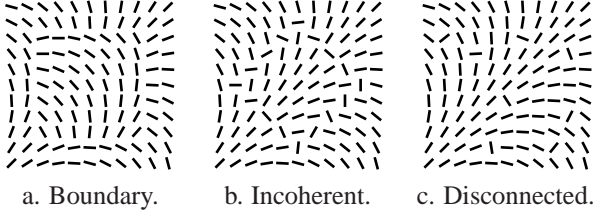
It has long been assumed that regions are first characterized by features which are homogeneous within the areas. These feature measures are then compared at neighbouring locations to locate boundaries between regions [12]. This view of feature discrimination for grouping is supported

Figure 1. Pre-attentive segmentation is to mark conspicuous image locations, which could be caused by a) region boundaries, b) smooth contours and c) popout 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.

by evidence in neurophysiology on elaborate feature detectors in visual cortex [6], in psychophysics on visual search [22] and in modeling on texture segmentation [10, 4, 14]. Some other approaches of texture segmentation go beyond the analysis of features obtained from image filters and model the interactions between filters [24]. These Markov Random Field models [8] capture context dependences and other statistical characteristics of texture features [12].

However, it has been shown [3, 11, 19, 15] that when feature similarity within an area and feature differences between areas are teased apart, the two aspects of perceptual organization, association and segregation, can contribute independently to grouping. 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 [15]. These results motivate models of preattentive vision where region boundaries are *directly* localized through lateral interactions between edge detectors [16, 12].

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.

Such contextual feature analysis for grouping can be 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. Gestalt grouping factors, such as proximity, similarity, continuity and symmetry, can be encoded and combined in pairwise similarity measures [23, 21, 18, 7, 20]. While Gestalt laws have always stressed the aspects of similarity of elements in grouping, the effect of saliency by local feature contrast cannot be described in a framework that models similarity grouping. Fig. 2c shows that, completely dissimilar elements that are spatially disconnected can be perceived as a figure simply because they are locally dissimilar to a common ground.

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 the normalized cuts criteria [21] to a multi-way partitioning on these dual measures. We derive the necessary and sufficient conditions on the graph weights for objects to be segmented in a variety of settings. We demonstrate that both salience detection and the asymmetry in visual search 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 presents experimental results. Section 4 concludes the paper.
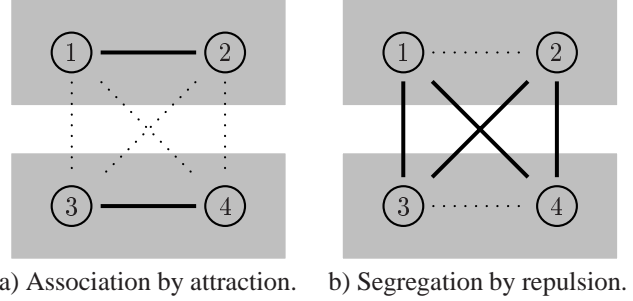
2. Method

In graph approaches, an image is described by an undirected 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 i and j is associated with weight W_{ij}. W is often assumed to be symmetric, $W = W^T$. Let the degree matrix D_W be a diagonal matrix, where $D_W(i, i) = \sum_j W(i, j)$, $\forall i$. In our model, we have two *nonnegative* weight matrices,

A and R, for attraction and repulsion respectively.

2.1. Criteria

For attraction, we desire the within-group association to be as large as possible; whereas for repulsion, we desire the between-group segregation to be as large as possible (Fig. 3). We extend the bipartitioning in normalized cuts [21] to a k-way partitioning based on both attraction and repulsion. A k-way vertex partitioning $\{V_l, l = 1, \ldots, k\}$ on graph $G = (V, E)$ has $V = \cup_{l=1}^k V_l$ and $V_i \cap V_j = \varnothing, \forall i \neq j$.

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, while minimizing those of dotted-lined weights. 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 weighted by their relative strengths of total connections.

A formal description of the criteria is as follows. Given nonnegative weight matrix W, vertex sets P and Q, let $\mathcal{C}_W(P, Q)$ denote the total W connections from P to Q, $\mathcal{D}_W(P)$ denotes the degree of W connections of P, and $\mathcal{S}_W(P, Q)$ denotes the connection ratio from set P to Q:

$$
\begin{aligned}
\mathcal{C}_W(P, Q) &= \sum_{j \in P, k \in Q} W(j, k), \\
\mathcal{D}_W(P) &= \mathcal{C}_W(P, V), \\
\mathcal{S}_W(P, Q) &= \frac{\mathcal{C}_W(P, Q)}{\mathcal{D}_W(P)}.
\end{aligned}
$$

Note that $\mathcal{S}_W(P, P) + \mathcal{S}_W(P, V \setminus P) = 1$. For each of the k partitions, we measure the goodness of association by its within-group attraction ratio and the goodness of segregation by its between-group repulsion ratio, resulting in the goodness of grouping as the linear summation of the two ratios, weighted by their relative strengths:

$$
\begin{aligned}
\epsilon_a(\Gamma_G^k) &= \sum_{l=1}^k \left[\frac{\mathcal{S}_A(V_l, V_l) \mathcal{D}_A(V_l)}{\mathcal{D}_A(V_l) + \mathcal{D}_R(V_l)} + \frac{\mathcal{S}_R(V_l, V \setminus V_l) \mathcal{D}_R(V_l)}{\mathcal{D}_A(V_l) + \mathcal{D}_R(V_l)} \right], \\
\epsilon_c(\Gamma_G^k) &= \sum_{l=1}^k \left[\frac{\mathcal{S}_A(V_l, V \setminus V_l) \mathcal{D}_A(V_l)}{\mathcal{D}_A(V_l) + \mathcal{D}_R(V_l)} + \frac{\mathcal{S}_R(V_l, V_l) \mathcal{D}_R(V_l)}{\mathcal{D}_A(V_l) + \mathcal{D}_R(V_l)} \right].
\end{aligned}
$$

A good partitioning maximizes ϵ_a while minimizes ϵ_c. Since $\epsilon_a + \epsilon_c = k$, both goals can be satisfied simultaneously. 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.

2.2. Computational solution

We introduce a few symbols to turn the above criterion into a computable form. Let X_l be a binary membership indicator vector for group l, $l = 1, \ldots, k$, which assumes the value of 1 if a node belongs to this group and 0 otherwise. Let

$$
\begin{aligned}
W_{eq} &= A - R + D_R, \\
D_{eq} &= D_A + D_R.
\end{aligned}
$$

It can be verified that

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

where $Y = X(X^T D_{eq} X)^{-\frac{1}{2}}$ 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, in the form of k column vectors of Y, is given by the k largest generalized eigenvectors of (W_{eq}, D_{eq}). Based on Gershgorin's theorem, we can estimate $|\lambda_l(W_{eq}, D_{eq})| \leq 2, \forall l$, where $\lambda_l(W_{eq}, D_{eq})$ is the l-th largest eigenvalue of (W_{eq}, D_{eq}).

Let's examine two extreme cases. If there is no repulsion, we have $W_{eq} = A$ and $D_{eq} = D_A$, which reduces to the conventional normalized cuts [21], where the all-one vector 1 is the eigenvector of (W_{eq}, D_{eq}) with the largest eigenvalue of 1. If there is no attraction, we have $W_{eq} = D_R - R$ and $D = D_R$, where 1 is the eigenvector of (W_{eq}, D_{eq}) 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_{eq}, D_{eq}). Indeed, attraction tends to bind elements together, while repulsion tends to break elements apart. The optimal partitioning results from the balance of these two forces.

2.3. Partitioning with a symmetrical matrix

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

$$
W = W_+ - (W_-) = A - R,
$$

where W_+ and W_- contain the absolute values of all non-negative and negative entries of W respectively. We regard W_+ as the attraction of A and W_- as the repulsion of R, and interpret the normalized cuts on A and R as that on W. The eigensystem (W_{eq}, D_{eq}) is thus

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

With the introduction of repulsion, we no longer require weight matrices to be nonnegative in graph partitioning algorithms. This opens up the possibilities of encoding negative correlations among constraints in optimization problems, e.g. those formulated in an energy function on Markov Random Fields [9, 5], whereby globally optimal solutions through simple and robust eigendecompositions become available.

2.4. Regularization

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

$$
W = (W_+ + \Delta) - (W_- + \Delta) = A - R,
$$

where Δ could be any nonnegative matrix. The previous case corresponds to $\Delta = 0$. If we interpret W using $A = W_+ + \Delta$ and $R = W_- + \Delta$, 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} has near-zero values 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 connection weight. However, such a technique lacks any theoretical justification and alters the measurements of pairwise affinity. In our current framework, we can introduce any constant baseline connection to the attraction matrix W_+, and its effect is cancelled by its 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. However, as the degree of total connections for each node increases, the grouping sensitivity to nodes of sparse connections is alleviated. The regularized solutions can thus reveal the underlying stable structures.

2.5 Conditions for popout

To aid our understanding of repulsion and regularization, we study a simple 4-node graph (Fig. 3). Let

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

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 determine the conditions on x, y and z such that figure-ground as given by $\Gamma_G^2 = \{\{1, 2\}, \{3, 4\}\}$ can be guaranteed.

Since scaling W does not change the grouping, we assume $z = 1$. The feasible set of x and y can be found (after a lengthy derivation!) by requiring ϵ_a for the first column of V to be larger than that for any other column. The closed-form feasible solutions are given in Table 2. How feasible sets change with regularization is illustrated in Fig. 4.

As shown in Fig. 4, repulsion and regularization expand the regions of affinity strengths that cause correct grouping. These effects are summarized in Table 1: with negative figure-ground connections such as those in figures defined by local feature contrast, repulsion allows an object of weak figure-figure connections to be segmented, while with negative ground-ground connections such as those in fragmented or incoherent background cases, regularization allows moderately coherent foreground to stand out.

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

Table 1. Popout through the 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 popout from a random ground.
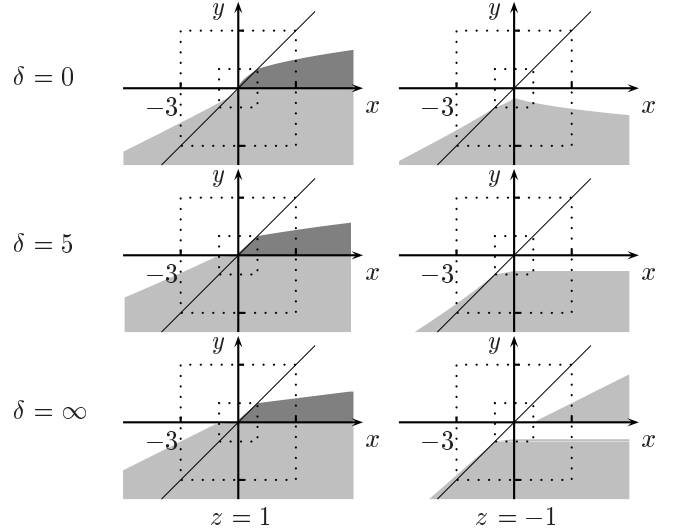
The problem of fragmented background has led [17] to adopt an unbalanced criterion which emphasizes figural (but not ground) coherence. However, an unbalanced criterion tends to favor small local clusters and thus miss global structures. Here we show that the same goal can be achieved with a balanced criterion in the attraction-repulsion framework. We demonstrate these effects in the results section.

a)

z	y	x
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	$[-1, -\frac{1}{2}]$	$(-y - 2y^2, \frac{-1+2y+8y^2}{2})$

b)

z	y	x
1	$(-\infty, 0)$	$(-1 + 2y, +\infty)$
1	$[0, 1]$	$(\max(0, \frac{-1+8y}{7}), +\infty)$
1	$(1, +\infty)$	$(-7 + 8y, +\infty)$
-1	$(-\infty, -1)$	$(\frac{1+8y}{7}, +\infty)$
-1	$[-1, -\frac{7}{8}]$	$(7 + 8y, +\infty)$
-1	$[0, +\infty)$	$(1 + 2y, +\infty)$

Table 2. Feasible sets of parameters for Fig. 3. a) No regularization: $\delta = 0$. b) Regularization at a limit: $\delta = \infty$.

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 regions for figural popout. The darker areas are those with attraction alone. When $z = 1$, the ground is made of similar elements. When $y > 0$, x has to increase rapidly (quadratic). However, if $y < 0$, 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, no coherent figure ($x > 0$) can popout. If y is repulsion, then a figure pops out 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 a random ground, which would be otherwise impossible.
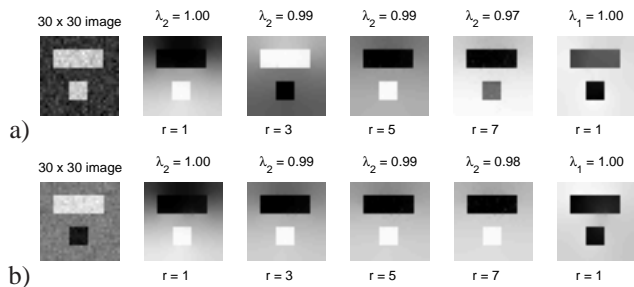
3. Results

To calculate the affinity between two features, we use a Mexican hat function of their difference X:

$$h(X; \Sigma_1, \Sigma_2) = g(X; \Sigma_1) - g(X; \Sigma_2),$$

where $g(X; \Sigma)$ is a Gaussian function with zero mean and covariance Σ. Let $\Sigma_2 = \beta^2 \Sigma_1$. The evaluation signals pairwise attraction if positive, repulsion if negative and neutral if zero. With this simple change from Gaussian functions [21, 13, 17] measuring attraction to Mexican hat functions measuring both attraction and repulsion, we show that repulsion plays a very unique and effective role in grouping.
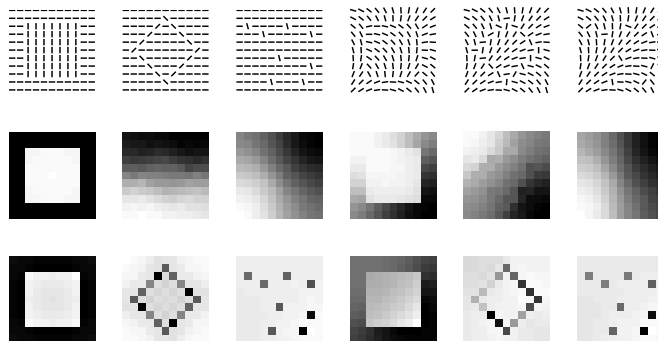
We compute pairwise affinity in a local neighborhood around each pixel. We denote the neighborhood radius by r. Larger radii require more computation, but allow for the discovery of larger structures through our grouping engine.

Fig. 5 shows that repulsion not only binds objects regardless of their contrasts, but also requires fewer local feature comparisons. For attraction, since zero could mean either two pixels are highly dissimilar or they are not neighbours, the result with $r = 1$ has graded valuation over the ground and the larger object. With larger r's, zero attraction becomes disambiguated and both objects come out as different groups until $r = 7$. If the objects have opposite contrasts, attraction cannot possibly group them, whereas repulsion capturing local feature contrast readily unites them against a common enemy.
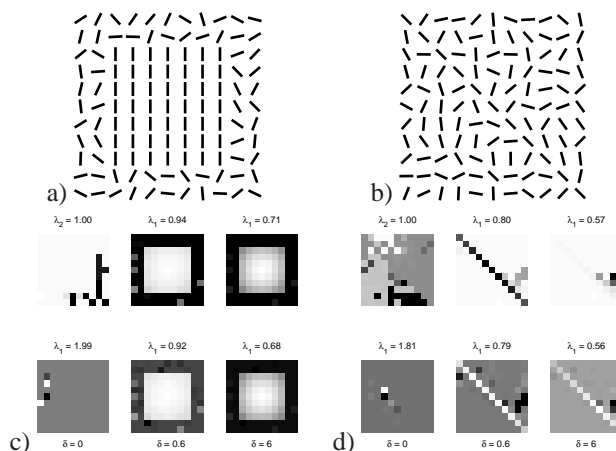

Figure 5. Repulsion can bind multiple objects with less computation. We use $\sigma_1 = 0.1$ to evaluate affinity by intensity, 3% of which become negative with $\beta = 5$. Column 2 to 5 are the results with attraction measured by $g(X; \sigma_1^2)$ with increasing r's. The rightmost is the result with affinity measured by $h(X; \sigma_1^2, 25\sigma_1^2)$. 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. Even with a large r, the two objects cannot be united by attraction.

Fig. 6 are results on bar configurations. Attraction is good for grouping similar elements, but poor at salience detection. When the ground is coherent, repulsion between figure-ground can greatly reduce the pressure on figural coherence, thus dissimilar elements are grouped together.

Figure 6. 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 $\beta = 2$.
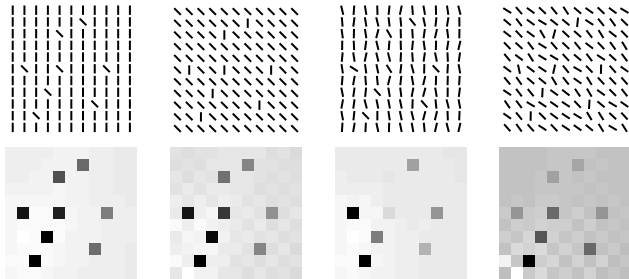
Fig. 7 shows coherent figures against random grounds. Since ground-ground connections are weak, no coherent figure can popout without regularization. Regularization improves the stability of grouping, the global figure-ground organization can thus be discovered.

Figure 7. Regularization helps coherent figures popout from random grounds. a) Region. b) Contour. Results in c) and d) are organized row-wise for attraction and repulsion, with $\sigma_1 = 5°$, $\sigma_2 = 5$, $r = 2$, $\beta = 2$. Across the columns varies the degree of regularization. It becomes saturated as eigenvalue $\lambda \to 0.5$.

Finally, Fig. 8 shows that the asymmetry in visual search can be accounted for by the asymmetry in ground-ground connections resulting from contextual influence (collinearity). The figure-ground connections are comparable, but ground-ground connections are weaker for vertical bars among $45°$ bars, leading to smaller figure-ground contrast.

Figure 8. Asymmetry in visual search. The images contain $0°$ and $45°$ bars. The first two are noiseless conditions, the last two are added with the same $±15°$ noise field. The asymmetry is reflected in different figure-ground contrasts.

4. Conclusions

We developed a grouping method unifying dual procedures of association by attraction and segregation by repulsion. Within this framework, we provided a theoretical basis for solution regularization in the normalized cuts algorithms.

We showed that all popout phenomena can be modelled with a balanced criterion, with attraction measuring feature similarity, repulsion measuring local feature contrast and regularization improving grouping stability.

We expanded graph partitioning to weight matrices with negative values, which provides a representation for negative correlations in constraint satisfaction problems and simple solutions to such formulations are thus possible.

Acknowledgements

References

[1] A. Amir and M. Lindenbaum. Ground from figure discrimination. In *International Conference on Computer Vision*, pages 521–7, 1998.

[2] A. Amir and M. Lindenbaum. Grouping-based nonadditive verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 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 *International Conference on Computer Vision*, 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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–41, 1984.

[9] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.

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

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

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

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

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

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

[16] 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.

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

[18] J. Puzicha, T. Hofmann, and J. Buhmann. Unsupervised texture segmentation in a deterministic annealing framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):803–18, 1998.

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

[20] E. Sharon, A. Brandt, and R. Basri. Fast multiscale image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 70–7, 2000.

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

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

[23] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1101–13, 1993.

[24] 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.