Affinity CNN: Learning Pixel-Centric Pairwise Relations for Figure/Ground Embedding II

Liangjie Cao
4 June 2018

1. Spectral Embedding & Generalized Affinit

Angular Embedding [6] addresses this optimization problem by minimizing error ε:

$$\varepsilon = \sum_{p} \frac{\Sigma_q C(p,q)}{\Sigma_{p,q} C(p,q)} \cdot |z(p) - \bar{z}_0(p)|^2 \tag{1}$$

where C(p,q) accounts for possibly differing confidences in

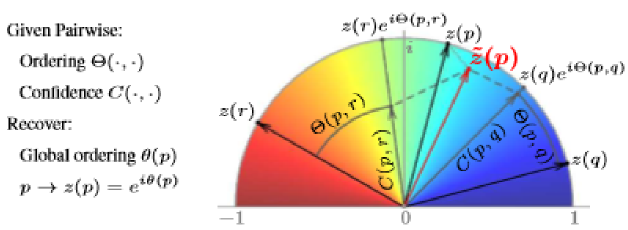

Figure 1. **Angular Embedding [6]**

the pairwise estimates and θ(p) is replaced by z(p)=$e^{i\theta(p)}$. As Figure 1 shows, this mathematical convenience permits interpretation of z(.) as an embedding into the complex plane, with desired ordering θ(.) corresponding to absolute angle. $\bar{z}$(p) is defined as the consensus embedding location for p according to its neighbors and θ:

$$\bar{z}(p) \sum_{p} \bar{C}(p,q) \cdot e^{i\theta(p,q)} \cdot z(q) \tag{2}$$

$$\bar{C}(p,q) = \frac{C(p.q)}{\Sigma_q C(p.q)} \tag{3}$$

Relaxing the unit norm constraint on z(.) yields a generalized eigenproblem:

$$Wz = \lambda Dz \tag{4}$$

with D and W defined in terms of C and θ by:

$$D = Diag(C1_n) \tag{5}$$

$$W = C \cdot e^{i\theta} \tag{6}$$

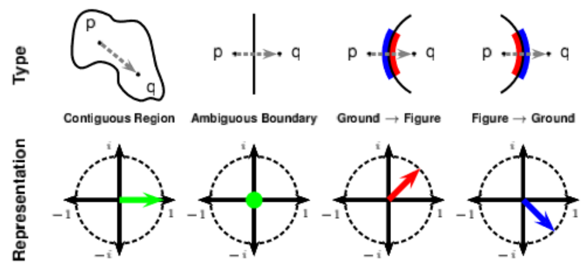

Figure 2. **Complex affinities for grouping and figure/ground.**

where n is the number of pixels, 1n is a column vector of ones, Diag(.) is a matrix with its vector argument on the main diagonal, and "." denotes the matrix Hadamard product.For θ everywhere zero (W = C), this eigenproblem is identical to the spectral relaxation of Normalized Cuts [5], in which the second and higher eigenvectors encode grouping [4, 5]. With nonzero entries in θ, the first of the now complex-valued eigenvectors is nontrivial and its angle encodes rank ordering while the subsequent eigenvectors still encode grouping [2]. They use the same decoding procedure as [1] to read off this information.

Figure 2 illustrates how W_B(shown in green), W_F(red), and W_G(blue) cover the base cases in the space of pairwise grouping relationships. Combining them into a single energy model(generalized affinity) spans the entire space:

$$W(p,q) = W_B(p,q) + W_F(p,q) + W_G(p,q) \tag{7}$$

One can regard W(p,q) as a sum of binding, figure transition, and ground transition forces acting between p and q. Figure 3 plots the onfiguration space of W(p,q) in which each component force is strong do not overlap, W behaves in distinct modes, with a smooth transition between them through the area of weak affinity near the origin.

Learning to predict e(p),b(p,q), and f(p,q) suffices to determine all components of W. For computational efficiency, they predict pairwise relationships between each pixel and its immediate neighbors across multiple spatial scales.

This defines a multiscale sparse W. As an adjustment prior to feeding W to the Angular Embedding solver of [3], they enforce Hermitian symmetry by assigning:

$$W \leftarrow (W + W^*)/2 \tag{8}$$

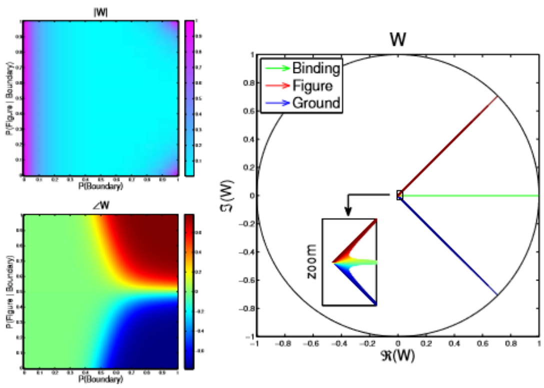

Figure 3. **Generalized affinity**

References

[1] X. He and A. Yuille. Occlusion boundary detection using pseudo-depth. In *ECCV*, 2010.

[2] M. Maire. Simultaneous segmentation and figure/ground organization using angular embedding. In *ECCV*, 2010.

[3] M. Maire and S. X. Yu. Progressive multigrid eigensolvers for multiscale spectral segmentation. In *ICCV*, 2013.

[4] A. P, M. M, F. C, and M. J. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 2011.

[5] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE TPAMI*, 2015.

[6] S. X. Yu. Angular embedding: A robust quadratic criterion. *IEEE TPAMI*, 2012.