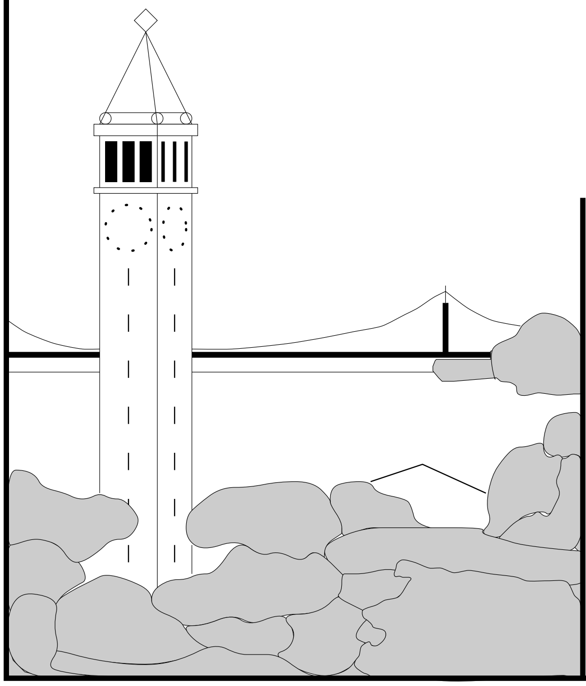
How Much Does Globalization Help Segmentation?

Charless Fowlkes and Jitendra Malik

{fowlkes,malik}@eecs.berkeley.edu



Report No. UCB/CSD-4-1340

July 2004

Computer Science Division (EECS)
University of California
Berkeley, California 94720

Abstract

This paper quantifies the information gained in integrating local measurements using spectral graph partitioning. We employ a large dataset of manually segmented images in order to learn an optimal affinity function between nearby pairs of pixels. Region cues are computed as the similarity in brightness, color, and texture between image patches. Boundary cues are incorporated by looking for the presence of an "intervening contour", a large gradient along a straight line connecting two pixels. We then use spectral clustering to find an approximate minimizer of the normalized cut, partitioning the image into coherent segments.

We evaluate the power of local measurements and global segmentations in predicting the location of image boundaries by computing the precision and recall with respect to the human groundtruth data. The results show that spectral clustering is successful in suppressing noise and boosting weak signals over a wide variety of natural images.

1. Introduction

Research on early vision problems such as edge detection and image segmentation has traditionally been critiqued on the grounds that quantitative measurements of performance are rare. It is therefore difficult to evaluate the effect of different design choices and the superiority (or inferiority) of various novel heuristics that have been proposed in the literature. Recently the availability of the Berkeley Segmentation DataSet [13, 4] has allowed the quantitative measurement of performance on boundary finding [12] and the relative power of various pairwise similarity cues [6]. While this is, of course, not the first example of quantitative measurement in segmentation (see for example [1, 3, 9]) the availability of this large data set containing a wide variety of images and segmentations by multiple human observers (11,0000 segmentations of 1000 images), allows one to draw conclusions with greater "statistical confidence". For example, understanding the relative importance of color and texture in grouping, or the finding [6] that proximity provides no significant additional information about the correct grouping once texture and color have been measured.

In recent years, a major trend in early vision has been the development of approaches to segmentation based on pairwise clustering such as spectral partitioning [26, 27, 17, 23, 24, 16], deterministic annealing [18], or stochastic clustering [7, 14]. Typically some global cost criterion is formulated which is composed of pairwise similarities measured between image pixels or regions. However, the most basic question in this context is, "Do these global formulations help?". The intuition is that local measurements are noisy and integrating them globally might give much better results, but is this actually true? The purpose of this paper is to use ground-truth human segmentations to provide some answers to this question.

Here we have a fundamental challenge, global approaches typically output a "hard" partitioning of the image into disjoint pieces while local measurements are usually some "soft" measurement of the edginess at a pixel or the similarity of two pixels. Our solution is to work in the framework of edge detection and project the much more complicated structures that are segmentations down into this space, finessing round peg into square hole.

A "soft" edge map can be evaluated by choosing a range of thresholds and recording the trade-off between *Precision*, the probability that an above threshold pixel is on a true boundary, and *Recall* the probability that a true boundary pixel is detected. The resulting precision-recall curve characterizes the performance of a detector independently of parameterization and shows the regime in which a particular detector performs well relative to others. In the case of a segmentation, we have only a single point but if the boundaries given by successive segmentations in a sequence $\{S_1, S_2, \ldots\}$ are nested (e.g. S_2 is a strict refinement of S_1) then the sequence index t gives a discrete parameterization and hence a curve of sorts.

In this paper we will focus on evaluating the benefit of globalization in the normalized cut framework [26] using multiple eigenvectors of the normalized Laplacian of the affinity matrix in order to find partitions of the image. The paper is organized as follows. In Section 2, we develop the pairwise cues used for segmenting natural images. In Section 3, we review spectral clustering and discuss how to treat intermediate results as a soft boundary map. Finally, in Section 4 and 5 we evaluate the performance gains in going from raw boundary cues to segmentations. Section 6 concludes.

2. Local to Global using Spectral Clustering

Let the symmetric matrix $W \in \mathbb{R}^{N \times N}$ denote the weighted adjacency matrix for a graph $G = (V, E)$ with vertices V and edges E. For disjoint vertex sets A and B, define $\text{cut}(A, B) = \sum_{i \in A, j \in B} W_{ij}$. Let $\text{vol}(A)$ of a vertex set A be defined as $\text{vol}(A) = \sum_{i \in A} \sum_j W_{ij}$. The k-way normalized cut for a disjoint partition of the vertex set $V = V_1 \sqcup V_2 \sqcup V_3 \sqcup \ldots \sqcup V_k$ is then given by

$$\text{KNCut}(V_1, V_2, V_3, \ldots, V_k) = \sum_{i=1}^{k} \frac{cut(V_i, V)}{vol(V_i)}$$

Shi and Malik [26] showed that for bi-partitioning, an approximate solution may be obtained by scaling and thresholding the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian $\mathcal{L} = D^{-1/2}(D - W)D^{-1/2}$ where D is the diagonal matrix with entries $D_{ii} = \sum_j W_{ij}$. Yu and Shi [22] and Bach and Jordan [2] show that the trailing k eigenvectors of $\mathcal{L}$ similarly provide

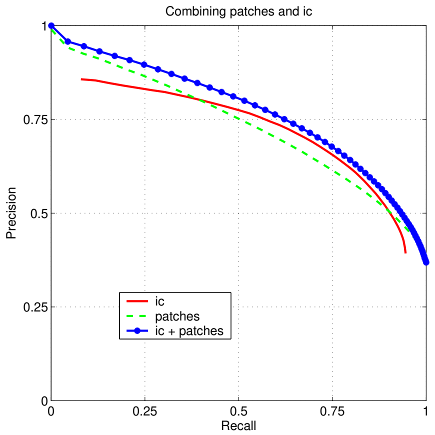


Figure 1: This precision-recall curve shows the performance of three different classifiers on the task of predicting when a pair of pixels lie in the same segment. Combining the intervening contour cue (ic) and the region similarity cue (patches) results in everywhere-better performance. This implies there is independent information in each cue which the classifier is able to exploit.

a relaxation of the k-way criterion. Each element of V is associated with an entry in each of the k eigenvectors giving an embedding of the graph in $\mathbb{R}^k$ (if the columns of E are the eigenvectors of $\mathcal{L}$ then the rows of E are the embedding vectors). The hope is that the distances between these embedded points provide a denoised version of their raw affinities making them much easier to cluster. The question of "how much easier" is exactly what we wish to quantify.

To apply this technique to real images requires developing a local model of similarity between pixels. We assume that the entries in W_{ij} should be a monotonic function of the likelihood that two pixels like in the same segment S_{ij}. We consider the functional form

$$W_{ij} = \exp\left(\frac{-\hat{S}_{ij}}{\sigma}\right) \tag{1}$$

where σ is a free parameter set by cross-validation and S is fit by our local probabilistic model $\hat{S}$ which we describe next.

3. Optimizing Local Cues

For each pair of pixels, we would like to estimate the probability that they belong to the same segment. Based on the work of [6], we use low-level measurements that describe the local neighborhood around each pixel as well as the existence of contours in the image plane separating the two pixels. Our novel contribution here is to adapt the support of local measurements to the estimated boundary map.

The pairwise similarity between the i-th and j-th pixel is given by a function whose arguments are the similarity between the i-th and j-th local descriptors and the gradients along the path from i and j. Both the patch and gradient

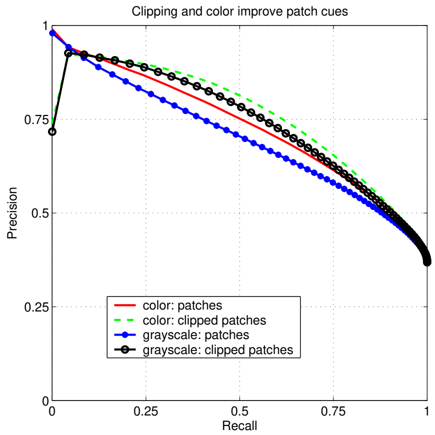


Figure 2: Clipping the support of a patch using the local boundary estimate allows a significant improvement over using isotropic support. This is because patches no longer span locally detectable boundaries and hence are less likely to be "poisoned" by texture and color from more than one segment. This figure also documents benefit of color. Clipping grayscale patches actually allows them perform better than unclipped color patches.

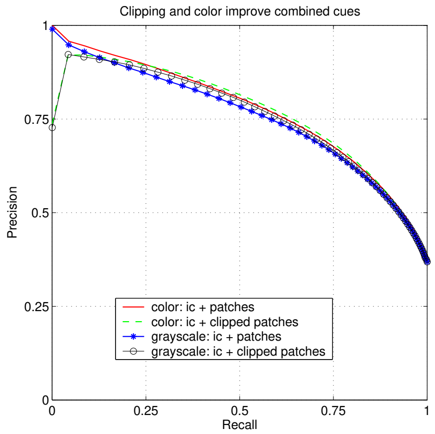


Figure 3: When the intervening contour cue (ic) is included, the jump in performance from clipped to unclipped patches is significantly less but still present.

cues have free parameters such as the the number of histogram bins, scale etc. We choose these parameters using coordinate ascent to optimize performance of each cue individually on the test data. We now describe how the cues are computed.

3.1. Gradient Features

Given a pair of pixels, consider the straight-line path connecting them in the image plane. If the pixels lie in different segments, then we expect to find, somewhere along the line, a photometric discontinuity or *intervening contour* [10]. If no such discontinuity is encountered, then the similarity between the pixels should be large.

In order to compute the intervening contour cue, we require a boundary detector that works robustly on natural images. For this we employ the gradient-based boundary detector of [12] (available here [4]). The output of the detector is a P_b image that provides the posterior probability of a boundary at each pixel for both color and grayscale images. The gradient features used by [12] to predict P_b are based on the histogram difference between the two halves of a single disc, similar to [21, 20]. The orientation of the dividing diagonal sets the orientation of the gradient, and the radius of the disc sets the scale. All of the parameters of the gradients have been tuned by [12] on the same dataset to optimally detect the boundaries marked by the human subjects.

3.2. Patch Features

Given a pair of pixels, we also wish to measure the brightness, color, and texture similarity between circular neighborhoods of some radius centered at each pixel.

We model brightness and color distributions with histograms constructed by binning kernel density estimates of marginals in the 1976 CIE L*a*b* color space. Histograms are compared with the χ^2 histogram difference operator [19]. The χ^2 histogram difference does not make use of the ground distance between the bin centers. Because the distance between points in CIELAB space is perceptually meaningful in a local neighborhood, binning a kernel density estimate whose kernel bandwidth σ matches the scale of this neighborhood means that perceptually similar colors will have similar histogram contributions. Beyond this scale, where colors are incommensurate, χ^2 will regard them as equally different.

For the patch-based texture feature, we compare the distributions of filter responses in the two discs. Our filter bank contains elongated quadrature pair filters—Gaussian second derivatives and their Hilbert transforms—at six orientations, along with a center-surround filter. The empirical distribution of filter responses has been shown to be a powerful feature for both texture synthesis [8] and texture discrimination [18].

We compare filter distributions using the texton approach developed in [11]. The joint distribution of filter responses is estimated using adaptive bins, which are computed with k-means. The texture descriptor for a pixel is therefore a k bin histogram over the texton labels in a disc of radius r centered on the pixel. Here again, comparing histograms using the χ^2 difference measure provided optimal performance on the training data.

Patches computed in this way accurately capture surface appearance as long as the patch center is far from any image boundary. In order to prevent patch histograms from being contaminated when they overlap two or more regions, we consider adapting the patch support to respect image gradients. Clipping patches is accomplished by weighting the histogram contribution of each pixel in the circular window with the intervening contour from the center point.

3.3. Evaluating Pairwise Features

We would like to fit pairwise similarities independently of later processing. To this end, we formulate the problem of optimizing the pixel similarity function as a classification problem of discriminating same-segment pixel pairs from different-segment pairs. Let S_{ij} be the true same-segment indicator so that $S_{ij}{=}1$ when pixels i and j are in the same segment, and $S_{ij}=0$ when pixels i and j are in different segments.

To evaluate proposed pairwise similarities, we use human segmentations for 200 color images as a training set. Each image has been segmented by at least 5 subjects, so the groundtruth S_{ij} is defined by a set of human segmentations. We declare two pixels to lie in the same segment only if all subjects declare them to lie in the same segment.

Given a classifier's estimate $\hat{S}_{ij}$, we compare with the human segmentations using precision-recall. In this task, precision measures the probability that two pixels declared by the classifier to lie in the same segment were marked as such by the humans, i.e. $\mathbb{P}(S_{ij}{=}1|\hat{S}_{ij}{>}t)$. Recall measures the probability that a pair of pixels marked by the humans as lying in the same segment where declared as such by the classifier, i.e. $\mathbb{P}(\hat{S}_{ij}{>}t|S_{ij}{=}1)$.

Figures 1,2,and 3 document the performance of our best pairwise cues for classifying pixel pairs which are within a distance of 20 pixels. In all cases, we fit a simple logistic function to the data. Figure 1 shows that patch and intervening contour cues in combination dominate each cue individually, showing that each contains independent information. Figure 2 shows that clipping patches greatly improves patch performance. Figure 3 shows that the performance gain from clipping when intervening contour is included is less but still present.

4. Evaluating Eigenvectors

Once we have pairwise affinities, we can compute eigenvectors of the normalized Laplacian as described in Section

2. We would like to utilize the embedding given by these eigenvectors to partition the image. Several techniques have been proposed such as thresholding the second eigenvector [25], recursive bi-partitioning, or running k-means on the embedded points [11, 16]. These proposed techniques for rounding the eigenvectors to a discrete solution all rely on points in different clusters being placed "far apart" by the eigenvector embedding.

To evaluate the eigenvectors in a way that is independent of losses or gains due to a particular rounding scheme, we reshape the set of eigenvectors E into a vector-valued image and measure the extent to which boundaries between human marked segments correspond to high gradients. For each pixel, we compute the direction in which the vector-valued image given by the eigenvectors is changing most quickly. The gradient magnitude in a direction θ at a point (x, y)

$$M(x,y,\theta) = \sum_i \left(\sin(\theta)\partial_x E_i(x,y) + \cos(\theta)\partial_y E_i(x,y)\right)^2$$

is a quadratic function of the derivatives of each embedding vector in the x and y directions. The θ for which M is maximal can be found in closed form. Let $g_x = \sum_i (\partial_x E_i)^2$, $g_y = \sum_i (\partial_y E_i)^2$, and $g_{xy} = \sum_i (\partial_x E_i)(\partial_y E_i)$. Then the maximal magnitude is given by

$$\begin{aligned} M^\star(x,y) &= \max_\theta M(x,y,\theta) \\ &= \frac{1}{2}\left(g_x + g_y + \sqrt{(g_x - g_y)^2 + 4g_{xy}^2}\right) \end{aligned}$$

The direction is given (modulo π) by

$$\Theta^\star = \frac{1}{2}\tan^{-1}\left(\frac{2g_{xy}}{g_x - g_y}\right)$$

In order to evaluate how well this gradient image $M^\star(x, y)$ describes the image segments, we use the boundary detection benchmark located here [4]. The benchmark first thins the detected boundaries for a given image and then attempts to put them in 1-to-1 correspondence with boundaries marked by human observers. The aggregate goodness of this match can again be displayed as a precision-recall curve. In this case, precision measures the probability that a pixel declared "on boundary" for a given threshold t is in fact on a true boundary, i.e. $\mathbb{P}(B(x, y) = 1 | M^\star(x, y) > t)$ where B is the indicator of human marked boundaries. Recall measures the probability that a true boundary pixel is detected, i.e. $\mathbb{P}(M^\star(x, y) > t | B(x, y) = 1)$.

Figure 4 shows the effect of the number of eigenvectors on performance. The result shows some trade-off between decreased precision and increased recall which appears to level off at 8 eigenvectors, after which more eigenvectors only introduce noise. We choose $nvec = 8$ for the remainder of the paper.

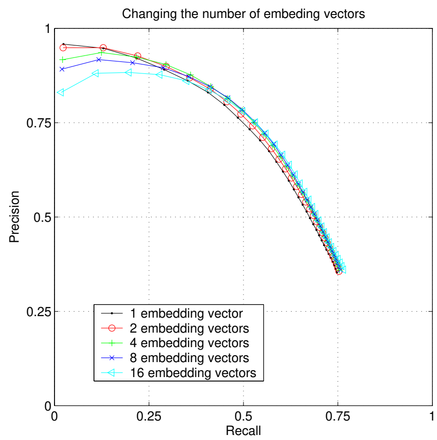


Figure 4: Increasing the number of embedding vectors used increases the recall but eventually hurts the precision.

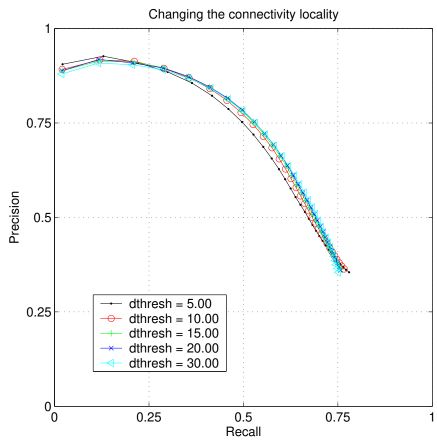


Figure 5: Increasing the connectivity radius improves performance

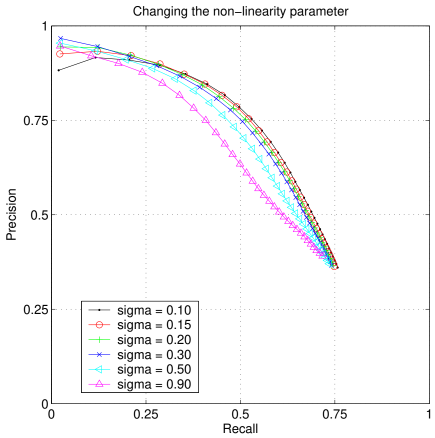


Figure 6: Decreasing σ parameter improves performance

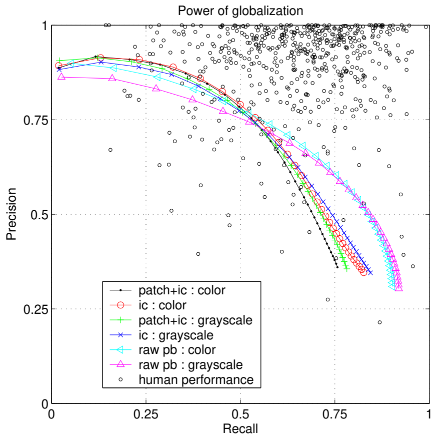


Figure 7: The globalized boundaries present in the eigenvectors provide higher precision than Pb in the high-precision / low-recall regime. It's also interesting to note that the difference between globalized color and grayscale is much less than between the raw color and grayscale boundary detector. Humans are similarly able to perform consistent segmentation with or without the benefit of color.

Figure 5 shows the effect of connectivity radius. Computational constraints demand that we use only sparse connections in the matrix W. Some have argued that only very local connections should be used since cues are unreliable, however, this results in large homogeneous regions being penalized. The result shows that increased connection density generally improves the contrast of boundaries in the embedding but comes at the cost of computation time. We choose $dthresh = 20$ for the remainder of the paper.

Figure 6 shows the effect of changing the σ in equation 1. With some surprise, we were unable to find a local minima. It appears that the smaller we make σ the better performance, up to the point where the eigensolver fails to converge in some fixed number of iterations. We choose $\sigma = 0.1$ for the remainder of the paper

Figure 7 shows the gain in performance over low level

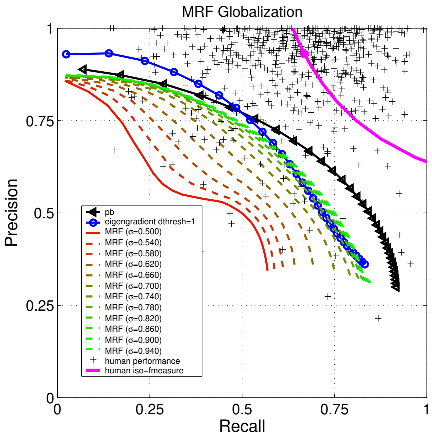


Figure 8: This figure shows the lack of useful "globalization" we observed in a MRF model of pairwise segmentation. Here hidden variables took on a discrete set of values and pairwise potentials between nearest neighbors i and j were given by W_{ij} using the intervening-contour cue. Generalized Belief Propagation (as in [14]) yielded estimates of the pairwise correlations between neighboring pixels which were interpreted as the probability of non-boundary. We were unable to find a temperature σ at which the pairwise marginals gave better estimates than the local boundary detector. The blue curve, plotted for comparison, shows the availability of useful information in the leading eigenvectors using the same nearest-neighbor graph and pair-wise affinities.

cues as well as the gain in combining cues. The principle conclusion is that in the low-recall, high-precision regime there is a clear improvement as result of the globalization provided by computing the eigenvectors of the normalized Laplacian. In this regime, normalized cuts is successful in suppressing high-contrast clutter which would mislead a local operator. However, the precision-recall curve based on eigenvectors crosses that curve given by the local boundary operator (P_b) so that in the high-recall/low-precision the performance of the local boundary detector is better. Intuitively this crossover can be explained by the bias that normalized cuts has against very small regions. When these exist in images, they will lead to edges which are more easily found by the local boundary detector.

A more subtle observation can also be made from these graphs. The use of color significantly improves the performance of a local boundary finder. However, when these measurements are piped through the machinery of spectral clustering this difference essentially disappears.

4.1. Comparison with Markov Random Fields

Shental et al recently proposed [14, 15] a probabilistic model for pairwise clustering in which a discrete random variable X_i corresponding to the segment label of the ith pixel is linked to neighboring pixels by image dependent pairwise potential functions $\psi_{ij}(X_i, X_j)$. The probability of a segmentation is then given by the MRF:

$$\mathbb{P}(X) = \frac{1}{Z} \prod_{\langle ij \rangle} \psi_{ij}(X_i, X_j)$$

This model naturally accommodates the same pairwise affinities we use for normalized cuts. [1] We set

$$\psi_{ij}(X_i, X_j) = \begin{cases} 1 & X_i = X_j \\ \exp(-\hat{S}_{ij}/\sigma) & X_i \neq X_j \end{cases}$$

where the parameter σ is commonly referred to as the *temperature*. As proposed in [14], we use generalized belief propagation to infer the marginal correlations between neighboring nodes. These correlations are easily treated as a soft boundary map which can then be compared to the local boundary detector.

Figure 8 shows the resulting performance over a range of temperatures σ. We find that the correlations estimated by belief propagation are generally worse at predicting human marked boundaries than the low-level, "un-globalized" measurements. Figure 8 shows that the eigensolver is able to squeeze out additional information even when restricted to nearest neighbor connections $dthresh = 1$.

5. Evaluating Segmentations

Now we turn to the task of using multiple eigenvectors to partition an image into regions. Several techniques have been proposed including thresholding the eigenvectors [25], recursive bi-partitioning, or running k-means on the embedded points [11, 16].

In this paper, we have evaluated a two stage strategy similar to that of [11]. First, use k-means or some variant to produce an oversegmentation (we call these segments *superpixels*). Then contract the graph so that each node in the contracted graph corresponds to a segment. Finally, recursively partition the contracted graph.

The results of this procedure are evaluated in Figure 9 by treating the segment boundaries as edges to be compared to the human ground-truth boundaries. The cloud of points on the graph show the precision and recall of the oversegmentation for each of 200 images. The desired property of an oversegmentation is that it has high recall so that few boundaries are missed.

Theoretical considerations [16, 22] suggest that the embedded points should be projected down onto the sphere before clustering. We found that sphering had no significant effect on performance for the task of oversegmentation.

[1]It is important to note that the $\hat{S}_{ij}$ we learn treats the pixel pairs as iid samples. This does not necessarily assign the maximal likelihood to the human segmentations under the MRF model. Instead, the logistic parameters should be iteratively re-estimated with respect to the expectations given by the MRF in order maximize the joint likelihood. This tuning may give improvements over the performance we document here.

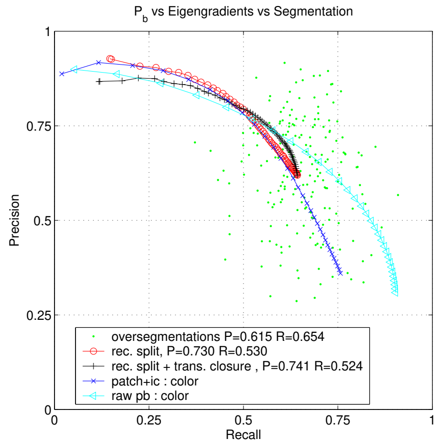


Figure 9: This figure shows that in passing from gradients of eigenvectors to segmentations, we are able to maintain high precision in identifying true boundaries. The cloud of points correspond to the initial oversegmentations for the set of images. The red (○) curve shows the results of recursively splitting the graph corresponding to this oversegmentation. The black curve (+) shows the results of the same procedure using weights which have been extended by setting undefined edges between a pair of superpixels to be the average of the defined weights.

5.1. Extending W

As mentioned earlier, computational constraints make it necessary to artificially set the connection between faraway pixels to 0 in order to introduce sparsity. It is well known [5, 14] that this introduces artifacts into segmentations. Intuitively, the k-way normalized cut consists of terms which, for local connection patterns, scales as the perimeter over the area. In the spectral relaxation, this manifests itself as eigenvectors which oscillate (like the modes of a drum head).

We consider the following procedure for *extending* W_{ij} by filling in values for long range connections. For each pair of pixels i and j which are not connected by an edge, we add an edge whose weight is the mean of the assigned connections between the superpixels containing i and j respectively. This is computationally feasible because all these newly introduced edges are immediately collapsed into the contracted edge.

Figure 9 shows the curves for recursively partitioning both the contracted graph and the extend contracted graph. They show that extending W improves performance in the mid-recall/mid-precision regime.

The last two pages show our results on the first 120 images from the BSDS. Here the recursive partitioning was terminated when no segment could be split with a normalized cut score of less than 1.15

6. Conclusions

In this paper we have presented a quantitative evaluation of the benefit of global information in the spectral partitioning framework over a purely local approach to boundary detection. The advantage is small but significant. However, it is primarily in the high-precision/ low-recall regime. In the high-recall regime, local boundary detection is actually superior, perhaps as a consequence of the bias in the normalized cut criterion towards balanced segments. This suggests research on multi-scale segmentation strategies as a way to provide “the best of both worlds”.

The other major take-home lesson is that there do exist segmentation strategies that make close to optimal use of the embedding provided by the eigenvectors. This is important because it opens up the possibility of high-level processing techniques that require a disjoint partitioning of an image into segments.

7. Acknowledgements

We would like to thank Noam, Assaf, Tomer and Yair for discussion and providing the GBP code.

References

[1] I. Abdou and W. Pratt. Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proc. of the IEEE*, 67(5):753–763, May 1979.

[2] F. Bach and M. I. Jordan. Learning spectral clustering. *Neural Information Processing Systems*, 2003.

[3] K. Bowyer, C. Kranenburg, and S. Dougherty. Edge detector evaluation using empirical ROC curves. 1999.

[4] Berkeley Segmentation Dataset, 2002. http://www.cs.berkeley.edu/projects/vision/grouping/segbench/.

[5] C. Fowlkes, S. Belongie, and J. Malik. Efficient spatiotemporal grouping using the nyström method. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, December 2001.

[6] C. Fowlkes, D. Martin, and J. Malik. Learning affinity functions for image segmentation: Combining patch-based and gradient-based approaches. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, June 2003.

[7] Y. Gdalyahu, D. Weinshall, and M. Werman. Stochastic image segmentation by typical cuts. In *CVPR*, 1999.

[8] D. Heeger and J. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH*, 1995.

[9] S. Konishi, A. L. Yuille, J. Coughlan, and S. C. Zhu. Fundamental bounds on edge detection: an information theoretic evaluation of different edge cues. *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 573–579, 1999.

[10] T. Leung and J. Malik. Contour continuity in region-based image segmentation. In *ECCV*, 1998.

[11] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *IJCV*, 43(1):7–27, June 2001.

[12] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using brightness and texture. In *NIPS*, 2002.

[13] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001.

[14] Shental N., Zomet A., Hertz T., , and Weiss Y. Learning and inferring image segmentations with the gbp typical cut algorithm. In *ICCV 03*, 2003.

[15] Shental N., Zomet A., Hertz T., , and Weiss Y. Pairwise clustering and graphical models. In *NIPS 03*, 2003.

[16] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Neural Information Processing Systems*, 2002.

[17] P. Perona and W. Freeman. A factorization approach to grouping. In *ECCV*, 1998.

[18] J. Puzicha, T. Hofmann, and J. Buhmann. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. 1997.

[19] J. Puzicha, Y. Rubner, C. Tomasi, and J. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. In *ICCV*, 1999.

[20] Y. Rubner and C. Tomasi. Coalescing texture descriptors. *ARPA Image Understanding Workshop*, 1996.

[21] M. Ruzon and C. Tomasi. Color edge detection with the compass operator. In *CVPR*, 1999.

[22] Yu S. and Shi J. Multiclass spectral clustering. In *ICCV 03*, 2003.

[23] S. Sarkar and K. Boyer. Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. In *CVPR*, 1996.

[24] G. Scott and H. Longuet-Higgins. An algorithm for associating the features of two images. *Proc. R. Soc. London*, B-244:21–26, 1991.

[25] J. Shi and J. Malik. Normalized cuts and image segmentation. In *CVPR*, 1997.

[26] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):888–905, August 2000.

[27] Y. Weiss. Segmentation using eigenvectors: a unifying view. *ICCV*, 1999.