Spectral Embedding S-T Mincuts for Hyperspectral Image Segmentation

Gavin Portwood
Alexander Nichols

1. Introduction

The Spectral Embedding Mincut (SE Mincuts) methodology (Estrada 2004) optimizes a standard S-T Mincut segmentation algorithm by finding salient seed regions to use for source and sink nodes in an unsupervised manner. A standard S-T Mincut segmentation performs well for images where the source and sink nodes have been interactively defined. The proposal regions are motivated by the three seed region criteria, namely,

1. source and sink regions are reasonably large with respect to the image resolution,
2. source and sink regions should incorporate disjoint unions of natural image segments, and
3. there must be significant inter-region variance and intra-region homogeneity.

A spectral embedding technique established proposal seed regions with respect to these criteria procedurally via analogies between Markov random walks and anisotropic smoothing. In this report we will evaluate an implementation of the algorithm for hyperspectral image segmentation, noting the deviations we've made to the original methodology and evaluating the technique.

(Motivation behind proposing good seed regions for the ST mincuts algorithm)

2. Seed Region Proposals

The outline of the unsupervised algorithm is provided as a reference, where an asterisk denotes a deviation from the methodology as implemented in Estrada (2004) to either accommodate hyperspectral segmentation, adapt for large images or to improve the methodology.

1. *Create a linear mapping from the hyperspectral images to a scalar image
2. *Normalize the mapping
3. Construct the affinity matrix, A, and normalize to construct the single-hop transitional probability matrix M
4. Take d eigenvalues of the transitional probability matrix M
5. Project the transitional probability matrix onto the spectral approximation with the eigendecomposition to form the weight matrix W
6. Construct the approximate blur kernel, $\widehat{P}$, at time t with d eigenpairs
7. Perform a coordinate transformation of the approximation mapped onto the lower dimensional subspace
8. *Perform K-Means clustering on the transformed coordinates and classify the image
9. *Threshold the classification based on τ_0 to eliminate loosely bound pixels
10. *Throw out sufficiently small clusters with cardinality less than τ_1
11. Classify to form seed regions and calculate the center of mass of each seed region
12. Perform Delaunay triangulation on the seed regions
13. Calculate the inter-region affinity measure $f_{i,j}$

In the proceeding section, each step will be evaluated with reference to Estrada (2004) and our deviations from the original algorithm.

2a. Constructing the Transition Probability Matrix (Steps 1-3)

The scalar affinity matrix is defined based on a scalar intensity I of the image as weighted by a Gaussian kernel, i.e

$$A_{ij} = exp(-(I(x_i) - I(x_j))^2/2\sigma^2),$$

where σ is a parameter defined by implementation. In the algorithm as implemented by Estrada 2004, and as considered here, we only consider adjacent pixel in the construction of the affinity matrix, that is $\left\|x_i - x_j\right\|_1 = 1.$ The scalar intensity mapping $I(x)$, has been defined by inspection as a linear combination of hyperspectral bands as to elicit the most apparent distinguishing features within the image. That is,

$$I(x_i) \equiv \sum_{\vee b} w_b I_h(x_i, b), \text{ where } \sum_{\vee b} w_b = 1,$$

b is a frequency band, w_b is a band weight and $I_h(x_i, b)$ is the hyperspectral intensity of x_i at band b. The scalar intensity map is then shifted by the mean and normalized by the variance such that the value of σ is consistent between different images. For the intensity map considered here, w_b were determined by trial and inspection and is featured in Figure 2. The affinity matrix is then normalized by the total transition matrix D, where

$$D_{jj} \equiv \sum_{\vee k} A_{kj}$$

and zero otherwise, to form the transition probability matrix for for a Markov random walk,

$$M \equiv AD^{-1},$$

where M_{ij} gives the probability of a random walk originating at x_i at $t = 0$ and moving to x_j at $t = 1$. More generally, the probability of a random walk originating at x_i and terminating at x_j at $t = t_c$, where t_c is parameter defined by the implementation, is given by

$$P \equiv M^{t_c}. \quad (1)$$

The value of t_c in this context effectively yields a length scale desired for seed regions. Again, note that the probability mass of a random walk originating at x_1 is given by $p_{1,t} = P_{1,j}$. Estrada 2004 suggests a value of $t_c = 600$ which we have used in the implementation discussed here.

2b. Spectral Embedding (Steps 4-6)

The principal concern of (1) is that the transition probability matrix M is matrix of size $nm \times nm$, where n and m are the dimensions of the image. Therefore the direct calculation of P is impractical for real-world applications. However, we may approximate P by making use of the spectral properties of the eigenvalue decomposition of M. Estrada (2004) makes use of this property by using the approximate $\widehat{P}$, where

$$P \approx \widehat{P} \equiv D^{1/2} U_d W^T$$

where U_d is the matrix of the first d columns of U, the eigenvector matrix from M. The weights of the spectral decompositions are given by $W \equiv \Delta_d^t U_d^T D^{1/2}$, where Δ_d is the diagonal matrix of the the first d eigenvalues of M and we have sorted U and Δ by in descending order of the the eigenvalues.

The value of d must be sufficiently large to properly resolved the detail of random the random walks. Estrada (2004) provides the metric

$$\left|\lambda_{d+1}\right|^t \leq 1/3,$$

where λ_{d+1} is the $d+1$ largest eigenvalue of M, which we have made use of in this work.

It is important to note the significance of the $\widehat{P}$, or P, in the context of image analysis. The key this this algorithm is that, since the transition probability of the Markov random walk is related to the affinity between adjacent pixels, a walk originating in pixels of a homogenous segment is expected to be have a similar probability distribution. That is, if a natural segment k includes pixels $k_1, k_2 \ldots k_r$ then

$$p_{k_1,t} \approx p_{k_2,t} \approx ... \approx p_{r,t} .$$

As such, Estrada (2004) creates an analogy between the probability mass $p_{i,t}$ and a convolution of I with an anisotropic filter kernel, where the filter kernel is a some function of A. The probability masses are therefore referred to as *blur kernels* in reference to this analogy.

2c. Construct Proposal Regions (Steps 7-10)

We may then perform a coordinate transform to simplify the distribution distribution of blur kernels with respect to our lower dimensional approximation. The transformed coordinates are given by the columns of Z, where

$$Z \equiv Q^{1/2} W^T, \text{ and}$$
$$Q \equiv U_d{}^T D U_d .$$

We note that Q is only positive semi-definite so that $Q^{1/2}$ cannot be calculated directly with a Cholesky decomposition.

We must then perform cluster analysis on the column of Z normalized their L2 norm, that is

$$s_i = z_i / \left\| z_i \right\| ,$$

in its d-dimensional space. Estrada (2004) defines a mixture model which is fitted to the data in order to find cluster centers then a linear classifier is utilized to classify the each column of S. The mixture model required an additional constant be defined in the algorithm. Instead, it was decided to use a k-means algorithm with k centers, where k was determined from the *elbow* criteria featured in Figure 1. The use of k-means was validated in personal correspondence with Professor Estrada.

Recall that the criteria for appropriate proposal seed regions dictates that there must be some high degree of confidence in all proposal regions. A thresholding is performed on the classification based on the inner product of the columns of z and the cluster centers m to form the proposal regions S_k, that is

$$S_k = \left\{ x_j | z_j^T m_k \geq \tau_0 \left\| z_j \right\| \right\} \quad .$$

The value of τ_0 is chosen large enough to eliminate ambiguous classification, where ambiguity is defined as a large distance from cluster centers. This procedure effectively adds an additional component to the linear classifier. Estrada (2004) proposes that $\tau_0 = 0.985$ is a good choice, but we observed that a smaller value ($\tau_0 = 0.95$) exhibited better results.

In our implementation with large images, it was observed that some S_k exhibited low cardinality, which contradicts our initial criteria for appropriate proposal regions. So it was decided to eliminate regions with low cardinality, that is

$$\left| S_k \right| > \tau_1 ,$$

where $|\cdot|$ indicates cardinality. For the image considered here, we chose $\tau_1 = 36$ which indicates that all proposal regions will span at least 36 pixels. This is a step added to the methodology in Estrada (2004), which has been added as an adaption for larger images.

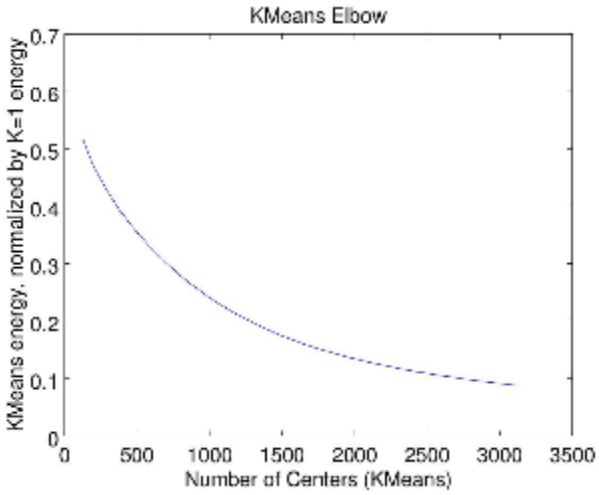


Figure 1: The criteria for establishing a critical K value in the KMeans algorithm. The elbow was chosen at K=3160, where subsequent cluster centers yielded sufficiently slow KMeans energy decay.

2d. Define Proposal Region Affinity (Steps 11-13)

The ST mincuts algorithm requires that each combination of adjacent sink and source regions be applied to the algorithm. Seed region adjacency is determined in Estrada (2004) with Delaunay triangulation, an algorithm common for meshing. The triangulation is performed in regard to seed region centroids and is shown in figure 2.

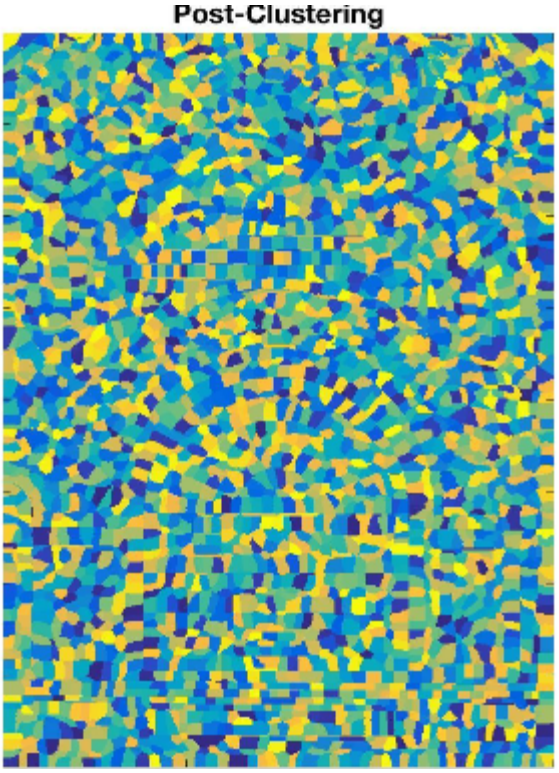
In applying the mincuts algorithm, it is essential that regions that exhibit some similarity be merged prior to segmentation. For instance, a pair of seed regions with a high-degree of similarity should not be proposed as a source-sink pair in ST mincuts. Estrada (2004) defines a similarity measure, $f_{i,j}$, between two seed regions S_i and S_j in z-space with regard to proximity to linear combinations of the two cluster centers, that is,

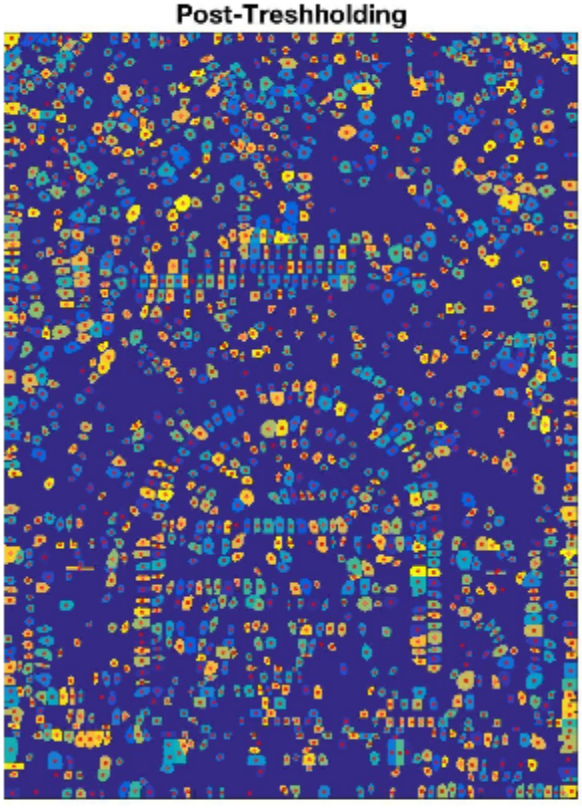
$$f_{k,j} \equiv min_{r\ in\ [0,1]} \left| \left\{ x_i | z_i^T m_{k,j}(r) \geq \tau_0 \left\| z_i \right\| \right\} \right| / \left(\left| S_k \right| r + \left| S_k \right| (1-r) \right), \text{ where}$$

$$m_{k,j} = n_{k,j}(r) / \left\| n_{k,j}(r) \right\| \text{ and}$$

$$n_{k,j}(r) = r m_k + (1-r) m_j .$$

Recall that $|\cdot|$ indicates cardinality. By intuition we may interpret $f_{i,j}$ as a measure of the change in cardinality of seed regions for all linear combinations of two cluster centers. We identify the minimum degree of increase overall linear combination as to elicit the 'worst case' combination of the two, where lower values of $f_{i,j}$ indicate discontinuity between seed regions. Estrada (2004) classifies two regions as being continuous if $f_{i,j} > 0.3$, ambiguous if $0.2 < f_{i,j} < 0.3$ and discontinuous if $f_{i,j} < 0.2$.





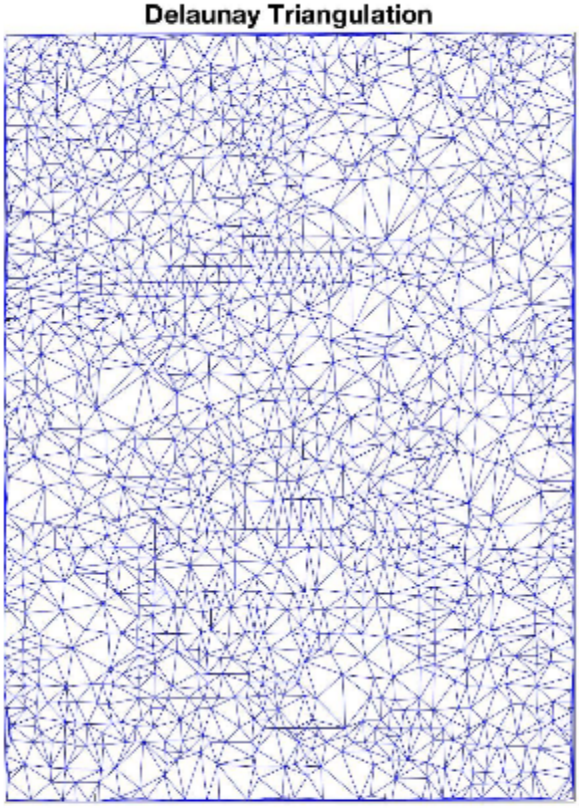


Figure 2: Illustration of the methodology. (From left to right) The linear combination of bands in step (1), the result of KMeans clustering in (8), the result of thresholding in step, where the red circles represent seed region centroids (9) and the Denaunay triangulation in step (12).

3. ST Mincuts

In practice, the valid seed region proposals which can be extracted with confidence in **2** often span a minority of pixels in the original image. This is where ST Mincuts comes in.

ST Mincuts is a special implementation of the Max-Flow Min-Cut algorithm which notionally separates vertices in a flow graph into three subgraphs: those connected strongly to the source and not at all to the sink (vertices in subgraph S), those connected strongly to the sink and not at all to the source (vertices in subgraph T), and those disconnected entirely from both the sink and the source. After any optimal cut is made, vertices in S will stay connected to the source, those in T will stay connected to the sink. Those vertices that were originally neither in S nor in T will be divvied up between these two 'strong' subgraphs.

This idea gels well with the sort of information given by the seed region proposals: every seed region has a small number of neighbors in a Delaunay Triangulation, which can take on 3 classifications, as described at the end of **2d**. Relative to every seed region, it is good to think of a continuous neighbor as cohabitating the source region of an incipient ST Mincut formulation, a discontinuous neighbor as inhabiting the sink region in such a formulation, and any ambiguous neighbors as existing in the 'weak' region which is the complement of $S \cup T$. Possible most importantly, it is good to include the 'interstitial' set of pixels which are not members of any seed region in every single such formulation, in an attempt to segment it as much as possible. This formulation maintains what is already known about the seed regions' relationships to each other, while giving structure to the interstitial region in the image, and possibly providing more certainty about ambiguous regions, if they are segmented as a result of inclusion in the 'weak' region.

To get a good picture of what these cuts could mean, Estrada recommends taking the union of all possible cut formulations for every seed region with those of every other seed region. Possible formulations for any given seed region take into account every combination of a seed region's continuous neighbors being either included in S or not.

This provides a rigorous formulation for our graph topologies, but how are these edges weighted? Estrada ses the affinity seen earlier:

$$A_{ij} = exp(-(x_i - x_j)^2/2\sigma^2)$$

But he is working with only grayscale images. The additional complexity duly associated with hyperspectral data needs to be considered. First, define our hyperspectral affinity function:

$$A^H_{ij} = exp(-H(x_i - x_j))$$

$H(\cdot)$ should fulfill two important constraints: it should make the $L-1$ norm of any hyperspectral image resemble that of a grayscale image, and it should take into account that not all bands are created equal. This is to say that some bands are going to be more important to us when calculating affinity between two adjacent pixels. To accomplish this, normalize by the global maximum value in the image, and create a special norm for spectra in the image which puts a greater value on closeness in a band with greater variance.

$$H(\cdot) = \|\cdot\|_H/2\delta^2$$

Define the hyperspectral norm on a difference between two vectors:

$$\|x_i - x_j\|_H = \left(\sum_{m=1}^{M} (x_i - x_j)_m^{\sigma^2_m/\sigma^2_k}\right)^{1/2}$$

Where M is the number of bands in the image, σ^2_m is the variance of band m, and σ^2_k is defined by $argmin_k(\sigma^2_k)$. Additionally, δ is substituted here to avoid confusion with the actual variances in the above equation

4. Implementation

The C implementation of the code, as well as post-processing script, are provided in the attached archive. The eigenpair calculation was performed with a sparse matrix calculation using the Feast eigenvalue solver, with input from Professor Eric Polizzi, on supercomputing resources provided by the DoD High Performance Modernization Project. With the criteria proposed by Estrada (2004), the ~150 largest eigenvalue pairs were selected, a calculation which took ~20 hours on a single 32-core node.

K-Means was implemented with the VLFeat library on the same computing resources. The elbow identification in Figure 1 took ~12 hours on a the same computing resources, where K=3160 was determined to be a good values for the number of centers.

Delaunay triangulation was performed in post-processing using a builtin matlab routine and the calculations of equation (2) for every set of adjacent nodes took 12 hours on supercomputing hardware.

ST Mincuts was implemented and run in C. Output is in the form of a segment-colorized PNG file. RGB values for individual segments were generated with a modulo hash on serial segment numbers. It took ~5 minutes to run through all the possible ST Mincut formulations and produce a fully-segmented image. The affinities were calculated with the equation given in **3**, with a value of $\delta = 0.025$ being used.

References

Estrada, Francisco J., Allan D. Jepson, and Chakra Chennubhotla. "Spectral Embedding and Min Cut for Image Segmentation." *BMVC*. 2004.

Estrada, Francisco J. *Advances in computational image segmentation and perceptual grouping*. Diss. University of Toronto, 2005.

Estrada, Francisco J., and Allan D. Jepson. "Benchmarking image segmentation algorithms." *International Journal of Computer Vision* 85.2 (2009): 167-181.

Polizzi, Eric, and James Kestyn. "FEAST Eigenvalue Solver v3. 0 User Guide." *arXiv preprint arXiv:1203.4031* (2012).