LA-UR- 10-02273

*Approved for public release; distribution is unlimited.*

*Title:* Patch-based image segmentation of satellite imagery using minimum spanning tree construction

*Author(s):* Skurikhin, A.N.

*Intended for:* GEOBIA-2010 Conference: Geographic Object-Based Image Analysis, Ghent, Belgium, 06/29-07/02/2010.

Los Alamos NATIONAL LABORATORY EST. 1943

# PATCH-BASED IMAGE SEGMENTATION OF SATELLITE IMAGERY USING MINIMUM SPANNING TREE CONSTRUCTION

A. N. Skurikhin

MS D436, Space and Remote Sensing Group, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA - alexei@lanl.gov

**KEY WORDS:** Segmentation, Laplacian, Delaunay Triangulation, Minimum Spanning Tree

**ABSTRACT:**

We present a method for hierarchical image segmentation and feature extraction. This method builds upon the combination of the detection of image spectral discontinuities using Canny edge detection and the image Laplacian, followed by the construction of a hierarchy of segmented images of successively reduced levels of details. These images are represented as sets of polygonized pixel patches (polygons) attributed with spectral and structural characteristics. This hierarchy forms the basis for object-oriented image analysis. To build fine level-of-detail representation of the original image, seed partitions (polygons) are built upon a triangular mesh composed of irregular sized triangles, whose spatial arrangement is adapted to the image content. This is achieved by building the triangular mesh on the top of the detected spectral discontinuities that form a network of constraints for the Delaunay triangulation. A polygonized image is represented as a spatial network in the form of a graph with vertices which correspond to the polygonal partitions and graph edges reflecting pairwise partitions relations. Image graph partitioning is based on the iterative graph contraction using Boruvka's Minimum Spanning Tree algorithm. An important characteristic of the approach is that the agglomeration of partitions is constrained by the detected spectral discontinuities; thus the shapes of agglomerated partitions are more likely to correspond to the outlines of real-world objects.

## 1. INTRODUCTION

Graph theory provides a general framework for image segmentation. According to the graph-theoretic approach, image elements such as pixels or regions (a.k.a. blobs, superpixels, pixel patches, polygons) are described using a weighted graph. Image segmentation is formulated as a graph-partitioning problem, where produced graph partitions have a semantic meaning in the real world. Although numerous graph partitioning based image segmentation approaches were developed, analyzed and showed good results for processing of general outdoor imagery, their scalability to the analysis of very large datasets of satellite imagery remains a challenge. The reduction of computational complexity is often achieved by agglomeration of pixels into larger patches, superpixels, prior to image partitioning. This effectively reduces the number of image data elements to be processed. A variety of approaches of grouping pixels into superpixels have been proposed. The used techniques include watershed transform (Vincent and Soille, 1991), Mean Shift, that is a kernel based density estimation technique (Cheng, 1995; Comaniciu and Meer, 2002; Fukunaga and Hostetler, 1975), and Delaunay triangulation. The normalized cuts or other method can then be applied to partition the graph of the pixel patches produced by the watershed transform, Mean Shift, or triangles and their groups (polygons) produced by Delaunay triangulation (DT). Bock et al. (Bock et al., 2005), Monteiro and Campilho (Monteiro and Campilho, 2008) presented combinations of watershed transform and the normalized cuts. Tao et al. (Tao et al., 2007) investigated integration of Mean Shift and the normalized cuts. It was also proposed to use constrained Delaunay triangulation (CDT) to replace pixels with a set of triangles. Cobzas and Zhang (Cobzas and Zhang, 2001) construct the triangular mesh via CDT that is based on the detected edges. Then, generated triangles are iteratively merged using their spectral similarity. Wu and Yu (Wu and Yu, 2003) applied normalized cuts to group triangles of the triangular mesh that is also generated by CDT over the detected edges. Similar to last two approaches, Prasad and Skourikhine (Prasad and Skourikhine, 2006) combined edge detection and CDT. They augmented this combination with heuristic criteria of perceptual organization to create initial polygonized image partitioning. Skurikhin (Skurikhin, 2008) presented a hierarchical image segmentation method that builds upon this polygonized image representation and evaluated the method using general outdoor imagery from the Berkeley Segmentation Dataset (the Berkeley Segmentation Dataset and Benchmark). In this paper we extend the work of (Skurikhin, 2008) by (1) combining edge detection with the image Laplacian to refine the initial edge dataset, and (3) evaluating an approach using satellite imagery.

The paper is structured as follows. The second section presents related research in the area of image segmentation using proximity graphs. The third section presents our method for image segmentation. The fourth section presents an evaluation of our approach by using satellite imagery. Finally conclusions are presented in the last section.

## 2. RELATED WORK

### 2.1 Proximity Graphs Based Segmentation

Delaunay triangulation and related proximity graphs such as MST, Relative Neighborhood Graph, and Gabriel Graph have been widely used in image analysis and spatial modeling. Much work has been devoted to the optimal image hierarchic partitioning through a sequence of local computations based on proximity graphs, specifically MST. The MST based image segmentation seeks image partitioning by iteratively linking image elements through the lowest cost tree edges, which represent similarity of neighbouring elements. One of the earliest applications of tree-based data clustering to visual like point data sets analyzed histogram of MST edges and investigated tree characteristics such as MST "relative compactness", tree diameter, and point densities was proposed in (Zahn, 1971). Horowitz and Pavlidis (Horowitz and Pavlidis, 1976) applied a tree-based concept image segmentation. It was suggested to use global homogeneity criterion to control construction of an irregular pyramid starting from a regularly sampled pixel grid. Morris et al. (Morris et al., 1986) proposed hierarchic image segmentation approach based on Kruskal's MST construction algorithm (Kruskal, 1956), starting from a regular pixel grid. Hierarchic merging of pixel patches is controlled by updated intensity dissimilarities between the agglomerated patches. Montanvert et al. (Montanvert et al., 1991), Jolion and Montanvert (Jolion and Montanvert, 1992) used irregular tessellations to generate an adaptive multi-scale image representation. The approach employs an irregular sampling of the pixel grid to build the initial (fine scale) image representations. The irregular sampling hierarchy is then recursively built from the lower scales. The result depends on the stochastic nature of the sampling procedure. Xu and Uberbacher (Xu and Uberbacher, 1997) used Kruskal's algorithm to construct MST of the image from a regularly sampled pixel grid. The tree is then partitioned by an optimization algorithm into subtrees based on the subtrees' spectral similarities. A set of produced subtrees represents a sought image partition. Similar to the work of (Xu and Uberbacher, 1997), Felzenszwalb and Huttenlocher (Felzenszwalb and Huttenlocher, 1998) started from a regular pixel grid and used Kruskal's algorithm to construct MST of the image. However, MST construction is based on thresholding a ratio of the variation between neighboring pixel patches and the variation within the patches. To avoid image over-fragmentation, the approach adjusts the measure of variation using the ratio of the sizes of neighboring pixel patches. The extent of this adjustment controls how easily small patches are merged with the larger neighbors. The approach proposed in (Kropatsch et al., 2007; Haximusa and Kropatsch, 2004) controls the grouping of pixels into patches based on image variation in similar fashion how it is done in (Felzenszwalb and Huttenlocher, 1998). The difference between two approaches is that the approach of (Kropatsch et al., 2007; Haximusa and Kropatsch, 2004) uses Boruvka's MST construction algorithm (Boruvka, 1926; Nešetřil et al., 2001) instead of Kruskal's algorithm that is used in (Felzenszwalb and Huttenlocher, 1998). Computational complexity of Kruskal's algorithm is $O(ElogE)$, and the complexity of Boruvka's algorithm is $O(ElogN)$, where $E$ is the number of edges in the graph. In contrast with Kruskal's MST construction algorithms, that build the MST one edge at a time, Boruvka's algorithm adds several MST edges at each stage. Skurikhin (Skurikhin, 2008) uses Boruvka's MST construction algorithm as well. In contrast with the stated approaches that start from regular or irregular pixel grids the method of (Skurikhin, 2008) builds an irregular hierarchy of image partitions starting from triangular tessellation of the image. Iterative agglomeration is controlled by the cost function based on spectral similarity and strength of edges. Termination of the agglomeration is controlled by global thresholding.

## 3. IMAGE SEGMENTATION

In a polygon-based image pyramid, each level represents a polygonal tessellation of the image. The pyramid is built iteratively from bottom-up using only local interactions of the neighboring polygons. On the lowest level ($l=0$, fine level of detail) of the pyramid the polygons are constructed from an irregular triangular tessellation of the image; they are unions of triangles. On higher level ($l>0$, coarser level of detail) of the pyramid the polygons are unions of neighboring polygons on a lower finer level ($l-1$). The polygons on level $l$ of the pyramid are considered as the vertices of an undirected graph $G_l$. The edges of the graph describe the adjacency relations between the polygons on level $l$. Thus $G_l = (V_l, E_l)$, where $V_l$ is the set of vertices, and $E_l$ is the set of edges. The derivation of $G_{l+1}$ from $G_l$ is formulated as construction of an MST of $G_l$. MST construction is based on Boruvka's algorithm. The built pyramid $P$ is described as a stack of graphs $G_l$ representing the image in a fine-to-coarse hierarchy.

### 3.1 Initial Image Partitioning

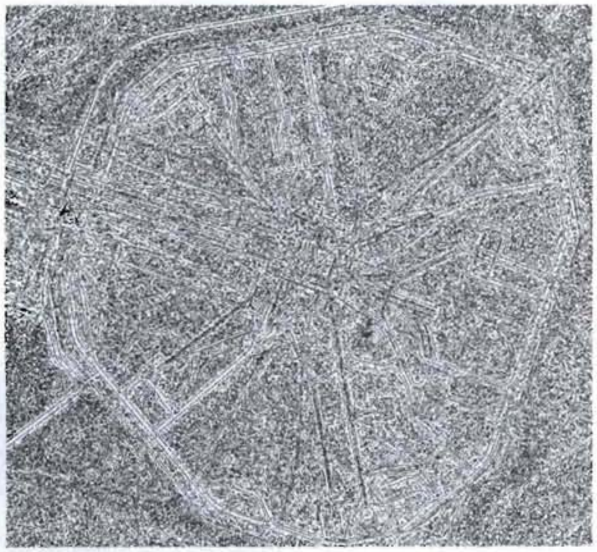
Polygons on the lowest (fine) level of a pyramid are built upon the triangular tessellation of the image. The tessellation is based on CDT of the detected edges. In contrast with (Skurikhin, 2008) that applies the triangulation directly to the edges, such as those detected with Canny detector (Canny, 1986), we introduce filtering of the detected edges using the image Laplacian and morphological processing prior to the triangulation. The basic idea is that the Laplacian is close to zero in regions with relatively uniform intensity and where the edges are. By differentiating these two cases, it is possible to filter out the edges detected inside "flat" regions. This effectively reduces the number of edges for further analysis. At the same time, because we use edges detected at fine scale, the localization accuracy of the kept edges remains high. Given the detected edges and computed image Laplacian, the filtering steps are:

1. Threshold the image Laplacian to detect "flat" regions, which have the Laplacian magnitude close to zero. The threshold is set to 4% of the absolute range of the image Laplacian. The result is binary Laplacian image.
2. Apply morphological erosion to the binary Laplacian image to delete small "flat" regions.
3. Apply morphological closing to the binary Laplacian image. This steps keeps those regions, where the edges are detected.
4. Delete edges inside "flat" regions.

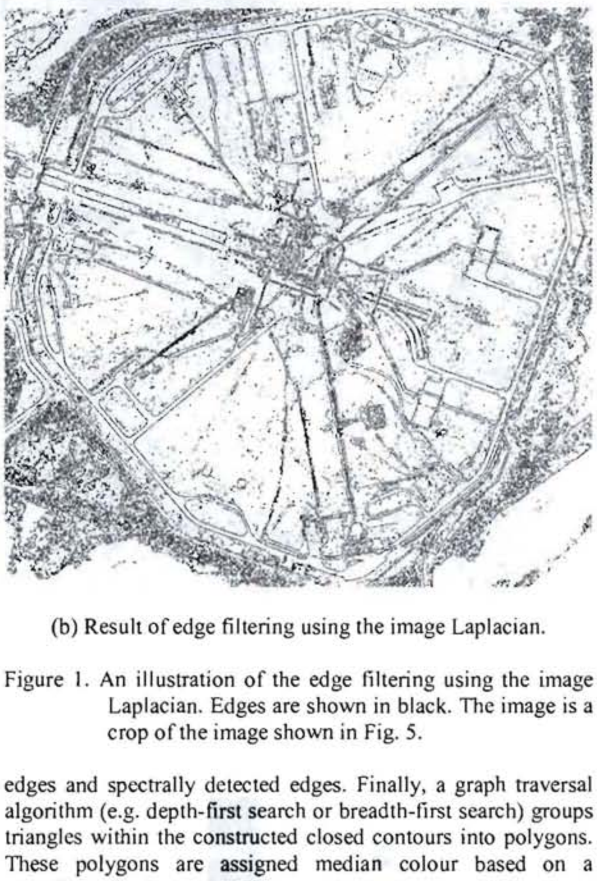
The example result of the edge filtering processing is shown in Fig.1. The filtered edge set is then triangulated based on CDT (Schewchuk, 1996). Therefore, the generated mesh is adapted to the image content, because the constraints are the detected edges. The generated triangle edges link the (Canny) detected edges which are often fragmented. In order to create initial set of polygons (superpixels), which serves as fine level-of-detail (LOD) image representation and is used as a starting point for hierarchical segmentation, the generated triangle edges are also subject to filtering process. The used triangle edge filtering is similar to the one in (Prasad and Skourikhine, 2006). Steps of triangle edges processing are:

1. Delete triangle edges based on their length. The threshold is set to three times median triangle edge length.
2. Process the remaining triangle edges using proximity and closure principles (Fig. 2) inspired by the visual psychology and perception studies (Wertheimer, 1958).

Proximity keeps the shortest triangle edge that connects end point of an edge chain to the other edge chain end point or interior (Fig. 2a). The closure rule is responsible for filtering out triangle edges which are bounded by the same detected edge chain (e.g., "U"-shape) or the same pair of the detected edges (Fig. 2b). This triangle edge filtering results in a set of closed contours consisting of a combination of the preserved triangle



(a) Result of Canny edge detection; no edge filtering applied.



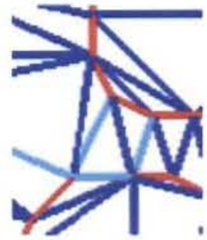
(b) Result of edge filtering using the image Laplacian.

Figure 1. An illustration of the edge filtering using the image Laplacian. Edges are shown in black. The image is a crop of the image shown in Fig. 5.

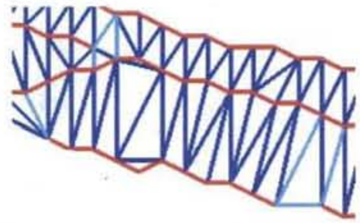
edges and spectrally detected edges. Finally, a graph traversal algorithm (e.g. depth-first search or breadth-first search) groups triangles within the constructed closed contours into polygons. These polygons are assigned median colour based on a sampling of triangles forming polygons. Thus the image is segmented in a set of spectrally attributed polygons.

### 3.2 Hierarchical Image Segmentation

Once the polygon-based image representation on the fine (lowest) level of a pyramid is produced, we iteratively group neighbouring polygons on level $l$ into larger polygonal chunks, producing level $(l+1)$ of the image pyramid. Polygon



(a) Triangle edge filtering based on proximity. Shortest edges closing gaps are kept. Cyan edges are kept, blue edges are deleted.



(b) Triangle edge filtering based on closure.

Figure 2. Triangle edge filtering based on closure and proximity. Canny edges are shown in red, triangle edges are shown in cyan and blue.

agglomeration is based on Boruvka's algorithm to construct MST. Boruvka's algorithm proceeds in a sequence of stages, and in each stage it identifies a forest $F$ consisting of the minimum-weight edge incident to each vertex in the graph $G$, then forms the graph $G_1 = G\backslash F$ as the input to the next stage. $G\backslash F$ denotes the graph derived from $G$ by contracting edges in $F$. This agglomeration iteratively goes until dissimilarity threshold is exceeded.

The quality of segmentation depends on the pairwise polygon adjacency matrix, containing $E_l$. The attributes of edges are defined using two features: colour similarity in CIELab space, $\Delta C_{ij}$, and strength of the contour fragment separating polygons, $S_{ij}$. We evaluate the affinity $w_{ij}$ between two neighbouring polygons $i$ and $j$:

$$w_{ij} = \Delta C_{ij} \times \left(1 + S_{ij}\right) \tag{1}$$

$$S_w = \sum_{k=1}^{N} \frac{l_k}{L} \cdot s_k \tag{2}$$

where $\Delta C_{ij}$ is the distance between two polygons in CIELab space,
$N$ is the number of edge fragments shared by two neighboring polygons (Fig. 3),
$L$ is the length of the shared contour segment,
$l_k$ is the length of the shared edge fragment belonging to a given contour fragment,
$s_k$ is the strength of the shared edge fragment.

The spatial layout of the polygons changes after each agglomeration iteration. Therefore the adjacency matrix corresponding is re-evaluated each iteration.

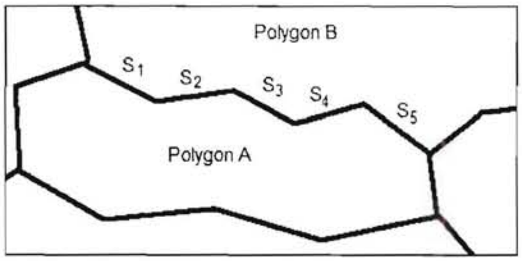



Figure 3. An example of the contour fragment shared by two neighbouring polygons A and B. The shared contour fragment consists of five edge fragments that can be spectrally detected (Canny) edge and triangle edges.

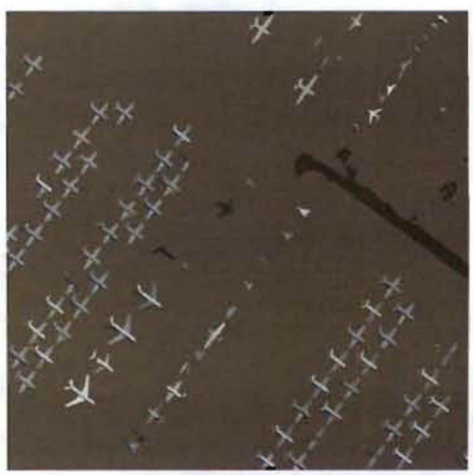
## 4. EXAMPLES

Figures 4 through 7 illustrate segmentation results produced by our approach. It should be noted that we have used the same settings in all experiments. Fine level of detail of a pyramid is constructed using Canny edge detector with the $\sigma_{Canny} = 1$, hysteresis low threshold = 5, and hysteresis high threshold = 15. We use the Triangle code (Schewchuk, 1996) to generate triangular tessellation over the detected edge map. Threshold for the image Laplacian is 4%. Color images were processed using CIELab space. Global threshold was set to 15 because the perceptually significant difference in color space is estimated in the range (15, 30). A method produces consistent outlines of the real-world objects. However, note that that some objects that are present at finer LODs are missing at coarser LODs. For instance, Fig. 5 shows an example of segmenting most pixels of the plane in the top left corner of Fig.5 together with surrounding background, though at finer LODs the plane chunks are distinct from the background (Fig. 6d). This is the subject of further investigation.

Figure 4. Segmentation example corresponding to 5(c). Contours of the constructed polygons (in white) are superimposed on the top of the original image.

(a) ©Digital Globe, Inc.. Original image, 867 × 867 pixels.

(b) Segmentation result at LOD = 5. Number of produced polygons = 884.



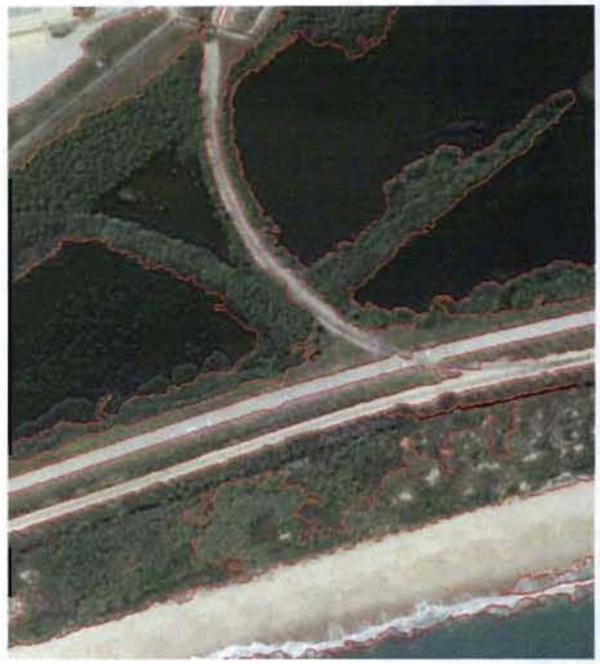
(c) Segmentation result at LOD = 12. Number of produced polygons = 286.

Figure 5. Segmentation example.

(a) ©Digital Globe, Inc.. Original image of 2,897 × 2,891 pixels.

(b) Result of the image segmentation on the 12th LOD. The number of constructed polygons is 602. Each polygon is attributed with the colour averaged over the colours of the pixels covered by polygon. Data reduction factor (number of pixels divided by number of polygons) is 13,912. Triangulation tessellation generated by the Constrained Delaunay triangulation over the edge set contained 1,722,514 triangles.

Figure 6. An example of the hierarchical image segmentation.



(c) Zoomed in region that is located in the middle of the original along the coast line. Contours (in red) of the constructed polygons are superimposed on the top of the original image.

(d) Zoomed in region corresponding to that in (c). Constructed polygons are outlined with the contours in red.

Figure 7. Details of the segmentation shown in Fig.6(b).

## 5. CONCLUSIONS

We have presented a segmentation method to construct a fine-to-coarse hierarchy of image partitions by means of edge detection, the exploitation of image derivatives and minimum spanning tree construction. The method yields good results that can be used by the search for objects of interest across the generated image hierarchy. Further improvement of the edge set that can preserve semantically salient but weak edges might require combination with region-based approaches. A future area of work will also be scale-space texture analysis, in particular, adding textures characteristics to the feature vectors describing polygons and their pairwise relations.

## REFERENCES

Bock, J. D., Smet, P. D., Philips, W., 2005. Image segmentation using watershed and normalized cuts, *Proc. SPIE*, 5675.

Borůvka, O., 1926. O jistém problému minimálnim. (about a certain minimal problem) Práce Moravské Přírodvédecké Společnosti v Brně, *Acta Societ. Scienc. Natur. Moravicae*, 3(3), pp. 37-58 (In Czech.).

Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6), pp. 679-698.

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17, 790-799.

Cobzas, D., Zhang, H., 2001. Mobile robot localization using planar patches and a stereo panoramic model. *Vision Interface*, pp. 94-99.

Comaniciu, D., Meer, P., 2002. Mean-Shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5), pp. 603-619.

Felzenszwalb, P. F., Huttenlocher, D. P., 1998. Image segmentation using local variation, In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 98-104.

Fukunaga, K., Hostetler, L. D., 1975. The estimation of the gradient of a density function. *IEEE Trans. Info Theory*, 21, pp. 32–40.

Haxhimusa, Y., Kropatsch, W. G., 2004. Segmentation graph hierarchies, Lecture Notes in Computer Science, 3138, pp. 343-351.

Horowitz, S. L., Pavlidis, T., 1976. Picture segmentation by a tree traversal algorithm. Journal of the Association for Computing Machinery, 23(2), pp. 368-388.

Jolion, J.-M., Montanvert, A., 1992. The adaptive pyramid, a framework for 2D image analysis, *CVGIP: Image Understanding*, 55(3), pp. 339-348.

Kropatsch, W. G., Haxhimusa, Y., Ion, A., 2007. Multiresolution image segmentation in graph pyramids. In: *Applied Graph Theory in Computer Vision and Pattern Recognition*, Kandel, A., Bunke, H. H., Last, M. (Eds.) Series: Studies in Computational Intelligence, 52, pp. 3-42.

Kruskal, J. B. L., 1956. On the shortest spanning subtree of a graph and the traveling salesman problem, *Proc. Am. Math. Soc.*, 7, pp. 48-50.

Montanvert, A., Meer, P., Rosenfeld, A., 1991. Hierarchical image analysis using irregular tessellations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(4), 307-316.

Monteiro, F. C., Campilho, A., 2008. Watershed framework to region-based image segmentation. In: *Proc. Intern. Conference on Pattern Recognition*.

Morris, O. J., Lee, M. J., Constantinides, A. G., 1986. Graph theory for image analysis: an approach based on the shortest spanning tree. *IEE Proc. F*, 133(2), pp. 146-152.

Nešetřil, J., Milková, E., Nešetřilová, H., 2001. Otakar Borůvka on minimum spanning tree problem. Translation of both the 1926 papers, comments, history. *Discrete Mathematics*, 233, pp. 3-36.

Prasad, L., Skourikhine, A. N., 2006. Vectorized image segmentation via trixel agglomeration. *Pattern Recognition*, 39(4), pp. 501-514.

Shewchuk, J. R., 1996. Triangle: engineering a 2D quality mesh generator and Delaunay triangulator. *Lecture Notes in Computer Science*, 1148, pp. 203-222.

Skurikhin, A. N., 2008. Proximity graphs based multi-scale image segmentation. *Lecture Notes in Computer Science*, 5358, pp. 298-307.

Tao, W., Jin, H., Zhang, Y., 2007. Color image segmentation based on Mean Shift and Normalized Cuts. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(5), pp. 1382-1389.

The Berkeley Segmentation Dataset and Benchmark. http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/ (accessed 4 Apr. 2010)

Vincent, L., Soille, P., 1991. Watersheds in digital spaces an efficient algorithm based on immersion simulations. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 13(6), pp. 583-598.

Wertheimer, M., 1958. Principles of perceptual organization, In: *Readings in Perception*, Beardslee, D., Wertheimer, M., (Eds.), D. Van Nostrand, Princeton, NJ, 115-135.

Wu, Q., Yu, Y., 2003. Two-level image segmentation based on region and edge integration. In: *Proc. 7th Digital Image Computing: Techniques and Applications*, pp. 957-966.

Xu, Y., Uberbacher, E. C., 1997. 2D image segmentation using minimum spanning trees, *Image and Vision Computing*, 15, 47-57.

Zahn, C. T., 1971. Graph-theoretic methods for detecting and describing gestalt clusters. *IEEE Trans. Comp.*, 20, pp. 68-86.