LA-UR- 08-4684

Title: Proximity Graphs Based Multi-Scale Image Segmentation

Author(s): Skurikhin, A.N.

Intended for: the National Reconnaissance Office's Director's Innovation Initiative (DII) program, the Advanced Systems and Technologies Directorate, National Reconnaissance Office, Intern. Symposium on Visual Computing-2008.

Los Alamos
NATIONAL LABORATORY
EST. 1943

Proximity Graphs Based Multi-Scale Image Segmentation

Alexei N. Skurikhin

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

Abstract. We present a novel multi-scale image segmentation approach based on irregular triangular and polygonal tessellations produced by proximity graphs. Our approach consists of two separate stages: polygonal seeds generation followed by an iterative bottom-up polygon agglomeration into larger chunks. We employ constrained Delaunay triangulation combined with the principles known from the visual perception to extract an initial irregular polygonal tessellation of the image. These initial polygons are built upon a triangular mesh composed of irregular sized triangles and their shapes are adapted to the image content. We then represent the image as a graph with vertices corresponding to the polygons and edges reflecting polygon relations. The segmentation problem is then formulated as Minimum Spanning Tree extraction. We build a successive fine-to-coarse hierarchy of irregular polygonal grids by an iterative graph contraction constructing Minimum Spanning Tree. The contraction uses local information and merges the polygons bottom-up based on local region- and edge- based characteristics.

1 Introduction

The problem of image segmentation remains one of great challenges for computer vision. Many image segmentation algorithms have been developed. Early in the 20^{th} century the Gestalt psychology has shown the importance of perceptual organization for image segmentation and interpretation. Wertheimer approached the problem by postulating principles that affect perceptual grouping, such as proximity, similarity, good continuation, symmetry, that can be used for image segmentation [17]. Inspired by the visual psychology a great number of image segmentation methods have been developed, e.g. [1, 2, 8, 11]. Many of them try to partition the image by optimizing a suitable cost function that combines different criteria of image element grouping.

Recently there has been a progress achieved in the area of graph-theoretic approach to the image segmentation and perceptual grouping. According to this approach, image structures such as pixels or edges are described using graph, and the segmentation is formulated as a graph-partitioning problem. Typically segmentation is achieved by minimizing a criterion that takes into account the similarity between possible partitions relative to the similarity within each partition [3, 16, 19]. Building a global cost criterion, capturing salient relationships among the image elements, and making it's optimization computationally tractable continues to be a difficult problem.

Another category of algorithms seeks optimal image hierarchic partitioning through a sequence of local computations. We emphasize approaches that are based on the use of proximity graphs, specifically Delaunay triangulation and Minimum Spanning Tree (MST). In contrast with global optimization approaches, the MST based image segmentation seeks image partitioning by iteratively linking image elements through the lowest cost tree edges. The tree edges represent similarity of neighboring elements. One of earliest applications of tree-based data clustering to visual like point data sets analyzed MST edges histogram and investigated tree characteristics such as MST "relative compactness", tree diameter, and point densities [20]. In [9] the tree-based concept was applied to image segmentation. It was suggested to use global homogeneity criterion to control construction of an irregular pyramid starting from a regularly sampled pixel grid. [12, 10] use irregular tessellations to generate an adaptive multi-scale image representation. The approach employs an irregular sampling of the pixel grid to build the initial (lower) 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. [18] uses Kruskal's algorithm to construct MST of the image from a regularly sampled pixel grid. The tree is then partitioned into subtrees, representing different homogeneous regions. [5] uses Kruskal's algorithm to construct MST by evaluating local variations of the neighboring pixel patches. The approach uses global constant that controls how the pairwise regions variation and regions' internal variations impact the region grouping. The approach [6, 7, 11] is similar to [5] in how it controls the grouping of pixel patches into larger ones. At the same time, in contrast with [5], the approach in [6, 7, 11] uses Boruvka's MST construction algorithm and produces an irregular hierarchy of image segmentations. The last property is especially useful for pattern recognition. It makes it possible to look for objects of interest across multiple layers of the built hierarchy, employing either top-down or bottom-up approach.

Our framework belongs to the image segmentation approaches, producing an irregular image pyramids. The framework derives a hierarchy of irregular polygonal patches adapted to the image content. The hierarchy is built based on the edges detected in the image. This adaptive polygon-based hierarchic image segmentation distinguishes our method from the previous tree-based image segmentation approaches. The framework combines constrained Delaunay triangulation, the Gestalt principles of visual perception, such as proximity and closure, exploits structural information on spectrally detected image edges and their spatial relations. This produces initial set of polygonal patches; they are iteratively grouped bottom-up using Boruvka's MST algorithm. We show experimental results and discuss opportunities to improve the proposed approach.

2 Polygon-Based Multi-Scale Image Segmentation

In a polygon-based pyramid, each level represents a polygonal tessellation of the image. The pyramid is built iteratively 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. The built pyramid P is described as a set of graphs G_l representing the image in a fine-to-coarse hierarchy.

2.1 Construction of Fine Level of Detail

Polygons on the lowest level of a pyramid are built upon the triangular tessellation of the image. We employ the image vectorization approach [13] to process the generated triangle grid. First, we detect edges in the image, e.g using Canny edge detector [4] (Figs. 1a, 1b). It is followed by constrained Delaunay triangulation (CDT) [15] where the detected edges are used as constraints for the triangulation (Fig. 1c). Thus, the CDT tessellation grid is adapted to the image content, since triangle vertices and edges reflect the structure and spatial adjacency of the detected edges, such as Canny edges. CDT generated triangular mesh is processed by edge filtering. The filtering keeps constraints (such as Canny edges) and selectively deletes generated triangle edges. Triangle edge filtering is based on a prespecified set of rules inspired by the principles of visual perception, such as proximity and closure. Proximity filters out triangle edges using thresholding based on edge length (Fig. 1e). As a result, the detected edges that are spatially close to each other are linked by the triangle edge. Otherwise, the spectral edges remain separated. The closure rule is responsible for filtering out triangle edges which are bounded by the same spectral edge (e.g., having "U"-shape) or the same pair of spectral edges (e.g., "| |" configuration) (Fig. 1e). This results in a set of closed contours consisting of combination of the generated triangle 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 (Fig. 1d). These polygons are assigned median color based on a sampling of pixels covered by the grouped triangles. Thus the image is segmented in a set of spectrally attributed polygonal patches (Fig. 2). The constructed polygon boundaries are built upon the spectral edges, and thus reflect important discontinuities in the image characterization. In turn, agglomeration of polygons across multiple levels of the image pyramid P will be implicitly directed in the sense that boundaries of agglomerated polygons will also be authentic to the image spectral discontinuities. This is in contrast with other approaches, where selection of good seed pixels or pixel patches is quite challenging. The problem is due to the fact that pixel itself does not carry any object-oriented information.

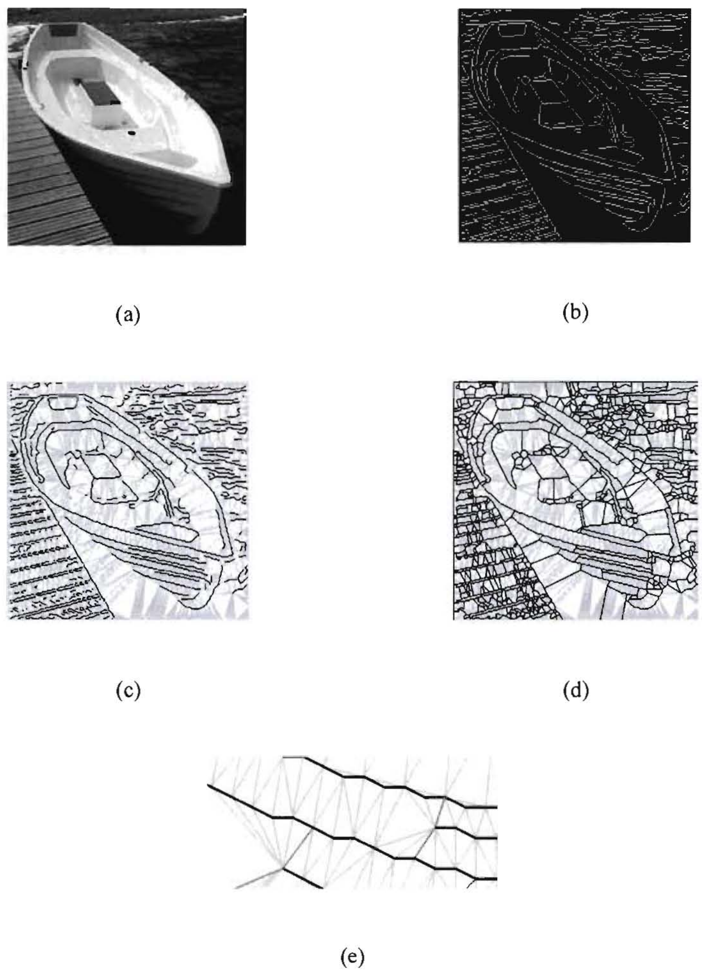


Fig. 1. Construction of a polygon-based image representation on fine level of detail of a pyramid. (a) an original image. (b) edge detection using Canny edge detector. (c) generation of an irregular triangular mesh based on Constrained Delaunay triangulation (Canny edges are shown in black color, triangle edges are shown in gray color). (d) polygon creation by filtering the triangle edges and creating closed contours (contours are shown in black color, deleted triangle edges are shown in gray color). (e) an example of triangle edges filtering. Canny edges are shown in black color, deleted triangle edges are shown in light gray color, and the kept triangle edges (closing the gaps between Canny edges) are shown in darker gray color

(a) (b)

Fig. 2. Result of segmentation of the image shown in Fig. 1(a): (a) Contours of the created polygons are superimposed on the original image and shown in gray. (b) Created polygons are shown with their colors estimated during the segmentation process

2.2 Pyramid Construction

Once the polygon-based image representation on the lowest level of a pyramid is produced, we iteratively group polygons, sharing their contour fragments, on level l into larger polygonal chunks, producing level $(l+1)$ of the image pyramid. Polygon agglomeration is based on Boruvka's algorithm to construct Minimum Spanning Tree. 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. Boruvka's algorithm takes $O(E\log V)$ time, where E is number of edges and V is number of vertices.

The overall quality of segmentation depends on the pairwise polygon adjacency matrix, containing E_l. The attributes of edges are defined using two features: color similarity, ΔC_{ij}, and strength of the contour segment separating polygons, P_w. We evaluate the affinity w_{ij} between two neighboring polygons i and j:

$$w_{ij} = k \cdot \Delta C_{ij} \cdot \exp\left({P_w}/{\sigma} \right), \tag{1}$$

$$P_w = \sum_{k=1}^{N} \frac{s_k}{S} \cdot m_k . \tag{2}$$

where contour segment shared by two neighboring polygons consists of N edge fragments (Fig. 3), S is the length of the shared contour segment, s_k is the length of the shared edge fragment belonging to a given contour fragment, k and σ control the scale of polygons similarity, and m_k is the magnitude of the shared edge segment. m_k is 0 for

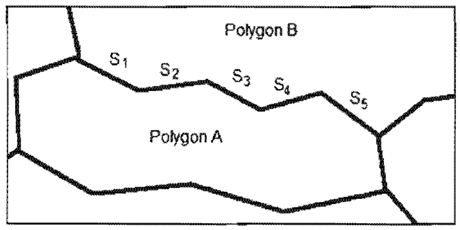
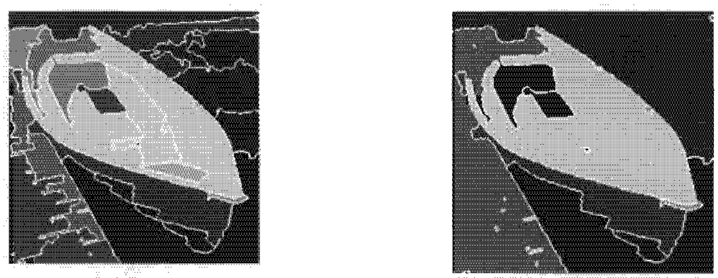


Fig. 3. An example of the contour fragment shared by two neighboring polygons *A* and *B*. The shared contour fragment consists of five edge fragments of corresponding lengths s_1 through s_5.



(a) (b)

Fig. 4. Result of the multi-scale segmentation of the image shown in Fig. 1(a). Contours of the created polygons are shown in white color. Created polygons are shown with their estimated colors. (a) level of detail # 4, 39 polygons. (b) level of detail # 6, 22 polygons.

triangle edge and non-zero for spectrally detected edge (such as Canny edges). Thus the cost of merging two polygons separated only by triangle edges is less than the cost of merging polygons separated by spectrally detected edges.

The algorithm constructs level $(l+1)$ of a pyramid containing coarser image partitioning by running one Boruvka's stage on level l using the evaluated color and contour relations between the polygons. Once a coarser level is constructed, color characterization of agglomerated polygons on level $(l+1)$ is evaluated:

$$C_i^{l+1} = \sum_{k=1}^{M} \frac{A_k^l}{A_i^{l+1}} \cdot C_k^l . \tag{3}$$

where M is the number of polygons merged into polygon i on level $(l+1)$, C_i^l is the color of a polygon i on level l, A_i^l i is the area of a polygon i on level l.

Spatial layout of the polygons (their contours) has changed as well. As a result, the adjacency matrix corresponding to level $(l+1)$ is evaluated. The described process of generation of coarser level based on finer level of a pyramid iteratively goes (Fig. 4) while spectral dissimilarity between polygons is less than a predefined spectral dissimilarity threshold.

3 Experimental Results

For experiments we have used the Berkeley segmentation dataset[1]. Figures 5 and 6 show the results of image segmentation based on our method. For comparison we have included the results produced by normalized cuts based segmentation. We used publicly available normalized cuts software[2]. We have also included integrated representation of human sketches as a reference, since different subjects produced different segmentations for the same image. It can obviously be argued that to produce meaningful segmentation it is necessary to utilize object-based knowledge in addition to bottom-up information agglomeration. In this paper we focus on the bottom up approach to image segmentation.

Fine level of detail of a pyramid is constructed using Canny edge detection with the $\sigma_{Canny} = 1.$, hysteresis low threshold = 2.5, and hysteresis high threshold = 5; and we use the Triangle code[3] to generate triangular tessellation over the detected edge map. Color images were processed using either RGB space, or converting them into gray scale or HSI representations. The results are shown for RGB space. Global spectral threshold was set to 30. Scale parameters k and σ for adjacency relations were both set to 0.5. On average the produced hierarchies contained 6-11 levels of detail. We have also experimented with generalization of the adjacency relations by incorporating convexity measure of the neighboring polygons to give preference to a construction of more convex objects though it did not produce consistent results at current stage.

4 Conclusions

We have introduced a polygon based method to construct a fine-to-coarse hierarchy of irregular image partitions. Experimental results prove the validity of the proposed method. It uses spectral and contour relations between polygons as criteria of their agglomeration. Computational complexity of the algorithm makes it possible to use it for processing large images. Our method is different from other tree based image segmentation approaches because it builds an irregular hierarchy of image partitions based on triangular and polygonal tessellations of the image. These tessellations are built upon the spectrally detected edges. Thus the method preserves information about salient spectral discontinuities in the image, and the constructed image partitions are adapted to the image content. The built pyramid is unique if polygon relations are unique. When number of relations included into segmentation grows, it will more likely be the case. Additional research is required to generalize the method to process texture. The problem to address would be an adaptive evaluation of local texture properties, which could be associated with polygons. Another extension will include integration of the proposed method with top-down analysis to cue an analysis on polygonal patches, meeting a prespecified set of criteria (e.g., on shape).

---

[1] http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/

[2] http://www.cis.upenn.edu/~jshi/software/

[3] http://www.cs.cmu.edu/~quake/triangle.html

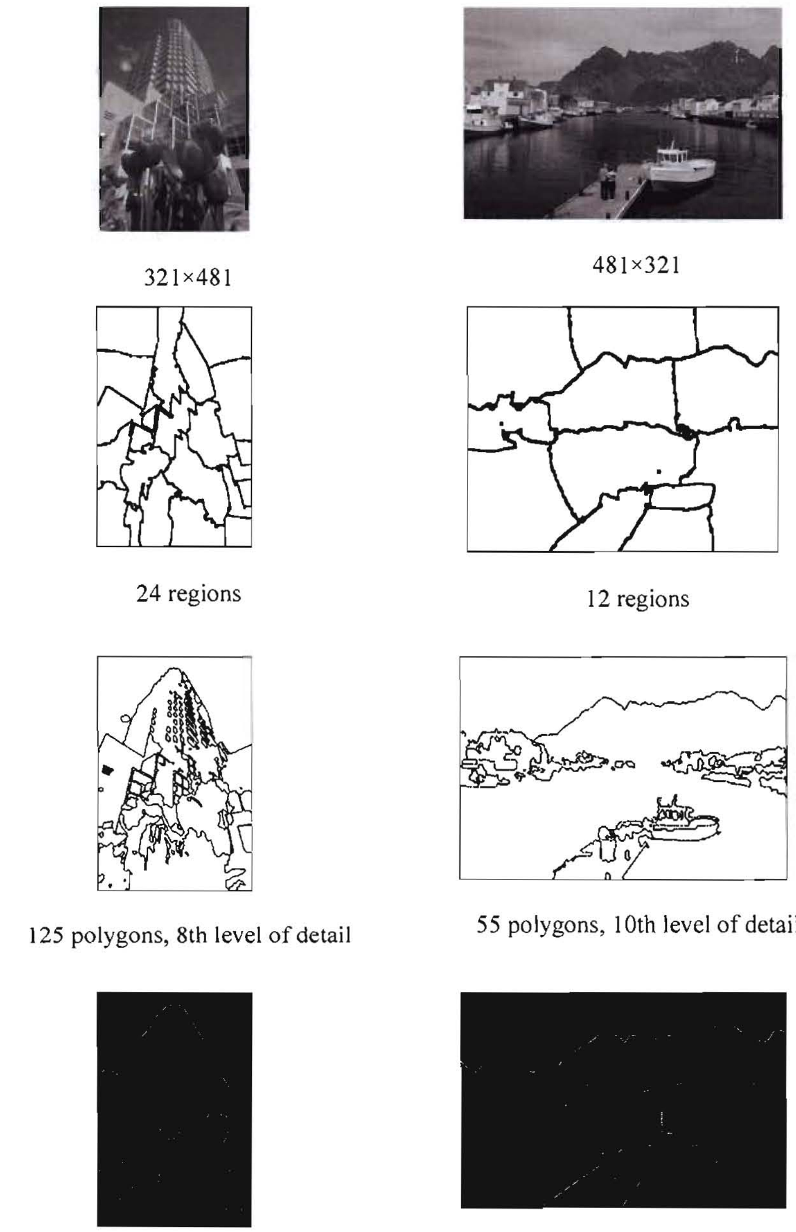


Fig. 5. Segmentation results. 1st row: input image (originally color images). 2nd row: results of the normalized cuts segmentation. 3rd row: results of our approach. 4th row: human sketches

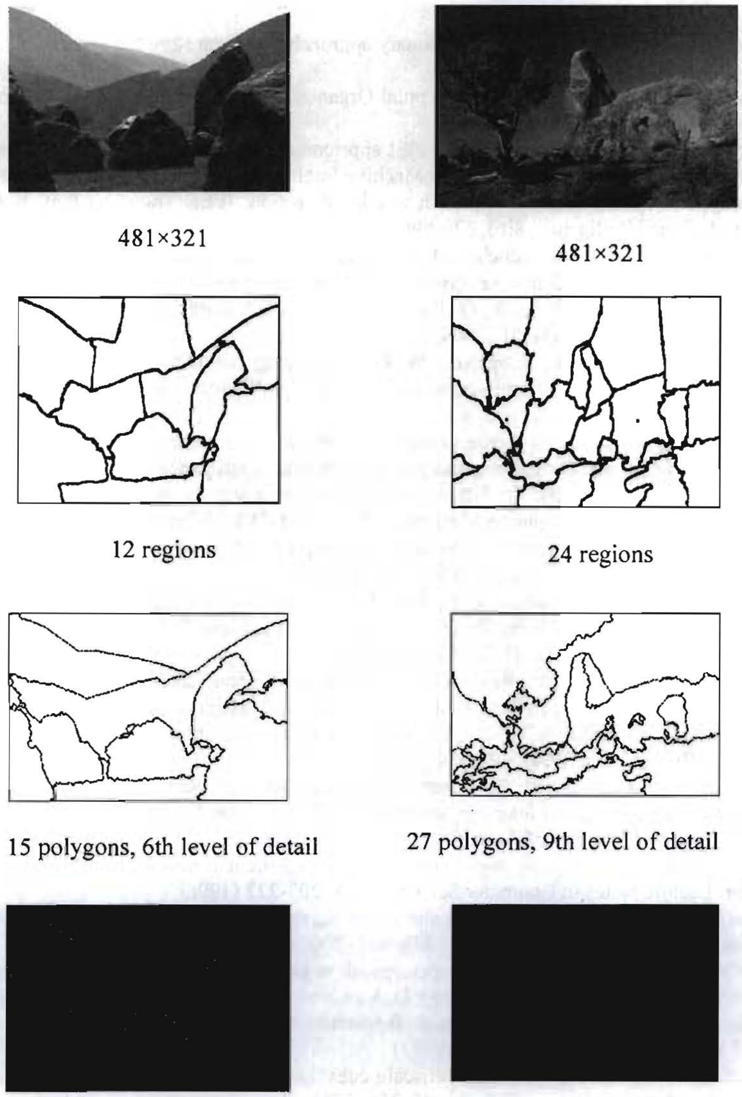


Fig. 6. Segmentation results (cont.). 1st row: input image (originally color images). 2nd row: results of the normalized cuts segmentation. 3rd row: results of our approach. 4th row: human sketches

References

1. Bhandarkar, S. M., Zeng, X.: Evolutionary approaches to figure-ground separation. Applied Intelligence, 11, 187-212 (1999)
2. Boyer, K.L., Sarkar, S., (Eds.): Perceptual Organization for Artificial Vision Systems, Kluwer Acad Publ. (2000)
3. Boykov, Y., Veksler, O., Zabin, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. On Pattern Analysis and Machine Intelligence, 23(11), 1222-1239 (2001)
4. Canny, J.: A computational approach to edge detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(6), 679-698 (1986)
5. Felzenszwalb, P. F., Huttenlocher, D. P.: Image segmentation using local variation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 98-104 (1998)
6. Haxhimusa, Y., Kropatsch, W. G.: Segmentation graph hierarchies. Lecture Notes in Computer Science, 3138, 343-351 (2004)
7. Haxhimusa, Y., Ion, A., Kropatsch, W. G.: Comparing hierarchies of segmentations: Humans, normalized cut, and minimum spanning tree. In: Digital Imaging and Pattern Recognition, 95-103 (2006)
8. Herault, L., Horaud, R.: Figure-ground discrimination: a combinatorial optimization approach. IEEE Trans. On Pattern Analysis and Machine Intelligence, 15(9), 899-914 (1993)
9. Horowitz, S. L., Pavlidis, T.: Picture segmentation by a tree traversal algorithm. Journal of the Association for Computing Machinery. 23(2), 368-388 (1976)
10. Jolion, J.-M., Montanvert, A.: The adaptive pyramid, a framework for 2D image analysis. CVGIP: Image Understanding, 55(3), 339-348 (1992)
11. Kropatsch, W. G., Haxhimusa, Y., Ion, A.: Multiresolution image segmentation in graph pyramids. In: Applied Graph Theory in Computer Vision and Pattern Recognition, Kandel, A., Bunke, H. H., Last, M. (Eds.) 52, 3-42 (2007)
12. Montanvert, A., Meer, P., Rosenfeld, A.: Hierarchical image analysis using irregular tesselations. IEEE Trans. on Pattern Analysis and Machine Intelligence, 13(4), 307-316 (1991)
13. Prasad, L., Skourikhine, A. N.: Vectorized image segmentation via trixel agglomeration. Pattern Recognition, 39(4), 501-514 (2006)
14. Sarkar, S., Soundararajan, P.: Supervised learning of large perceptual organization: graph spectral partitioning and learning automata. IEEE Trans. On Pattern Analysis and Machine Intelligence, 22(5), 504-525 (2000)
15. Shewchuk, J. R.: Triangle: engineering a 2D quality mesh generator and Delaunay triangulator. Lecture Notes in Computer Science, 1148, 203-222 (1996)
16. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(8), 888-905 (2000)
17. Wertheimer, M.: Principles of perceptual organization. In: Readings in Perception, Beardslee, D., Wertheimer, M., (Eds.), D. Van Nostrand, Princeton, NJ, 115-135 (1958)
18. Xu, Y., Uberbacher, E. C.: 2D image segmentation using minimum spanning trees. Image and Vision Computing, 15, 47-57 (1997)
19. Yu, S.X.: Segmentation using multiscale cues. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 1, 247-254 (2004)
20. Zahn, C. T.: Graph-theoretic methods for detecting and describing gestalt clusters. IEEE Transactions on Computers, 20(1), 68-86 (1971)