SUPER PIXEL EXTRACTION VIA CONVEXITY INDUCED BOUNDARY ADAPTATION

H. Emrah Tasli [1,2] *Cevahir Cigla* [1] *Theo Gevers* [2] *A. Aydin Alatan* [1]

[1] Electrical & Electronics Engineering Department, Middle East Technical University, Ankara, Turkey
[2]Faculty of Science, University of Amsterdam, The Netherlands

ABSTRACT

This study presents an efficient super-pixel extraction algorithm with major contributions to the state-of-the-art in terms of accuracy and computational complexity. Segmentation accuracy is improved through convexity constrained geodesic distance utilization; while computational efficiency is achieved by replacing complete region processing with boundary adaptation idea. Starting from the uniformly distributed rectangular equal-sized super-pixels, region boundaries are adapted to intensity edges iteratively by assigning boundary pixels to the most similar neighboring super-pixels. At each iteration, super-pixel regions are updated and hence progressively converging to compact pixel groups. Experimental results with state-of-the-art comparisons, validate the performance of the proposed technique in terms of both accuracy and speed.

***Index Terms*—** Super-pixel, image segmentation, convexity constraint

1. INTRODUCTION

Pixel representing of an image is often redundant due to the spatial similarity in the image. In order to reduce this redundancy, a preprocessing stage is introduced by Ren and Malik [1]. This method groups pixels into homogeneous image regions, called superpixels (SPs). The utilization of SPs has become important for image processing applications by providing an efficient representation. SP regions on the image possess similar color and texture characteristics. This supports the assumption that pixels in the same SP belong to the same semantic object or region. Inspiring from this idea, all the pixels in a SP can be assigned to specific models representing motion, depth or segmentation structures. This representation replaces the use of pixels in various applications [2] [3]. Moreover, by the utilization of SPs for image representation, the aim is to capture the inter-pixel details and preserve the local variations in the image. The proposed SP structure is also crucial for graph-based approaches. When the graph nodes are constructed with SPs instead of pixels, graph complexity and computation time would substantially reduce.

SP extraction involves four main challenges. Firstly, a successful method should preserve local structure by considering and adapting to the local object and region boundaries. Secondly, under-segmentation of the regions should be avoided for realizing an expressive image representation. Thirdly, regular region identification is targeted with quasi-uniform SP regions. Finally, computational complexity should be kept at minimum. The first two challenges are related with the local information encapsulation of the method that enforces adaptation of SP boundaries to the object boundaries. Uniform localization and compactness are required to form regular grid structure among graph models with unbiased neighbor relations. This property has an influence on the precision and accuracy of graph based solutions, especially in image segmentation problem. Computational efficiency is also important for practical usability of the method.

In this study, a novel and efficient SP extraction algorithm is presented addressing the four fundamental constraints. Local structure is preserved with the selected energy function; it is discussed in section 2.1. Adaptation on the object boundary is satisfied with a color based similarity measurement and the proposed distance metric takes care of the convexity constraint by penalizing irregular shaped regions. Computational efficiency is achieved by processing only pixels on the region boundaries.

The organization of the paper is as follows; related work is discussed in section 1.1, details of the proposed algorithm are presented in section 2. Section 3 is devoted to experimental results and the final part 4 concludes the study with final remarks and restatement of the contributions.

1.1. Related Work

The main challenges discussed for SP extraction have been previously addressed by various methods. As provided in [4], these methods can be classified in two groups: Graph-based [5], [6] and gradient-based solutions [7], [4]. In graph-based approaches, SP extraction is achieved by partitioning the graph where nodes correspond to individual pixels and edge weights are assigned according to a cost function relating inter-pixel similarities. In [5], the graph is partitioned recursively, as in Normalized Cuts segmentation [8], in order to minimize a global cost function based on color and texture cues until desired number of SPs is achieved. This approach satisfies the compactness constraint required for SPs in order to provide efficient graph representation. However, it suffers from computational complexity. In [9], SP extraction is improved in terms of complexity by grouping nodes of the graph via greedy decisions through pair-wise region comparisons on edge measures of minimum spanning-trees. This method, on the other hand, does not enforce a control on region compactness and number of SPs. In [10], a lattice structure is enforced by finding horizontal and vertical seams that cut the image optimally via graph cuts. The seams determine the SP boundaries considering region compactness and total SP number. A recent study [6] proposes a novel method to generate 2D SPs and 3D super-voxels in an energy minimization framework utilizing graph cuts. It provides various controls on the SP structure and distribution; however, it suffers from computational complexity during the optimization stage.

Gradient-based approaches start from initial seeds of rough SPs. Pixel groupings are refined iteratively, depending on the local similarities. Mean-Shift [7], which is one of the well known methods in image segmentation, is adapted for SP extraction by the use of recursive smoothing kernel over pixel feature space. The weakness of this method is that it does not have a control on the SP properties, such as compactness, distribution and total region number. In [11], an image is considered as a topographic structure and intensity gradient vectors are utilized to form pixel groups. This approach also lacks control on SP properties. TurboPixels [12] introduces geometric-flow over initial seeds which are considered as the starting points of the SPs. Level set method is exploited to update and refine SPs based on local image gradients. This approach enables regular distribution of compact SPs with less complexity compared to graph-based approaches. In [13], geodesic distance [14] is exploited to iteratively group neighboring pixels starting from the initial seeds as proposed in TurboPixels [12]. Utilization of geodesic distance enables higher structure sensitivity compared to geometric-flow with almost similar complexity. Initial seed placement in [12]-[13] is refined in [15] by rectangular shaped initial SPs. Instead of geometric-flow, boundary pixels are re-assigned to SPs iteratively, based on color similarity and spatial distance. In [4], a similar method is proposed, where all pixels are updated during the refinement rather than only boundary pixels. A recent study that aims to preserve the image topology is proposed in [16]. However, in this study quantitative experiments for evaluating the performance with the conventional metrics are missing, hence no comparison with this method has been presented.

2. PROPOSED METHOD

This study addresses the main challenges towards a successful SP extraction method; quasi-uniform distribution on the image, adaptation on the object boundary, and fast execution capability. For this purpose, iterative boundary refinement approach in [15] is improved by constructing a general framework that utilizes color and locational similarity for pixel label assignment.

Proposed approach involves three main steps: Initialization, boundary update and structure update, as illustrated in Figure 1. In the first step, the image is divided into equal-size regions according to the desired number of SPs. Each region initially starts with a rectangular shape and the centers are equally spaced among the image. The regular placement of SPs has been previously proposed where center pixels are considered as the seeds of pixel groups. Starting from these seeds, SPs are enlarged and the boundaries are constructed. However, this study approaches the problem from a different perspective. Instead of enlarging from the seed locations, SPs are refined through boundary pixels based on specific energy cost functions. The refinement is achieved iteratively through the boundary and structure update steps.

In the boundary update step, a greedy search is conducted only on the SP boundaries. During the update, the cost function relating similarity of the pixels to the corresponding SP candidates is minimized. This approach aims to keep SPs connected without any sub-detachment. Computational efficiency is realized by performing a search between boundary pixels and neighboring SP candidates. Label assignment of each boundary pixel is conducted in an eight-neighborhood search. Pixel to SP assignment is performed according to the formula given in (1), where $L(p)$ is the SP label of the pixel p, $S_i(p, Q_i)$ is the similarity cost between the corresponding pixel p and SP Q_i. N is the number of neighboring candidate SPs. Thus, starting from the initial SP distribution, boundary pixels are reassigned to the most similar neighboring SPs. After all the boundary pixels are visited, SP centers and mean color values are updated.

$$\mathcal{L}(p) = argmax_{i=1:N} \left(S^i(p, Q_i) \right) \tag{1}$$

During the structure update, the SP model (i.e. mean color values and SP centers) is recalculated based on the added or removed boundary pixels. This update provides pixel groups to adapt changes along the boundaries and converge to compact SP models. The boundary and structure update steps are iterated several times until a stopping criteria is met. Termination criteria can be set as a fixed number of iteration or it can be computed by the ratio of updated boundary pixels over the unchanged ones.

The changes in SP boundaries at different stages of the iteration are shown in Figure 2. SP boundaries are represented with blue lines, while yellow pixels denote the SP center locations. The initial distribution of the square shaped SPs with a uniform spacing is given in Figure 2.a. A greedy search is conducted on the region boundary among the neighboring SPs with respect to the update rule in (1). An intermediate SP distribution is shown in Figure 2.b. It is observed that the SP boundaries, as well as center locations, show powerful adaptation to local edges without losing their connectedness. At

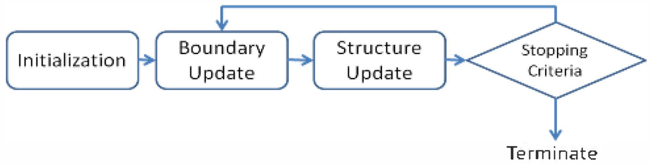

Fig. 1. Algorithmic flow of the proposed SP generation algorithm

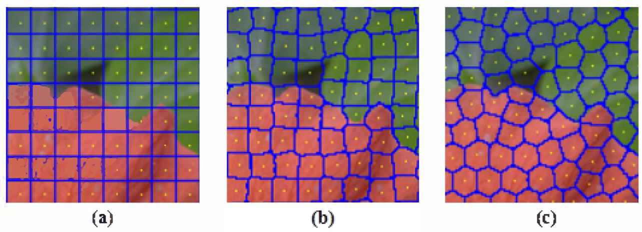

Fig. 2. Boundary and SP center update at different iterations; a) 0 b) 4 and c) 10

the final step in Figure 2.c, the iterative procedure is terminated due to small number of updated boundary pixels.

2.1. Energy Function for SP Generation

The optimization rule given in (1) updates the SP boundary. Each boundary pixel is visited and the cost of assigning each one to the neighboring SP is computed. The boundary pixel is assigned to the region that provides minimum cost. The proposed cost function used in this study is composed of two main energy terms (2). The first term relates the color similarity of the boundary pixel to its neighboring SPs. Second term defines the spatial distance of the pixel to the SP centers.

$$E(p, Q) = \lambda\, C(p, Q) + (1 - \lambda) D(p, Q^c) \tag{2}$$

The term λ in (2) is a trade off parameter to be tuned depending on the content and user selection. In this study the value of 0.5 is selected and used in all experiments. The first term in the cost function is used so that the pixels in the same SP region present a color similarity. In this study Lab color space is selected due to its perceptual uniformity. Color distance is computed over the individual color channels i as shown in (3).

$$C(p, Q) = \sum_{i=1}^{3} |\, p_i - Q_i|^2 \tag{3}$$

The spatial distance between the boundary pixel p and the SP centroid Q^c is computed using the geodesic metric. Geodesic distance is defined as the length of the shortest path from p to Q^c, as given in (4) [17]. Suppose $P = p_1, p_2, ...p_n = Q^c$ is a path between the pixels p_1 and $p_n = Q^c$ where p_i and p_{i+1} are connected neighbors. The

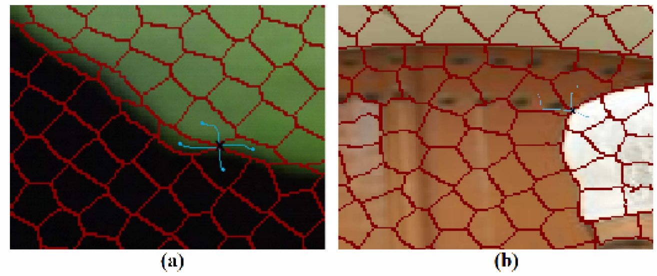

Fig. 3. Illustration of the shortest (geodesic) path between the SP centroid and the boundary pixel

path length $l(P)$, as defined in (5), is the sum of individual neighbor distances $d_N(p_i, p_{i+1})$ between adjacent points in the path.

$$D(p, Q^c)_G = min_{P=p_1,p_2,..,p_n} l(P) \tag{4}$$

$$l(P) = \sum_{i=1}^{n-1} d_N(p_i, p_{i+1}) \tag{5}$$

For the computation of adjacent pixel distance d_N, three color channel (Lab) distance is utilized (6). No significant performance difference has been observed in the selection of k, hence, it is selected as 1 in all the experiments due to its computational efficiency.

$$d_N(p, q) = \sum_{i=1}^{3} (p_i - q_i)^k \quad k = 1, 2 \tag{6}$$

Figure 3 illustrates the update procedure of a boundary pixel x on the junction of different SP neighborhoods. Neighboring SP centroids are marked with blue dots and the path printed in blue connecting the SP centroid to the boundary pixel indicates the shortest path. Computation of the shortest path from the boundary pixel to the SP centroid is performed via the shortest path algorithm provided by [18]. At each iteration, the shortest paths from the neighboring boundary pixels to the SP centroid are computed. Since the termination criteria for path computation is at the boundary, calculation of the shortest paths over the whole image is avoided.

The value of λ in (2) has a major impact on the SP size and shape. Visual results corresponding to different λ values are presented in Figure 4. As the contribution of the distance term is increased, SPs converge to a quasi-uniform distribution with increased convexity. If this ratio is further increased as in Figure 4.d, the resulting distribution becomes almost uniform and color homogeneity within SPs is violated. According to the visual interpretation of Figure 4, equal color and spatial distance weights are utilized throughout the experiments. However, a different weight selection might be preferred depending on the application and content.

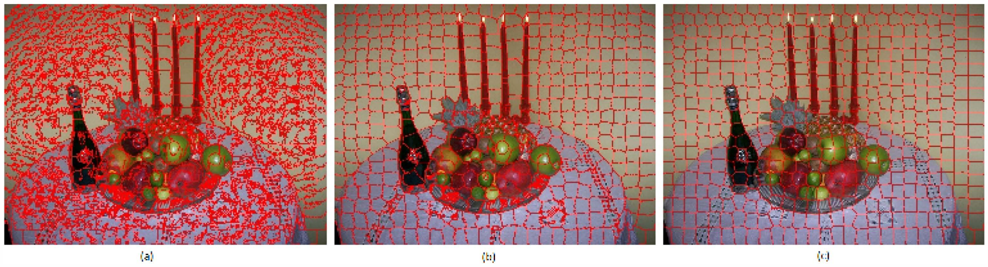

Fig. 4. SP boundaries under different convexity weights (a) λ=0.9, (b) λ=0.5, (c) λ=0.1

3. EXPERIMENTAL RESULTS

This section presents quantitative and qualitative results regarding the proposed SP extraction method in comparison with the state-of-the-art. The known methods in the literature are selected for performance evaluations: Graph-based [9], TurboPixels [5], Structure Sensitive Geo [13] and SLIC [4]. The performance of the extracted SPs is measured in terms of under-segmentation error, E_{UnSeg} and boundary-recall statistics. Under-segmentation error is calculated by measuring the ”bleeding” of the segment boundaries with respect to the ground truth (human) segmentation. Bleeding is measured by the formula in (7), where N corresponds to the number of pixels, L is the number of ground truth segments G_l, and S_j is the extracted SP. In (7), pixel area of a SP intersecting with the G_l is computed. B is selected to be equal to 5% throughout this study in order to compensate for small errors in ground truth segmentation data.

$$E_{UnSeg} = \frac{1}{N}\left(\sum_{l=1}^{L}\left(\sum_{[S_j|S_j\cap G_l>B]} Area(S_j)\right) - N\right) \tag{7}$$

E_{UnSeg} measures how well the extracted SPs fit the ground truth segment boundaries. The experiments are conducted on the Berkeley segmentation database [1] with the human segmentation results over 300 different images with a resolution of 481x321. All the presented under-segmentation errors are the average error over the whole dataset. The second error metric is the boundary recall and it is used to measure the percentage of overlap between the ground truth boundary pixels and the generated SP boundaries within one or two pixel neighborhood. Although this metric is itself inconclusive, it is widely used and clearly gives an idea about the boundary precision of the SP extraction algorithm. Different number of SPs are tested for performance evaluation to observe the performance of the algorithm. Measurements are performed on a 3.06GHz Intel Core i7 CPU with a 6 GB RAM. The source code of the proposed implementation will be made public in the authors' web page [1].

[1] Authors' Web page

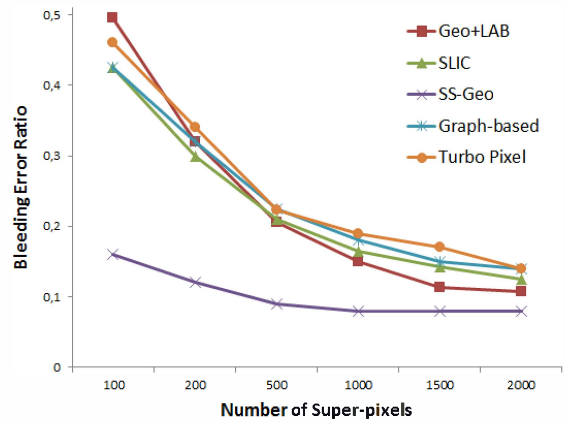

Fig. 5. Bleeding error comparison for different number of SPs

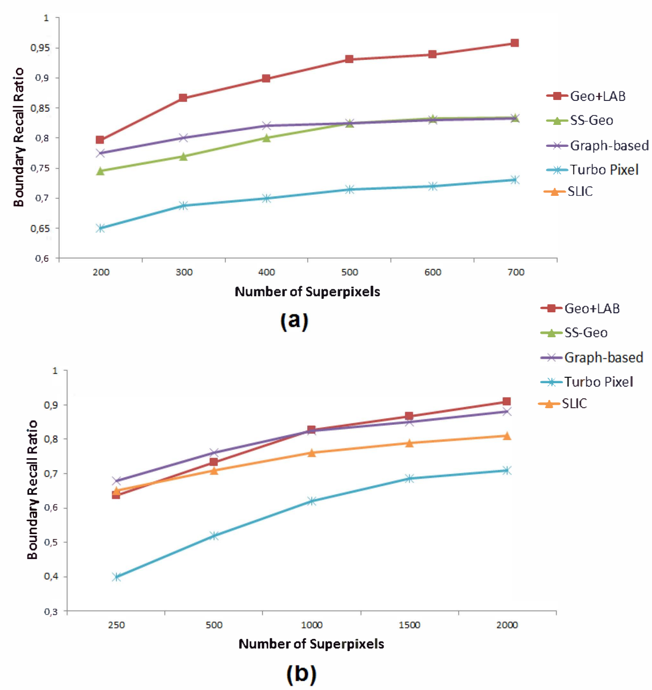

Fig. 6. Boundary-recall ratio for various number of SPs in (a) 2-pixel neighborhood, (b) 1-pixel neighborhood

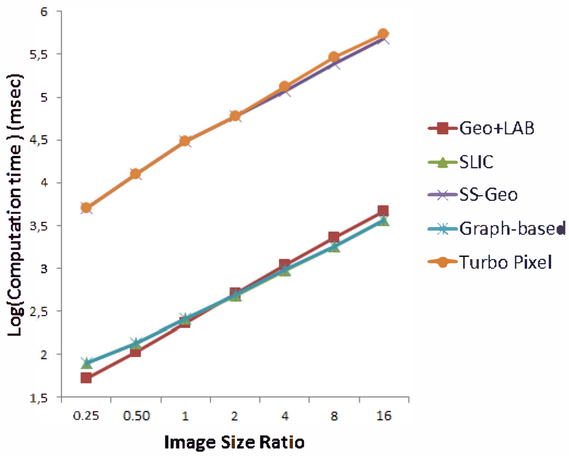

Fig. 7. Computation time comparison at different image sizes in logarithmic scale

3.1. Comparison against state-of-the-art Techniques

The quantitative performance evaluation of the proposed method against the state-of-the-art is done using the bleeding and boundary-recall metrics. Moreover, generated boundaries are also presented to have a visual evaluation. At this point, it is important to note that segmentation accuracy results of the state-of-the-art techniques are taken from their related references. The bleeding error ratio of the generated SPs provided by Graph-based [9], TurboPixels [12], Structure Sensitive Geo [13] and SLIC [4] and the proposed method is presented in Figure 5. It is observed that Structure Sensitive Geo algorithm [13] has the best performance in terms of under-segmentation error, which is followed by our proposed method, especially when the number of SPs is sufficiently high (≥ 500). SLIC is observed to perform better than the proposed method for smaller number of SPs.

Boundary-recall ratio measures the amount of match between the super pixel boundaries and the ground truth segmentation boundaries. This metric is prone to errors, since it is quite difficult to distinguish the actual boundaries in pixel precision. Hence, two versions of the metric are tested for measuring the ratio of boundary fit. The first one checks, whether the indicated SP boundary is within two-pixel neighborhood of the actual boundary. Similarly one-pixel neighborhood test is also conducted. According to the results presented in Figure 6.a and Figure 6.b the proposed method outperforms the state-of-the-art techniques.

Final quantitative comparison is conducted in terms of the computational times of the corresponding methods. In this case, the number of SPs is kept constant at 1000 and the images are scaled up and down using bi-cubic interpolation for different ratios of the original size. Seven different scales of the original image are used for measuring average running time of the methods. According to the results presented in logarithmic scale in Figure 7.a, TurboPixels [12] and Structure Sensitive Geo [13] require orders of magnitude longer execution times compared to SLIC [4], Graph-based [9] and the proposed approach.

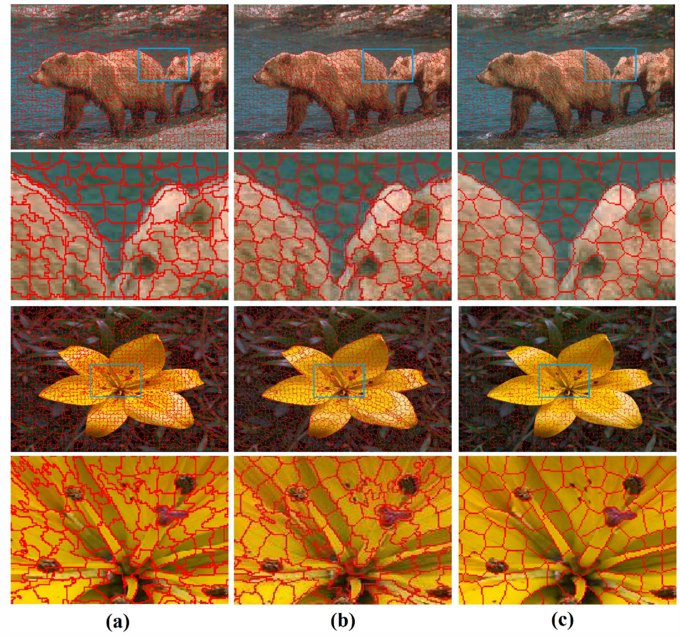

Fig. 8. SP boundaries (a) Proposed, (b) SLIC [4], (c) Turbo Pixel [12]

Apart from the quantitative comparisons, SP boundaries with the proposed method is presented in Figure 8 and Figure 9 for visual interpretation. It can be clearly observed that extracted SP boundaries fit object boundaries very well with inherent convexity and quasi-uniform distribution; and, hence providing competitive performance against the state-of-the-art.

4. CONCLUSION

This study presents a novel SP extraction method and improves the state-of-the-art results with respect to computational efficiency and segmentation accuracy. SP regions are iteratively updated based on color and spatial similarity. The boundary adaptation idea and energy function selection are the two main contributions of the proposed method. *Lab* color space is chosen for its perceptual uniformity and has proven success in combination with the proposed geodesic distance. According to the extensive quantitative evaluation with state-of-the-art, it can be concluded that the proposed scheme yields a remarkable alternative for SP extraction with faster execution time and competitive segmentation performance.

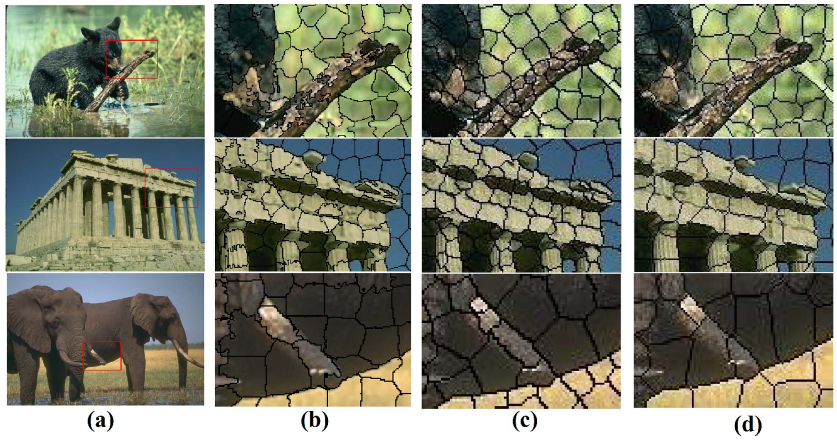

Fig. 9. (a) Original image, SP boundaries of (b)Proposed, (c)SS-Geo [17], and (d)Turbo Pixel [12]

5. REFERENCES

[1] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Computer Vision, 2003. Proceedings. 9th IEEE International Conference on*, Oct. 2003.

[2] H. E. Tasli and A. A. Alatan, "Interactive object segmentation for mono and stereo applications: Geodesic prior induced graph cut energy minimization," *ICCV Workshop Human Interaction on Comp. Vision*, 2011.

[3] Soatto S. Ayvaci, A., "Motion segmentation with occlusions on the superpixel graph," *Proc. of the Workshop on Dynamical Vision, Kyoto, Japan, 2009*.

[4] A. Lucchi, K. Smith, R. Achanta, V. Lepetit, and P. Fua, "A fully automated approach to segmentation of irregularly shaped cellular structures in em images," *International Conference on Medical Image Computing (MICCAI) 2010*.

[5] A. Levinshtein, S. J. Dickinson, and C. Sminchisescu, "Motion segmentation with occlusions on the superpixel graph," *IEEE International Conference on Computer Vision, 2009*.

[6] Olga Veksler, Yuri Boykov, and Paria Mehrani, "Superpixels and supervoxels in an energy optimization framework.," in *Perspectives in neural computing*, 2010.

[7] Dorin Comaniciu, Peter Meer, and Senior Member, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.

[8] Jianbo Shi and J. Malik, "Normalized cuts and image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

[9] P. Felzenswalb and D. Huttenlocher, "Efficient graph based image segmentation," *International Journal on Computer Vision, 2004.*

[10] Prince S. Warrell J. Mohammed U. Jones G Moore, A., "Efficient graph based image segmentation," *International Conference on Computer Vision and Pattern Recognition, 2008*.

[11] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, no. 6, pp. 583 –598, June 1991.

[12] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast Superpixels Using Geometric Flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[13] G. Zeng, P. Wang, J. Wang, R. Gan, and H. Zha, "Structure-sensitive superpixels via geodesic distance," *International Conference on Computer Vision, 2011*.

[14] G. Borgefors, "Distance transformations in digital images," *Computer Vision, Graphics and Image Processing,vol. 34, no. 3, pp. 344371, 1986*.

[15] C. Cigla and A. A. Alatan, "Efficient graph-based image segmentation via speeded-up turbo pixels," in *IEEE International Conference on Image Processing, 2010.*

[16] Dai Tang, Huazhu Fu, and Xiaochun Cao, "Topology preserved regular superpixel," *2012 IEEE International Conference on Multimedia and Expo.*

[17] Antonio Criminisi, Toby Sharp, and Andrew Blake, "Geos: Geodesic image segmentation," in *European Conference on Computer Vision*. 2008, Springer-Verlag.

[18] E. W Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik 1959*.