Pacific Graphics (2015)
N. J. Mitra, J. Stam, and K. Xu (Editors)

Short Paper

Superpixels Generation of RGB-D Images Based on Geodesic Distance

Xiao Pan[1], Yuanfeng Zhou[1]†, Shuwei Liu[2] and Caiming Zhang[1,3]

[1]School of Computer Science and Technology, Shandong University
[2]Center for Sectional and Imaging Anatomy, Shandong University
[3]Shandong Key Lab. of Digital Media Technology, Shandong University of Finance and Economics

Abstract

A novel algorithm for generating superpixels of RGB-D images is presented in this paper. A regular triangular mesh is constructed by the depth and a local geometric features sensitive initialization method is proposed for initializing seeds by a density function. Over-segmentation of the vertices on mesh can be generated by minimizing a new energy function defined by weighted geodesic distance which can be used for measuring the similarity of vertices with color information. At last, superpixels are generated by re-mapping the mesh over-segmentation to 2D image. During energy optimizing, we will check the topology correctness of the superpixels and refine the topology of the superpixels. Experiments on a large RGB-D images database show that the superpixels generated by the new method can adhere to the object boundaries well and outperform the state-of-the-art methods.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

As the rapid advancement of the depth information capturing equipment such as Time-of-flight (TOF) camera and Kinect, the acquisition of RGB-D image which contains a depth image and a color image becomes very easy. Over-segmentation of an RGB-D image can be a preprocessing of many applications, such as segmentation of RGB-D image [SF11, SSPW14], RGB-D video segmentation [WSC13] etc. Over-segmentation of RGB-D image is clustering the pixels on image to small regions. The important properties of superpixels are that it should be adhere to the objects boundaries. Some applications which construct the superpixels to graph [MB14] need the superpixels have fine topological structure like that there should be no hole in a superpixel and the barycenter of the superpixel should be inside the superpixel.

So far, many algorithms have been adopted to generate superpixels on 2D image. There are clustering-based methods like SLIC [ASS*12] and VCells [WW12], and graph-based approaches like graph-cut [VBM10] and N-cut [Mor05, MBLS01]. There are also some geometric based methods like Turbopixels [LSK*09] and Structure-sensitive method [WZG*13].There are also over-segmentation algorithms based on the RGB-D images. Some of them over-segment the RGB-D image by extending the 2D superpixel generation methods to the RGBD image, like depth adaptive method [WGB12]. The VCCS method [PASW13] segments the point cloud which generated based on the RGB-D image to supervoxels. The over-segmentation of image is also extended to mesh by Patricio Simari.et al [SPDF14] recently, which generates superfacets on triangular mesh. In this paper, we present a new over-segmentation method of RGB-D image. We construct triangular mesh on RGB-D image, and define a weighted geodesic distance on mesh. A new energy function is defined based on the geodesic distance and optimized by a Lloyd based iteration algorithm [Llo82] to get the over-segmentation of the mesh vertices. We also check and refine the superpixels' topological structure during the iteration process. The final superpixels are generated by re-mapping the over-segmentation results of mesh vertices to 2D image. Fig.1 shows an overview of our method. The main contributions of our paper include:

† Corresponding author: yfzhou@sdu.edu.cn

DOI: 10.2312/pg.20151284

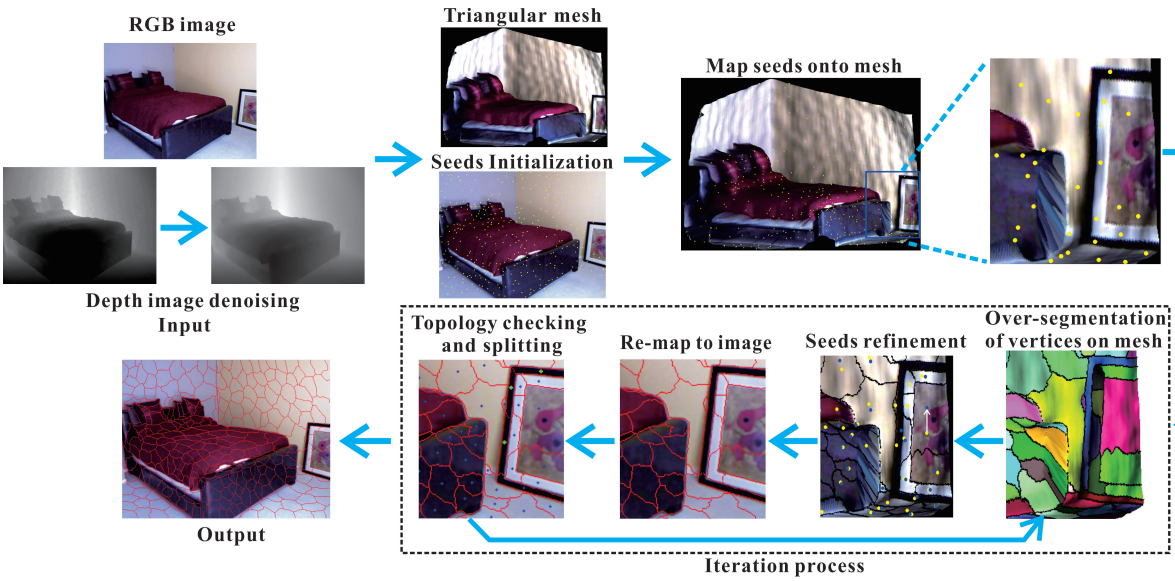


Figure 1: *Pipeline of the proposed method.*

- We define a weighted geodesic distance and an energy function to generate superpixels on the corresponding triangular mesh of RGB-D image;
- We present a simple and rapid splitting algorithm to refine the superpixels with bad topological structure;
- We propose a new seeds initialization scheme to generate geometric feature sensitive superpixels.

2. Pre-processing of superpixels generation

There is strong noise and corruption in the depth images captured by depth camera. We denoise the depth image using a smoothing strategy based on the point cloud's geometric structure combining with color information.

After getting the denoised depth image, we compute the 3D points based on the denoised depth image and construct the triangle mesh by the connected relationship between pixels in the 2D image. That is we connect every four adjacent vertices into two triangles. The edge belongs to two adjacent triangles should have opposite directions when we travel the three points in the triangles deasil or withershins.

3. Superpixels generation based on weighted geodesic distance of RGB-D images

Given a mesh $M(X)$ generated from a RGB-D image with color image $I(x)$ and depth image $D(x)$. X_i is the vertex on mesh corresponding to x_i in $I(x)$. $L(X)$ is the region label of vertex X on the mesh and $L(x)$ is the superpixel label of pixel x. The goal is to over-segment the mesh vertices to $\{R_l|l = 1...N_c\}$ such that $M(X) = \bigcup_{l=1}^{N_c} R_l$, and for $l \neq k, R_l \bigcap R_k = \emptyset$. And the corresponding superpixels on image is $\{S_l|l = 1...N_c\}$, N_c is the total number of seeds.

We define a weighted energy function based on the weighted geodesic distance. The triangular mesh is over-segmented by minimizing the energy function based on the Lloyd method.

3.1. Energy function definition

We define the weighted geodesic distance on mesh M as follows:

$$D_g(X_i,X_j) = \min_{P_{X_i,X_j}} \int_0^1 \omega(P_{X_i,X_j}(t))\|\dot{P}_{X_i,X_j}(t)\|dt \quad (1)$$

$$\omega(P_{X_i,X_j}(t)) = e^{E(x)/\sigma_1} \cdot e^{C(X)/\sigma_2} \quad (2)$$

where $P_{X_i,X_j}(t)$ is a path parameterized by $t = [0,1]$, connecting X_i and X_j respectively, $\dot{P}_{X_i,X_j}(t)$ is the first derivative of P_{X_i,X_j} at t. Similar to [LSK*09], $E(x) = \frac{\|\nabla I\|}{G_\sigma + \gamma}$, where $\|\nabla I\|$ is the color gradient of image $I(x)$, G_σ is a Gaussian smoothing function with standard deviation σ, and γ is a constant. $C(X) = 1 - \cos(\overline{\mathbf{n}_l}, \mathbf{n}_X)$, where $\mathbf{n}_X$ is the normal of X and $\overline{\mathbf{n}_l}$ is the seed's normal which is computed by averaging all the vertices' normal in R_l.

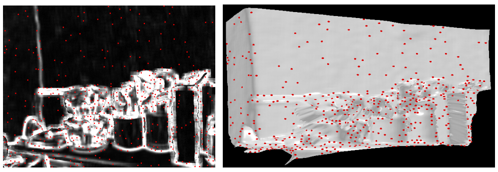
To reduce the influence of the noise, we define the energy function based on the weighted k-means model similar to [WZG*13]. The energy function is defined as:

$$E = \sum_{l} \int_{L(X)} W_{X,l} D_g(r_l, X) dX \tag{3}$$

where $W_{X,l} = e^{-\nabla I(x,l)/\sigma_1} \cdot e^{-C(X)/\sigma_2}$, $\nabla I(x,l)$ is the color distance between x and the average color of R_l. $C(X)$ is defined the same as in Eq.2 and r_l is the seeds position of R_l.

3.2. Local geometric feature sensitively seeds initialization

For every pixel and its neighbor pixels, we use a principal component analysis based method to estimate the best fitting plane. We define a density function $\rho(x)$ by the ratio between the minimum value of the three eigenvalues and the sum of other two eigenvalues generated by the principal component analysis. And using a threshold value to suppress the maximal density value to avoid placing too many seeds in the regions with sharp feature. We sample on the image using a method similar to Poisson disk sampling method [DW85] according to the probability density function as shown in Fig.2. And the minimum radius of the Poisson disk is defined as $R = 1/4\sqrt{N/N_c}$. N is the total number of the pixels in the image.



(a) (b)

Figure 2: *Initial seeds scheme. (a) Density map and initial seeds. (b) Map the seeds to the mesh .*

3.3. Over-segmentation based on Geodesic distance

We use the fast-marching method [KS98] to compute the numerical solution of the geodesic distance between the seeds and vertices. The basic idea of the fast-marching method is solving a Eikonal equation. Based on the definition of geodesic distance in Eq. 1, the weighted geodesic distance can be computed by the following Eikonal equation:

$$|\nabla D_g(r_l, X)| F(x) = 1 \text{ with } D_g(r_l, r_l) = 0, \forall l. \tag{4}$$

where F is the velocity function can be defined as:

$$F(x) = \omega(x)^{-1} \tag{5}$$

The label of vertex X is the label of the seed which is the first one to arrive it.

3.4. Seeds refinement

After getting the labels of every vertex, we refine the seeds' position. The $\nabla D_g(r_l, X)$ cannot be written explicitly, so we approach the $D_g(r_l, X)$ by $\nu\|X - r_l'\|$, where the $\|X - r_l'\|$ is the 3D Euclidean distance in camera coordinate system, and ν is a constant. Thus the new position of the seed point r_l' can be computed by minimizing the energy function through the Lloyd's algorithm:

$$r_l' = \frac{\sum_{X \in R_l} \frac{W_{X,l}}{\|X - r_l\|} X}{\sum_{X \in R_l} \frac{W_{X,l}}{\|X - r_l\|}} \tag{6}$$

Here r_l' computed by Eq. 6 may be not on the mesh. So we search the over-segmentation regions and find the vertex which has the minimal Euclidean distance to r_l' as the new seed point.

Superpixel splitting: According to the mesh construction process, we can easily project the over-segmentation result to image to get the superpixels. There may be holes in the superpixels or it doesn't have a good convex shape which lead to the result that it's barycenter is not inside the superpixel. We correct these superpixels' topological structure by splitting it.

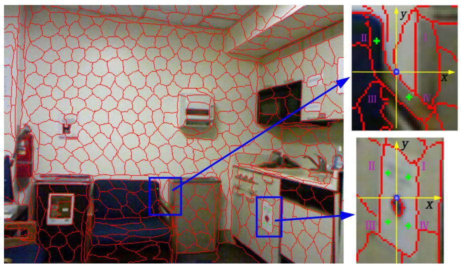


Figure 3: *Splitting of superpixels. Upper-right corner: the superpixel whose barycenter is not inside the superpixel is split to two new superpixels. Bottom-right corner: superpixel which has holes will split to four superpixels.*

For the superpixel which satisfies the splitting condition, We split it as follows: we set the barycenter of the superpixel as the original point, set the horizontal direction as the x-axis the vertical direction as the y-axis. Then the image space is divided to four quadrants as shown in Fig.3. We use S_l^i to denote the subset of S_l in the ith quadrant. If $|S_l^i| > \eta|S_l|$, we add the barycenter of S_l^i to the seed point list as a new seed point, where $|S_l|$ and $|S_l^i|$ are the number of pixels in S_l and S_l^i respectively. η is a constant, and here we set it as $3/16$. The splitting process is shown in Fig.3.

3.5. Termination condition and complexity analysis

We compute the difference value ∇E of the energy function with the previous iteration. So we terminate the optimization iterations if ∇E is smaller than a threshold or the number of iteration exceed the maximum iteration number and there is no new seed inserted in this iteration. The maximum iteration number is set to 10 in this paper.

The time complexity of the whole algorithm is $O(Nlog(N)+kN)$. Where $O(N)$ is for the Pre-processing and seeds updating process, and $O(Nlog(N))$ is for the fast-marching process.

4. Experimental results and comparisons

In order to evaluate the quality of the superpixels generated by the new method, we perform comparisons with the state-of-art RGB-D image over-segmentation methods using depth information such as VCCS [PASW13] and DASP [WGB12], and the superpixels generation algorithms on image plane such as SLIC [ASS*12], turbopixels [LSK*09] and structure-sensitive superpixels [WZG*13]. In order to illustrate the effectiveness of the new seed point initialization method, we compare both the uniformly seeds superpixels and the geometric information sensitive superpixels with other five methods. We also compare the superfacets generated on the constructed triangular mesh with the algorithm [SPDF14]. All the RGB-D data used in this paper come from NYU Depth Dataset V2 of Silberman et al. [NSF12], which contains 1449 pairs of aligned RGB and depth images and human annotated densely labeled ground truth.

4.1. Parameters setting

Our algorithm is not sensitive to most of the parameters. The σ_1 and σ_2 in Eq. 2 are set to be 0.1 and 0.05 respectively, and the σ_1 and σ_2 in Eq. 3 are set to be the same as they are in Eq. 2. The σ and γ in Eq. 2 are set to be 11 and 0.25.

4.2. Quantitative evaluation

Boundary recall: Boundary recall [LSK*09] is computed by calculate the fraction of the ground truth within a small disk shaped neighborhood of the superpixel's boundary, and in our experiments the disk radius is set to 2 pixels. The boundary recall comparisons between our method and other five methods are shown in Fig.4(a).

Under-segmentation Error: The under-segmentation error [LSK*09] is defined as:

$$U=\frac{1}{N}\left[\sum_{k=1}^{K}\left(\sum_{\{S_l\|S_l\cap G_k|>B\}} Area(S_l)\right)-N\right] \tag{7}$$

where $Area(S_l)$ is the area of the superpixel S_l, G_k is the ground truth segmentation and N is the total number of pixels. B is the minimum area of overlapping and is set to be 5% of $Area(S_l)$. The comparisons are shown in Fig.4(b).

Achievable Segmentation Accuracy: The achievable segmentation accuracy [NGL10] gives the highest accuracy achievable for object segmentation that utilizes superpixels as units. We plot the achievable segmentation accuracy of our method and the other five methods as shown in Fig.4(c).

4.3. Qualitative evaluation

Compare with superpixels generation method

Fig.5 shows a visual comparison of new method with the two RGB-D image over-segmentation algorithms VCCS [PASW13] and three algorithms which generate superpixels directly on 2D image SLIC [ASS*12], turbopixels [LSK*09] and structure sensitive superpixels [WZG*13]. From Fig.5, we can see that the superpixels generated by the new method can adhere to the objects boundaries better and are more compact.

Compare with the superfacets generation method

The new method can be used for over-segmenting the vertexes on mesh. We can generate superfacets by a postprocessing. If the three vertexes of a triangle have the same over-segmentation label l, the triangle is labeled to l. If the three vertexes of a triangle have different labels, then the triangle is labeled as one of the three vertices' labels which has the minimal normal distance with it. We test the effectiveness of the geometric information in the weight function of Eq. 2 by generating superpixels using only the geometric information. We compare the superpixels generated by this weight function with a superfacets generation algorithm [SPDF14], which over-segments the mesh to superfacets with only the aid of geometric information. Fig.6 shows both the superfacets results on mesh and the superpixels results on 2D image.

5. Conclusion and limitation

An efficient superpixels generation method of RGB-D images is presented in this paper. To utilize the spatial information of RGB-D images effectively, we map the pixels to 3D space based on the denoised depth image and construct the regular triangular mesh. Based on color and geometry information, a weighted geodesic distance is defined to measure the similarity of vertices on the mesh. Lloyd based optimization method is used for updating seeds of superpixels. From the experimental results, we can see the new method not only outperforms the existing methods on RGB images and RGB-D images, but also the superfacets generation method which is used on triangular mesh.

As we give a preprocessing process and a superpixel splitting scheme in the algorithm, the new method has a slightly higher time complexity than the existing RGB-D images over-segmentation methods. We intend to reduce the time complexity of the algorithm by accelerating the optimization process such as using GPU computation in the future.

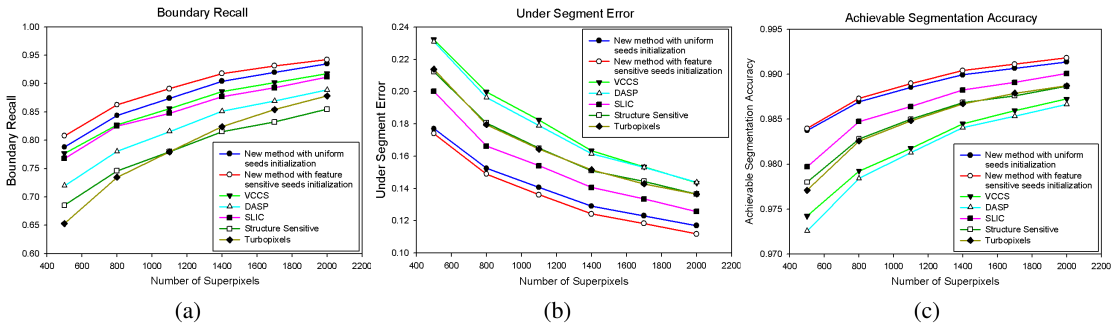


Figure 4: *Quantitative comparisons with other methods.*

Figure 5: *Comparisons with other superpixels results. The superpixels' number is 1100 (upper left) and 500 (low right) respectively. From left to right: DASP [WGB12], VCCS [PASW13], SLIC [ASS* 12], Turbopixels [LSK* 09], structure sensitive superpixels [WZG* 13], and new method.*

6. Acknowledgements

This work was supported by National Natural Science foundation of China (No.61202148 and No.61373078), the scientific research foundation of Shandong province of Outstanding Young Scientist Award (No.BS2013DX041), and the Fundamental Research Funds of Shandong University (2014JC028).

References

[ASS*12] Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Susstrunk S.: Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 34*, 11 (2012), 2274–2282. 1, 4, 5

[DW85] Dippé M. A., Wold E. H.: Antialiasing through stochastic sampling. *ACM Siggraph Computer Graphics 19*, 3 (1985), 69–78. 3

[KS98] Kimmel R., Sethian J. A.: Computing geodesic paths on manifolds. *Proceedings of the National Academy of Sciences 95*, 15 (1998), 8431–8435. 3

[Llo82] Lloyd S. P.: Least squares quantization in pcm. *Information Theory, IEEE Transactions on 28*, 2 (1982), 129–137. 1

[LSK*09] Levinshtein A., Stere A., Kutulakos K. N., Fleet D. J., Dickinson S. J., Siddiqi K.: Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 31*, 12 (2009), 2290–2297. 1, 2, 4, 5

[MB14] Muller A. C., Behnke S.: Learning depth-sensitive conditional random fields for semantic segmentation of rgb-d images. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on* (2014), IEEE, pp. 6232–6237. 1

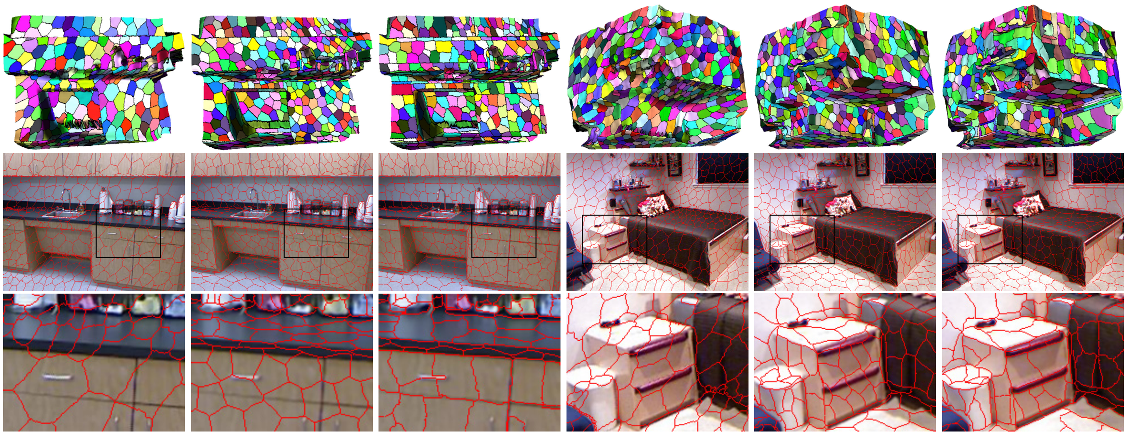


Figure 6: *Comparisons with the superfacets generation algorithm. First row: Results on the mesh. Second Row: Map the superfacets result to images and the third row: regions of interest by the black rectangles in the second row. (a)and(d): the results of the superfacets generation algorithm [SPDF14], (b) and (e) results of the new method with only geometric information in geodesic distance weight function, (c) and (f) results of the new method using color and geometric information.*

[MBLS01] MALIK J., BELONGIE S., LEUNG T., SHI J.: Contour and texture analysis for image segmentation. *International journal of computer vision 43*, 1 (2001), 7–27. 1

[Mor05] MORI G.: Guiding model search using segmentation. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (2005), vol. 2, IEEE, pp. 1417–1423. 1

[NGL10] NOWOZIN S., GEHLER P. V., LAMPERT C. H.: On parameter learning in crf-based approaches to object class image segmentation. In *Computer Vision–ECCV 2010*. Springer, 2010, pp. 98–111. 4

[NSF12] NATHAN SILBERMAN DEREK HOIEM P. K., FERGUS R.: Indoor segmentation and support inference from rgbd images. In *ECCV* (2012). 4

[PASW13] PAPON J., ABRAMOV A., SCHOELER M., WÖRGÖTTER F.: Voxel cloud connectivity segmentation - supervoxels for point clouds. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on* (june 2013). 1, 4, 5

[SF11] SILBERMAN N., FERGUS R.: Indoor scene segmentation using a structured light sensor. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (2011), IEEE, pp. 601–608. 1

[SPDF14] SIMARI P., PICCIAU G., DE FLORIANI L.: Fast and scalable mesh superfacets. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 181–190. 1, 4, 6

[SSPW14] STEIN S. C., SCHOELER M., PAPON J., WORGOTTER F.: Object partitioning using local convexity. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on* (2014), IEEE, pp. 304–311. 1

[VBM10] VEKSLER O., BOYKOV Y., MEHRANI P.: Superpixels and supervoxels in an energy optimization framework. *In Proc. European Conference on Computer Vision* (2010), 211–224. 1

[WGB12] WEIKERSDORFER D., GOSSOW D., BEETZ M.: Depth-adaptive superpixels. In *21st International Conference on Pattern Recognition* (2012). Accepted for publication. 1, 4, 5

[WSC13] WEIKERSDORFER D., SCHICK A., CREMERS D.: Depth-adaptive supervoxels for rgb-d video segmentation. In *ICIP* (2013), pp. 2708–2712. 1

[WW12] WANG J., WANG X.: Vcells: Simple and efficient superpixels using edge-weighted centroidal voronoi tessellations. *Pattern Analysis and Machine Intelligence 34*, 6 (2012), 1241–1247. 1

[WZG*13] WANG P., ZENG G., GAN R., WANG J., ZHA H.: Structure-sensitive superpixels via geodesic distance. *International journal of computer vision 103*, 1 (2013), 1–21. 1, 2, 4, 5