VCells: Simple and Efficient Superpixels Using Edge-Weighted Centroidal Voronoi Tessellations

Jie Wang
Adviser: Dr. Xiaoqiang Wang
Department of Scientific Computing
Florida State University

Abstract : Centroidal Voronoi tessellations (CVTs) are special Voronoi tessellations for which the generators of the tessellation are also the centers of mass (or means) of the Voronoi cells or clusters. Recently, we generalize the CVTs to the Edge-Weighted Centroidal Voronoi Tessellations (EWCVTs) by limiting the length of cluster boundaries. We apply this newly developed EWCVT for generating superpixels, which are in fact an oversegmentation of the image. EWCVT can segment the image and the clusters nicely preserve image local color difference. Moreover, the undersegmentation errors can be effectively limited in a controllable manner. It is well known that k-means algorithm can be used to generate Centroidal Vononoi tessellation. The simplicity and efficiency of k-means are well inherited by our EWCVT model. Meanwhile our new model is capable to generate high-quality superpixels on very complicated images. Even for megapixel sized images, EWCVT is able to generate the superpixels in a matter of seconds. We will provide the segmentation results on a wide range of complicated images. The simplicity and efficiency of our models will be demonstrated by complexity analysis, time and accuracy evaluations.

What is "SuperPixels"?

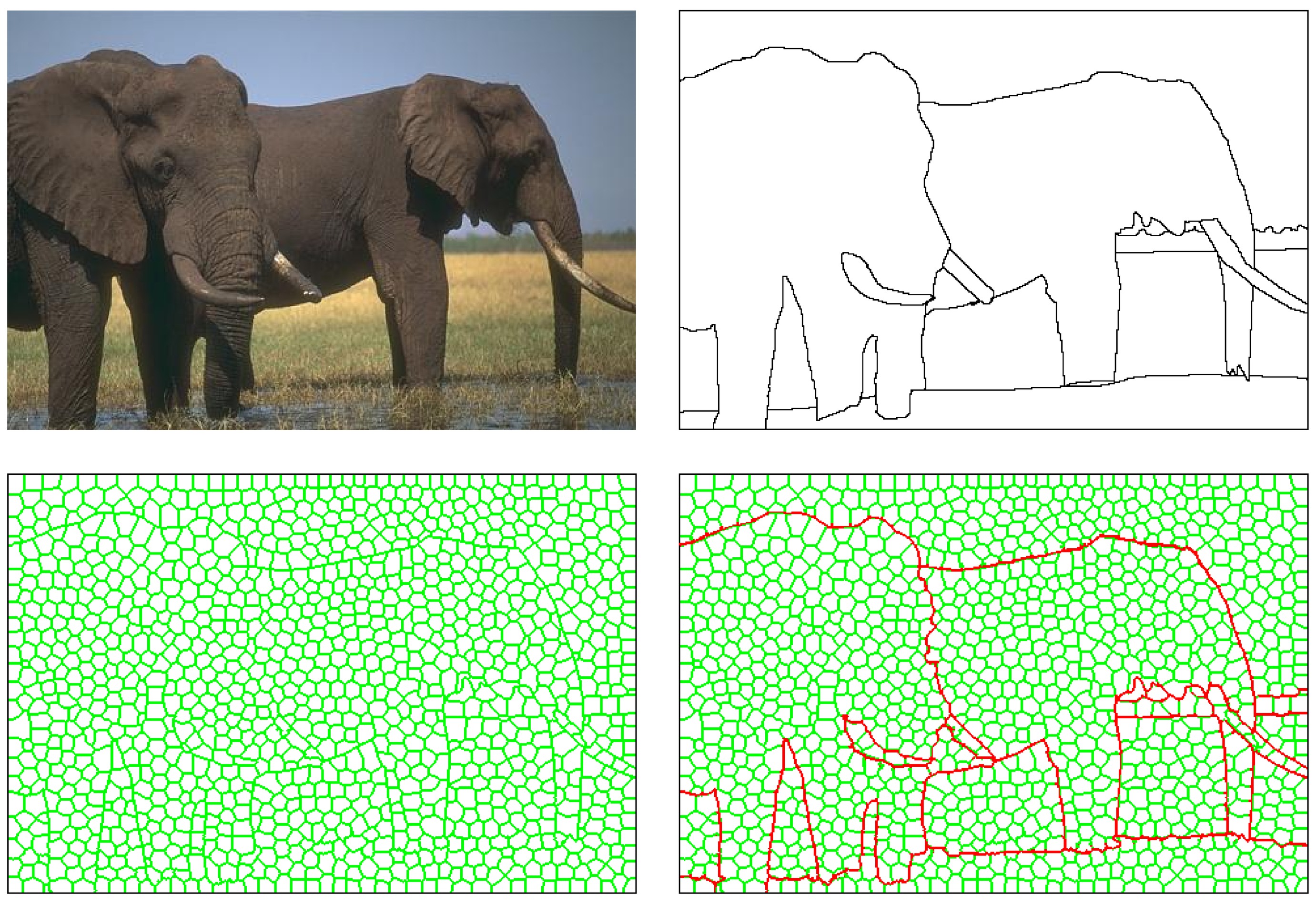

Figure 1: (elephant) Illustration of Superpixels. The upper left image is the original image. The upper right image is segmented by human. The lower left image is the superpixels obtained by our VCells algorithm. The lower right image shows a reconstruction of the human segmentation from the superpixels.

Each superpixel is in fact a group of individual pixels. Therefore, superpixels are in fact an oversegmentation of an image. Compared with the pixel-grid representation of the real scenes, X. Ren and J. Malik [1], suggests us that small segments obtained by pixel grouping process(*Superpixels*) might be a more natural and perceptually meaningful representation of the real scenes.

There are several basic principles to measure the quality of superpixels, i.e., *approximately uniform size and coverage*, *roughly constant intensities of each superpixel*, *connectivity*, *compactness*, *ability to extract and preserve boundaries* and *no overlapping*.

Another important and very useful advantage of superpixel is that the computational cost can be greatly reduced since the superpixel graph is obtained by contracting/grouping the pixel graph. Although superpixels greatly reduce the complexity of the image, we need the final segmentations in some sense conserve the local structure of the original image, i.e., avoiding or limiting the undersegmentation errors.

VCells Algorithm

Suppose we have a digital image $\mathbb{U} = \{u(i,j)\}_{(i,j)\in D}$ and a partition $\mathcal{P} = \{P_i\}_{i=1}^{n_c}$ of $\mathbb{U}$. The centroid of each P_k is given by

$$\tilde{g}_k^e = \frac{1}{\sharp P_k} \sum_{(i,j)\in P_k} u(i,j) \quad (1)$$

where $\sharp P_k$ is the number of pixels belong to P_k.

Given a weight λ we can define the so called *Edge-Weighted Distance* as a generalization of the classic Euclidean distance.

Definition 1. Suppose we have a digital image $\mathbb{U}$ and we are given a partition $\mathcal{P} = \{P_i\}_{i=1}^{n_c}$ and a set of colors $\mathcal{G} = \{g_i\}_{i=1}^{n_c}$ as generators. The *edge-weighted distance* from a pixel (i,j) to a generator g_k is defined as:

$$\text{dist}((i,j), g_k) = \sqrt{|u(i,j) - g_k|^2 + 2\lambda \tilde{n}_k(i,j)} \quad (2)$$

where $\tilde{n}_k(i,j) = \sharp\mathbb{N}_\omega(i,j) \setminus (P_k \cup (i,j))$, the number of pixels within $\mathbb{N}_\omega(i,j) \setminus (P_k \cup (i,j))$.

Algorithm EWCVT. Assume we have a digital image $\mathbb{U}$, we are given an integer n_c and an arbitrary partition $\{P_i\}_{i=1}^{n_c}$. Calculate the centroids $\{g_i\}_{i=1}^{n_c}$ of $\{P_i\}_{i=1}^{n_c}$ and take them as the generators.

1. For each pixel $(i,j) \in \mathcal{D}$,
   (a) evaluate the edge weighted distance defined in (2) from the pixel (i,j) to all the generators $\{g_i\}_{i=1}^{n_c}$;
   (b) transfer (i,j) to the cluster whose generator has the shortest distance to it, say, from P_l to P_m. Note the pixel transfer is equivalent to an update of $\mathcal{P}$;
   (c) replace g_l and g_m with the centroids of the modified clusters P_l and P_m respectively.
2. if no pixel transfer occured, return $\left(\{g_i\}_{i=1}^{n_c}; \{P_i\}_{i=1}^{n_c}\right)$ and exit the loop; otherwise, go to Step 1.

The VCells algorithm has two stages. The first stage is to divide the image into small segments with uniform size and shape; the second stage is to apply the EWCVT algorithm.

VCells Superpixel Algorithm: Suppose we have a digital image $\mathbb{U}$ and the number of superpixels n_c.

1. First Stage: Apply the classic k-means algorithm to segment the domain D of image $\mathbb{U}$ into n_c small regions which are hexagons with approximately the same size. These hexagons form a partition of $\mathbb{U}$. Denote these hexagons as $\mathcal{P} = \{P_i\}_{i=1}^{n_c}$.
2. Second Stage: Adopt $\mathcal{P}$ as the initial clusters and then set the centroids $\mathcal{G} = \{g_i\}_{i=1}^{n_c}$ of $\mathcal{P}$ as the generators. Apply the EWCVT algorithm by using $\mathcal{P}$ and $\mathcal{G}$ as initial configuration.

Undersegmentation Error and Boundary Recall

The *undersegmentation error* is calculated by using the formula

$$\frac{\left[\sum_{\{s_j | s_j \bigcap g_i \neq \phi\}} Area(s_j)\right] - Area(g_i)}{Area(g_i)}. \quad (3)$$

in which $\{g_i\}_{i=1}^{K}$ represents the segmentation of the ground truth image and $\{s_j\}_{j=1}^{L}$ denotes the superpixels produced by our algorithm.

The *boundary recall* is calculated by using the standard measure, i.e., the fraction of the ground truth boundaries which fall within a small disk shape(radius is set to be 2 pixels) neighborhood of the superpixels' boundaries.

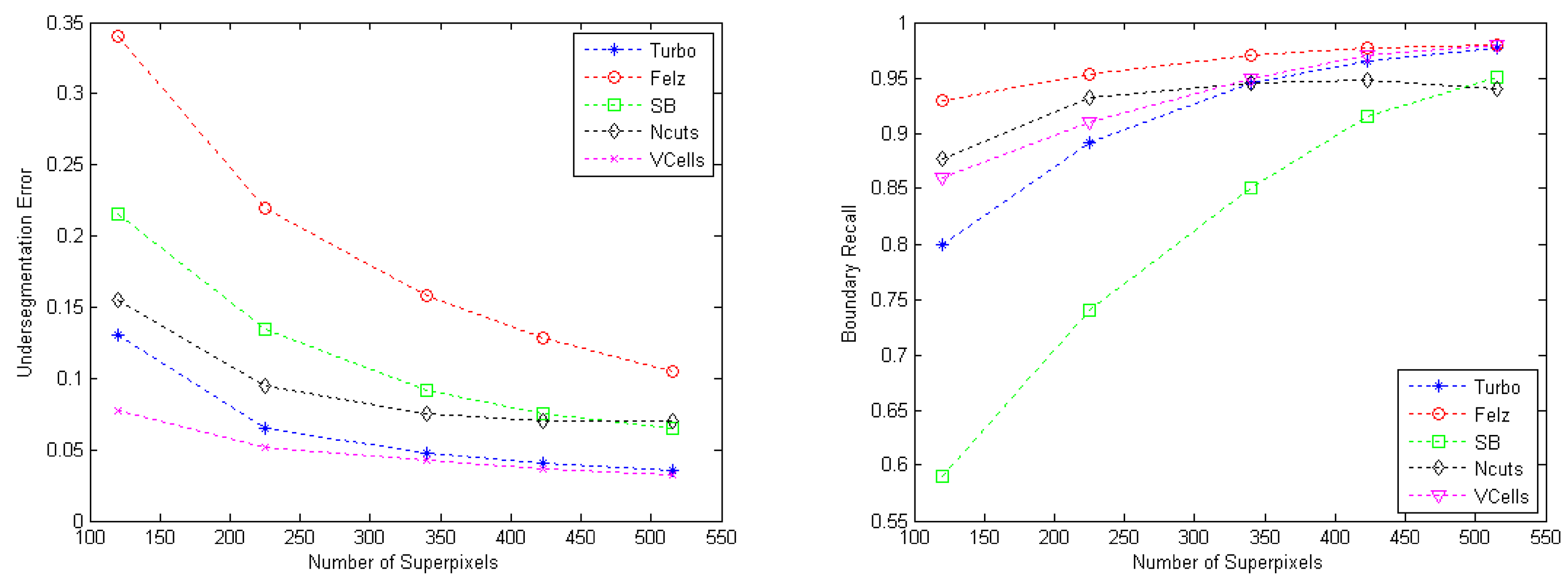

Figure 2: The left image is the undersegmentation error and the right one is the boundary recall. Both of the figures are plotted as functions of superpixel densities.

Time Evaluation

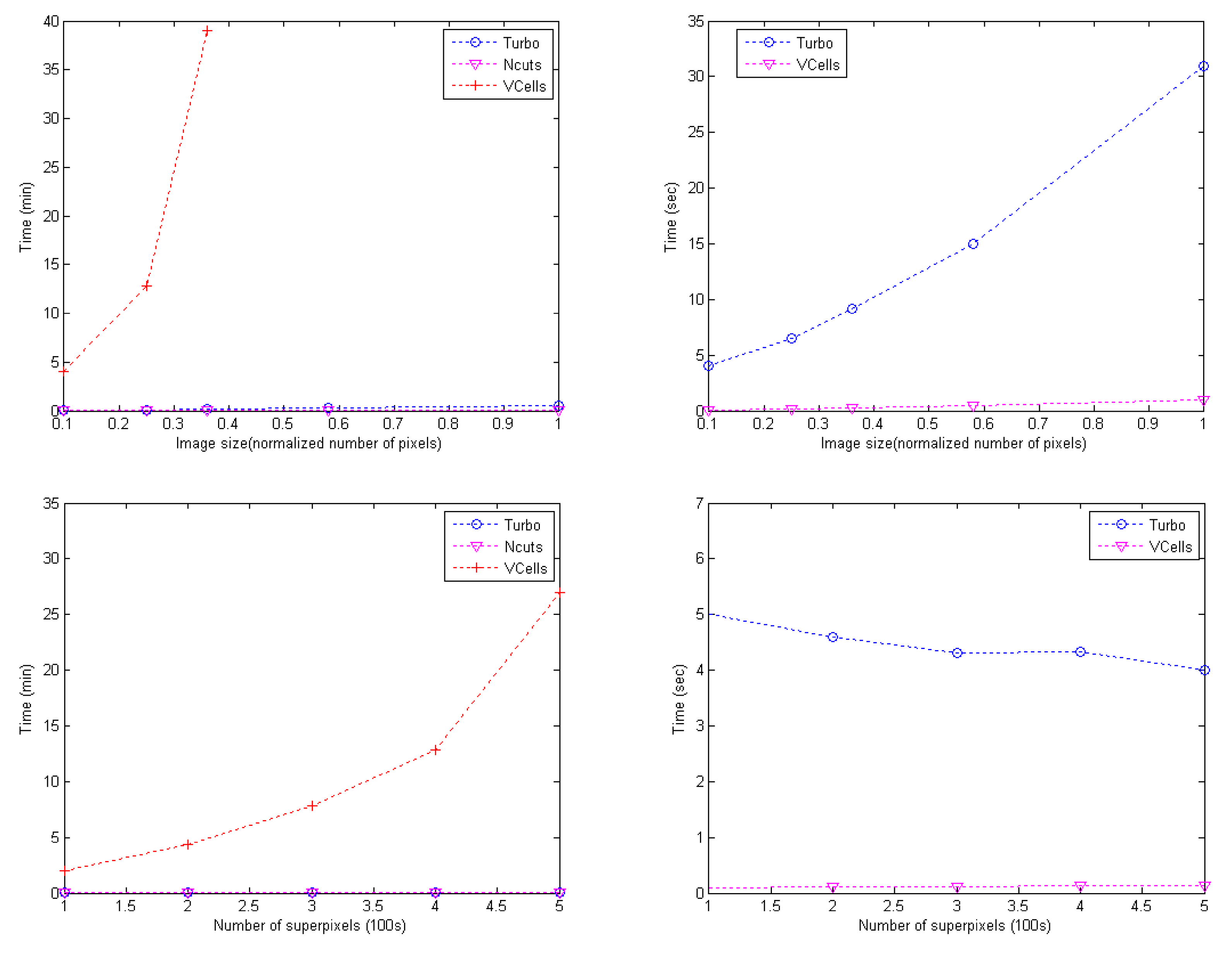

Figure 3: Comparison of running time of TurboPixels, N-Cuts and VCells. Right column is a close comparison of TurboPixels and VCells.

VCells for Texture Images

Texture images appear very frequently in the real world. In most of the cases, texture means some regular pattern which is of great interest to researchers. Thus, correctly extracting the texture from the image is usually desirable.

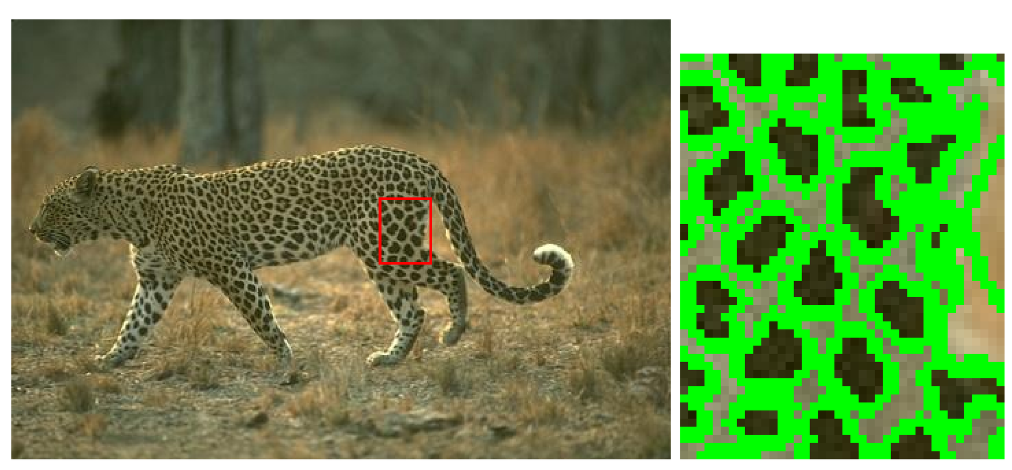

Figure 4: Superpixel results obtained by VCells for Texture Images.

For more information about *VCells*, please refer to [3] and [2].

References

[1] X. Ren and J. Malik, Learning a classification model for segmentation, *Proceedings of 9th international conference in Computer Vision*, **Vol 1**, pp. 10-17, 2003.

[2] J. Wang, L. Ju, and X. Wang, An Edge-Weighted Centridal Voronoi Tessellation Model For Image Segmentation, *IEEE Trans. Image Process.*, **18**, pp. 1844-1858, 2009.

[3] J. Wang and X. Wang, VCells: Simple and Efficient Superpixels Using Edge-Weighted Centroidal Voronoi Tessellations, *submitted to IEEE Trans. Pattern Analysis and Machine Intelligence*