Numer. Math. Theor. Meth. Appl. Vol. **3**, No. 2, pp. 223-244
doi: 10.4208/nmtma.2010.32s.7 May 2010

Edge-Weighted Centroidal Voronoi Tessellations

Jie Wang and Xiaoqiang Wang*

Department of Scientific Computing, Florida State University, Tallahassee, FL 532306-4120, USA.

Received 25 September 2009; Accepted (in revised version) 6 January 2010

Available online 8 March 2010

Abstract. Most existing applications of centroidal Voronoi tessellations (CVTs) lack consideration of the length of the cluster boundaries. In this paper we propose a new model and algorithms to produce segmentations which would minimize the total energy — a sum of the classic CVT energy and the weighted length of cluster boundaries. To distinguish it with the classic CVTs, we call it an Edge-Weighted CVT (EWCVT). The concept of EWCVT is expected to build a mathematical base for all CVT related data classifications with requirement of smoothness of the cluster boundaries. The EWCVT method is easy in implementation, fast in computation, and natural for any number of clusters.

AMS subject classifications: 52B55

Key words: Centroidal Voronoi tessellations, cluster boundary, edge detection, clustering, image processing.

1. Introduction

Centroidal Voronoi tessellations (CVTs) are special Voronoi tessellations whose *generators* are also the *centroids* of the associated Voronoi regions, with respect to a given density function. In the past few years, CVT-based methodologies have been applied successfully to diverse disciplines, including but not limited to, image processing and analysis [8, 14, 32], vector quantization and data analysis [17, 18], model reduction [12], high-quality point sampling [23], meshless computing [13], mesh generation and optimization [1, 16, 20], numerical partial differential equations [11, 21], and computer graphics and vision [18, 27]. The application list is still growing.

Lots of applications of clustering require the cluster boundaries to be smooth, while keeping the total length of the boundaries to be as small as possible. For example, we often obtain zigzag cluster boundaries when the basic CVT technique is applied to classify data sets. Especially in image segmentation, those zigzag boundaries are mainly due to noises or natural properties of the images, smoothing the boundaries can help us reduce or even eliminate the noises or unnecessary details [32].

*Corresponding author. *Email addresses:* `jw05j@fsu.edu` (J. Wang), `wwang3@fsu.edu` (X. Wang)

The popular level set method provides us a classic method to partially solve this problem [2–7, 9, 25, 28, 31]. However, it becomes very complicated when dealing with more than two clusters since multi-phase level sets have to be considered [31] in this case. In contrast, the CVT based techniques do not cause significant increase of computational cost due to increment of the number of clusters. In fact, we successfully developed an improved CVTs model for image segmentation recently and obtained very satisfactory results [32]. Our new model produces smooth boundaries in a controllable manner. More important, our model can be easily generalized to handle the cases of more than two clusters without introducing much difficulty in both theoretical and computational considerations. The key idea of our new model is to introduce a new energy term related to the *cluster boundaries*. And thus we call our model an Edge Weighted Centroidal Voronoi Tessellations (EWCVT) model. In this paper, we apply the similar idea to general data spaces, not restricted to image and even not restricted to 2D data set. The main contribution of our work here is to build a mathematical base for all CVT related data classification/clustering with requirement of smoothness of the cluster boundaries.

Here we would like to emphasize the differences between the method proposed in this paper and the one discussed in our another paper [32]. These two methods have the same name — EWCVT, because they both from the same idea: adding the weighted edge length to the classic CVT energy. In [32] the edge length considered is from physical space while the CVT energy is specifically from the color space (the image intensity), however, this paper handles the edge length and the CVT energy in the same physical space. Another major difference is that the method in [32] is designed only for 2D images, while the method in this paper can handle any dimensional data clustering problems. One may think about these two papers in this way: suppose we have a box filled by several soap bubbles, paper [32] takes a photo and tries to divide them by the color difference, this paper tries to simulate them by calculating their occupancy and surface tension.

We organize our paper as follows. First, we give a brief review of the classic CVT models and related algorithms in Section 2. In Section 3, the new EWCVT model and corresponding implementation algorithms are carefully developed. Together with some analysis and discussions, extensive numerical examples are presented to demonstrate special features of the EWCVT model in Section 4. Concluding remarks are finally given in Section 5.

2. Review of Classic Centroidal Voronoi Tessellations

2.1. Basic definition

Generally speaking, the computational domain is an open subset of $\mathbb{R}^n$, say Ω. A tessellation of Ω is in fact a non-overlapping covering $\mathscr{V} = \{V_l\}_{l=1}^{L}$ of Ω. Rigorously, we require $V_i \cap V_j = \emptyset$ if $i \neq j$ and $\overline{\Omega} = \cup_{l=1}^{L} \overline{V}_l$. The *Voronoi* region V_k of Ω can be easily computed once we are given a set of points $\mathscr{Z} = \{z_l\}_{l=1}^{L}$ according to

$$V_k = \Big\{x \in \Omega : |x - z_k| \leq |x - z_l|, \ \text{for } \ l = 1, \cdots, L\Big\}, \qquad k = 1, \cdots, L. \tag{2.1}$$