Nonparametric Density Estimation on A Graph: Learning Framework, Fast Approximation and Application in Image Segmentation*

Zhiding Yu Oscar C. Au Ketan Tang Chunjing Xu†
Dept. of Electronic and Computer Engineering
Hong Kong University of Science and Technology
{zdyu, eeau, tkt}@ust.hk
†Shenzhen Inst. of Advanced Technology
Chinese Academy of Sciences
cj.xu@siat.ac.cn

Abstract

We present a novel framework for tree-structure embedded density estimation and its fast approximation for mode seeking. The proposed method could find diverse applications in computer vision and feature space analysis. Given any undirected, connected and weighted graph, the density function is defined as a joint representation of the feature space and the distance domain on the graph's spanning tree. Since the distance domain of a tree is a constrained one, mode seeking can not be directly achieved by traditional mean shift in both domain. we address this problem by introducing node shifting with force competition and its fast approximation. Our work is closely related to the previous literature of nonparametric methods. One shall see, however, that the new formulation of this problem can lead to many advantages and new characteristics in its application, as will be illustrated later in this paper.

1. Introduction

Nonparametric density estimation provides a versatile tool for feature space analysis such as clustering and local maxima detection. The rationale behind, as pointed out by Comaniciu et al., is that "feature space can be regarded as the empirical probability density function (pdf) of the represented parameter." Finding local estimated density maxima (or mode seeking) results in the computational module of mean shiftv [1], an old pattern recognition technique. The robust nature of mean shift leads to wide applications in low level computer vision, including edge preserved smoothing, image segmentation and object tracking. Recent works tries to improve its performance by introducing asymmetric biased kernels in specific tasks, or seeks to reduce its complexity with fast algorithms.

*This work has been supported in part by the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China. (GRF 610109)

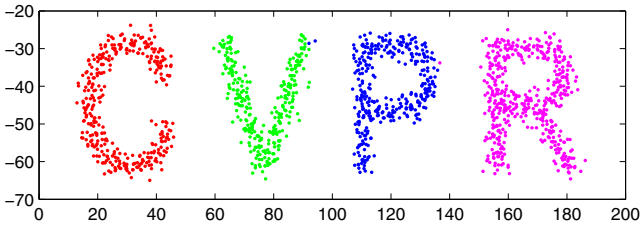


Figure 1. Example of data clustering using the proposed mode seeking algorithm with $h_1 = 180$ and $h_2 = 40$.

We investigate the problem of tree-structure embedded density estimation, providing a novel angle looking into this problem. Our method introduces metrics learned from a spanning tree into mode seeking. In particular, we adopt minimum spanning tree (MST) to learn compact structures in the feature space or on a connected graph. On one hand, the inclusion of MST helps to find manifold structures for feature space analysis and data clustering. On the other hand, the graph-based attribute works compatibly with regional level image operations in computer vision. A wide range of computer vision problems in principle requires regional support, where relation between image regions are typically depicted with a weighted graph and graph-based methods have consequently become a powerful tool. Such characteristic offers several intuitionally reasonable advantages. First, region-wise operation allows one to investigate and design more versatile and powerful features, as a region often contains much more information than a single pixel. Second, adopting region as basic processing unit can largely alleviate the computational burden.

In the paper, we only illustrate the applications of our method in data clustering and region-based image segmentation, due to the limit of page length. Figure 1 shows one example of data clustering using our proposed method. The potential application of this algorithm, however, is considerable, as mode seeking has diverse applications.

This paper is organized as follows: In Section 2, we briefly introduce the background and closely related works.

Readers already familiar with nonparametric density estimation and mean shift may jump to Section 3, where we describe the proposed method and discuss its important properties. Some experimental results regarding clustering and application of our method in image segmentation are illustrated in Section 4, showing that the method is an effective one. Finally, conclusions are made in the last Section.

2. Background and related works

Given a set of independent and identically distributed data points, nonparametric density estimation seeks to approximate its pdf. Instead of representing the pdf by a single parametric model or a mixture model, the method finds a small number of nearest (or most similar) training instances and interpolate from them. To obtain smooth pdf estimation, gaussian kernel is commonly utilized as the kernel density estimator, also known as Parzen window.

The paradigm of density estimation and clustering includes a family of mode seeking algorithms with Parzen density estimation. More recently, several works have explored the improvement of traditional mean shift algorithm. In [2], the author introduced asymmetric kernel to mean shift object tracking. The scale and orientation of the kernel is automatically and adaptively selected, depending on the observations at each iteration. In [3], A new mode seeking algorithm called the medoid shift was proposed. The purpose of medoid shift is to extend mode seeking to general metric spaces. The method, however, requires huge computational load and tends to result in over-fragmentation. It essentially becomes a finite point searching problem and is quite different from our method in terms of both purpose and algorithmic process. In [4], the authors proposed the quick shift algorithm which is considerably faster than mean shift and medoid shift. Their emphasis tends to concentrate on algorithm acceleration while preserving its performance. The GPU implementation of quick shift was discussed in [5] to further speed up the algorithm from the hardware perspective. There has also been other works trying to improve the efficiency of mode seeking [8].

Considering the nearest neighbor property of MST, our method to some extent are related to previous works that generalize mean shift to non-linear manifolds [9], or introduce nonlinear kernelized or manifold metrics [3, 4]. Our method can achieve some similar goals but the idea remains very different. We also notice there exist a great many works concerning MST based graph segmentations [10]. Even though our method have also utilized MST, we generally think it belongs to the family of mode seeking methods where the algorithm characteristics are quite different from many graph based segmentation methods. Hence these methods may not fall within the scope of comparison in this paper. In fact our work presents a general framework of embedding tree structures into the mode seeking process. Therefore it is straight forward for one to plug in many other trees and bring in additional algorithm characteristics.

3. Graph-based density estimation

We propose to perform density estimation on a joint domain represented by the node feature space and the distance space defined by the minimum spanning tree of that graph. There are several advantages operating on an MST-based structure. First, tree-based structure helps to uniquely define distances for any node pair, as a tree does not have circles. Of course, one could directly define the pairwise node distances in the Euclidean space, resulting in the traditional mean shift. But this basically discards the structural information preserved by a graph. In applications such as image segmentation, spatial information preserved by a graph can be very important. Second, an MST is the connected graph structure where all nodes are connected with least edges numbers and weights. In other words, an MST can be regarded as a "compact" structure that preserves important information about the cluster structure in a feature space. Although the introduction of a tree structure in practice could possibly be problematic - as it faces the risk of large tree structure variation induced by noise points, especially for those important tree roots - one shall see, the proposed method works pretty well and robustly in real image segmentation tests. In addition, such formulation helps to improve mode seeking performances for many manifold-shaped clusters.

There are several existing methods extracting an MST. In this paper, we adopt the Kruskal's Algorithm to obtain the MST structure from the graph. We then define the density function and describe its mode seeking process in the following part of this section.

3.1. Proposed density estimator

Given N samples represented by the set $\mathbf{V} = \{\mathbf{v}_i | i = 1, \ldots, N, \mathbf{v}_i \in \mathbf{R}^d\}$ and the undirected weighted graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, the minimum spanning tree $\mathbf{S} = (\mathbf{V}, \mathbf{E_S})$ is a connected graph of G with $\mathbf{E_S} \subseteq \mathbf{E}$, $|\mathbf{E_S}| = N - 1$. For any node pair (i, j) where $i \neq j$, there exists a unique path $\mathbf{E}_{ij}$ such that $\mathbf{E}_{ij} \subseteq \mathbf{E_S}$, i and j is connected by $\mathbf{E}_{ij}$ and deleting any element of the set results in the disconnection of i and j. In addition, we define $\mathbf{E}_{ij}$ to be $\emptyset$, if $i = j$.

Property 3.1 *For any given node pair (i, j), the set of connecting edges $\mathbf{E}_{ij}$ is unique.*

The above attribute comes directly from the tree structure. The proof is simple: if there is more than one $\mathbf{E_{ij}}$ then there exists at least one circle, which contradicts with the proposition. The unique distance definition on an MST facilitates the definition of density for a given location.

We propose to use a joint representation of the MST distance space (or MST space for short) and the feature space

to define the density estimator. Consider the simplest case where the MST space kernel center is located exactly at a tree node $\mathbf{v}_j$, then the density estimator can be written as follows:

$$f(\mathbf{v}) = c_0 \sum_i k\left(\frac{d(\mathbf{v}_j, \mathbf{v}_i)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right), \quad (1)$$

where $d(\mathbf{v}_j, \mathbf{v}_i) = \sum_{(\mathbf{v}_{k1}, \mathbf{v}_{k2}) \in \mathbf{E_{ij}}} ||\mathbf{v}_{k1} - \mathbf{v}_{k2}||$ is the cumulative weight of edges that connects the two nodes, $\mathbf{v}$ is the feature space kernel center, h_1 and h_2 are the bandwidth parameters controlling the window size and c_0 is a constant term determined by the sample size and bandwidth. $k(x)$ is the profile of a normal kernel:

$$k(x) = \exp(-\frac{1}{2}x). \quad (2)$$

To define a density estimator for any location on the MST space, we have to first define the branch of an MST node. Here by saying "any location" we actually allow the MST space kernel center to be located on an MST edge between neighboring nodes. In other words, the kernel can shift on the constrained space defined by MST. Suppose $\mathbf{v}_{neigh}$ is a neighboring node of $\mathbf{v}_i$, we have the following definition:

Definition 3.1 *The branch of a given tree node $\mathbf{v}_i$ with respect to its connected edge $(\mathbf{v}_i, \mathbf{v}_{neigh})$ is a set of nodes and edges $\mathbf{B} = (\mathbf{V_B}, \mathbf{E_B})$, such that $\mathbf{V_B} = \{\mathbf{v}_j | j \neq i, (\mathbf{v}_i, \mathbf{v}_{neigh}) \in \mathbf{E_{ij}}\}$, $\mathbf{E_B} = \{(\mathbf{v}_i, \mathbf{v}_j) | i \neq j, (\mathbf{v}_i, \mathbf{v}_{neigh}) \in \mathbf{E}_{ij}\}$.*

The branch of a node is an "induced subgraph" rooted at $\mathbf{v}_i$, and descending from its referenced connected edge. There exist at least one corresponding MST edge - denoted as e_{ref} - where the MST space kernel center is located on. If the center is located exactly on a tree node, then one may choose any edge connecting this node to one of its neighboring nodes as e_{ref}. Suppose that the two nodes connected by e_{ref} are respectively $\mathbf{v}_{ref1}$ and $\mathbf{v}_{ref2}$, and that the distances from the kernel center to $\mathbf{v}_{ref1}$ and $\mathbf{v}_{ref2}$ are respectively x_1 and x_2 ($x_1 + x_2 = d(\mathbf{v}_{ref1} - \mathbf{v}_{ref2}) = ||\mathbf{v}_{ref1} - \mathbf{v}_{ref2}||$), then the density estimator defined with respect to $\mathbf{v}_{ref1}$ can be written as:

$$\hat{f}_{eref,vref1}(\mathbf{v}, x_1) = c_0 \sum_{i, \mathbf{v}_i \in \mathbf{V}_{ref1}} k\left(\frac{(d(\mathbf{v}_{ref1}, \mathbf{v}_i) - x_1)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right) + c_0 \sum_{i, \mathbf{v}_i \notin \mathbf{V}_{ref1}} k\left(\frac{(d(\mathbf{v}_{ref1}, \mathbf{v}_i) + x_1)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right). \quad (3)$$

where $\mathbf{V}_{ref1}$ is the set of branch nodes with respect to $\mathbf{v}_{ref1}$ and e_{ref}. Similarly, we can define the density estimator with respect to $\mathbf{v}_{ref2}$:

$$\hat{f}_{eref,vref2}(\mathbf{v}, x_2) = c_0 \sum_{i, \mathbf{v}_i \in \mathbf{V}_{ref2}} k\left(\frac{(d(\mathbf{v}_{ref2}, \mathbf{v}_i) - x_2)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right) + c_0 \sum_{i, \mathbf{v}_i \notin \mathbf{V}_{ref2}} k\left(\frac{(d(\mathbf{v}_{ref2}, \mathbf{v}_i) + x_2)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right). \quad (4)$$

where $\mathbf{V}_{ref2}$ is defined in a similar way. Associated with the above density estimator are some good properties that facilitates the mode seeking process:

Property 3.2 $\hat{f}_{eref,vref1} = \hat{f}_{eref,vref2}, \forall e_{ref} \in \mathbf{E}$

The above equality holds in the sense that $\mathbf{V}_{ref1} \cup \mathbf{V}_{ref2} = \mathbf{V}$ and $\mathbf{V}_{ref1} \cap \mathbf{V}_{ref2} = \emptyset$, which indicates $\{\mathbf{v}_i | \mathbf{v}_i \in \mathbf{V}_{ref1}\} = \{\mathbf{v}_i | \mathbf{v}_i \notin \mathbf{V}_{ref2}\}$. In addition, since $d(\mathbf{v}_{ref1}, \mathbf{v}_i) - x_1 = d(\mathbf{v}_{ref1}, \mathbf{v}_{ref2}) + d(\mathbf{v}_{ref2}, \mathbf{v}_i) - x_1 = d(\mathbf{v}_{ref2}, \mathbf{v}_i) + x_2$ when $\mathbf{v}_i \in \mathbf{V}_{ref1}$, we obtain the following equality:

$$\sum_{i, \mathbf{v}_i \in \mathbf{V}_{ref1}} k\left(\frac{(d(\mathbf{v}_{ref1}, \mathbf{v}_i) - x_1)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right) = \sum_{i, \mathbf{v}_i \notin \mathbf{V}_{ref2}} k\left(\frac{(d(\mathbf{v}_{ref2}, \mathbf{v}_i) + x_2)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right).$$

The equality relation between the second term of (3) and the first term of (4) can be proved similarly. Property 3.2 states that the estimated density does not depend on the choice of reference point.

Property 3.3 *If e_{ref1} and e_{ref2} are two edges that connects the same node $\mathbf{v}_{ref}$, $\hat{f}_{eref1,vref}(\mathbf{v}, 0) = \hat{f}_{eref2,vref}(\mathbf{v}, 0), \forall \mathbf{v}_{ref} \in \mathbf{V}$.*

Property 3.3 states that the estimated density does not depend on the choice of reference edge when the MST space kernel is located on a tree node. Here we consider the special situation where the MST space kernel is shifting from one edge to another. When the kernel is located on $\mathbf{v}_{ref}$, the density estimator degenerates to (1), as $x = 0$. The same condition also holds when we define the density estimator with respect to any other edge connecting to $\mathbf{v}_{ref}$, which indicates the above property.

Property 3.4 *The kernel defined on the MST distance space is continuous and is piecewise differentiable.*

According to the definition of density estimator, one is easy to verify the piecewise continuity and differentiability given the MST space kernel is located on the same edge. Together with Property 3.3, we can obtain Property 3.4. The above property also infers the continuity and piecewise differentiability of the density estimator since it is a linear combination of continuous and piecewise differentiable kernels.

3.2. Mode seeking with force competition

We seek the mode by maximizing the density estimator with respect to $\mathbf{v}$ and x simultaneously. The step is to piecewisely estimate the density gradient, which is similar to mean shift. Taking the derivative of the density estimator with respect to $\mathbf{v}$, one get the estimated density gradient:

$$\begin{aligned}&\frac{\partial \hat{f}_{eref,vref}(\mathbf{v},x)}{\partial \mathbf{v}}\\&=\frac{2c_0}{h_2^2}\sum_i(\mathbf{v}_i-\mathbf{v})K_i g\Big(\Big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\Big\|^2\Big)\\&=\frac{2c_0}{h_2^2}\Big[\sum_i K_i g\Big(\Big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\Big\|^2\Big)\Big]\Big[\frac{\sum_i K_i g\big(\big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\big\|^2\big)\mathbf{v}_i}{\sum_i K_i g\big(\big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\big\|^2\big)}-\mathbf{v}\Big]\end{aligned} \tag{5}$$

where $g(x)=-k'(x)$, K_i is the MST space kernel function:

$$K_i=\begin{cases}k((d(\mathbf{v}_{ref},\mathbf{v}_i)-x)^2/h_1^2) & \text{if } \mathbf{v}_i\in\mathbf{V}_{ref1}\\ k((d(\mathbf{v}_{ref},\mathbf{v}_i)+x)^2/h_1^2) & \text{otherwise}\end{cases}$$

The second term in (5) is the well known mean shift vector for the feature space kernel center $\mathbf{v}$:

$$\mathbf{m}(\mathbf{v})=\frac{\sum_i K_i g\big(\big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\big\|^2\big)\mathbf{v}_i}{\sum_i K_i g\big(\big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\big\|^2\big)}-\mathbf{v}. \tag{6}$$

[1] has already developed a sound theoretical basis for mean shift algorithm concerning its physical meaning, convergence analysis and relation to other feature space analysis methods. Here we will not extend the discussion. Now consider the second variable. Taking the derivative of $\hat{f}_{eref,vref}(\mathbf{v},x)$ with respect to x, we have:

$$\begin{aligned}&\frac{\partial \hat{f}_{eref,vref}(\mathbf{v},x)}{\partial x}=\\&\frac{2c_0}{h_1^2}\sum_{i,\mathbf{v}_i\in\mathbf{V}_{ref}}(d(\mathbf{v}_{ref},\mathbf{v}_i)-x)K_{joint,i}\\&+\frac{2c_0}{h_1^2}\sum_{i,\mathbf{v}_i\notin\mathbf{V}_{ref}}(-d(\mathbf{v}_{ref},\mathbf{v}_i)-x)K_{joint,i},\end{aligned} \tag{7}$$

where $K_{joint,i}$ is the product of the feature space kernel and the negative derivative of the MST space kernel profile:

$$K_{joint,i}=\begin{cases}-k'\big(\frac{(d(\mathbf{v}_{ref},\mathbf{v}_i)-x)^2}{h_1^2}\big)k\big(\big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\big\|^2\big) & \text{if } \mathbf{v}_i\in\mathbf{V}_{ref}\\ -k'\big(\frac{(d(\mathbf{v}_{ref},\mathbf{v}_i)+x)^2}{h_1^2}\big)k\big(\big\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\big\|^2\big) & \text{otherwise}\end{cases}$$

Equation (7) can be further rewritten as:

$$\begin{aligned}&\frac{\partial \hat{f}_{eref,vref}(\mathbf{v},x)}{\partial x}=\\&\frac{2c_0}{h_1^2}\Big[\sum_i K_{joint,i}\Big]\Big[\Big(\sum_{i,\mathbf{v}_i\in\mathbf{V}_{ref}}K_{joint,i}d(\mathbf{v}_{ref},\mathbf{v}_i)\\&-\sum_{i,\mathbf{v}_i\notin\mathbf{V}_{ref}}K_{joint,i}d(\mathbf{v}_{ref},\mathbf{v}_i)\Big)/\sum_i K_{joint,i}-x\Big]\end{aligned} \tag{8}$$

The last term of (8) results in the displacement of the MST space kernel, which is the so called *force competition*. Force competition can also be regarded as a special case of univariate mean shift with $\mathbf{v}_{ref}$ representing the origin. One could imagine it as a tug of war where data points weighted by K_{joint} are tugging along each side of $\mathbf{v}_{ref}$. The shifting step size, however, should be chosen carefully since $\hat{f}_{eref,vref}$ is only piecewise differentiable. Suppose we use the ms to denote the last term of (8), the displacement of the MST space kernel is defined as:

$$\mathbf{m}(x)=\max(-x,\min(|e_{ref}|-x,ms)) \tag{9}$$

The above term generantees that the MST space kernel is always shifted along the same reference edge. Here we seek to provide more intuition by discussing some properties of the density gradient estimation:

Property 3.5 *The estimation of density gradient does not depend on the choice of reference node* $\mathbf{v}_{ref}$.

Since the density estimator is piecewise differentiable on the edge, according to Property 3.2 we can verify the above property. The estimated density gradient, however, does depend on the choice of reference edge when the MST space kernel reaches a tree node with more than two connecting edges. Difference in the choice of the reference edge results in the following inequality:

$$\mathbf{V}_{vref,eref1}\cup\mathbf{V}_{vref,eref2}\neq\mathbf{V},$$

where $\mathbf{V}_{vref,eref1}$ is the branch node set with respect to node $\mathbf{v}_{ref}$ and its connecting edge e_{ref}, and similar for $\mathbf{V}_{vref,eref2}$. Such inequality leads to the sudden jump of estimated density gradient at some tree nodes.

Theorem 3.1 *Given any node* $\mathbf{v}_{ref}$ *where the MST space kernel is located and there are more than two connecting edges, the number of reference edge* e_{ref} *with positive MST space kernel displacement is no more than 1.*

Proof: Without loss of generality, suppose the MST space kernel is located on node $\mathbf{v}_{ref}$ with three connecting edges e_{ref1}, e_{ref2} and e_{ref3}, and $D_{eref1} > D_{eref2} > D_{eref3}$, where D_{eref} is defined as follows:

$$D_{eref} = \sum_{i,\mathbf{v}_i \in \mathbf{V}_{vref,eref}} -k'\left(\frac{d(\mathbf{v}_{ref},\mathbf{v}_i)^2}{h_1^2}\right) k\left(\left\|\frac{\mathbf{v}-\mathbf{v}_i}{h_2}\right\|^2\right) d(\mathbf{v}_{ref},\mathbf{v}_i).$$

The force competition term $ms_{vref,eref}$ equals to the estimated density gradient with respect to $\mathbf{v}_{ref}$ and e_{ref} times a positive scalar:

$$\begin{aligned} ms_{vref,eref1} &= c\frac{\partial \hat{f}_{eref,vref}(\mathbf{v},x)}{\partial x}\Big|_{x=0} \\ &= D_{ref1} - D_{ref2} - D_{ref3}. \end{aligned}$$

Similarly, we have $ms_{vref,eref2} = D_{ref2} - D_{ref1} - D_{ref3}$ and $ms_{vref,eref3} = D_{ref3} - D_{ref1} - D_{ref2}$. Since $D_{eref1} > D_{eref2} > D_{eref3}$ and $D_{eref} > 0$, $ms_{vref,eref2}$ and $ms_{vref,eref3}$ can not possibly be larger than 0. The only positive $ms_{vref,eref}$ comes when $D_{ref1} > D_{ref2} + D_{ref3}$ and the above proof can be easily extended to nodes with multiple edges. Thus we have proved the above Theorem.

3.3. Algorithmic description

Theorem 3.1 states that when the MST space kernel is located on any tree node, either this node is a local maxima, or there is only one edge to which shifting the kernel results in the increase of the density. The conveyed intuition here is important: each time the MST space kernel is shifting from one edge to another, one does not face the problem of multiple selectable paths since there is at most one edge that increases the estimated density. Such property leads to the basis of our implemented algorithm and its fast approximation method. The mode seeking algorithm is a step size controlled gradient ascent:

1. For each data point $\mathbf{v}_i, i = 1, 2, ..., N$, initialize the its feature space kernel position as the data point itself. Select $\mathbf{v}_i$ as $\mathbf{v}_{ref}$ and initialize the MST space kernel on the reference node.

2. Compute the MST space kernel shift with the following rules:

   If the MST space is exactly located on any tree node, calculate $\mathbf{m}_j(x)|_{x=0}$ with respect to all its connecting edges e_j.

   If There exists one positive $\mathbf{m}_j$, select the corresponding edge e_j as the reference edge e_{ref}. $\mathbf{m}(x) = \mathbf{m}_j$ as the MST space kernel shift.

   Else $\mathbf{m}(x) = 0$.

   Else calculate $\mathbf{m}(x)$ with respect to $\mathbf{v}_{ref}$ and e_{ref}.

3. Calculate the step control factor α:

   If $\mathbf{m}(x) = 0$, $\alpha = 1$.

   Else $\alpha = |\mathbf{m}(x)|/|ms|$.

4. Compute the feature space kernel shift and scale it with α: $\mathbf{m}'(\mathbf{v}) = \alpha\mathbf{m}(\mathbf{v})$.

5. Simultaneously shift the MST space kernel and the feature space kernel with respect to the kernel shifts calculated in Step 2 and Step 4. The MST space kernel is shifted with the following rule:

   If the MST space kernel is exactly located on a node

   If $\mathbf{m}(x) = |e_{ref}|$, shift the MST space kernel to the neighboring node connected by e_{ref} and select the neighboring node as the new reference node.

   Elseif $\mathbf{m}(x) = 0$, the MST space kernel stays on the current node.

   Else update the kernel position on the edge: $x = \mathbf{m}(x)$.

   Elseif the MST space kernel is located on an edge

   If $\mathbf{m}(x) == -x$, shift the MST space kernel to the reference node.

   Elseif $\mathbf{m}(x) = e_{ref} - x$, shift the MST space kernel to the neighboring node connected by e_{ref} and select the neighboring node as the new reference node.

   Else update the kernel position on the edge: $x = \mathbf{m}(x) + x$.

6. Repeat Step 2 to Step 5 until convergence.

3.4. Fast approximation

Due to the piecewise differentiability and step control, the above algorithm gives the best mode seeking performance but requires more iterations before convergence. In addition, the algorithm contains numerous "if-then-else" conditions, which is not friendly to hardware implementation. Here we also propose a fast approximation to the original algorithm by iteratively shifting the MST space kernel and the feature space kernel. The method is straight forward:

1. For each data point, initialize the MST space kernel and the feature space kernel.

2. Shift the feature space kernel according to (6).

3. If there exist neighboring nodes that increase the estimated density, shift the MST space kernel to the nearest one. Otherwise, stop shifting.

4. Repeat Step 2 and 3 until convergence.

In all of the following experiments, we only implement the above fast algorithm.

4. Experimental results

We show three sets of experiments using our proposed algorithm. The first set of experiments demonstrates the performance of the method in the task of data clustering. Figure 2(a) shows a character shaped distribution containing 934 data points and its clustering result. The bandwidth parameters h_1 and h_2 were respectively set to 150 and 40 for this experiment. Figure 2(b) shows the mixture of 4 gaussian distributions with a total of 1500 data points. Here we set h_1 to 700 and h_2 to 150. From the two experiments one could observe that the method works reasonably well for both arbitrarily shaped and regularly shaped cluster of data. The real challenge comes when we want to cluster the spiral-like data distribution with highly nonlinear cluster separation boundaries. The example of spiral-like data given in [3] was reproduced with the Matlab code kindly available at http://www.cs.cmu.edu/∼new_medoid.htm. In this experiment h_1 and h_2 are respectively set to 150 and 300. Note that we have achieved the clustering performance that approximates the one given in [3] without using any non-Euclidean metric, while mean shift or Euclidean medoid shift usually will fail on such task.

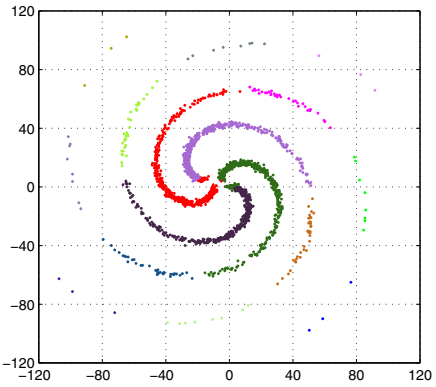
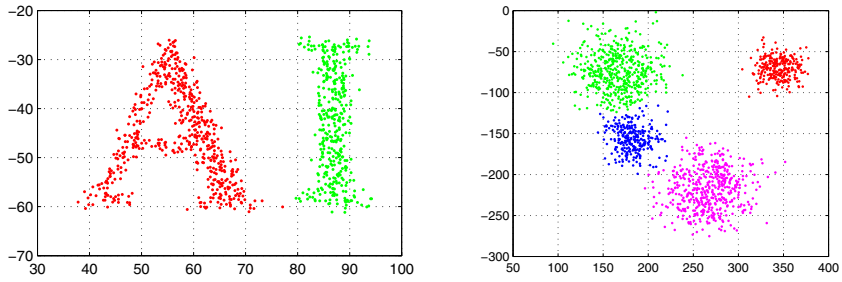


Figure 3. Clustering with spiral-like cluster of data using the proposed method

The second set of experiments address the problem discontinuity preserved smoothing with superpixelized images. As discussed in previous section, region-wise operation significantly reduces the required computation power, thus greatly accelerates the image smoothing and segmentation process. The introduction of MST space kernel works in compatible with the region adjacency graph and in addition, further improves the smoothing and segmentation performance. Figure 4 shows the images and their smoothing results using different methods in the RGB color space. The images are first superpixelized using normalized cut[6, 7]. The corresponding Matlab code is kindly provided at http://www.cs.sfu.ca/∼mori/research/superpixels/. We set the number of coarse superpixels N_sp to 200, the number of fine superpixels N_sp2 to 400 and the number of eigenvectors N_ev to 40. Each superpixel is then represented by the mean RGB value and the whole image is mapped to an undirected, weighted region adjacency graph where edges corresponds to the eight-connectivities of two regions and edge weights are defined as the Euclidean distances between the region means. We extract the minimum spanning tree from the region adjacency graph using Kruskal's Algorithm and perform mode seeking using our proposed method. Here we fixed h_1 as 30 and h_2 as 50 for all the test images. The obtained results are illustrated in the second column of figure 4. To demonstrate the improvement of algorithm performance by introducing the MST space kernel, we compare the results with medoid shift smoothing where each super pixel is represented by the 5D joint representation of the RGB mean and spatial coordinate mean. The distance matrix is obtained by calculating the Euclidean distances between each pair of super pixels and the parameter $Sigma$ is set to 2000. We also compare our results with quick shift which is a fast mode seeking algorithm. We run the quick shift algorithm with the VLFeat Matlab package which is publicly available at http://www.vlfeat.org/. The parameters $ratio$, $kernelsize$ and $maxdist$ are respectively set to 0.3, 12 and 30. The results illustrated in figure 4 indicates the advantage of using our proposed method for image smoothing.

We illustrate the potential application of image segmentation using our method in the last set of experiments. Note that the segmentation performance depends largely on the defined feature. With superpixelized images, the definition of image feature becomes much more versatile than pixel based methods. Such framework allows one to improve the segmentation performance by defining the feature in a sophisticated way, using textons, texture detectors or other region statistics. For simplicity we only adopt region color histogram in this paper. Each region is represented by a 24-D concatenated histogram with each RGB channel returning a histogram of 8 bins. We then use principal component analysis (PCA) to perform dimensionality reduction on the obtained histograms. The percentage of preserved variance for PCA is set to 0.9, a typical rule of thumb value for PCA. For most of the images, the reduced dimension after performing PCA often lies in between 4-8, which is much smaller than the original dimension number. By running PCA we reduces the computational complexity and effec-



(a) (b)

Figure 2. Data clustering using the proposed method. (a) Clustering with linearly separable data. (b) Clustering with mixture of gaussians

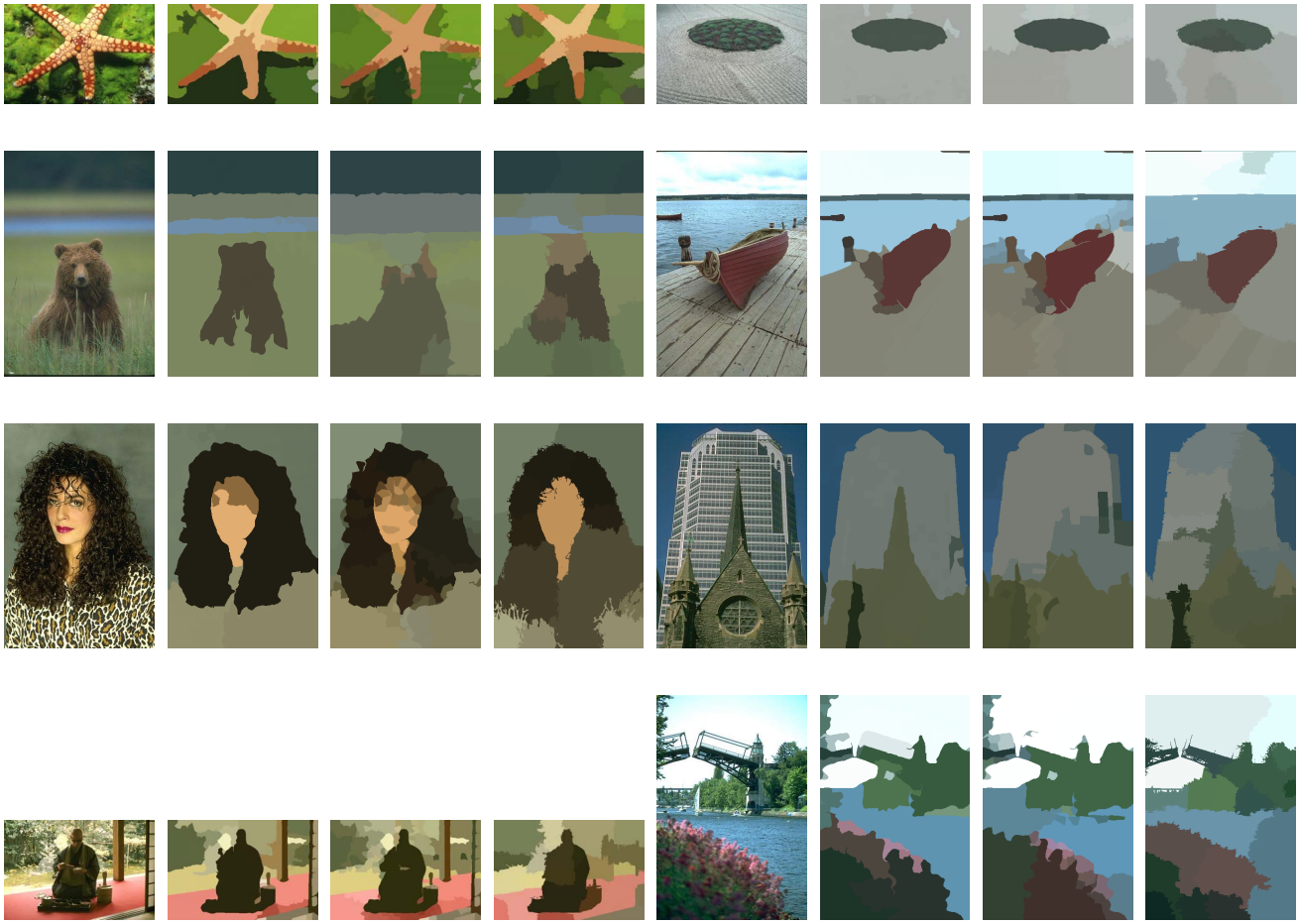


Figure 4. Discontinuity preserved smoothing with superpixelized images: The first column contains the original images. The second column corresponds to the smoothing results using the proposed method. The second column contains the smoothing results using medoid shift. The last column are the results obtained by quick shift.

tively avoids from suffering the ”curse of dimensionality”. The segmentation results are shown in figure 5. One could observe that the proposed method is effective and produces reasonably good segmentations.

5. Conclusion

In this paper, by introducing the MST space kernel, we have proposed a novel mode seeking method that can improve mode seeking performance on manifold-structured data and can work compatibly with region-wise image pro-

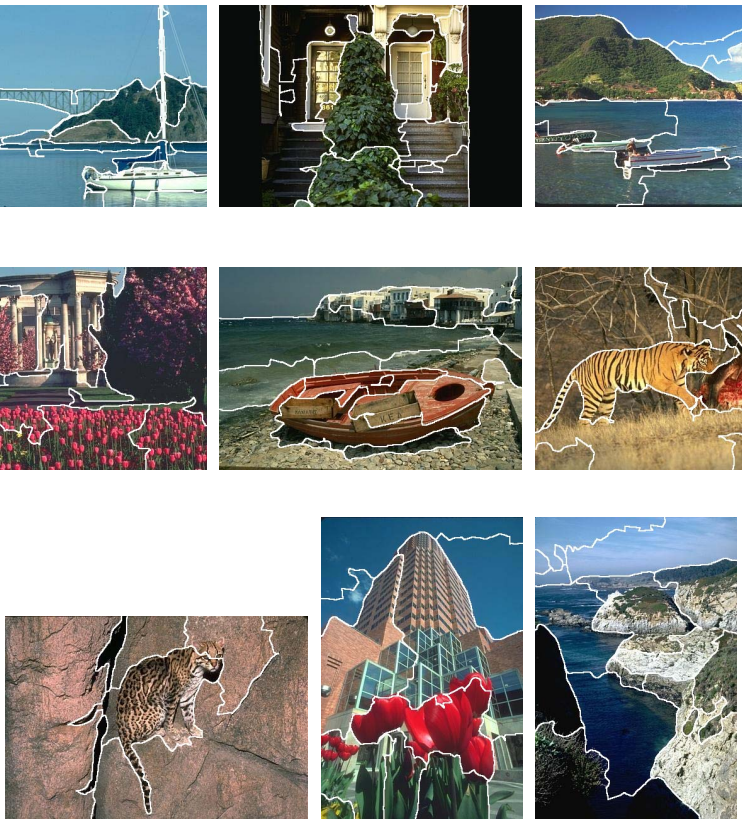


Figure 5. Image segmentation experiments with region histogram

cesing operations. We achieved good algorithm performance in clustering data with highly nonlinear separation boundaries without using any manifold distance or some other non Euclidean metrics, which is of considerable challenge. The advantage of using the proposed method for image smoothing and segmentation is also supported by our experiments.

References

[1] D. Comaniciu and P. Meer. "Mean shift: A robust approach toward feature space analysis." *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603-619, 2002.

[2] A. Yilmaz, "Object tracking by Asymmetric kernel mean shift with automatic scale and orientation selection." In *CVPR*, 2007.

[3] Y. A. Sheikh, E. A. Khan and T. Kanade. "Mode-seeking by Medoidshifts." In *ICCV*, 2007.

[4] A. Vedaldi and S. Soatto. "Quick shift and kernel methods for mode seeking." In *ECCV*, 2008.

[5] A. Vedaldi and S. Soatto. "Really quick shift: Image segmentation on a GPU." In *Workshop on Computer Vision using GPUs, held with ECCV*, 2010.

[6] J. Shi and J. Malik. "Normalized cuts and image segmentation." *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888-905, 2000.

[7] X. Ren and J. Malik. "NLearning a classification model for segmentation." In *ICCV*, 2003.

[8] K. Zhang, J. T. Kwok and M. Tang. "Accelerated convergence using dynamic mean shift." In *ECCV*, 2006.

[9] R. Subbarao and P. Meer. "Nonlinear mean shift for clustering over analytic manifolds." In *CVPR*, 2006.

[10] O. J. Morris, M.de J. Lee, and A.G. Constantinides. "Graph theory for image analysis: An approach based on the shortest spanning tree," In *IEE Proc. F., Communications. Radar & Signal Processing*, 133:146-152, 1986.