Directional Mean Shift and its Application for Topology Classification of Local 3D Structures

Mehran Kafai   Yiyi Miao   Kazunori Okada
Computer Science Department, San Francisco State University
1600 Holloway Avenue, San Francisco, CA 94132
{mkafai, yiyim, kazokada}@sfsu.edu

Abstract

In this study, we introduce a new directional nonparametric clustering algorithm for 3D medical structure topology classification. This paper proposes directional mean shift (DMS) which extends the well known mean shift-based clustering, for handling directional statistics, toward analyzing directional/circular-domain data with phase-wraparound boundary conditions. Our overall approach transforms the 3D topology classification problem into a clustering analysis of a 2D image, following the work by Bahlmann et al. [2] in the context of computer-aided diagnosis (CAD). The proposed DMS replaces the expectation-maximization (EM) algorithm for Gaussian mixture model (GMM) fitting used in the previous method addressing the shortcomings of the Bahlmann's method. Results from our experiments demonstrate the effectiveness of DMS in contrast to the original EM-based approach in solving the clustering problem with a 2D image unwrapped from a 3D spherical data, leading to better accuracy in the topology classification task.

1. Introduction

Cluster analysis [21] is widely used in medical image analysis due to the necessity to economically model the statistical nature of the great amount of data held by medical images. The analysis divides data into meaningful and/or useful clusters/groups. Various clustering techniques have played an important role for determining key features in many high-level analysis tasks, such as segmentation and registration. Mean shift (MS) [9, 3, 4] is a well known adaptive step-size mode seeking algorithm for kernel density function, which has been widely adapted to various vision and pattern analysis problems. The advantage of MS-based clustering is its ability to handle non-Gaussian shaped clusters, as well as to analyze data with arbitrary number of clusters.

In this paper, we propose directional mean shift (DMS): an extension of the mean shift to directional statistics. Directional statistics [13] handles directional data defined over a circular domain. Directional data is very common in vision applications where angle measurements (e.g., SIFT de-scriptor [12]) and the hue component of color images are examples of such data. Furthermore, they also result from unfolding a 3D space/manifold into a 2D image representation, such as omni camera, cartographic, and radar images. Any regular statistical analyses of such data, defined over a linear vector space, become erroneous, failing to capture the proximity of points at the two ends of the data domain. Thus special treatments [8, 1] for handling statistics over such data are required. Extending the mean shift toward such directional data is, therefore, non-trivial and favorable in widening its application scope.

Recently, the MS algorithm, originally defined over Euclidean vector space, has been extended toward non-Euclidean domains, such as Riemannian [22] and Grass-man [20] manifolds. While these extended MS algorithms provide a principled way to cluster data points within some non-linear manifolds, they cannot be applied to data points that are already embedded into a Euclidean vector space. For example, these methods cannot correctly cluster locations on a 2D world map because, after a cartographic transformation, we no longer have an access to the original 3D coordinates required by these methods. The proposed DMS is designed for such data sets where data points are defined over a linear space but have an underlying and inaccessible non-linear distribution. Angular distributions, such as Von Mises, are also commonly used when angular variables are present, but such distributions cannot be applied to domains if they are defined by a combination of linear and non linear dimensions [1] (e.g. HSV color space).

By exploiting the proposed DMS, this paper also addresses a medical image analysis task for classifying topological types of local 3D structures. A variety of local anatomical and pathological structures are present in 3D medical images (e.g, CT, MRI, etc). For example, in a chest CT, there exist tubular-shaped vessels/airways, branches of such tubular structures, blob-shaped nodules indicating tumors, and attachments of such nodules to other structures. The goal of the 3D topology classification is to determine the type of a given local 3D structure and to provide useful information to post processes such as chest CAD. For instance, the differentiation of nodules from vessel-branches can be used to reduce typical false positive cases of chest CAD, leading to better overall nodule detection accuracy.

A challenge of structure type classification in dense 3D images is due to the high variability of the said structure's orientation, co-articulation and size, as well as the inherent difficulty in visualizing such data. Solutions that have been previously proposed can be principally categorized in two distinct approaches. The first approach fits a set of flexible intensity models of specific structure type to data directly [6, 5, 17, 15]. In a 3D domain, such an approach becomes inefficient due to the high structural variability that can in turn make the model parameter space exceedingly large. The second approach analyzes eigenvalues of the Hessian matrix as a way to characterize structure types [16, 7, 10]. Although the second approach has been found successful in vasculature analysis in 3D volumetric data, its flaw rested in its inability to differentiate branching cases, commonly found in medical data. Overall a comprehensive analysis of structure topology classification is seldom reported, despite clinical significance and potential. Bahlmann *et al.* [2] proposed a promising approach to solve the 3D topology classification problem by transforming it into a clustering analysis within a 2D image and performing an EM algorithm for fitting a GMM to the data point set. This approach has two major defects. First, it tends to over-estimate the number of clusters. Second, the EM-based GMM fitting requires that the number of clusters to be known a priori. One of the major contributions of this work is to replace the EM-based GMM fitting by the proposed DMS. The resulting extended classification algorithm drastically improves the overall accuracy by overcoming the two shortcomings of the previous method.

This paper is organized as follows. Sec. 2 and 3 review the mean shift and directional statistics theories. The proposed DMS algorithm is then described in Sec. 4. The convergence proof of DMS is provided in the appendix. Sec. 5 describes our approach to 3D structure topology clustering using DMS. Finally Sec. 6 concludes this paper by discussing our experimental results and future work.

2. Mean Shift Theory: Overview

2.1. Kernel Density Estimation

Given n data points $\mathbf{x}_i, i = 1, \ldots, n$ in d-dimensional Euclidean space R^d, and a bounded kernel $K(\mathbf{x})$, the kernel density estimation of random variable $\mathbf{x}$ with kernel $K(\mathbf{x})$ and bandwidth h (smoothing parameter or window size) is defined [19, 18] as

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) \qquad (1)$$

where $K(\mathbf{x})$ is usually chosen with properties $\int_{-\infty}^{+\infty} K(x)\, dx = 1$ and $\int_{-\infty}^{+\infty} x K(x)\, dx = 0$.

Clearly, $\hat{f}(\mathbf{x})$ has the form of a density function in that it is non-negative and its integral is 1. Further, $\hat{f}(\mathbf{x})$ is smooth. The kernel density estimation(KDE) is the most widely used method of nonparametric density estimation.

2.2. Mean Shift Procedure

Mean shift can be derived from the estimate for the gradient of probability density functions following [9]. Such density gradient naturally helps to seek the mode(s) of a density function in the gradient-ascent sense. From (1) the estimate of density gradient can be approximated by the gradient of density estimator which gives

$$\hat{\bigtriangledown} f_{h,K}(\mathbf{x}) \equiv \bigtriangledown \hat{f}(\mathbf{x}) = \frac{c}{nh^d} \left[\sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)\right] \times$$
$$\left[\frac{\sum_{i=1}^{n} \mathbf{x}_i K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)} - \mathbf{x}\right] \quad (2)$$

The second term $\frac{\sum_{i=1}^{n} \mathbf{x}_i K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)} - \mathbf{x}$ in the rhs of the above formula defines the mean shift step. A convergent gradient-ascent like iterative algorithm is then given by the sequence of successive applications of (2) until convergence.

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}_i K\left(\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right)}, \quad j = 1, 2, \ldots \quad (3)$$

Under certain regularity conditions, we can show that gradient is asymptotically unbiased and uniformly consistent [3, 4].

3. Directional Statistics

Directional statistics is commonly defined on circular observations $z = e^{i\theta}$ on a unit radius circle or as a unit vector $\mathbf{x} = (\cos\theta, \sin\theta)^T$ [13]. All arithmetic that can be performed on the circle will be modulo 2π so that the angles θ and $\theta + 2\pi$ represent the same point on the circle and that θ can be treated as in a range $[0, 360]$. For a given directional data ϕ and ω the directional distance between them can be defined as

$$1 - \cos(\phi - \omega) \qquad (4)$$

3.1. Directional Distance

While the definition given in (4) provides a correct distance for directional data θ in the range $[0, 360]$, it fails when the variable range becomes different from $[0, 360]$. For instance, if, for a certain application, the radius of support is defined on [-45,30] then (4) would not result in the correct distance. This would also cause a problem when working in multivariate domains where data from each dimension has a different radius of support. For example, a cartographic transformation from a glove to a 2D map typically produces a directional image within a circular domain whose support is given by $[0, \pi]$ for latitude and $[0, 2\pi]$ for longitude. To overcome this problem we redefine the Euclidean distance of any 2 real numbers ϕ and ω ($\phi < \omega$)

under an underlying directional domain with arbitrary domain support as

$$\mathcal{D}(\phi, \omega) = \sqrt{s_{\phi,\omega}^2} \qquad (5)$$

where the *directional difference vector* $s_{\phi,\omega}$ is defined in a wraparound fashion as

$$s_{\phi,\omega} =$$
$$\begin{cases} sgn(\phi - \omega) \times r - (\phi - \omega), & |\phi - \omega| > \frac{r}{2} \\ \\ \phi - \omega & \text{otherwise} \end{cases} \qquad (6)$$

The domain support of directional data ϕ and ω is generalized as

$$\phi, \omega \in [r_\alpha, r_\beta] \qquad (7)$$

where r_β and r_α are respectively the maximum and minimum values of the domain, and the radius r of data support is defined as

$$r = r_\beta - r_\alpha \qquad (8)$$

We call $\mathcal{D}(\phi, \omega)$ the *directional distance* [11]. It universally represents the shortest path between ϕ and ω on the directional space. Two points on a circle circumference divide the circle into two arcs and directional distance is length of the shorter arc.

Applying equation (6) to any linear space can be interpreted as applying geodesic distance to a corresponding underlying circular domain with arbitrary domain support.

3.2. Directional Mean

The directional mean proposed in [13] for two given unit vectors ϕ and ω, with corresponding angles θ_1 and θ_2, is defined as

$$\bar{\theta} = \tan^{-1}\left(\bar{S}/\bar{C}\right) \qquad (9)$$

where the Cartesian coordinates of the center of mass are given as

$$\bar{C} = \frac{1}{2}\left(\cos\theta_1 + \cos\theta_2\right) , \ \bar{S} = \frac{1}{2}\left(\sin\theta_1 + \sin\theta_2\right) \quad (10)$$

Similar to the discussion in 3.1, varying range of domain is sometimes problematic for calculating the mean. Therefore, we introduce another definition for directional mean with arbitrary domain support as

$$\mathcal{D}_{mean}(\phi, \omega) =$$
$$\begin{cases} \frac{\phi + sgn(r_\beta + r_\alpha - (\phi+\omega))*r + \omega}{2}, & |\phi - \omega| > r/2 \\ \\ \frac{\phi + \omega}{2} & \text{otherwise} \end{cases} \qquad (11)$$

3.3. KDE with Directional Data

Combining all the aforementioned concepts of directional statistics, we can define a KDE function estimated from directional data in circular domain. Introducing the directional difference vector definition (6) into (1) results in

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h}\right) \qquad (12)$$

Figure 1 compares kernel density estimation between conventional Euclidean and the proposed directional difference vector, demonstrating how directional statistics affects the KDE result given the same input. The directional KDE shown in figure 1(c) captures the periodic wraparound nature of the data. The bivariate Gaussian kernel $K(\mathbf{x}) = \frac{1}{2\pi}e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}$ is applied to obtain results presented in figure 1.
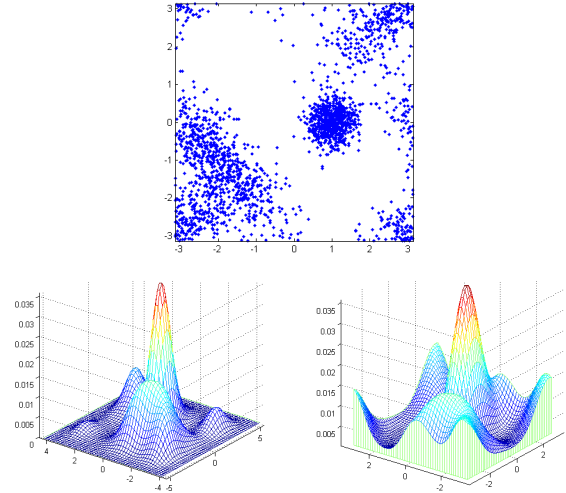

Figure 1. (a) 2D input using a sample point set representation in $[-\pi, \pi]$ (b) KDE with conventional Euclidean distance measure showing five modes (c) KDE with directional difference vector exhibiting two correct modes

4. Directional Mean Shift

4.1. Deriving DMS

Exploiting the directional difference vector, any arbitrary convex and bounded kernel $K(\mathbf{x})$ can be defined by using the profile kernel k

$$K(\mathbf{x}) = c_{k,d} k\left(\left\|\frac{s_{\mathbf{x},\mathbf{x}_i}}{h}\right\|^2\right) \qquad (13)$$

where $\mathbf{x}, \mathbf{x}_i \in [r_\alpha, r_\beta]$, h is the bandwidth(window size), and $c_{k,d} > 0$ is a constant to make certain that K is a density function. From (12) and (13) we have

$$\hat{f}_{h,K}(\mathbf{x}) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\|\frac{s_{\mathbf{x},\mathbf{x}_i}}{h}\right\|^2\right) \qquad (14)$$

The estimate of density gradient is given by the gradient of density estimator,

$$
\begin{aligned}
\hat{\triangledown} f_{h,K}(\mathbf{x}) &\equiv \triangledown \hat{f}(\mathbf{x}) \\
&= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} s_{\mathbf{x},\mathbf{x}_i} k' \left(\left\| \frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right\|^2 \right).
\end{aligned} \quad (15)
$$

Given $g(\mathbf{x}) = -k'(\mathbf{x})$, we have

$$
\triangledown \hat{f}(\mathbf{x}) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (-s_{\mathbf{x},\mathbf{x}_i}) g \left(\left\| \frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right\|^2 \right). \quad (16)
$$

Having (6), (16) branches into $\hat{f}_1(\mathbf{x})$ and $\hat{f}_2(\mathbf{x})$ which correspond to the two possible definitions defined in (6). Therefore

$$
\begin{cases}
\triangledown \hat{f}_1(\mathbf{x}) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{x} \pm r) g \left(\left\| \frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right\|^2 \right), \\
\triangledown \hat{f}_2(\mathbf{x}) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (\mathbf{x} - \mathbf{x}_i) g \left(\left\| \frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right\|^2 \right).
\end{cases} \quad (17)
$$

Similar to (2), $\triangledown \hat{f}_2(\mathbf{x})$ can be rewritten as

$$
\triangledown \hat{f}_2(\mathbf{x}) = \frac{c}{nh^d} \left[\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right) \right] \times
$$
$$
\left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)}{\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)} - \mathbf{x} \right]. \quad (18)
$$

Defining $\gamma_i = \mathbf{x}_i \pm r$ and some algebra yields

$$
\triangledown \hat{f}_1(\mathbf{x}) = \frac{c}{nh^d} \left[\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right) \right] \times
$$
$$
\left[\mathbf{x} - \frac{\sum_{i=1}^{n} \gamma_i g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)}{\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)} \right]. \quad (19)
$$

Since $k(x)$ is bounded and convex, the term $\frac{c}{nh^d} \left[\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right) \right]$ is strictly positive. To locate the mode(s), the points where $\triangledown \hat{f}(\mathbf{x}) = 0$ should be found. Hence, from (18) and (19) we respectively have

$$
\mathbf{x} = \frac{\sum_{i=1}^{n} \mathbf{x}_i g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)}{\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)} \quad (20)
$$

and

$$
\mathbf{x} = \frac{\sum_{i=1}^{n} \gamma_i g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)}{\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)} \quad (21)
$$

Merging γ_i and $\mathbf{x}_i$ into one variable results in the overall repositioning definition defined as

$$
\mathbf{x}'_i = \begin{cases} \mathbf{x} + s_{\mathbf{x},\mathbf{x}_i}, & |\mathbf{x} - \mathbf{x}_i| > r/2 \\ \mathbf{x}_i & \text{otherwise} \end{cases} \quad (22)
$$

Thus (20) and (21) can be integrated as

$$
\mathbf{x} = \frac{\sum_{i=1}^{n} \mathbf{x}'_i g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)}{\sum_{i=1}^{n} g \left(\frac{s_{\mathbf{x},\mathbf{x}_i}}{h} \right)} \quad (23)
$$

where the successive locations of the kernel are defined as

$$
\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}'_i g \left(\left\| \frac{s_{\mathbf{y}_j,\mathbf{x}_i}}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g \left(\left\| \frac{s_{\mathbf{y}_j,\mathbf{x}_i}}{h} \right\|^2 \right)}, \quad j = 1, 2, \dots \quad (24)
$$

Determining $\mathbf{y}_{j+1}$ using (24) may result in $\mathbf{y}_{j+1} \notin [r_\alpha, r_\beta]$. From (6), (22), and (24) we know $\mathbf{y}_{j+1} \in [r_\alpha - \frac{r}{2}, r_\beta + \frac{r}{2}]$. Therefore, after each iteration of calculating the successive location, we reposition $\mathbf{y}_{j+1}$ by

$$
\mathbf{y}'_{j+1} = \begin{cases} \mathbf{y}_{j+1} - r, & \mathbf{y}_{j+1} > r_\beta \\ r - \mathbf{y}_{j+1}, & \mathbf{y}_{j+1} < r_\alpha \\ \mathbf{y}_{j+1} & \text{otherwise} \end{cases} \quad (25)
$$

therefore (24) is rewritten as

$$
\mathbf{y}'_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}'_i g \left(\left\| \frac{s_{\mathbf{y}'_j,\mathbf{x}_i}}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g \left(\left\| \frac{s_{\mathbf{y}'_j,\mathbf{x}_i}}{h} \right\|^2 \right)}, \quad j = 1, 2, \dots \quad (26)
$$

4.2. DMS Convergence

To show the correctness of DMS, we prove the following theorem for the convergence of (26). A proof of this theorem will be provided in the appendix.

Theorem 1. If the kernel $K(x)$ is convex and bounded, the sequence $\{\hat{f}(\mathbf{y}'_j)\}_{j=1,2,\dots}$ is monotonically increasing and $\{\mathbf{y}'_j\}_{j=1,2,\dots}$ and $\{\hat{f}(\mathbf{y}'_j)\}_{j=1,2,\dots}$ converge.

4.3. Algorithm Overview

Let X denote the set of data points, $\{\mathbf{x}_i\}_{i=1,\dots,n}$ and th the predefined distance threshold. The following describes the DMS algorithm

DMS(X,th)
    for all $\mathbf{x}_i \in X$
        set current point cp to $\mathbf{x}_i$
        loop
            calculate $s_{cp,X}$ (6)
            determine repositioned data points, $\mathbf{x}'_i$ (22)
            calculate the new point mp by DMS (24)
            reposition mp (25)
            exit loop if $s_{cp,mp}$ (6) is less than th
            reset cp by mp
        set cp as the convergence point $\mathbf{x}^*_i$

4.4. Example with HSV Color Segmentation

In order to demonstrate the effectiveness of the proposed DMS algorithm, we apply it for color segmentation using HSV color space. Figure 2 demonstrates the merit of our method in a few illustrative examples.

All three components of HSV color space: hue, saturation, and value do not share the same domain properties. Saturation and value are defined in linear space with values $[0, 1]$ while hue is defined circularly on $[0, 360]$. This property of the HSV color space, in many cases, prevents linear space segmentation algorithms from producing desired HSV color segmentation results. For example, the standard mean shift segmentation can fail to extract a foreground object by falsely separating parts with similar colors that are placed at the different sides of the origin of the circular hue component. On the other hand, the proposed DMS correctly groups pixels with similar colors across the color space origin, mitigating this shortcoming.

In the case of DMS-based HSV color segmentation, $\hat{f}(\mathbf{x})$ would be a multivariate density function built upon a combination of both linear and directional feature components. Our directional mean shift can be readily extended to this combined feature domain.

5. 3D Medical Structure Classification

As described briefly in introduction, Bahlmann *et al.* [2] proposed a novel approach to solve the structure type classification problem by transforming the problem of classifying 3D topology types into a clustering analysis within a 2D image. The following describes the main idea of this algorithm [14]. Consider a 3D sphere that encompasses the target 3D structure shown in a 3D scan such as CT or MRI. Assume that any non-target regions have near-zero intensity values and that any target's part that protrudes the spherical surface would create a compact region of high-intensity values. Notice that the number of such regions on the spherical surface explicitly determines the type of 3D topology: nodule (0), nodule attached to an other structure (1), vessel (2), and vessel branch (≥ 3). The location and the size of these regions change according to various articulation and geometry of these structures. Thus counting the number of high-intensity islands on the 3D spherical surface domain yields the wanted classification.

The algorithm by [2] breaks down this process into two parts. First, creating a set of 2D images by unfolding the 3D bounding spheres with linearly increasing radius values. The optimal radius value is then selected by calculating the local entropy minimum for this 2D image set. Second, counting the number of high-intensity islands by fitting Gaussian mixture models (GMMs) with various number of components using the EM algorithm and then picking the model that fits best to the data. Figure 3 illustrates this process. After arbitrarily choosing an origin, the 3D manifold is first unfolded into a 2D image using the following standard 3D spherical to Cartesian coordinate transformation

$$x = \sin\theta \cdot \cos\phi \cdot r \ , \ y = \sin\theta \cdot \sin\phi \cdot r \ , \ z = \cos\theta \cdot r$$

(a) Original image
(b) MS segmentation (h=60)
(c) MS segmentation (h=70)
(d) DMS segmentation(h=13)

(e) Original image
(f) MS Segmentation(h=60)
(g) MS Segmentation(h=70)
(h) DMS Segmentation(h=70)

(i) Original image
(j) MS Segmentation(h=60)

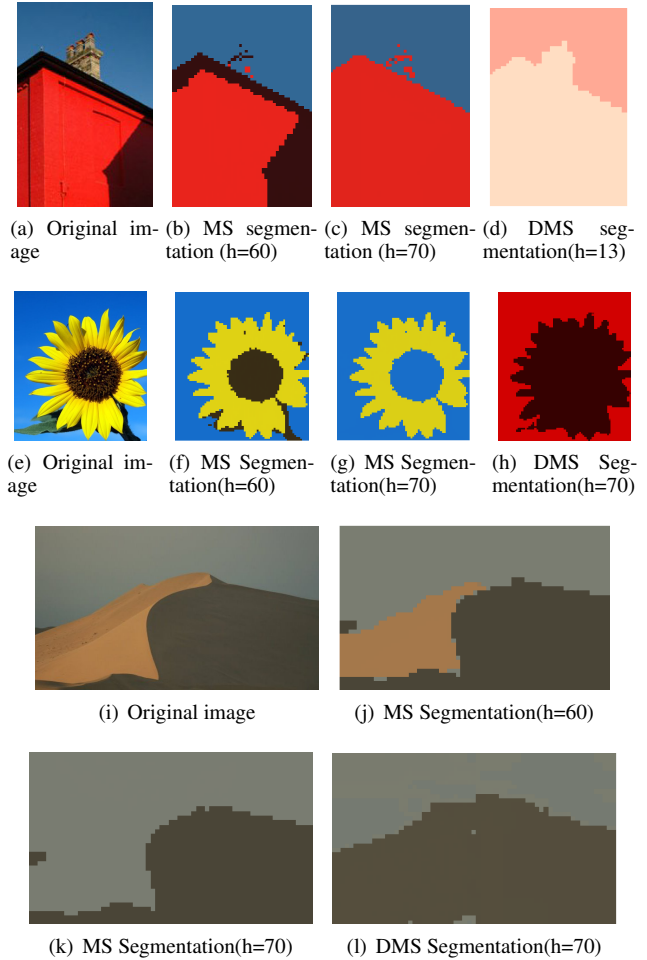(k) MS Segmentation(h=70)
(l) DMS Segmentation(h=70)

Figure 2. Image Segmentation with MS and DMS.

where x, y, and z represent the original 3D Cartesian coordinates, r is the radial distance from each point to the origin, θ is the zenith angle, and ϕ is the azimuth angle. Each 2D unfolded image will then be a 2D directional data in a domain (θ,ϕ). This directional image is then subjected to an intensity thresholding and normalization. Considering it as a probability distribution, a set of points are drawn. Bahlmann's algorithm performs an EM algorithm to fit a GMM on this point set. To handle the directional data, the GMM and the corresponding EM algorithm was extended using the similar idea to ours [1].

In this study, we propose a new directional non-parametric clustering algorithm, replacing the EM-based GMM fitting in [2] by the proposed DMS. A directional clustering algorithm is derived by following the standard method in [3] except to use the proposed DMS instead of the original mean shift. Each sample point in the 2D unfolded image is considered as a unit vector (θ,ϕ) represented in radians. The directional distance of any paired sample points can then be modeled by the directional distance definition. The resulting clustering algorithm does not require the com-

ponent number *a priori* and can robustly handle irregular-shaped regions due to its non-parametric nature. Figure 3(c) compares the result of the proposed and original methods for the periodic wraparound attached lung nodule case.
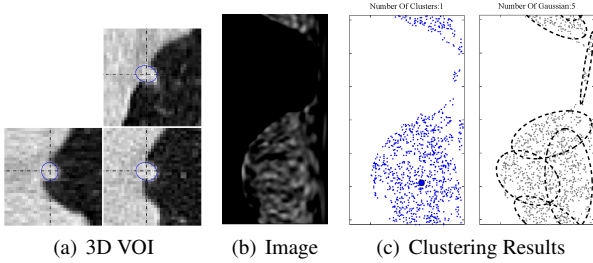

(a) 3D VOI   (b) Image   (c) Clustering Results

Figure 3. Illustration of 3D structure topology classification. (a) input 3D CT volume: lung nodule attached to a wall in three orthogonal viewing planes. (b) 2D unfolded image in the (θ, ϕ)-domain after thresholding. (c) lhs: DMS result presented in color markers, rhs: fitted GMM shown in dashed ellipses.

5.1. Experimental Results

We performed quantitative experiments in order to compare the proposed DMS algorithm and the original EM algorithm by [2]. For the bandwidth value we used values ranging from 0.3 to 0.5 in radians. Typically, for smaller cluster size which usually appears in vessel and vessel-branch cases, we use values close to 0.3; for bigger cluster size which is often generated by attached-nodule and nodule cases, we use values close to 0.5. These kernel values are derived from our pilot study with all data input. We use a data set of 33 local volumetric regions whose size is 32 cubic voxel and taken manually from high-resolution lung CT scans. There are 9 attached nodules, 10 vessels, and 14 vessel branches among the test set. Results by the algorithm in [2] are created by using the software provided by the authors. Our proposed directional clustering with the DMS was implemented in MATLAB. For these data, our DMS-based method resulted in no error while the original EM-based method resulted in 87.8% success rate with 4 failure cases. Two of the failures were branch cases that were classified as vessels by combining two islands as one Gaussian. The other two failures were attached nodule cases classified as vessel cases. This was caused by periodicity in the unfolded data, although our DMS-based method was successful for integrating the multiple directional islands into one island correctly.

Figure 4, 5 and 6 display 48 illustrative cases including vessels, vessel-branches and wall-attached nodule for comparing the two algorithms. In each presented image, the left column shows the results by our proposed DMS-based clustering method. Estimated cluster centers are marked with circles whose location corresponds to the DMS's convergence. The estimated number of clusters are shown at the top of each figure and cluster membership is denoted by different symbols and colors. The right column shows the results by the original EM-based clustering method. Dashed ellipses illustrate Gaussian mixture models fitted to the data while the number of fitted Gaussians is shown at the top of each figure. Figure 5(e) and (f) denote the two branch fail-

ure cases while figure 6(e) and (f) show the two attached nodule failure cases by the original method. Note also that the proposed method successfully clusters irregular-shaped islands (e.g., (a) and (b) in Figure 4, (a), (b) and (c) in Figure 5) while the original method heavily over-estimates the number of components. Also note that directional statistics have been successfully handled with proposed method (e.g., (c) and (d) in Figure 4, (d) in Figure 5).

6. Conclusion

This paper proposes DMS, a new directional nonparametric clustering algorithm used for 3D medical structure topology classification. DMS generalizes the linear space mean shift to directional data. We show its correctness by providing a formal convergence proof and demonstrate its usefulness by applying it to a clinically-relevant medical image analysis application. Our experimental results demonstrated its effectiveness with significant improvement over the original EM algorithm in handling directional data. Our algorithm suggested a significant 12.2% performance improvement for solving the 3D topology type classification problem.

In future work, we plan to perform further quantitative performance evaluation of the proposed algorithm and consider applying the proposed method to more vision problems with directional data, such as omni camera and cartographic image analyses. Another standard way to define a KDE function with directional data is with the Von Mises-Fisher distribution [13]. It is of our theoretical interest to derive an algorithm that is convergent to modes of such directional KDEs and compare/unify with the proposed DMS method.

A. Proof of Theorem 1.

Proof. The directional difference vector (6) for $\mathbf{x}$ and $\mathbf{x}_i$ is defined as

$$s_{\mathbf{x},\mathbf{x}_i} = \begin{cases} sgn(\mathbf{x} - \mathbf{x}_i) \times r - (\mathbf{x} - \mathbf{x}_i), & |\mathbf{x} - \mathbf{x}_i| > \frac{r}{2} \\ \mathbf{x} - \mathbf{x}_i & \text{otherwise} \end{cases} \quad (27)$$

Introducing $s_{\mathbf{x},\mathbf{x}_i}$ into (22), we obtain

$$\mathbf{x}'_i = \begin{cases} sgn(\mathbf{x} - \mathbf{x}_i) \times r + \mathbf{x}_i, & |\mathbf{x} - \mathbf{x}_i| > \frac{r}{2} \\ \mathbf{x}_i & \text{otherwise} \end{cases} \quad (28)$$

The definition for repositioned data point locations in (28) is a function of $\mathbf{x}$ and independent from $s_{\mathbf{x},\mathbf{x}_i}$, therefore it can be introduced into $s_{\mathbf{x},\mathbf{x}_i}$ without changing the properties of it which gives $s_{\mathbf{x},\mathbf{x}_i} = \mathbf{x} - \mathbf{x}'_i$. Thus, the directional mean shift equation (26) could be rewritten as

$$\mathbf{y}'_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}'_i g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}'_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}'_i}{h}\right\|^2\right)}, \quad j = 1, 2, ... \quad (29)$$

Comaniciu and Meer [4, Theorem 1:(A.6),(A.7)] show for an arbitrary kernel location $\mathbf{y}_j \neq 0$, the sequence
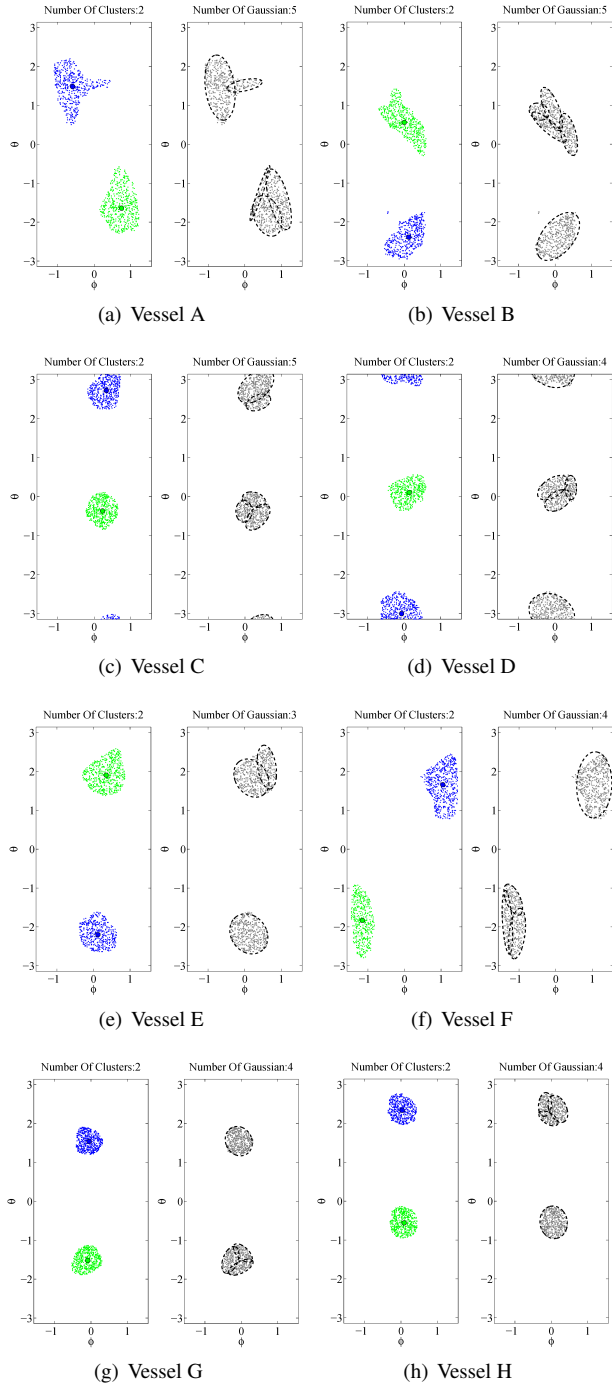
Figure 4. Illustration of DMS and EM clustering results on Vessel cases. (a) and (b) Irregular-shaped clusters (c) and (d) Clusters with directional statistics (e), (f), (g) and (h) Normal clusters.
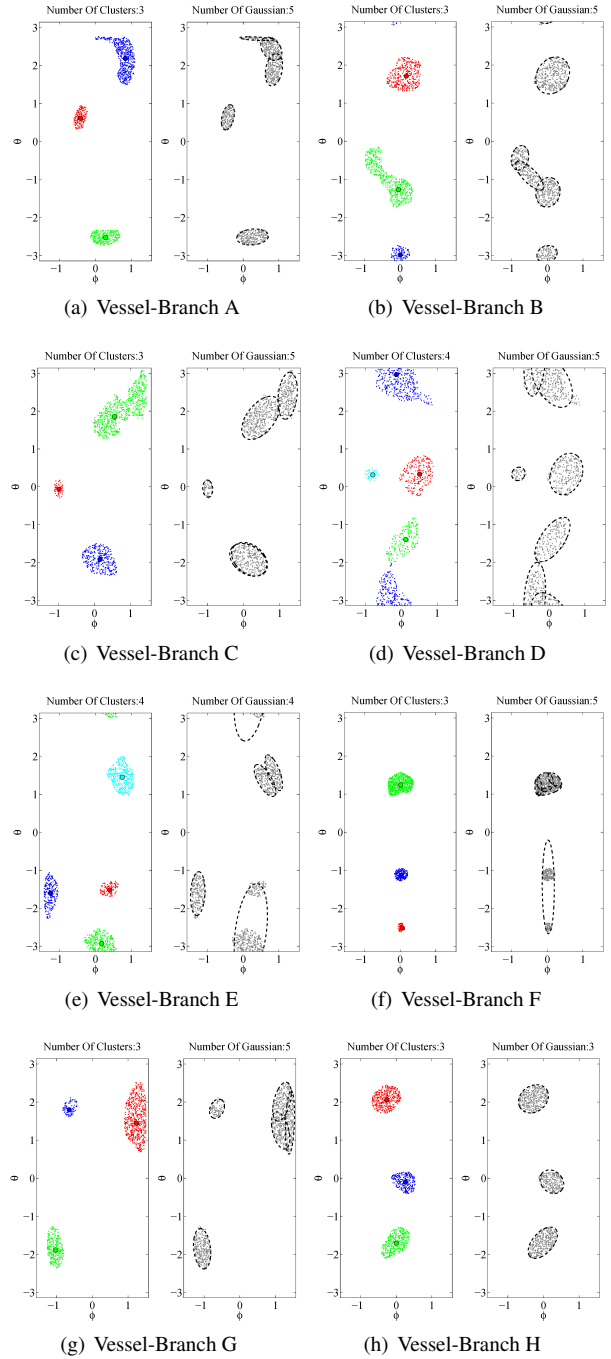

Figure 5. Illustration of DMS and EM clustering results on Vessel-Branch cases. (a), (b) and (c) Irregular-shaped clusters (d) Clusters with directional statistics (e), (f), (g) and (h) False prediction from EM-clustering.

$\{\hat{f}(\mathbf{y}_j)\}_{j=1,2,...}$ is monotonically increasing and converges. They also prove $\{\mathbf{y}_j\}_{j=1,2,...}$ is a Cauchy sequence and therefore convergent. Considering the linear transformation $t_d(\mathbf{x}_i) = \mathbf{x}_i \pm r$, we conclude without loss of generality that $\mathbf{x}_i$ can be recentered on a new coordinate origin and

transformed by $\pm r$ within radius of convergence of $\hat{f}$ which gives $\hat{f}(\mathbf{y}'_{j+1}) > \hat{f}(\mathbf{y}'_j)$. In other words $\{\hat{f}(\mathbf{y}'_j)\}_{j=1,2,...}$ is monotonically increasing and therefore converges. In addition to this, similar to the proof in [4, Theorem 1(A.7)] it is obvious that $\{\mathbf{y}'_j\}_{j=1,2,...}$ is also convergent. $\qquad\square$
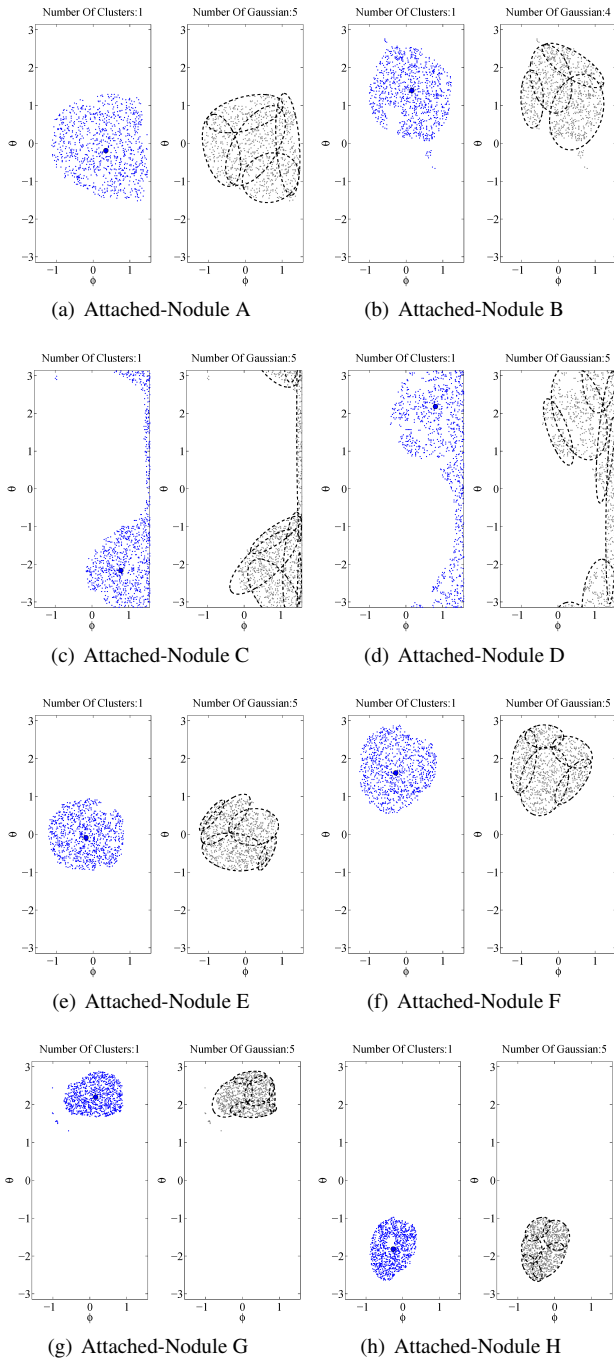
Figure 6. Illustration of DMS and EM clustering results on Attached-nodule cases. (a) and (b) Cluster size is relatively large (c) and (d) Bicubic interpolation results in unproportional cluster size along the edge (e), (f), (g) and (h) Normal clusters.

References

[1] C. Bahlmann. Directional features in online handwriting recognition. *Pattern Recognition*, 39(1):115–125, 2006. 1, 5

[2] C. Bahlmann, X. Li, and K. Okada. Local pulmonary structure classification for computer-aided nodule detection. In *Proc. SPIE Conf. Medical Imaging*, volume 6144, pages 1–5, 2006. 1, 2, 5, 6

[3] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. and Machine Intell.*, 17(8):790–799, 1995. 1, 2, 5

[4] C. Comaniciu and P. Meer. Mean shift: A Robust Approach Toward Feature Space Analysis. *IEEE Trans. Pattern Anal. and Machine Intell.*, 24(5):603–619, 2002. 1, 2, 6, 7

[5] G. Danuser and M. Stricker. Parametric model fitting: From inlier characterization to outlier detection. *IEEE Trans. Pattern Anal. and Machine Intell.*, 20(3):263–280, 1998. 2

[6] D. DeCarlo and D. Metaxas. Shape evolution with structural and topological changes using blending. *IEEE Trans. Pattern Anal. and Machine Intell.*, 20(11):1186–1205, 1998. 2

[7] A. F. Frangi, W. J. Niessen, R. M. Hoogeveen, T. van Walsum, and M. A. Viergever. Model-based quantitation of 3D magnetic resonance angiographic images. *IEEE Trans. Medical Imaging*, 18:946–956, 1999. 2

[8] J. Fujiki, A. Torii, and S. Akaho. Epipolar geometry via rectification of spherical images. In *MIRAGE*, pages 461–471, 2007. 1

[9] K. Fukunaga and L. Hostetler. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. *IEEE Trans. Information Theory*, 21:32–40, 1975. 1, 2

[10] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt. Shape representation and classification using the poisson equation. In *Computer Vision and Pattern Recognition, 2004.*, volume 2, pages 61–67, 2004. 2

[11] M. Kafai. Directional Mean Shift, 2009. MSc. master project, San Francisco State University. 3

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Computer Vision*, 60(2):91–110, 2004. 1

[13] K. Mardia and P. Jupp. *Directional Statistics*. Wiley, 1999. 1, 2, 3, 6

[14] Y. Miao. Pulmonary Structure Classification Using Bayesian Framework, 2008. MSc. master project, San Francisco State University. 5

[15] X. Qian, M. Brennan, D. Dione, W. Dobrucki, M. Jackowski, C. Breuer, A. Sinusas, and X. Papademetris. A nonparametric vessel detection method for complex vascular structures. *Medical Image Analysis*, 13(1):49–61, 2009. 2

[16] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis. Segmentation and visualization of curvilinear structures in medical images. *Medical Image Analysis*, 2:143–168, 1998. 2

[17] Y. Sato, C. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura, and R. Kikinis. Tissue classification based on 3D local intensity structures for volume rendering. *IEEE Trans. Visualization and Computer Graphics*, 6, 2000. 2

[18] D. Scott. *Multivariate Density Estimation*. New York: Wiley, 1992. 2

[19] B. Silverman. *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall/CRC, 1998. 2

[20] R. Subbarao and P. Meer. Nonlinear mean shift for clustering over analytic manifolds. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 1168–1175, Los Alamitos, CA, USA, 2006. IEEE Computer Society. 1

[21] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison Wesley, May 2005. 1

[22] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. *IEEE Trans. Pattern Anal. and Machine Intell.*, 30(10):1713–1727, 2008. 1