Hindawi
Mathematical Problems in Engineering
Volume 2017, Article ID 3696850, 11 pages
https://doi.org/10.1155/2017/3696850

Research Article

A Clustering Method for Data in Cylindrical Coordinates

Kazuhisa Fujita[1,2]

[1] *University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan*
[2] *Department of Computer and Information Engineering, National Institute of Technology, Tsuyama College, 654-1 Numa, Tsuyama, Okayama 708-8506, Japan*

Correspondence should be addressed to Kazuhisa Fujita; k-z@nerve.pc.uec.ac.jp

Received 7 April 2017; Revised 18 July 2017; Accepted 16 August 2017; Published 27 September 2017

Academic Editor: Ivan Giorgio

We propose a new clustering method for data in cylindrical coordinates based on the k-means. The goal of the k-means family is to maximize an optimization function, which requires a similarity. Thus, we need a new similarity to obtain the new clustering method for data in cylindrical coordinates. In this study, we first derive a new similarity for the new clustering method by assuming a particular probabilistic model. A data point in cylindrical coordinates has radius, azimuth, and height. We assume that the azimuth is sampled from a von Mises distribution and the radius and the height are independently generated from isotropic Gaussian distributions. We derive the new similarity from the log likelihood of the assumed probability distribution. Our experiments demonstrate that the proposed method using the new similarity can appropriately partition synthetic data defined in cylindrical coordinates. Furthermore, we apply the proposed method to color image quantization and show that the methods successfully quantize a color image with respect to the hue element.

1. Introduction

Clustering is an important technique in many areas such as data analysis, data visualization, image processing, and pattern recognition. The most popular and useful clustering method is the k-means. The k-means uses the Euclidean distance as coefficient and partitions data to k clusters. The Euclidean distance is a reasonable measurement for data sampled from an isotropic Gaussian distribution. We cannot always obtain a good clustering result using the k-means because not all data distributions are isotropic Gaussian distributions.

The present study focuses on data in cylindrical coordinates. Data in cylindrical coordinates have a periodic element, so clustering methods using the Euclidean distance will lead to an improper analysis of the data. Furthermore, a clustering method using the Euclidean distance may not be able to extract meaningful centroids. For example, if a distribution in cylindrical coordinates is remarkably curved crescent-shape, the centroid of the distribution calculated by the k-means may not be on the data distribution. However, there are no clustering methods optimized for data in cylindrical coordinates.

The cylindrical data are found in many fields such as image processing, meteorology, and biology. Movements of plants and animals and wind direction with another environmental measure are typical examples of cylindrical data [1]. The most popular example of data in cylindrical coordinates is color defined in the HSV color model. The HSV color has three attributes that are hue (direction), saturation (radius), and value that means brightness (height). The HSV color model can represent hue information and has a more natural correspondence to human vision than the RGB color model [2]. The clustering method for cylindrical coordinates is useful for many fields, especially image processing.

The purpose of this study is to develop a new clustering method for data in cylindrical coordinates based on the k-means. We first derive a new similarity for clustering data in cylindrical coordinates assuming that the data are sampled from a probabilistic model that is the product of a von Mises distribution and Gaussian distributions. We propose a new clustering method with this new similarity for data in cylindrical coordinates. Using numerical experiments, we demonstrate that the proposed method can partition synthetic data. Furthermore, we evaluate the performance of the proposed method for real world data. Finally, we apply the proposed method to color image quantization and demonstrate that it can quantize a color image according to the hue.

2. Related Works

The most commonly used clustering method is the k-means [3], which is one of the top 10 most common algorithms used in data mining [4]. We have applied the k-means to various fields because it is fast, simple, and easy to understand. It uses the Euclidean distance as a clustering criterion and assumes that the data is sampled from a mixture of isotropic Gaussian distributions. Thus, we can apply the k-means to data sampled from a mixture of isotropic Gaussian distributions, but the k-means is not appropriate for data generated from other distributions. Data in cylindrical coordinates have periodic characteristics, so the k-means will be inappropriate as a clustering method for the data.

We can cluster periodic data distributed on an n-dimensional sphere surface using the spherical k-means (sk-means). Dhillon and Modha [5] and Banerjee et al. [6, 7] have developed the sk-means for clustering high dimensional text data. It is a k-means based method that uses cosine similarity as the criterion for clustering. The sk-means assumes that the data are sampled from a mixture of von Mises-Fisher distributions with the same concentrate parameters and the same mixture weights. However, we cannot apply the sk-means to data that have direction, radius, and height. To appropriately partition these data, we need a different nonlinear separation method.

There are many methods for achieving nonlinear separation. One method is the kernel k-means [8], which partitions the data points in a higher-dimensional feature space after they are mapped to the feature space using a nonlinear function. The spectral clustering [9] is another popular modern nonlinear clustering method, which uses the eigenvectors of a similarity (kernel) matrix to partition data points. The support vector clustering [10] is inspired by the support vector machine [11]. These nonlinear clustering methods based on the kernel method can provide reasonable clustering results for non-Gaussian data. However, these methods can hardly provide significant statistics because they perform the clustering in a feature space. This is a problem when we also want to determine some features of data, such as color image quantization. Furthermore, we must experimentally select the optimal kernel functions and its parameters.

Clustering methods are frequently used for color image quantization. Color image quantization reduces the number of colors in an image and plays an important role in applications such as image segmentation [12], image compression [13], and color feature extraction [14]. A color quantization technique consists of two stages: the palette design stage and the pixel mapping stage. These stages can be, respectively, regarded as calculating the centroids and assigning a data point to a cluster. Many researchers have developed color quantization methods including median cut [15], the k-means [16], the fuzzy c-means [17, 18], the self-organizing maps [19–21], and the particle swarm optimization [22]. However, generally, color quantization is performed in the RGB color space although HSV color space is rarely adopted.

3. Methodology

3.1. Assumed Probabilistic Distribution. A data point in cylindrical coordinates, $\mathbf{x}$, is represented by $\mathbf{x} = (r, \mathbf{u}, z)$ with $\|\mathbf{u}\| = 1$ and $r \geq 0$, where $r, \mathbf{u}, z$ are called the radius, azimuth, and height, respectively. In this study, we represent the azimuth as a unit vector $\mathbf{u} = (u_x, u_y)$ to simply calculate the cosine similarity. Here, each element of $\mathbf{x}$ is assumed to be independent and identically distributed. Let a data point in cylindrical coordinates, $\mathbf{x} = (r, \mathbf{u}, z)$, be generated by a probability density function (pdf) of the form

$$p(\mathbf{x} \mid \theta) = \mathrm{vM}(\mathbf{u} \mid \boldsymbol{\mu}_u, \kappa) N(r \mid \mu_r, \sigma_r^2) N(z \mid \mu_z, \sigma_z^2), \tag{1}$$

where $\theta = \{\boldsymbol{\mu}_u, \mu_r, \mu_z, \kappa, \sigma_r^2, \sigma_z^2\}$ and $\mathrm{vM}(\cdot)$ and $N(\cdot)$ are pdfs of a von Mises distribution and an isotropic Gaussian distribution, respectively. A pdf of a von Mises distribution $\mathrm{vM}(\cdot)$ has the form

$$\mathrm{vM}(\mathbf{u} \mid \boldsymbol{\mu}_u, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \boldsymbol{\mu}_u^{\mathrm{T}} \mathbf{u}), \tag{2}$$

where $\boldsymbol{\mu}_u$ is the mean of the azimuth with $\|\boldsymbol{\mu}_u\| = 1$, κ is the concentrate parameter, and $I_0(\cdot)$ is the modified Bessel function of the first kind (order 0). A pdf of an isotropic Gaussian distribution $N(\cdot)$ has the form

$$N(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \tag{3}$$

where μ is the mean and σ^2 is the variance. μ_r and μ_z are the means of the radius and height, respectively. σ_r^2 and σ_z^2 are the variances of radius and height, respectively. Thus, the density $p(\mathbf{x} \mid \theta)$ can be written as

$$p(\mathbf{x} \mid \theta) = \frac{1}{4\pi^2 \sigma_r \sigma_z I_0(\kappa)} \cdot \exp\left(\kappa \boldsymbol{\mu}_u^{\mathrm{T}} \mathbf{u} - \frac{(r-\mu_r)^2}{2\sigma_r^2} - \frac{(z-\mu_z)^2}{2\sigma_z^2}\right). \tag{4}$$

We can estimate the parameters of density $p(\mathbf{x} \mid \theta)$ using maximum likelihood estimation. Let data set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be generated from density $p(X \mid \theta)$. The log likelihood function of $p(X \mid \theta)$ is

$$\ln p(X \mid \theta) = \sum_{i=1}^{N} \ln p(\mathbf{x}_i \mid \theta) = \sum_{i=1}^{N} \left(-\ln(4\pi^2 \sigma_r \sigma_z I_0(\kappa)) + \kappa \boldsymbol{\mu}_u^{\mathrm{T}} \mathbf{u}_i - \frac{(r_i-\mu_r)^2}{2\sigma_r^2} - \frac{(z_i-\mu_z)^2}{2\sigma_z^2}\right). \tag{5}$$

Maximizing this equation subject to $\boldsymbol{\mu}_u^T\boldsymbol{\mu}_u = 1$, we find the maximum likelihood estimates $\widehat{\boldsymbol{\mu}}_u$, $\widehat{\mu}_r$, $\widehat{\mu}_z$, $A(\widehat{\kappa})$, $\widehat{\sigma}_r$, and $\widehat{\sigma}_z$ obtained from

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_u &= \frac{\sum_{i=1}^{N}\mathbf{u}_i}{\left\|\sum_{i=1}^{N}\mathbf{u}_i\right\|}, \\ A(\widehat{\kappa}) &= \overline{R}, \\ \widehat{\mu}_r &= \frac{1}{N}\sum_{i=1}^{N} r_i, \\ \widehat{\sigma}_r^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(r_i - \widehat{\mu}_r\right)^2, \\ \widehat{\mu}_z &= \frac{1}{N}\sum_{i=1}^{N} z_i, \\ \widehat{\sigma}_z^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(z_i - \widehat{\mu}_z\right)^2, \end{aligned} \tag{6}$$

where $A(\widehat{\kappa})$ is

$$A(\widehat{\kappa}) = \frac{I_0'(\widehat{\kappa})}{I_0(\widehat{\kappa})} = \frac{\left\|\sum_{i=1}^{N}\mathbf{u}_i\right\|}{N} = \overline{R}. \tag{7}$$

It is difficult to estimate the concentrate parameter κ, because an analytic solution cannot be obtained using the maximum likelihood estimate and we can only calculate the ratio of the Bessel functions. We approximate κ using the numerical method proposed by Sra [23], because it produces the most accurate estimates for κ (compared to other methods).

We estimate κ using the recursive function

$$\begin{aligned} \kappa_n &= \kappa_{n-1} \\ &\quad - \frac{A_p(\kappa_{n-1}) - \overline{R}}{1 - A_p(\kappa_{n-1})^2 - ((p-1)/\kappa_{n-1}) A_p(\kappa_{n-1})}, \end{aligned} \tag{8}$$

where n is the iteration number. The recursive calculations terminate when $|\kappa_n - \kappa_{n-1}| < \varepsilon_\kappa$. In this study, $\varepsilon_\kappa = 0.001$. We calculate κ_0 using the method proposed by Banerjee et al. [6]. κ_0 is

$$\kappa_0 = \frac{\overline{R}\left(2 - \overline{R}^2\right)}{1 - \overline{R}^2}. \tag{9}$$

3.2. Cylindrical k-Means. The k-means family uses a particular similarity to decide whether a data point belongs to a cluster. The Euclidean distance (dissimilarity) is most frequently used by the k-means family, and, moreover, is derived using the log likelihood of an isotropic Gaussian distribution. Therefore, the k-means using the Euclidean distance will be able to appropriately partition data sampled from isotropic Gaussian distributions but not other distributions. We must develop a new similarity for data in cylindrical coordinates because the k-means family clusters by maximizing the sum of similarities between a centroid of a cluster and data points that belong to the cluster. In this study, we obtain the optimal similarity for partitioning data in cylindrical coordinates from an assumed pdf.

First, to develop a k-means based method for data in cylindrical coordinates (cylindrical k-means; cyk-means), we obtain a new similarity measure for data in cylindrical coordinates by assuming a probability distribution. Assume that a data point $\mathbf{x}$ in a cluster that has a centroid $\mathbf{m} = (\mu_r, \boldsymbol{\mu}_u, \mu_z)$ is sampled from the probability distribution $p(\mathbf{x} \mid \theta)$ denoted by (4) where $\theta = (\mathbf{m}, \kappa, \sigma_r, \sigma_z)$. The natural logarithm of $p(\mathbf{x} \mid \theta)$ is

$$\ln p(\mathbf{x} \mid \theta) = \kappa\boldsymbol{\mu}_u^{\mathrm{T}}\mathbf{u} - \frac{(r - \mu_r)^2}{2\sigma_r^2} - \frac{(z - \mu_z)^2}{2\sigma_z^2} + \ln C, \tag{10}$$

where C is a normalizing constant given by

$$C = \frac{1}{4\pi^2\sigma_r\sigma_z I_0(\kappa)}. \tag{11}$$

Here, we ignore the normalizing constant $\ln C$ to obtain

$$S(\mathbf{x}, \mathbf{m}) = \kappa\boldsymbol{\mu}_u^{\mathrm{T}}\mathbf{u} - \frac{(r - \mu_r)^2}{2\sigma_r^2} - \frac{(z - \mu_z)^2}{2\sigma_z^2}. \tag{12}$$

In this study, this equation is used as a similarity for the cyk-means. $S(\mathbf{x}, \mathbf{m})$ denotes the similarity between the data point $\mathbf{x}$ and the centroid $\mathbf{m}$. The terms in (12) consist of the cosine similarity and the Euclidean similarities, and the new similarity is a sum of these similarities weighed. The weights indicate the concentrations of distributions. This similarity can also be considered as a simplified log likelihood.

The cyk-means partitions data points in cylindrical coordinates into K clusters using the procedure same as the k-means. Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a set of data points in cylindrical coordinates. Let $\mathbf{m}_j = (\mu_{rj}, \boldsymbol{\mu}_{uj}, \mu_{zj})$ be the centroid of the jth cluster. Using the similarity $S(\mathbf{x}_i, \mathbf{m}_j)$, the objective function J is

$$\begin{aligned} J &= \sum_{i=1}^{N}\sum_{j=1}^{K}\gamma_{ij} S\left(\mathbf{x}_i, \mathbf{m}_j\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{K}\gamma_{ij}\left(\kappa\boldsymbol{\mu}_{uj}^{\mathrm{T}}\mathbf{u}_i - \frac{\left(r_i - \mu_{rj}\right)^2}{2\sigma_{rj}^2} - \frac{\left(z_i - \mu_{zj}\right)^2}{2\sigma_{zj}^2}\right), \end{aligned} \tag{13}$$

where γ_{ij} is a binary indicator value. If the ith data point belongs to the jth cluster, $\gamma_{ij} = 1$. Otherwise, $\gamma_{ij} = 0$. The aim of the cyk-means is to maximize the objective function J. The process to maximize the objective function is the same as that of the k-means and is described as follows.

(1) Fix K and initialize $\mathbf{m}_j$.

(2) Assign each data point to the cluster that has the most similar centroid.

(3) Estimate parameters of clusters.

(4) Return to Step (2) if the cluster assignment of data points changes or the difference in the values of the optimal function from the current and last iteration is more than a threshold ε_J. Otherwise, terminate the procedure.

```
Input: Set $X$ of data points in cylindrical coordinates
Output: A clustering of $X$
(1) Initialize $\theta_j,\ \ j = 1, \ldots, K$
(2) repeat
(3)     Set $\gamma_{ij} = 0,\ \ i = 1, \ldots, N,\ \ j = 1, \ldots, K$
(4)     {Assign data points to clusters}
(5)     for $i = 0$ to $N$ do
(6)         $j \longleftarrow \arg\max_{j'} \left( \kappa_{j'} \boldsymbol{\mu}_{uj'}^{\mathrm{T}} \mathbf{u}_i - \frac{(r_i - \mu_{rj'})^2}{2\sigma_{rj'}^2} - \frac{(z_i - \mu_{zj'})^2}{2\sigma_{zj'}^2} \right)$
(7)         $\gamma_{ij} = 1$
(8)     end for
(9)     {Estimate parameters}
(10)    for $j = 1$ to $K$ do
(11)        $N_j = \sum_{i=1}^{N} \gamma_{ij}$
(12)        $\mu_{rj} = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} r_i$
(13)        $\boldsymbol{\mu}_{uj} = \frac{\sum_{i=1}^{N} \gamma_{ij} \mathbf{u}_i}{\left\| \sum_{i=1}^{N} \gamma_{ij} \mathbf{u}_i \right\|}$
(14)        $\mu_{zj} = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} z_i$
(15)        $\sigma_{rj}^2 = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} (r_i - \mu_{rj})^2$
(16)        $\sigma_{zj}^2 = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} (z_i - \mu_{zj})^2$
(17)        $\overline{R}_j = \frac{\left\| \sum_{i=1}^{N} \gamma_{ij} \mathbf{u}_i \right\|}{N_j}$
(18)        $\kappa_j = \frac{\overline{R}_j (2 - \overline{R}_j^2)}{1 - \overline{R}_j^2}$
(19)        repeat
(20)            $\kappa_j \longleftarrow \kappa_j - \frac{A_p(\kappa_j) - \overline{R}_j}{1 - A_p(\kappa_j)^2 - ((p-1)/\kappa_j) A_p(\kappa_j)}$
(21)        until convergence
(22)    end for
(23) until convergence
```

Algorithm 1: cyk-means.

In this study, we use $\varepsilon_J = |0.001 \times J_n|$ where J_n is the objective function of the nth iteration. Algorithm 1 shows the details of the algorithm of the cyk-means. From (6), the elements of the centroid vector, $\mathbf{m}_j = (\mu_{rj}, \boldsymbol{\mu}_{uj}, \mu_{zj})$, of the jth cluster are

$$\mu_{rj} = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} r_i,$$
$$\boldsymbol{\mu}_{uj} = \frac{\sum_{i=1}^{N} \gamma_{ij} \mathbf{u}}{\left\| \sum_{i=1}^{N} \gamma_{ij} \mathbf{u} \right\|}, \qquad (14)$$
$$\mu_{rj} = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} z_i,$$

where N_j is the number of data points in the jth cluster (which has the form $N_j = \sum_{i=1}^{N} \gamma_{ij}$). The other values used to calculate the objective function are

$$\overline{R}_j = \frac{\left\| \sum_{i=1}^{N} \gamma_{ij} \mathbf{u}_i \right\|}{N_j},$$
$$\sigma_{rj}^2 = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} (r_i - \mu_{rj})^2, \qquad (15)$$
$$\sigma_{zj}^2 = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} (z_i - \mu_{zj})^2.$$

κ_j is approximated by Sra's method using the ratio of the Bessel function $\overline{R}_j$.

```
Input: Set $X$ of data points in cylindrical coordinates
Output: A clustering of $X$
Initialize $\theta_j,\ j = 1, \ldots, K$
repeat
    Set $\gamma_{ij} = 0,\ i = 1, \ldots, N,\ j = 1, \ldots, K$
    {Assign data points to clusters}
    for $i = 0$ to $N$ do
        $j \longleftarrow \arg\max_{j'} \left( \kappa_{j'} \boldsymbol{\mu}_{uj'}^{\mathrm{T}} \mathbf{u}_i - \frac{(r_i - \mu_{rj'})^2}{2\sigma_{rj'}^2} - \frac{(z_i - \mu_{zj'})^2}{2\sigma_{zj'}^2} \right)$
        $\gamma_{ij} = 1$
    end for
    {Estimate parameters}
    for $j = 1$ to $K$ do
        $N_j = \sum_{i=1}^{N} \gamma_{ij}$
        $\mu_{rj} = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} r_i$
        $\boldsymbol{\mu}_{uj} = \frac{\sum_{i=1}^{N} \gamma_{ij} \mathbf{u}_i}{\left\| \sum_{i=1}^{N} \gamma_{ij} \mathbf{u}_i \right\|}$
        $\mu_{rj} = \frac{1}{N_j} \sum_{i=1}^{N} \gamma_{ij} z_i$
    end for
until convergence
```

Algorithm 2: Fixed cyk-means.

The cyk-means method has many parameters. The k-means method for data in three-dimensional Cartesian coordinates has only $3K$ parameters, which are multiples of the number of centroid vectors and dimensions. However, the cyk-means has $7K$ parameters, which are multiples of the number of clusters and the number of parameters of a cluster. The parameters of the jth cluster are μ_{rj}, $\boldsymbol{\mu}_{uj}$ (two dimensions), μ_{rj}, σ_{rj}, κ_j, and σ_{zj}. Because the cyk-means has more degrees of freedom, the dead unit problem (i.e., empty clusters) will frequently occur if the initial k is not optimal.

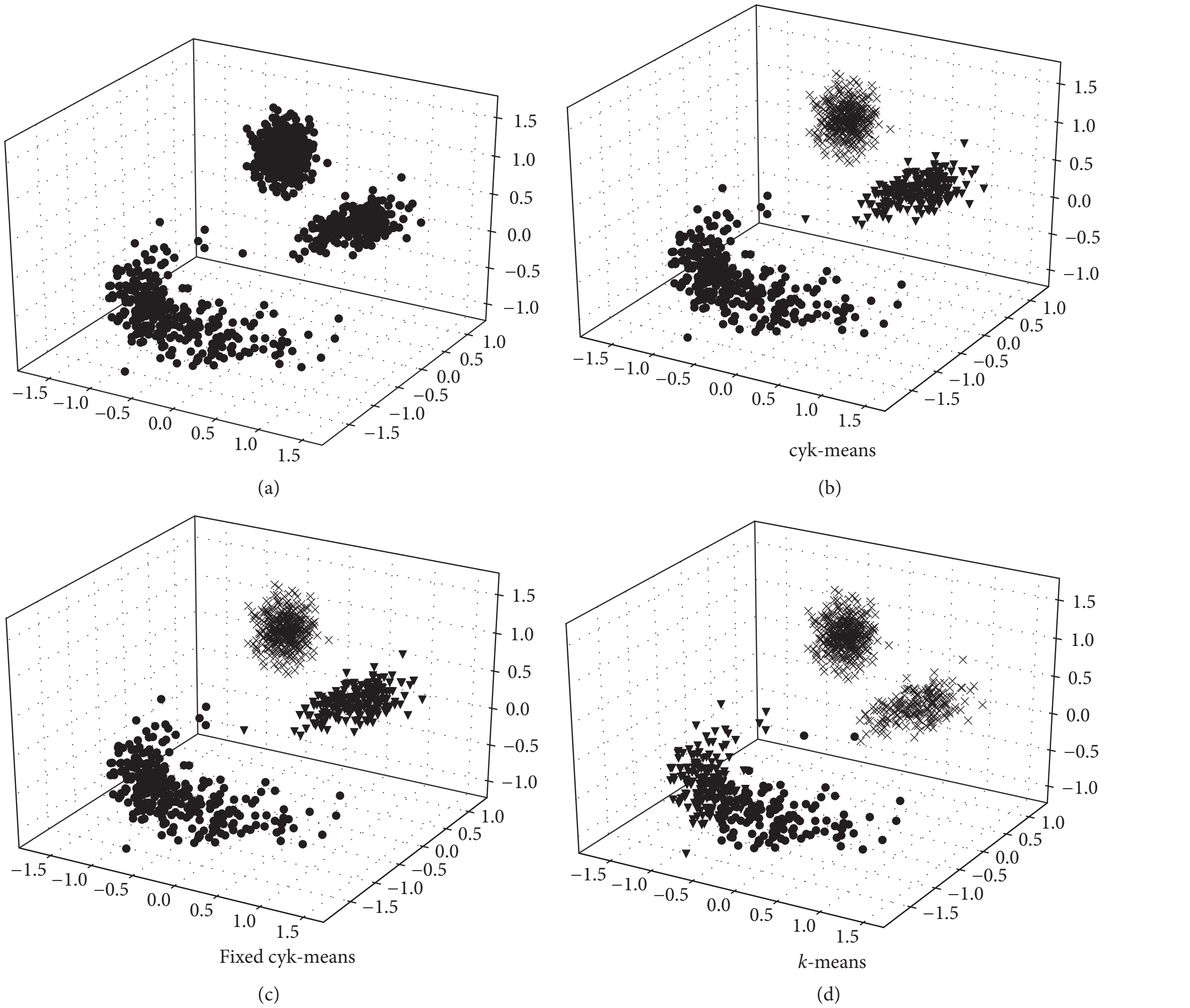
3.3. Fixed cyk-Means. Model based clustering methods have various problems such as the dead units and initial value problems. One reason for this is that the log likelihood equation can have many local optima [9]. If a model has more parameters, these problems tend to be more frequent. In the fixed cyk-means, the concentrate parameter κ and the variances σ^2s are fixed for particular values. As a consequence, the fixed cyk-means has $4K$ parameters. Fixing the parameters decreases the complexity of the model and makes these problems less. Algorithm 2 indicates the fixed cyk-means algorithm.

3.4. Computational Complexity. Assigning data points to clusters has a complexity of $O(KN)$ per iteration. We must estimate six parameters. We obtain three μs, two σs, and κ in $O(6KN + Kt_{\kappa\max})$ time per iteration, where $t_{\kappa\max}$ is the convergence time of κ. Therefore, the total computational complexity per iteration is $O(7KN + Kt_{\kappa\max})$. The complexity of the fixed cyk-means is $O(5KN)$ per iteration, so the cyk-means is approximately 1.5 times as complex as the fixed cyk-means.

4. Experimental Results

In our experiments, we use Python and its libraries (NumPy, SciPy, and scikit-learn) to implement the proposed method.

4.1. Synthetic Data. In this subsection, we demonstrate that the cyk-means and the fixed cyk-means can partition synthetic data that is defined using cylindrical coordinates. The dataset used in this experience has three clusters, as shown in Figure 1(a). The data points in each cluster are generated from the probability distribution denoted by (4), with the parameters shown in Table 1. Figures 1(b), 1(c), and 1(d) show the clustering results of the cyk-means, the fixed cyk-means with $\kappa = 25$, $\sigma_r = 0.1$, and $\sigma_z = 0.1$, and the k-means, respectively. We can see that the cyk-means and the fixed cyk-means properly partition the dataset into each cluster. On the other hand, the k-means regards two upper right clusters as one cluster and unsuccessfully partitions the dataset. Table 2 shows the parameters estimated by the cyk-means, the fixed cyk-means, and the k-means. The cyk-means can only estimate the concentrate parameters and the variances. The values of the concentrate parameters and the variances estimated by the cyk-means are approximate to the true values. The cyk-means most appropriately estimates the



(a)

(b)

(c)

(d)

Figure 1: (a) Scatter plot of the synthetic dataset including three clusters. (b), (c), (d) Clustering results of the cyk-means, the fixed cyk-means, and the k-means. The three clusters are shown with circles, triangles, and crosses.

Table 1: Parameters of dataset.

j	N_j	$\mu_{\theta j}$	κ_j	μ_{rj}	σ_{rj}	μ_{zj}	σ_{zj}
1	200	0	8	1	0.2	0.5	0.1
2	300	−2	4	1.5	0.15	−0.5	0.2
3	400	2	12	0.5	0.1	1	0.2

j is a cluster number. $\mu_{\theta j}$ is the azimuth of the centroid of the jth cluster. $\mu_{\theta j} = \arctan(u_{yj}/u_{xj})$.

number of data points in each cluster. The fixed cyk-means most approximately estimates the all means and the cyk-means also approximately calculates the all means. These results show that the cyk-means and the fixed cyk-means sufficiently approximately estimate the all means.

In the next experiment, we examine the effectiveness of the proposed methods (the cyk-means and the fixed cyk-means with $\kappa = 15$, $\sigma_r = 0.1$, and $\sigma_z = 0.1$) compared to the k-means and the kernel k-means with a radial basis function. The parameter of the radial basis function is $\gamma = 0.1$. The synthetic data have K clusters and are defined in cylindrical coordinates. The number of data points in each cluster is 200. The mean azimuth of the jth cluster $\arctan(u_y/u_x)$ is a random number in $[0, 2\pi\}$. The concentrate parameter κ_j is a random number in $[5, 30\}$. The mean radius of the jth cluster μ_{rj} is a random number in $[1, 4\}$. The mean height of the jth cluster μ_{zj} is a random number in $[-1.5, 1.5\}$. The standard deviations of σ_{rj} and σ_{zj} are random numbers in $[0.05, 0.2\}$.

Figure 2 shows the relationship between the number of clusters and adjusted rand index (ARI). ARI evaluates the performance of clustering algorithms [24]. When ARI = 1, all data points belong to true clusters. The figure shows that the cyk-means has the largest ARI for almost all cases. The fixed cyk-means performs better than the kernel k-means and the k-means. The k-means performs the worst. In conclusion, the cyk-means most accurately partitions synthetic data defined in cylindrical coordinates, and the fixed cyk-means also performs well.

Table 2: Parameters estimated by the cyk-means, the fixed cyk-means, and the k-means.

Method	j	N_j	$\mu_{\theta j}$	κ_j	μ_{rj}	σ_{rj}	μ_{zj}	σ_{zj}
cyk-means	1	**200**	-1.960×10^{-3}	**7.523**	0.9746	**0.1942**	0.5065	**0.09417**
	2	**300**	−1.997	**3.979**	1.495	**0.1492**	−0.5070	**0.1866**
	3	**400**	1.986	**12.69**	0.4966	**0.1004**	1.005	**0.1997**
Fixed cyk-means	1	200.2	$\mathbf{-5.124 \times 10^{-5}}$	—	**1.001**	—	**0.4995**	—
	2	299.9	**−2.000**	—	**1.501**	—	**−0.5056**	—
	3	400	**1.996**	—	**0.4998**	—	**0.9995**	—
k-means	1	184.85	−0.6627	—	1.091	—	0.1871	—
	2	253.65	−1.974	—	1.358	—	−0.4974	—
	3	461.5	1.704	—	0.4406	—	0.9477	—

The results are the mean of twenty runs on randomly generated initial values. j is a cluster number. $\mu_{\theta j}$ is the azimuth of the centroid of the jth cluster. $\mu_{\theta j} = \arctan(u_{yj}/u_{xj})$. The best estimations are bold.

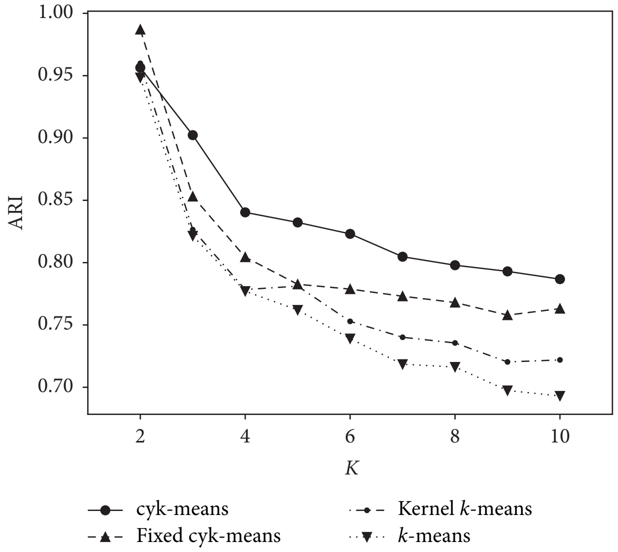


Figure 2: Relationship between the number of clusters K and adjusted rand index (ARI). The vertical and the horizontal lines indicate ARI and the number of clusters K, respectively. The results are the mean of 200 runs on randomly generated synthetic data for each K.

4.2. Real World Data. We show the performances of the proposed methods for the iris dataset (http://mlearn.ics.uci.edu/databases/iris/) and the segmentation benchmark dataset (http://www.ntu.edu.sg/home/asjfcai/Benchmark_Website/benchmark_index.html) [25]. The iris dataset has 150 data points of three classes of irises. The data point consists of the four attributes, sepal length in cm, sepal width in cm, petal length in cm, and petal width in cm. The segmentation benchmark dataset consists of 100 images from the Berkeley segmentation database [26] and ground-truths generated by manual labeling.

Table 3 depicts the ARI scores of the cyk-means, the fixed cyk-means, the k-means, and the kernel k-means for the iris dataset. The parameters of the fixed cyk-means are $\kappa = 15$, $\sigma_r = 0.1$, and $\sigma_z = 0.1$. γ of the radial basis function of the kernel k-means is 0.01. In this experiment, we use only three attributes of the iris dataset because the proposed methods are specialized for 3-dimensional data. Furthermore, we transform this dataset that has three attributes into zero mean dataset. In all cases, the performance of the cyk-means is lower than the other methods. Conversely, in almost all cases, the performance of the fixed cyk-means is the best. However, the difference in the performance between the fixed cyk-means, the k-means, and the kernel k-means is not large.

Table 4 shows the ARI scores of the cyk-means, the fixed cyk-means, and the k-mean for seven images in the segmentation benchmark dataset. The parameters of the fixed cyk-means are $\kappa = 15$, $\sigma_r = 0.1$, and $\sigma_z = 0.1$. To evaluate the performances of the cyk-means and the fixed cyk-means, we convert images from RGB color to HSV color. When we cluster the dataset by the k-means, we use images represented by RGB color and HSV color. In this experiment, we compare a clustering result with a ground truth using the ARI score. We set the number of clusters K to the number of segments in a ground truth. In all cases, the fixed cyk-means stably shows good performance. The cyk-means indicates much better or worse performances than the other methods. In other words, the cyk-means shows unstable performance. This instability will be caused by the cyk-means more easily trapping a local minimum because of more parameters.

4.3. Application to Color Image Quantization. We apply the cyk-means and the fixed cyk-means to color image quantization and compare the results to those using the k-means. We convert images quantized by the proposed methods from RGB color space to HSV color space before quantization, whereas an image processed by the k-means is represented using RGB. Figure 3 contains the four test images from the Berkeley segmentation database [26] and their quantization results. The original color images have sizes of 481 × 321 or 321 × 481 and are used as the test images to quantize into three colors. These quantization results are generated by the cyk-means, the fixed cyk-means with $\kappa = 25$, $\sigma_r = 0.1$, and $\sigma_z = 0.1$, and the k-means. The color of a pixel in the quantized image represents the value of the centroid of the cluster that contains the pixel.

Table 3: Comparison of performances of the four methods using the iris dataset.

Attributes	cyk-means	Fixed cyk-means	k-means	Kernel k-means
Sepal length, sepal width, petal length	0.5543	**0.6961**	0.6569	0.6614
Sepal width, petal length, petal width	0.5627	**0.7900**	0.7462	0.7590
Petal length, petal width, sepal length	0.4179	0.7114	0.7302	**0.7359**
Petal width, sepal length, sepal width	0.5171	**0.6121**	0.6104	0.6099

ARI are the mean of 200 runs with random initial values. The best estimations are bold. "Attributes" indicates three attributes used in clustering.

Table 4: Comparison of performances of the four methods using the segmentation benchmark dataset.

Number	cyk-means	Fixed cyk-means	k-means (HSV)	k-means (RGB)
12003 ($K = 4$)	**0.6397**	0.3744	0.3695	0.2277
69020 ($K = 2$)	0.1611	**0.5247**	0.07680	0.1406
80099 ($K = 2$)	**0.9205**	0.7960	0.6544	0.04890
103029 ($K = 3$)	0.4697	**0.5934**	0.1794	0.09143
106047 ($K = 2$)	**0.3210**	0.2988	0.04155	0.02332
310007 ($K = 7$)	0.1548	0.3338	0.3075	**0.3396**
353013 ($K = 4$)	**0.6125**	0.5993	0.5020	0.4683

ARI are the mean of 200 runs with random initial values. The best estimations are bold. "Number" indicates the number of the image. "k-means (HSV)" and "k-means (RGB)" indicate that we cluster HSV and RGB color images by the k-means, respectively.

For image 118035 in Figure 3, the colors of the background, the wall, and the roof are obviously different from each other. The cyk-means and the fixed cyk-means successfully segment this image, whereas the k-means extracts the shade from the wall and can not merge the wall to one color. Furthermore, the quantization results using the cyk-means and the fixed cyk-means are very similar.

Image 26098 consists of red and green peppers on a display table. The cyk-means merges the red peppers and the planks of the display table and divides the dark area into two colors. The fixed cyk-means successfully extracts the red peppers. The k-means assigns red to the planks and part of the green peppers.

Image 299091 consists of some sky with cloud, an ocher pyramid, and ocher ground. The cyk-means groups the ocher pyramid and white cloud into the same color, whereas the fixed cyk-means correctly segments the pyramid and the sky. The k-means is unsuccessful; it divides the pyramid into three regions (an ocher region, a highlight region, and a shade region).

The cyk-means did not perform well for image 295087. It segments the image into two colors even though we set the number of clusters to three. Thus, the cyk-means makes a dead unit. This is because the concentrate parameter and variances, respectively, become small and large if a distribution of data points is regarded to visually consist of a few clusters. Thus, a few clusters include all data points and dead units (empty clusters) appear, even if we fix the number of clusters to a large number. In contrast, the fixed cyk-means (which has fixed concentrate and variance values) appropriately partitions the ground and the blue and the deep blue regions of the sky. The k-means extracts shaded regions from the ground; that is, it can not group the ground into one region.

Furthermore, the initial parameters, κ and σs, of the fixed cyk-means can control the quantization results. Figure 4 shows the quantization results generated by the fixed cyk-means using the different parameters. The original image in Figure 4 consists of two objects: the red fish and the arms of an anemone. The fixed cyk-means with $\kappa = 25$, $\sigma_r = 0.1$, and $\sigma_z = 0.1$ can not extract the red fish shown in the middle image of Figure 4. However, the fixed cyk-means with $\kappa = 50$, $\sigma_r = 0.5$, and $\sigma_z = 0.5$ extracts the red fish in the left image of Figure 4. This is because a large κ and/or large variances relatively increase the cosine similarity term of (12), and consequently clustering is more focused on the hue element.

In conclusion, the fixed cyk-means is a more suitable method for color image quantization than the cyk-means. The fixed cyk-means quantizes color images with respect to the hue. The quantization results of the fixed cyk-means differ from that generated by k-means. That is because the Euclidean metric cannot consider the hue.

5. Conclusion and Discussion

In this study, we develop the cyk-means and the fixed cyk-means methods, which are new clustering methods for data in cylindrical coordinates. We derive a new similarity for the cyk-means from a probability distribution that is the product of a von Mises distribution and two Gaussian distributions (see (4)), because the Euclidean distance cannot properly represent dissimilarities between data points on periodic axes. Our experiments demonstrate that the cyk-means and the fixed cyk-means can properly partition synthetic data in cylindrical coordinates. Furthermore, the experimental results using real world data show that the fixed cyk-means has equal or better performance than the k-means and

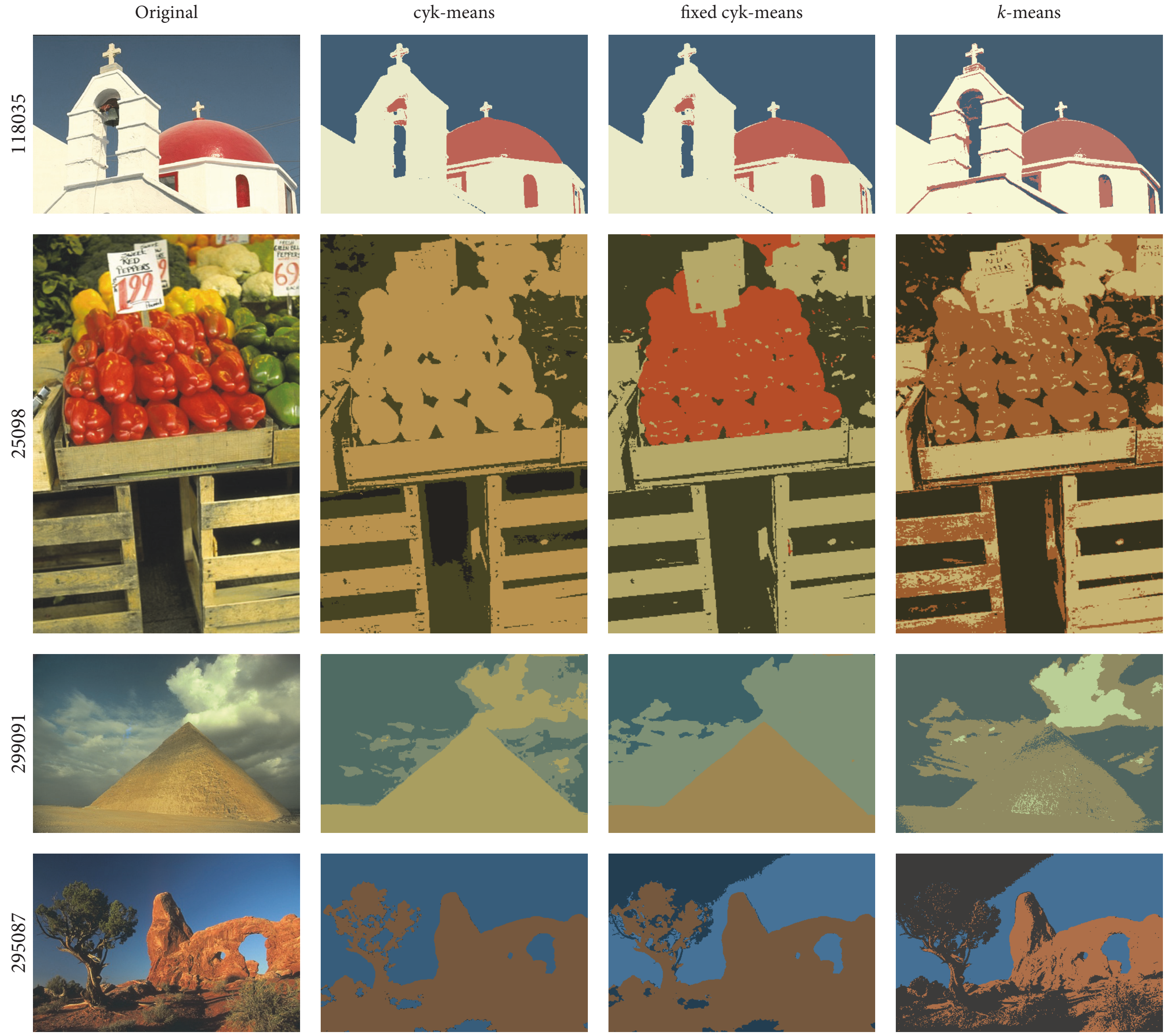


Figure 3: Quantization results. The first column contains the original images. The second, third, and fourth columns contain the quantization results generated by the cyk-means, the fixed cyk-means, and the *k*-means, respectively. All original images are clustered with $K = 3$.

the kernel *k*-means. In the final experiment, the proposed methods are applied to color image quantization and successfully quantize a color image with respect to the hue element.

The experiments that partitioned synthetic data demonstrate the effectiveness of the cyk-means. In the first experimental results, the cyk-means produces good estimates of the parameters and clustering data. The results of the second experiment show that the cyk-means performs the best when clustering synthetic data. However, in the experiment using real world data we find that the cyk-means did not provide good clustering results. Furthermore, the results of the color image quantization suggest that the flexibility of the cyk-means often produces dead units or a small cluster containing few data points. Thus, the cyk-means may not be appropriate for actual applications.

The fixed cyk-means will be an effective method for actual applications. The fixed cyk-means is stable and performs well when we apply it to clusterings of synthetic data, real world data, and color image quantization. Furthermore, the fixed cyk-means hardly makes dead units because the number of its parameters is smaller than the cyk-means. The fixed cyk-means requires less computational time than the cyk-means with similar results.

In future work, we will improve the performance of the proposed methods. The proposed methods are exposed to the ill-initialization problem and/or the dead unit problem caused by an incorrect initialization, similar to *k*-means.

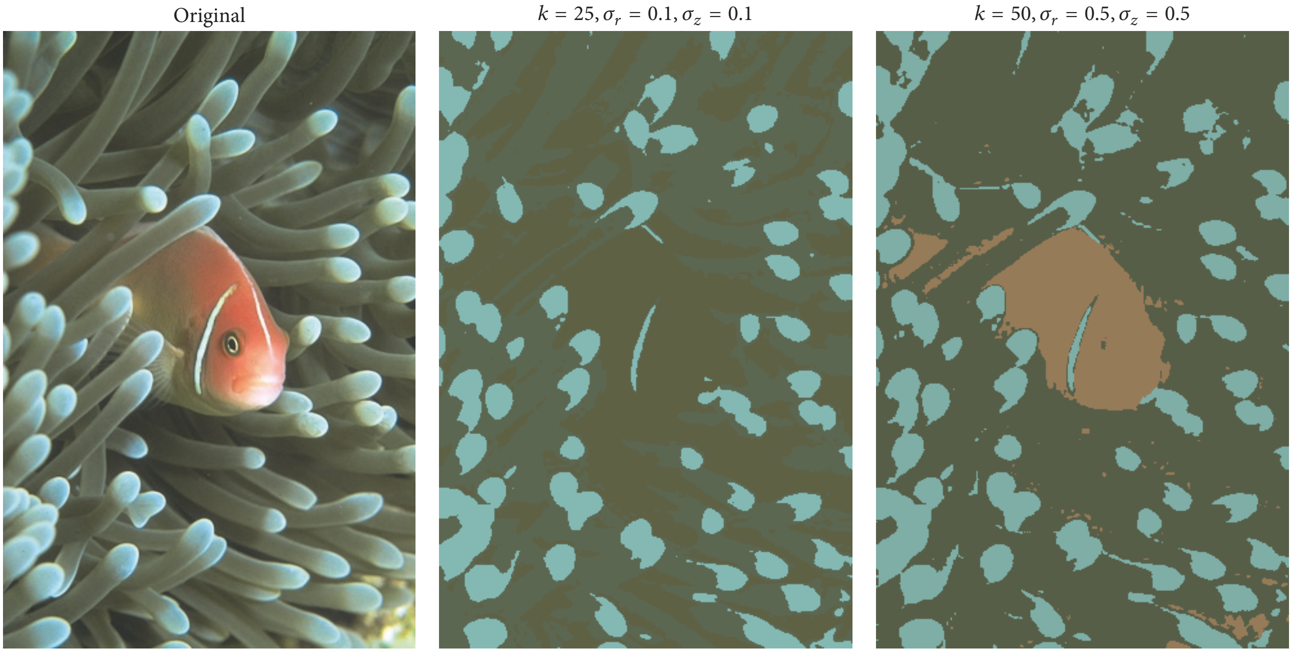

Original
$k = 25, \sigma_r = 0.1, \sigma_z = 0.1$
$k = 50, \sigma_r = 0.5, \sigma_z = 0.5$

FIGURE 4: Quantization results of the fixed cyk-means with different parameters. The left image is the original. The middle and the right images are quantized with $K = 3$.

The k-means++ method proposed by Athur and Vassilvitskii [27] solves the ill-initialization problem of k-means and improves the clustering performance by obtaining an initial set of cluster centers that is close to the optimal solution. The conscience mechanism improves the performance of competitive learning and clustering algorithms [28–30]. It inserts a bias into the competition process so that each unit can win the competition with equal probability. Xu et al. [31] proposed an algorithm based on competitive learning called rival penalized competitive learning [2, 32], which determines the appropriate number of clusters and solves the dead unit problem. The strategy of rival penalized competitive learning is to adapt the weights of the winning unit to the input and to unlearn the weights of the 2nd winner. By incorporating the approaches in these algorithms into the proposed methods, we will improve the performance and reduce the effect of the intrinsic problems.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The author would like to thank Toya Teramoto of the University of Electro-Communications for testing their algorithm.

References

[1] C. M. Anderson-Cook and B. S. Otieno, “Cylindrical data,” *Ecological Statistics*, 2013.

[2] N.-Y. An and C.-M. Pun, “Color image segmentation using adaptive color quantization and multiresolution texture characterization,” *Signal, Image and Video Processing*, vol. 8, no. 5, pp. 943–954, 2014.

[3] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, University of California Press, Berkeley, Calif, USA, 1967.

[4] X. Wu, V. Kumar, J. R. Quinlan et al., “Top 10 algorithms in data mining,” *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008.

[5] I. S. Dhillon and D. S. Modha, “Concept decompositions for large sparse text data using clustering,” *Machine Learning*, vol. 42, no. 1-2, pp. 143–175, 2001.

[6] A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, “Generative model-based clustering of directional data,” in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’03*, pp. 19–28, New York, NY, USA, August 2003.

[7] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, “Clustering on the unit hypersphere using von Mises-Fisher distributions,” *Journal of Machine Learning Research*, vol. 6, pp. 1345–1382, 2005.

[8] M. Girolami, “Mercer kernel-based clustering in feature space,” *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 780–784, 2002.

[9] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: analysis and an algorithm,” in *Advances in Neural Information Processing Systems*, pp. 849–856, MIT Press, Cambridge, Mass, USA, 2001.

[10] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, “Support vector clustering,” *Journal of Machine Learning Research*, vol. 2, pp. 125–137, 2002.

[11] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “Training algorithm for optimal margin classifiers,” in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT ’92)*, pp. 144–152, New York, NY, USA, July 1992.

[12] Y. Deng and B. S. Manjunath, “Unsupervised segmentation of color-texture regions in images and video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800–810, 2001.

[13] C.-K. Yang and W.-H. Tsai, “Color image compression using quantization, thresholding, and edge detection techniques all based on the moment-preserving principle,” *Pattern Recognition Letters*, vol. 19, no. 2, pp. 205–215, 1998.

[14] N.-C. Yang, W.-H. Chang, C.-M. Kuo, and T.-H. Li, “A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval,” *Journal of Visual Communication and Image Representation*, vol. 19, no. 2, pp. 92–105, 2008.

[15] P. Heckbert, “Color image quantization for frame buffer display,” *Computer Graphics*, vol. 16, no. 3, pp. 297–307, 1982.

[16] M. Emre Celebi, “Improving the performance of k-means for color quantization,” *Image and Vision Computing*, vol. 29, no. 4, pp. 260–271, 2011.

[17] D. Özdemir and L. Akarun, “A fuzzy algorithm for color quantization of images,” *Pattern Recognition*, vol. 35, no. 8, pp. 1785–1791, 2002.

[18] X. Zhao, Y. Li, and Q. Zhao, “Mahalanobis distance based on fuzzy clustering algorithm for image segmentation,” *Digital Signal Processing: A Review Journal*, vol. 43, pp. 8–16, 2015.

[19] A. Atsalakis, N. Papamarkos, N. Kroupis, D. Soudris, and A. Thanailakis, “Colour quantisation technique based on image decomposition and its embedded system implementation,” in *Proceedings of the Vision, Image and Signal Processing, IEE Proceedings*, vol. 151, pp. 511–524.

[20] C.-H. Chang, P. Xu, R. Xiao, and T. Srikanthan, “New adaptive color quantization method based on self-organizing maps,” *IEEE Transactions on Neural Networks*, vol. 16, no. 1, pp. 237–249, 2005.

[21] E. J. Palomo and E. Domínguez, “Hierarchical color quantization based on self-organization,” *Journal of Mathematical Imaging and Vision*, vol. 49, no. 1, pp. 1–19, 2014.

[22] M. G. Omran, A. P. Engelbrecht, and A. Salman, “A color image quantization algorithm based on particle swarm optimization,” *Informatica*, vol. 29, no. 3, pp. 261–269, 2005.

[23] S. Sra, “A short note on parameter approximation for von Mises-Fisher distributions: and a fast implementation of $I_s(x)$,” *Computational Statistics*, vol. 27, no. 1, pp. 177–190, 2012.

[24] K. Y. Yeung and W. L. Ruzzo, “Details of the adjusted rand index and clustering algorithms, supplement to the paper “Principal component analysis for clustering gene expression data”,” *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.

[25] H. Li, J. Cai, T. N. A. Nguyen, and J. Zheng, “A benchmark for semantic image segmentation,” in *Proceedings of the 2013 IEEE International Conference on Multimedia and Expo, ICME 2013*, July 2013.

[26] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *Proceedings of the 8th International Conference on Computer Vision*, pp. 416–423, July 2001.

[27] D. Arthur and S. Vassilvitskii, “k-means++: the advantages of careful seeding,” in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’07*, Society for Industrial and Applied Mathematics, pp. 1027–1035, Philadelphia, PA, USA, 2007.

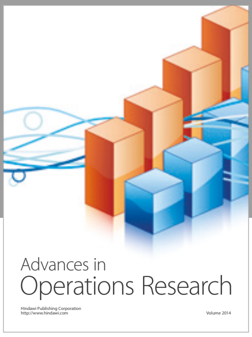
[28] D. DeSieno, “Adding a conscience to competitive learning,” in *Proceedings of 1993 IEEE International Conference on Neural Networks (ICNN ’93)*, pp. 117–124 vol.1, San Diego, CA, USA, March 1993.

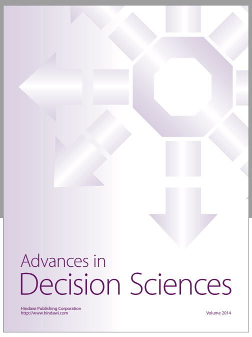
[29] C.-D. Wang, J.-H. Lai, and J.-Y. Zhur, “A conscience on-line learning approach for kernel-based clustering,” in *Proceedings of the 10th IEEE International Conference on Data Mining, ICDM 2010*, pp. 531–540, December 2010.

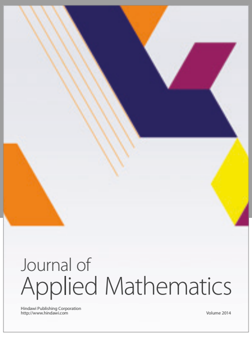
[30] C.-D. Wang, J.-H. Lai, and J.-Y. Zhu, “Conscience online learning: An efficient approach for robust kernel-based clustering,” *Knowledge and Information Systems*, vol. 31, no. 1, pp. 79–104, 2012.

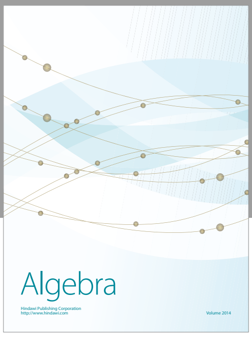
[31] L. Xu, A. Krzyżak, and E. Oja, “Rival penalized competitive learning for clustering analysis, rbf net, and curve detection,” *IEEE Transactions on Neural Networks*, vol. 4, no. 4, pp. 636–649, 1993.

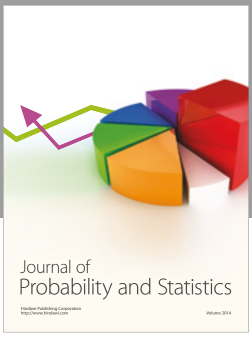
[32] T. M. Nair, C. L. Zheng, J. L. Fink, R. O. Stuart, and M. Gribskov, “Rival penalized competitive learning (RPCL): a topology-determining algorithm for analyzing gene expression data,” *Computational Biology and Chemistry*, vol. 27, no. 6, pp. 565–574, 2003.


Advances in
Operations Research

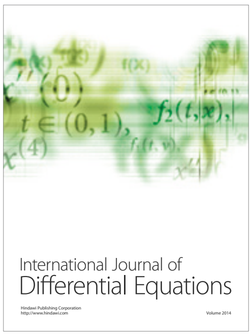

Advances in
Decision Sciences

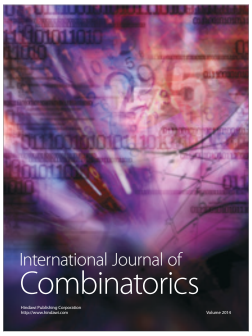

Journal of
Applied Mathematics

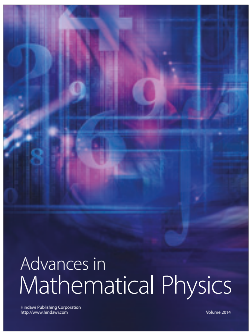

Algebra

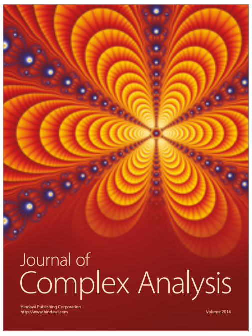

Journal of
Probability and Statistics

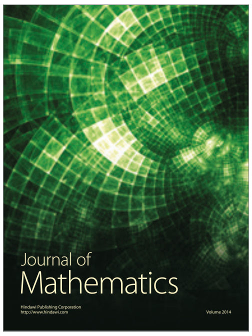
The Scientific
World Journal

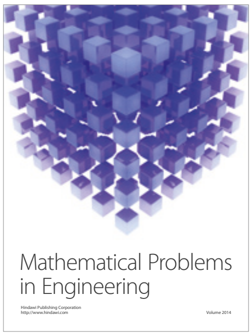

International Journal of
Differential Equations

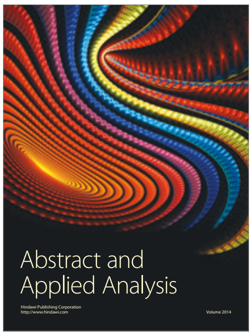

International Journal of
Combinatorics

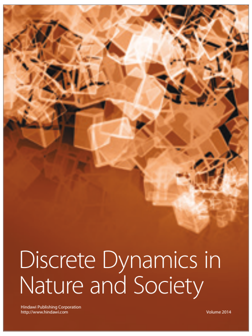

Advances in
Mathematical Physics

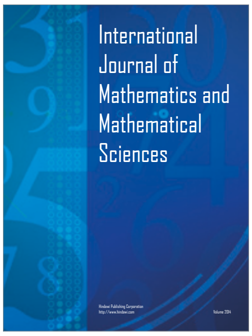

Journal of
Complex Analysis


Journal of
Mathematics


Mathematical Problems
in Engineering

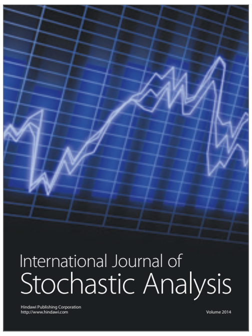

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


International Journal of
Stochastic Analysis

Journal of
Optimization