

ランダムイズドクラスタリングアルゴリズムに関する 実験結果について

稲葉真理*, 今井浩*, 加藤直樹†

*東京大学理学系研究科

†神戸商科大学

アブストラクト: クラスタリング問題とは与えられたデータをある評価基準に従いグループに分割する問題で、統計、コンピュータグラフィックス、データベースなど様々な分野に応用されている。与えられるデータの間には何らかの類似度が定義されており、評価基準としては、たとえば距離二乗和最小化といった類似度の高いデータが同じクラスターに分類されるものが使われる。類似度としては大別すると、各データ間の類似度がそれぞれ与えられるものと、各データの類似空間 (similarity space) での座標が与えられる幾何的なものに分けられる。後者は問題の持つ幾何構造から導出される制約により、より解きやすい性質を持っていると考えられる。本稿では d 次元空間上に与えられた n 点 $p_i = (x_i) (i = 1, \dots, n)$ からなる集合 S の距離二乗和の和最小を評価基準とする幾何的クラスタリング問題についてランダムイズドアルゴリズム (稲葉, 加藤, 今井 [3]) を実装し、以下のような計算機実験を行ない結果および考察を述べる。(1) 必要なサンプル数の評価。(2) 既存のアルゴリズム (Wan らのアルゴリズム [9]) との比較。(3) k -means 局所改良法 [5] の初期値としての性能比較。(4) 3次元空間への応用である色量子化問題への適用。

Experimental Results of Randomized Clustering Algorithm

Mary Inaba*, Hiroshi Imai* and Naoki Katoht

* Faculty of Science, University of Tokyo, Tokyo 113, Japan

†Department of Management Science, Kobe University of Commerce,
Kobe 651-21, Japan

Abstract: This paper describes computational results for k -clustering algorithm using random sampling technique [3]. By computational experiments, first, we show that small size of samples are actually enough for 2-clustering problem. Then, we apply this algorithm to k -clustering problem in a recursive manner and use the output as an initial solution of the existing local improvement technique, called k -means[5]. We compare the result with another variance-based algorithm [9] which is commonly used. By the experimental result, we see that, when k is small, the randomized algorithm performs better on the average and more robust than variance-based algorithm. As k increases and k -means iterations are performed, the difference among these two methods becomes smaller. We also apply our algorithm to color quantization problem.

1 はじめに

k -クラスタリング問題とは与えられた n 個のデータをある評価基準に従い k 個のグループに分割する問題で、統計、コンピュータグラフィックス、データベースなど様々な分野に応用されている。与えられるデータの間には類似度 (similarity) が定義され、評価基準としてはたとえば距離二乗和最小化といった類似度の高いデータを同じクラスタに分類するものが使われる。類似度の与え方としては、大別すると全データ対間の近似度をそれぞれ与えるもの、すなわち $n \times n$ の大きさの類似度行列 (similarity matrix) あるいは非類似度行列 (dissimilarity matrix) を与えるものと、各データについて類似空間 (similarity space) の座標が与えられる幾何的なものに分類される。後者から前者への変換は容易であり、また、後者は問題の持つ幾何構造から導出される制約により、より解きやすい性質を持っていると考えられ、これを幾何的クラスタリング問題と呼ぶ。より厳密に定義すると幾何的 k -クラスタリング問題とは d -次元空間 R^d に与えられた n 点 $p_i = (x_i)$ ($i = 1, \dots, n$) の集合 S をある評価基準の元で互いに素な k 個の空でない部分集合 S_1, \dots, S_k へと分割する“よい” k -クラスタリングを求める問題である。本稿では距離二乗和最小を評価基準とする幾何的 k -クラスタリング問題について考える。

クラスタリングの目的の一つとして、対象となるデータ集合のサイズが大きすぎるために個々のデータをそれぞれ取り扱うことが不可能な際に、データ集合をクラスタに分割しクラスタ内のデータは同じものと見做すことにより取り扱いを容易にすることがあげられる。すなわち各クラスタ S_j に対しそれぞれ代表点 r_j を定めクラスタ S_j 内の点をすべて r_j と同一視する。この時、 r_j とクラスタ内の点 $p_i \in S_j$ の距離二乗和を S_j のコストを $Cost(S_j) = \sum_{p_i \in S_j} \|x_i - r_j\|^2$ と定義する。集合 S のある k -クラスタリングのコストは $\sum_{j=1}^k Cost(S_j)$ 、最適な k -クラスタリング問題は

$$\min \left\{ \sum_{j=1}^k Cost(S_j) \mid S \text{ の } k\text{-クラスタリング } (S_1, \dots, S_k) \right\}$$

と定義する。集合 S_j の分散は $Var(S_j) = \frac{1}{|S_j|} \sum_{p_i \in S_j} \|x_i - \bar{x}(S_j)\|^2$ 。ここで $\bar{x}(S_j)$ は S_j の重心すなわち $\bar{x}(S_j) = \frac{1}{|S_j|} \sum_{p_i \in S_j} x_i$ とする。任意の r について $\frac{1}{|S_j|} \sum_{p_i \in S_j} \|x_i - r\|^2 \geq Var(S_j)$ なのでクラスタの代表点を R^d より任意に選べる場合、重心 $\bar{x}(S_j)$ を S_j の代表点に採るのが最適であり以下のことが示されている。

定理 1 (*E. Boros and P.L. Hammer [1]*) 距離二乗和最小を評価基準とする k -クラスタリング問題の最適解は、クラスタ内の各点の重心を母点とする ボロノイ図による n 点の分割として特徴づけられる。

d -次元空間上の k 点 $q_j = (r_j)$ に対し点 q_j のボロノイ領域 $Vor(q_j)$ は

$$Vor(q_j) = \bigcap_{l=1}^k \{p = (x) \mid \|x - r_j\| \leq \|x - r_l\|\}$$

と定義される。このボロノイ図によって、 n 点の集合 S は k 個のクラスタに分割することができ、これをボロノイ分割と呼ぶ。 k 個の母点によるボロノイ図による n 点のボロノイ分割は $O(n^{kd})$ 個しかないので ([3] 参照)、すべての可能なボロノイ分割を数えあげそのコストを求め、そのうちのコスト最小となるものを探すことにより k -クラスタリング問題の最適解を求めることができる。これは、 k, d を定数と見れば多項式時間とも言えるが、例えば色量子化におけるカラーマップテーブルの計算のように 10 万点 3 次元 256-分割 ($n = 100,000, d = 3, k = 256$) というような計算には役に立たないためランダムサンプリングを用いたアルゴリズム RC (稲葉, 加藤, 今井 [3]) が提案された。本稿では d -次元空間 n 点 $p_i = (x_i)$ ($i = 1, \dots, n$) からなる集合 S について以下の実験考察を行なった。

- 2-クラスタリング問題の 2RC アルゴリズムを実装し、実際に、サンプル数 m が小さくても良い解が得られ、また n がある程度大きい時必要とされるサンプル数は n によらないことを示した。
- 再帰的に 2-クラスタリングを適用する k -クラスタリングアルゴリズム RC を実装し同じく再帰的な手法を採り色量子化問題で利用される既存のアルゴリズム Wan のアルゴリズム VB [9] との比較を行ないより良い解ができることを示した。

3. 上で得られる k -クラスタリングを k -means 局所改良法 [5] の初期値として使用して局所最適解の比較および局所最適解へと収束するまでの反復回数の比較を行なった。
4. 3-次元空間への応用である色量子化問題に上のアルゴリズムを適用し画像の減色を行ない得られる画像の比較を行ない、また k -means 法での反復回数による変化を観察した。

なおアルゴリズムは C 言語により実装され SPARC station 20 で実行された。

2 2-クラスタリングアルゴリズム

2.1 アルゴリズム 2RC

n, k ともに大きいクラスタリング問題を解く際利用されるアルゴリズムは、大別してトップダウンに2分割を再帰的に繰り返すものとボトムアップに類似度の高いデータをグループ化していくものとに分類される。ここではまずランダムサンプリングを使用した近似2-クラスタリングアルゴリズム 2RCを示す。

アルゴリズム 2RC

INPUT: n 点の集合 S

OUTPUT: S の 2-クラスタリング

1. 復元抽出で n 点の集合 S より m 個の部分集合 T をランダムにサンプルする。
2. T のすべての線形分割可能な 2-クラスタリング (T_1, T_2) に関し以下の計算を行なう。
 - (a) T_1 と T_2 の重心 t_1, t_2 を求める。
 - (b) 点 t_1 と 点 t_2 を結ぶ線分を垂直二等分面によって分割される 2-クラスタリング (S_{t_1}, S_{t_2}) を求めそれぞれの重心 s_1 と s_2 を求める。
 - (c) 点 s_1 と 点 s_2 を結ぶ線分の垂直二等分面によって分割される 2-クラスタリング (S_{s_1}, S_{s_2}) を求める。
 - (d) $Cost(S_{s_1}) + Cost(S_{s_2})$ を計算し、これが最小となるものを保持する。
3. 上で得られた最小のものを出力する。

アルゴリズム 2RC については局所改良を行なっているステップ (2.c) を省略したアルゴリズムについて以下のことが証明されている [3]。

定理 2 S の最適な 2-クラスタリング (S_1^*, S_2^*) が定数 γ ($0 < \gamma \leq 1/2$) について $|S_1^*|, |S_2^*| \geq \gamma|S| = n$ である時、2RC はサンプル数 m に関して $O(m^d n)$ 時間で高い確率で最適解のコストに対し $1 + O(1/m)$ 倍以内のコストで納まるクラスタリングを出力する。□

たとえばコストが最適解の $1 + 1/3 = 4/3$ 倍以内に納まる 2-クラスタリングを、2RC は平面上の n 点の中から 10 点 サンプリングすることにより $O(10^2 n) = O(n)$ 時間で $\sum_{i=3}^7 \binom{10}{i} / 2^{10} \approx 0.89$ の確率で出力することができる。

2.2 必要なサンプル数に関する実験評価

アルゴリズム 2RC において必要な計算時間は $O(m^d n)$ なので実際にサンプル数 m が本当に小さくても 2RC が良い解をだすことが必須である。ここでは、アルゴリズム 2RC においてサンプル数の大きさと得られる解の質の間の相関を調べる。

図 1 は $n = 2^{10}$ 個の点を

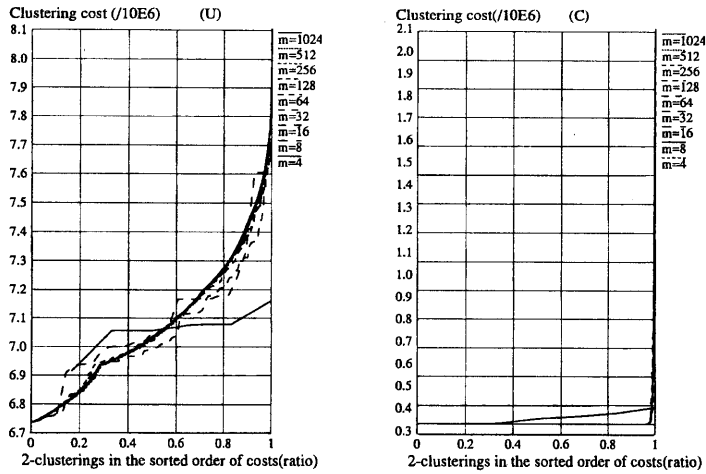


図 1: 線形分離可能な 2-クラスタリングのコスト分布

表 1: サンプル数について 2^{10} 個に対する $2^i, i = 2, \dots, 9$ の誤差の百分率

$ T $	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9
2^{10} (U)	2.6397%	0.1198%	0.0457%	0.0127%	0.0015%	0.0015%	0.0012%	0.0000%
(N)	0.9164%	0.2375%	0.0373%	0.0121%	0.0057%	0.0033%	0.0011%	0.0000%
(C)	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
2^{14} (U)	0.5809%	0.3527%	0.1442%	0.0019%	0.0019%	0.0002%	0.0000%	0.0000%
(N)	0.4721%	0.0546%	0.0099%	0.0008%	0.0006%	0.0000%	0.0000%	0.0000%
(C)	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
(M)	0.0862%	0.0527%	0.0082%	0.0040%	0.0000%	0.0000%	0.0000%	0.0000%

1. $[0, 255]^2$ の正方形¹ に一様分布させたもの (左)

2. $[0, 255]^2$ の範囲にランダムに選んだ二点を中心とした同心円内の一様分布 (右)

について、 $m = 2^i, (i = 2, \dots, 10)$ 個のサンプル数についてアルゴリズム 2RC の (2.d) おいて比較される重心 s_1, s_2 の垂直二等分面によって線形分離された 2 つの集合 S_{s_1}, S_{s_2} に関する $Cost(S_{s_1}) + Cost(S_{s_2})$ を計算しこれを整列させ縦軸にコストを横軸を分割総数により正規化した順序を取っている。サンプル数の多寡にかかわらずグラフはほぼ同じ形状を示しており最小値すなわち横軸=0 に近づくにつれ、グラフの傾きが緩やかになり横軸=0 においては理論的に得られている期待率 $O(1+1/m)$ よりも良い値がでている。

次に、 $Cost(S_{s_1}) + Cost(S_{s_2})$ の最小値についてサンプル数による得られる値の比較を示す。 $[0, 255]^2$ の正方形内にそれぞれ 2^{10} および 2^{14} 点を (U) 一様分布させたもの、(N) 正規分布させたもの、(C) ランダムに選んだ二点を中心とした同心円内に一様分布させたもの、および現実のデータとして、(M) 2,0726 点からなる関東地方の道路の交差点位置のデータについて、それぞれ $m = 2^i, (i = 2, \dots, 9)$ 個のサンプルを取った場合に得られる分割に関する最小コストを $m = 2^{10}$ 個のサンプルの最小コストと比較し、表 2.2 にその差を百分率で示している。 2^{10} 個のデータに関してはサンプル数を 2^{10} とした場合、復元抽出であるので完璧ではないがほぼ最適解が求められていると考えられる。

いずれの場合も与えられたデータの大きさに関係なくサンプル数 たかだか $m = 8$ で誤差は 1% 以下に納まり、また $m = 2^4 = 16$ 点で誤差は 0.2% に納まっており、理論的な期待値 $1/m$ すなわち $1/8 = 12.5\%$

¹ 実装にあたり色量子化への応用を考え RGB それぞれ 8 ビットすなわち一辺 256 の正方形格子上でクラスタリングの計算を行っている。

/16 = 6.25% より格段に良い値となっている。これにはいくつかの理由が考えられるが一つにはステップ (2.c) の局所改良が効いていると考えられる。

3 k -クラスタリングアルゴリズム

2-クラスタリングをトップダウンに再帰的に繰り返す k -クラスタリングアルゴリズム RC を示す。なお実装にあたっては現在の最大コストのクラスタを探すためにヒープを使用している。

アルゴリズム RC

INPUT: n 点の集合 S

OUTPUT: S の k -クラスタリング

$S_1 = S, Cost(S_1) =$ 集合 S のコスト

for $i = 2$ to k

$Cost(S_j)$ が最大なクラスタ S_j を探す。

S_j を $2RC$ により 2-クラスタリングし これを S_j, S_i とする。

同じくトップダウンに再帰的に 2 クラスタリングを繰り返すアルゴリズムとして Wan らのアルゴリズムが VB がよく利用される。これは各軸に鉛直な分割面を動かすことによって dn 個の候補から最適なものを選ぶアルゴリズムである。

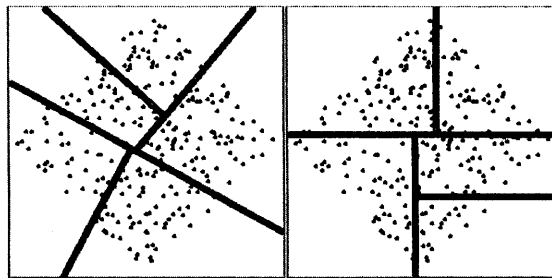


図 2: 5-クラスタリング

RC と VB の差が顕著に現れる例として図 2 に RC (左) と VB (右) による $[0, 255]^2$ の正方形を半分の大きさにして 45 度回転し 256 点の一樣分布させたデータの 5-分割の例を示す。得られた分割のコスト $\sum_i Cost(S_i)$ は 4-分割で RC は VB の 79.6%、5-分割で RC は VB の 83.7% となる。同時にこれは 2-クラスタリングを再帰的に繰り返すこととの限界も示している。 VB との比較に関しては次節の表 4 の初期値として示す。

4 k -means 局所改良法との関連

ここでは、 k -クラスタリングを与えられたときこれを局所最適解へと改良する k -means アルゴリズムを紹介する。 k -means アルゴリズムは初期値として k -クラスタを入力し各クラスタごとに重心を計算し重心を母点とするボロノイ分割によりクラスタを再構成することを収束するまで繰り返すアルゴリズムで、統計を初めとする様々な分野で非常によく使われている (たとえば [5] 参照)。

k -means アルゴリズム

INPUT: n 点の k -クラスタリング

OUTPUT: 局所最適な n 点の k -クラスタリング以下を収束するまで繰り返し、結果を出力する。

REPEAT

1. 各クラスタ S_j の重心 $\bar{x}(S_j)$ を計算する。
2. k 個の重心 $\bar{x}(S_j)$ ($j = 1, \dots, k$) を母点とする ボロノイ分割を計算し、これを S_1, \dots, S_k とする。

UNTIL 収束するまで

図 3,4 に k -means アルゴリズムの反復によるボロノイ分割の変化の例を示す。

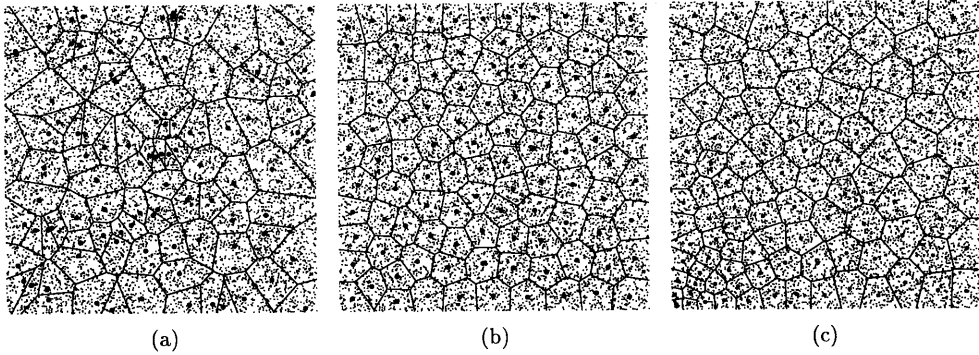


図 3: 正方形内一様分布している 10000 点に対して $k = 100$ で k -means アルゴリズムを適用する過程 (a から c へ)

k -means 法ではコストが単調減少するのでこのアルゴリズムはボロノイ分割の総数ですなわち $O(n^{kd})$ 以下の反復回数で収束することがわかる。この種類の局所改良を収束するまで続けるアルゴリズムでは与えられた初期値が重要な意味を持つことになる。初期値が悪いと収束に時間がかかり、また、出力されるクラスタの質も低下する。

ここでは k -means アルゴリズムの初期値としてランダムに k 点を選びこれを母点とするボロノイ分割による k -クラスタリング、アルゴリズム RC による k -クラスタリングアルゴリズム VB による k -クラスタリングをとった場合の比較考察を行なう。 RC に関してはランダムサンプリングを行なうため 10 回試行を行ない、初期値すなわち RC の出力と局所最適すなわち k -means アルゴリズムの出力について

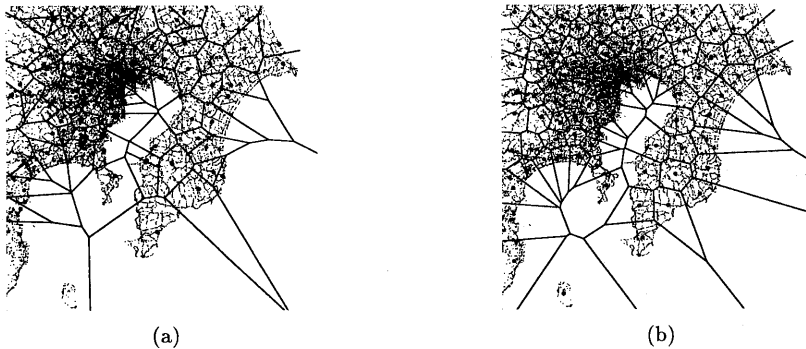


図 4: 20726 点からなる関東道路データに対して $k = 100$ で k -means アルゴリズムを適用した結果: (a) 初期解, (b) 収束解

表 2: k -means アルゴリズムの計算結果

		V の値	平均値	最小値	最大値	標準偏差
$k = 12$	初期値	1.65E7	98.51%	96.30%	100.70%	1.34%
	局所最適	1.52E7	99.05%	97.45%	100.20%	0.85%
	反復回数	10	18.2	12	31	6.09
正規分布	初期値	1.10E7	95.00%	94.60%	95.33%	0.22%
	局所最適	1.03E7	99.58%	99.50%	99.70%	0.06%
	反復回数	33	9.2	7	11	1.32
12 個の円	初期値	4.17E6	96.67%	96.50%	96.77%	0.08%
	局所最適	3.97E6	99.94%	99.90%	99.95%	0.01%
	反復回数	13	8.5	8	9	0.5
道路地図	初期値	4.96E6	93.84%	92.69%	95.59%	0.86%
	局所最適	4.85E6	90.93%	89.87%	93.39%	1.46%
	反復回数	7	9.3	8	13	1.41
$k = 25$	初期値	7.96E6	97.11%	95.60%	98.71%	0.83%
	局所最適	7.22E6	99.19%	98.19%	100.92%	0.83%
	反復回数	51	27.7	19.0	45	7.68
正規分布	初期値	5.61E6	103.7%	100.7%	106.1%	1.70%
	局所最適	5.20E6	100.5%	98.65%	102.1%	0.98%
	反復回数	28	28.8	14	42	9.29
25 個の円	初期値	3.35E6	97.61%	97.01%	98.81%	0.45%
	局所最適	3.14E6	99.68%	98.7%	100.9%	0.64%
	反復回数	20	17.8	14	25	3.76
$k = 50$	初期値	3.78E6	97.47%	97.35%	101.85%	1.31%
	局所最適	3.55E6	100.0%	98.59%	101.4%	0.90%
	反復回数	18	19.1	12	27	4.76
正規分布	初期値	2.80E6	98.92%	97.85%	99.99%	0.54%
	局所最適	2.65E6	99.8%	99.2%	100.3%	0.40%
	反復回数	20	17.1	10	26	4.81
50 個の円	初期値	2.30E6	100.4%	99.1%	102.6%	0.99%
	局所最適	2.21E6	99.1%	98.6%	101.3%	0.71%
	反復回数	10	16.3	8	23	4.34
$k = 100$	初期値	1.93E6	98.05%	96.66%	98.92%	0.70%
	局所最適	1.77E6	98.86%	97.76%	100.08%	0.67%
	反復回数	16	25.0	16	32	5.56
正規分布	初期値	1.46E6	97.55%	96.04%	99.00%	0.94%
	局所最適	1.36E6	97.89%	96.70%	99.27%	0.69%
	反復回数	19	19.1	11	32	5.62
100 個の円	初期値	1.27E6	97.57%	96.52%	99.78%	0.99%
	局所最適	1.18E6	98.78%	97.92%	99.63%	0.66%
	反復回数	17	15.4	17	25	4.47
道路地図	初期値	4.86E5	97.56%	95.77%	99.15%	0.96%
	局所最適	4.67E5	98.62%	97.45%	99.70%	0.72%
	反復回数	7	7.8	5	19	3.99

平均値、最小値、最大値、および標準偏差を計算する。クラスタリングされるデータは前章のデータと同じであり、サンプル集合の大きさ $m = 16$ とし、 $k = 12, 25, 50, 100$ に関して試行を行ない、アルゴリズム V との結果との比較を行なった。

k -クラスタリングについてアルゴリズム RC はアルゴリズム V より平均して良い成績を示しているが k -means アルゴリズムを行なうことにより両者の差は減少してゆく傾向にある。また k が大きくなるにつれ両者の差は減少してゆく傾向にある。これは、トップダウンに再帰的に 2 分割を繰り返していることによって起こる最適解とのずれが k が増加するにつれ大きくなり V かと RC との差が局所改良により吸収されてしまうためであると考えられる。

5 色量子化への応用

上のアルゴリズムを色量子化問題に適応した。

行なったのは、(A)1856 色の 238×260 画素の画像データ、(B)1876 色 824×643 画素の画像データを $k = 2, 4, 8, 16, 32, 64, 128, 240$ 色へと減色を行なった結果、以下のことが観察された。

1. k -means アルゴリズムと組み合わせることにより画質が向上する。 k が大きい時は人間の肌の色のように輝度の変化が不連続でその変化が誇張されて見えるマッハバンド効果 (Mach band effect、たとえば [7] 参照) の解消に有効である。 k が小さい時には色数の少なさによる形状の不鮮明さが改善される。
2. オーダード法あるいは誤差拡散法のようにディザリングにおいて特定のパターンが画像に現れることがない(たとえば [8] 参照)。
3. k -means の反復回数については特に最初の 3 回程度が目に見えて有効である。
4. 特に $k \leq 32$ 以下の減色に関してはアルゴリズム RC と k -means アルゴリズムの組み合わせは効果が大きい。
5. サンプル数を $m = 8$ で行なったがアルゴリズム RC は遅い。

6 まとめ

アルゴリズム $2RC$ においてサンプル集合の大きさは小さくても充分であり、また得られた解の質は与えられたデータのサイズにはよらず、また理論的に得られる期待値よりも格段に良い性能を持つことが示された。また 2 -クラスタリングについてアルゴリズム RC はアルゴリズム VB より平均して良い成績を示している。 k が大きくなり k -means アルゴリズムで反復が行なわれるうちに両者の差は減少してゆく傾向にある。アルゴリズム RC はより小さい k に対して有効であり、またトップダウンに再帰的に 2 分割を繰り返す限り理論的に誤差を押えることはできないため、たとえば k 個の代表点の候補をサンプル集合から直接選び再帰的なトップダウン 2 分割を行なわない方式を考えていくことが課題である。またアルゴリズム RC は $d = 100$ といった高次元空間を取り扱えないため高次元空間におけるクラスタリングを行なうためのアルゴリズムを開発することも課題である。また k -means アルゴリズムについて局所最適となる分割の数を N とするとランダムに選んだ初期値から始めた場合の反復回数期待値は $O(n^{kd}/N)$ で押えることができるので N の評価を行ない、 k -means 反復回数について考察していきたい。

謝辞

本研究の一部は文部省科学研究費の援助を受けて行なったものである。

参考文献

- [1] E. Boros and P. L. Hammer: On clustering problems with connected optima in Euclidean spaces. *Discrete Applied Mathematics*, Vol.75 (1989), pp.81-88.
- [2] S. Hasegawa, H. Imai, M. Inaba, and J. Nakano: Efficient algorithms for variance-based k -clustering. *Proceedings of the First Pacific Conference on Computer Graphics and Applications*, World Scientific, (1993), pp.75-89.
- [3] M. Inaba, H. Imai and N. Katoh: Application of weighted Voronoi diagrams and randomization to variance-based k -clustering. *Proceedings of the 10th ACM Symposium on Computational Geometry* (1994), pp.332-339.
- [4] M. Inaba, H. Imai and N. Katoh: Experimental results of randomized clustering algorithm. *Proceedings of the 10th ACM Symposium on Computational Geometry* (1996), pp.C1-C2.
- [5] S. Z. Selim, M. A. Ismail: K -means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Trans. Pattern Anal. Mach. Intell.*, (1984), pp.81-87.
- [6] L. Kaufman and P. J. Rousseeuw: *Finding groups in data*. Wiley Interscience, (1990).
- [7] N. M. Thalmann, D. Thalmann and H. T. Minh: Interphong shading. *Graphics Gems II*(J. Arvo, ed.), Academic Press, Boston, (1991), pp.232-241
- [8] S. W. Thomas, R. G. Bogart: Color dithering. *Graphics Gems II*(J. Arvo, ed.), Academic Press, Boston, (1991), pp.72-77
- [9] S. J. Wan, S. K. M. Wong and P. Prusinkiewicz: An algorithm for multidimensional data clustering. *ACM Trans. on Mathematical Software*, Vol.14, No.2 (1988), pp. 153-162.