

レーザ・レーダによるエアロゾル分布パターンの分類と判別
The classification and discriminant of aerosol distribution
patterns by Laser Radar

村上 吉繁 山元 正人
Y. Murakami M. Yamamoto

大阪大学工学部
Faculty of Engineering, Osaka University

[1] はじめに

レーザ・レーダによるエアロゾル観測を行なうことにより、局地的大気状況を3次元的にとらえる。レーザ・レーダにより得られたパターンを数個の代表的パターンに分類し、その分類されたパターンと気象状況等とを対応付け、大気汚染予測を行なうことが研究目的である。

そこで、そのアプローチとして、筆者らは R.H.I. (Range-Height-Indication) 方式で測定されたパターンから、明度推移行列を用いて特徴量抽出を行ない、主成分に圧縮した後、クラスター分析を用いてパターンの分類を行なった。また、新たに得られたパターンがどの代表的パターンに属するか決定するために判別分析を行なった。

[2] テクスチャ解析による特徴量抽出

Haralickらは、テクスチャの特徴を表わすための14種類の濃淡の推移を表わす計算式を提案した。ここでは、そのうちの4種類を適用している。その計算手順は次の通りである。

- i) 前処理として、直流成分の除去、距離補正の処理を行なった後、直角座標に変換する。その後パターンを 64×64 のブロックに分割し、32レベルに量子化する。これにより、ディジタル配列による2次元画像を得る。
- ii) このパターンから4方向 (0° および 180° 、 90° および 270° 、 45° および 225° 、 135° および 315°) において、隣接するブロック間の結合密度、すなわち明度推移行列 $P(i, j)$ を求める。 (i, j) 要素は、隣接するブロックの一方が i 、他方が j の明度をもつ場合の結合確率で、 $P(i, j)$ は、 32×32 の正方行列となる。
- iii) 4方向の明度推移行列から、次の4種類の特徴量を計算する。

$$\begin{aligned}
 f_1 \text{ Angular Secondary Moment (ASM)} & \sum_i \sum_j \{P(i, j)\}^2 \\
 f_2 \text{ Contrast (CONT)} & \sum_{n=1}^{N_g-1} n^2 \left\{ \sum_i \sum_j P(i, j) \right\}, |i-j|=n \\
 f_3 \text{ Correlation (CORR)} & \left\{ \sum_i \sum_j P(i, j) - \mu_x \mu_y \right\} / \sigma_x \sigma_y \\
 f_4 \text{ Entropy (ENT)} & - \sum_i \sum_j P(i, j) \log \{P(i, j)\}
 \end{aligned} \tag{1}$$

ただし、 μ_x, μ_y は周辺分布の平均、 σ_x, σ_y は周辺分布の標準偏差である。

- iv) パターンから、次式を用いてその重心を求める。

$$\begin{aligned}
 G_x &= \sum_j j L(i, j) / \sum_i \sum_j L(i, j) \\
 G_y &= \sum_i i L(i, j) / \sum_i \sum_j L(i, j)
 \end{aligned} \tag{2}$$

(ただし、 $L(i, j)$ は2次元画像の濃度)

[3] 主成分分析

主成分分析とは、もとのP個の変量の一次式で表わされ、主成分の係数 $\{l_{ki}\}$ はm個の主成分が互いに無相関で、かつもとのP個の変量の情報をできるだけ多く集めるように定める方法である。その計算手順は次の通りである。

- i) 変量をすべてのパターンに関して、平均0、分散1に標準化する。
- ii) その標準化された特徴量の相関行列から固有値を求め、大きい順に並べかえ、その固有値に対する固有ベクトルを主成分の係数 $\{l_{ki}\}$ として、次式に代入する。

$$Z_k = \sum_{i=1}^P l_{ki} x_i' \quad \text{----- (3)}$$

ただし、 Z_k は、第k主成分、 x_i' は標準化された特徴量である。

Table. 1は主成分番号、固有値およびその主成分までの累積寄与率を示している。なお、特徴量としては、各特徴量の4方向の平均およびパターンの重心の合計6個を用いている。この表から、第3主成分までで92.0%の累積寄与率があり、ほとんどP個の変量すなわちこの場合6個の変量の特徴を表わしている。

主成分 番号	固有値	累積寄与率
1	3.855	0.643
2	1.093	0.825
3	0.572	0.920
4	0.420	0.990
5	0.049	0.999
6	0.009	1.000

Table. 1 主成分と累積寄与率

[4] クラスター分析

多変量解析の一つに似たものの集団を求める数値的分類にクラスター分析がある。ここでは、類似度として、次式のマハラノビス距離によるモード法を適用した。

$$d_{\alpha\beta} = \sum_{i=1}^m \sum_{j=1}^m (Z_{\alpha i} - Z_{\beta i}) W^{ij} (Z_{\alpha j} - Z_{\beta j}) \quad \text{----- (4)}$$

ただし、 W^{ij} は、行列 (W_{ij}) の逆行列の i, j 要素で、 W_{ij} は

$$W_{ij} = \frac{1}{n-1} \sum_{\alpha=1}^n (Z_{\alpha i} - \bar{Z}_i)(Z_{\alpha j} - \bar{Z}_j) \quad (\text{ただし } \bar{Z} \text{ は平均})$$

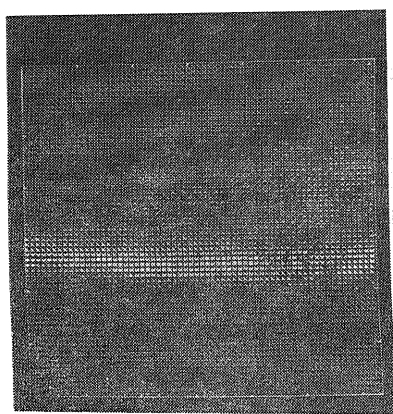
である。

モード法では、密度に関係する第k近接点のkと半径を表わすDとを2つのパラメータとして、kを固定して、Dを変化される方法を適用した。

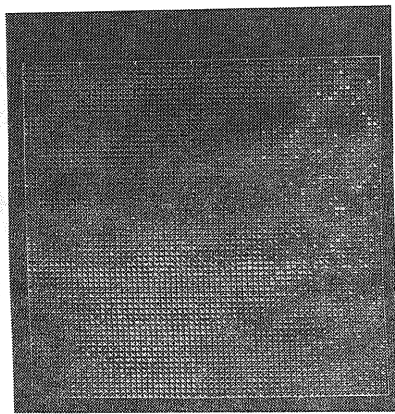
群番号	パターン番号
1	13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
2	31, 33, 34, 35, 36, 37, 38, 39, 40
3	24, 25, 26, 27, 28
4	1, 2, 3, 4, 41, 42, 43, 44
5	5, 6, 7, 8, 9, 11, 12

Table. 2 クラスター分析による分類

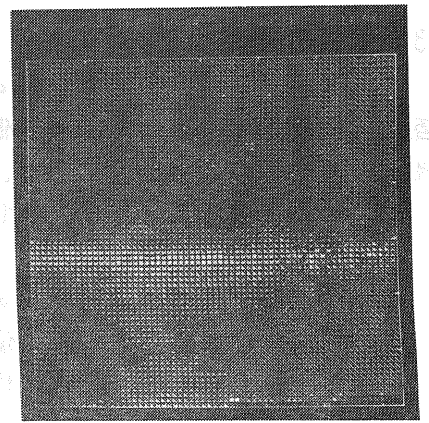
Table. 2は、44個のパターンについて、第3主成分までを用い、 $k=3$ として、90%まで分類された状態で止めたものである。なお、分類されなかった残り10%のパターンは、No. 10, 29, 30, 32である。視覚により、5種類に分類した場合と比較して、一致している。Fig. 1 (a)~(e)はその代表的パターンを示している。



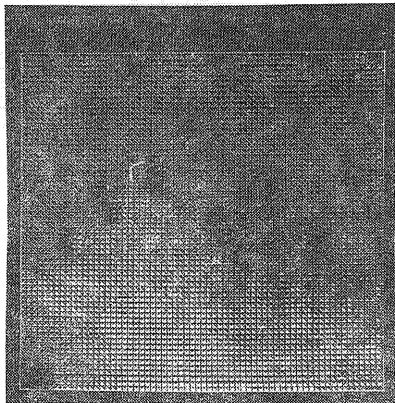
(a) 78/10/31 4:17:29



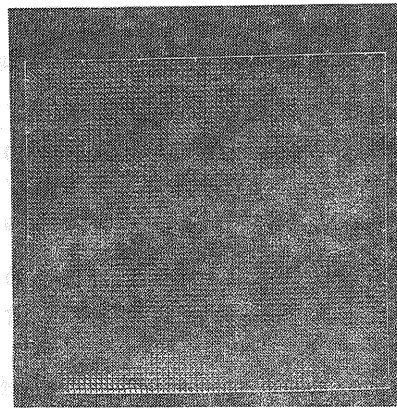
(b) 78/10/31 8:06:02



(c) 78/10/31 6:14:54



(d) 78/10/30 15:44:39



(e) 78/10/30 17:10:27

Fig. 1 RHI Patterns at KASUMIGAURA by N.I.E.S.

[5] 判別分析

判別分析とは、2つ以上の群から取り出された多変量データに基づいて、所属不明の新しいサンプルをそのいずれかの群に判別しようとする手法である。

クラスター分析で分類された各群では、分散・共分散および出現割合が異なっているため、新しい判定値 x は事後確率が最大となるような群に割り付けるという最通判別方式を適用した。すなわち、 k 群の判別得点を

$$S_k = \pi_k P_k(x) \quad \text{----- (5)}$$

として、 S_k を最大にすることである。ただし、 π_k は出現割合、 P_k は確率密度である。そこで、 x の分布がどの群でも P 変量正規分布に従うと仮定すれば、その確率密度は、

$$P_k(x) = (2\pi)^{-p/2} |\Sigma_k|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu_k)' \Sigma_k^{-1} (x - \mu_k) \right\} \quad \text{----- (6)}$$

と書ける。ただし、 μ_k は、母平均ベクトル、 Σ_k は、母分散・共分散行列である。この(6)式を(5)式に代入し、その自然対数を取り、どの k にも共通な項 $(2\pi)^{-p/2}$ を削除すれば、 k 群に対する(5)式と等価な判別得点として、

$$S_k = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (x - \mu_k)' \Sigma_k^{-1} (x - \mu_k) + \log \pi_k \quad \text{----- (7)}$$

を得る。すなわち、この値が最大となる群kに対象xが割り付けられる。

群番号	判別得点
1	-122.8
2	-24.0
3	-1033.5
4	-188.5
5	0.3

(a)

群番号	判別得点
1	-139.3
2	-13.0
3	-2655.6
4	-9.3
5	-57.9

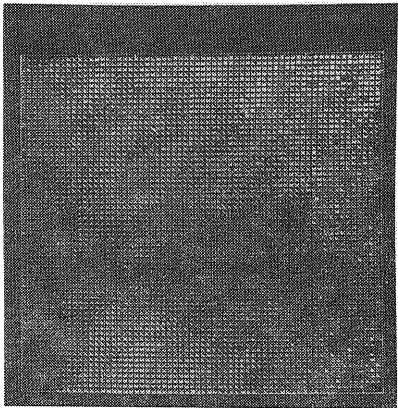
(b)

Table. 4 判別得点

Table. 4 (a)は、クラスター分析で分類されなかったパターン No.10 について判別分析を行なった結果である。この表から、オ5群の判別得点が最大となり、このパターンは、オ5群に割り付けられる。なお、測定値xとしては、オ3主成分までを用いている。

Fig. 2は Fig. 1のどの代表的パターンにも属さない未知のパターンで、Table. 4 (b)は、そのパターンについて判別分析を行なった結果である。この表から、オ4群の判別得点が最大となるが、Table. 4 (a)のオ5群の判別得点に比較して非常に小さいため、このパターンは新しい群を形成する。

Fig. 2. RHI Pattern at KASUMIGAURA (78/11/2 8:06:08)



[6] あとがき

レーザ・レーダにより観測されたエアロゾル分布パターンをHaralickらが提案した明度推移行列を用いたテクスチャ解析で特徴量抽出を行ない、主成分に圧縮した後、モード法によるクラスター分析を行なった結果、パターンの分類が可能となった。また、判別分析により新たに得られたパターンの判別が可能となった。今後、多くのパターンを蓄積し、気象状況等との対応付けからパターンの予測を行ないたい。

最後に、データを提供して載せました国立公害研究所に深く感謝致します。

参考文献

- 1) R.M.Haralick et al : Texture Features for Image Classification, IEEE Trans. Vol. SMC-3, No. 6 pp 610/620 (1973)
- 2) 奥野忠一他 : 多変量解析法, pp 159/321, 日科技連出版社 (1971)
- 3) 奥野忠一他 : 続多変量解析法, pp 77/146, 207/237, 日科技連出版社 (1976)