

階層的クラスタリングによる衛星画像圧縮に関する考察*

5G-01

古屋 亮 細村 幸 市野 学†

東京電機大学‡

1 はじめに

本研究では, Landsat 衛星に搭載されている TM (Thematic Mapper) により得られた 7 バンドのマルチバンド画像に対して, 階層的クラスタリングを用いたベクトル量子化手法により圧縮符号化を実現する.

クラスタリング手法として, サンプル間の距離を用いる事でサンプルの類似度を表す距離尺度が存在する. この手法は距離でしかサンプルの類似度を判断することができず, 形成されたクラスタがどのような特徴において類似しているか知ることができない. その為, クラスタ形成の根拠が分かりづらいと言える.

そこで本手法ではクラスタが形成される際, サンプルがどのような特徴に関して類似しているかを定義する尺度として, サンプルの近さを表す概念である「相対近隣」^[1]を用いる.

2 提案手法

2.1 サンプル含有数^[2]

ある特徴上で N 個のサンプル $P_k, (k=1,2,\dots,N)$ が存在すると仮定する. ここで, 任意の 2 つのサンプル P_i と P_j の k 番目の特徴 F_k に関して考える. 2 つのサンプルに挟まれた領域内に含まれる他のサンプルの数をサンプル含有数 β と定義する.

特徴 F_k 上のサンプル対 P_{ik}, P_{jk} によって形成される領域内に他のサンプルが存在しない場合, サンプル対 P_i, P_j は特徴 F_k について相対的な意味で性質が似ていると解釈することができる.

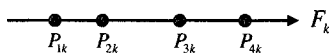
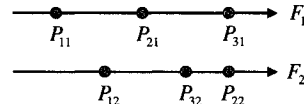


Fig 1. $\beta(P_{ik}, P_{jk})$ の例

Fig 1 において 2 つのサンプル P_{1k}, P_{2k} によって形成される領域には他のサンプルを含まない為 $\beta(P_{1k}, P_{2k})=0$ となる. つまり特徴 F_k においてサンプル対 P_{1k} と P_{2k} は他のサンプル対 (P_{3k}, P_{4k}) よりも特徴 F_k という性質が似ていると判断できる. また, 同様の事はサンプル対 P_{2k}, P_{3k} とサンプル対 P_{3k}, P_{4k} にも言える.

2.2 隣接行列

サンプル含有数 β を基に隣接行列 A_k を作成する. ここで A_k とは特徴 F_k におけるサンプル含有数が 0 となるサンプル間の隣接行列とする. Fig 2 に特徴 F_1, F_2 における隣接行列 A_k を示す.



$$A_1 = \begin{matrix} & P_{11} & P_{21} & P_{31} \\ \begin{matrix} P_{11} \\ P_{21} \\ P_{31} \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{matrix} \quad A_2 = \begin{matrix} & P_{12} & P_{22} & P_{32} \\ \begin{matrix} P_{12} \\ P_{22} \\ P_{32} \end{matrix} & \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{matrix}$$

(a) A_1 の隣接行列 (b) A_2 の隣接行列

Fig 2. 特徴 F_1, F_2 における隣接行列の例

サンプル P_2 と P_3 は特徴 F_1, F_2 において隣り合って存在しており, P_1 より互いに性質が似ていると言える. つまり, より多くの特徴に関して隣り合って存在するサンプル対は相対的に性質が似ていると解釈できる.

2.3 類似性の尺度

本提案法では最も類似性の高いサンプル対は最も多くの特徴で「近い存在」という概念を用いて類似性の尺度を「相対近隣」として定義する.

隣接行列 A_k よりサンプル対 P_i, P_j の類似性の尺度を (1) 式と定義する. d は全特徴数を表す.

$$S(i, j) = \sum_{k=1}^d (A_k)_{i,j} \quad (1)$$

$$0 \leq S(i, j) \leq d$$

2.4 提案手法のアルゴリズム

階層的クラスタリング手法のアルゴリズムを示す.

- (1) ベクトルとして d 次元ごとに画像データを区切り, それぞれをサンプル P_i とする.
- (2) ある特徴 (バンド) F_k における二つのサンプル対 P_{ik}, P_{jk} のサンプル含有数 β を求める.
- (3) すべての特徴におけるサンプル対 P_{ik}, P_{jk} がとりうるサンプル含有数 β を求める.
- (4) (3) を全てのサンプル対に対して行う.
- (5) サンプル含有数を基に隣接行列を作成し, 類似性の尺度を作成する.
- (6) 全てのサンプル対においてサンプル含有数 $\beta = 0$ を一番多く持つサンプル対に対応するサンプルを融和する.
- (7) 融和されたサンプルの平均をとり, 1 本のベクトルに置き換える. 得られたベクトルを新たなサンプルとして加える.
- (8) 最終的にサンプルが 1 本のベクトルに融和されるまで (2) ~ (7) の操作を繰り返す.

*A consideration about satellite image coding by hierarchical clustering

† Ryo Furuya, Tsukasa Hosomura, Manabu Ichino

‡ Tokyo Denki University

3 評価実験

3.1 対象とする衛星画像

実験で対象とする画像は Landsat /TM の 7 バンドからなるデータを利用した。この画像は金沢市とその周辺地域であり、1 画素あたり各 8bit、512×512 画素からなる。

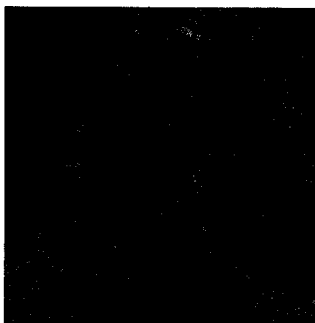


Fig 3 原画像 (全体)

3.2 実験方法

実験に使用する画像は、Fig 3 の原画像から選出した特性の異なる領域(a), (b)を使用する。各領域あたり 32×32 画素、7 バンドである。

提案手法は領域あたり 7 次元のベクトル、32×32 の 1024 サンプルとしてクラスタリングを行い、任意の融和階層数で代表ベクトルを選出した。

また、提案手法との比較に、従来手法^[3]である距離尺度を用いた非階層的クラスタリングによるベクトル量子化を用いる。

非階層的クラスタリングは原画像全体を 7 次元のベクトル 262144 本とし、256 本のベクトルで代表させる。得られた復号画像に対して、使用領域と同一領域を抜き出し比較対象とする。

ベクトル量子化後の画像と原画像との比較値には、画質の評価値として広く用いられている信号対雑音比 *PSNR* を用いる。

3.3 結果

原画像と復号画像の比較を Fig 4, Fig 5 に示す。対象とするバンドはクラスタリング結果の違いが顕著に表れているものを選別した。左からそれぞれ原画像、提案手法を用いた復号画像、従来手法を用いた復号画像である。それぞれ領域を 700[%]に拡大している。



Fig 4 領域(a)の原画像と復号画像の比較

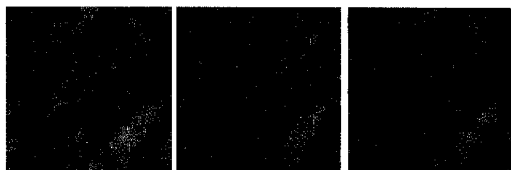


Fig 5 領域(b)の原画像と復号画像の比較

提案手法によるベクトル量子化は、領域(a)のように高周波成分を多く含み、境界がはっきりと目視できるような画像に対して高い復元結果を得ることが可能である。

逆に領域(b)のように低周波成分を多く含み、近隣関係にあるサンプルを多く含む画像では、融和回数を重ねるとサンプル同士は同一階調値で量子化されてしまう。そのため従来手法と視覚的に比較しても、原画像を再現できていないことが確認できる。

提案手法における代表ベクトル本数と信号対雑音比 *PSNR* の関係を Table 1 に、従来手法の各領域における *PSNR* を Table 2 に示す。

Table 1 階層的クラスタリングの結果

	代表ベクトル本数	<i>PSNR</i> [dB]
(a)牧草地	438	38.553512
(b)山腹	144	30.783111

Table 2 非階層的クラスタリングの結果

	<i>PSNR</i> [dB]
(a)牧草地	34.874241
(b)山腹	37.831963

提案手法を用いた領域(a)の *PSNR* は、従来手法と比較して高い数値を示している。しかし代表ベクトル本数は他の領域(b)と比較して、同サンプル数であるにも関わらず格段に多い。

これは選出した領域内に近隣関係にあるサンプル郡が非常に少ないため「相対近隣」を用いたクラスタリング処理の際、一度に融和されるベクトル本数が非常に少なくなり、結果として代表ベクトル本数が増加するためである。

4 おわりに

「相対近隣」を用いた階層的クラスタリングによるベクトル量子化方法を提案した。提案手法は高周波成分等を多く含む画像に対して高い復元結果を得ることが可能であり、有用な手法であるといえる。

提案手法は画像の性質に強く依存される方法である。既存の類似性尺度に他のサンプル含有数の概念をも考慮に入れた新たな類似性の尺度を導入することにより、近隣関係のサンプルを多く含んだ画像に対しても高い復元結果が得られると期待できる。

参考文献

- [1]市野学：“記述の一般性を考慮した近隣グラフの考察”東京電機大学内部メモ(2001)
- [2]中川清太郎：“シンボリックデータに対するクラスタリングの研究”東京電機大学大学院理工学研究科システム工学専攻修士論文(2000)
- [3]小出隆：“マルチバンド衛星画像に対する高効率圧縮符号化に関する研究”金沢工業大学大学院工学研究科情報工学専攻修士論文(1998)