

# 疎らな分布を持つ空間情報データの画像表現の 圧縮についての検討

作 小軍\*

北原 正樹\*

清水 淳\*

## 概要

ICTの普及に伴い、より高精細、よりリッチな膨大なデータが収集できるようになり、データの可視化をはじめ、様々な応用に期待が高まってきた。本稿では、画像表現を用いてこれらのデータを扱う手法を紹介するとともに、画像圧縮技術の効率的な適用手法を提案する。本手法では、データが空間的分布が疎らである特徴に着眼し、選択的に圧縮アルゴリズムを適用することで、高い圧縮率を実現することを示す。

To achieve quick querying or interactive visualization of large spatial data, using image representation as data container is considered as an efficient approach. Based on such approach, to improve data compressing ratio, we propose an algorithm by utilizing several image compression methods adaptively, according to the sparseness of the data. The effectiveness of the algorithm is shown by some experimental results.

## 1 空間情報データの画像表現

位置情報に関連付けされるデータは非常に価値が高いと言える。例えば、国勢調査[1]などは経済状況の把握に役立ち、IoTセンサによる広域モニタリングなどは防災・減災には欠かせない。我々は、このような位置情報に関連付けられるデータを空間情報データと呼び、画像化による高速検索・可視化の手法を提案した。基本的考え方を図1に示す。図中左側は人口値の立体縦棒グラフである。棒グラフの示す値を画像表現の画素値とすると、図中右側のような濃淡画像が得られる。画像表現にすると、画像圧縮・処理技術が容易に応用でき、特に膨大なデータセットに対し、非常に効率の高い処理が可能になる(文献[2])。

本稿のモチベーションは、この基本的考え方にに基づき、可視化用途を前提とした非可逆画像圧縮技術を応用することで、さらなるデータ量の削減を試みる。2節では、具体的な提案を述べる。3節では、評価実験の結果を示す。

## 2 非可逆圧縮における課題と提案

図1で示す画像表現と区別するため、カメラなどの撮影で得られる画像を自然画像と呼ぶ。JPEGをはじめ、自然画像の持つ空間的連続性を前提に、多くの圧縮アルゴリズムが考案されてきた。特に非可逆圧縮の場合、鑑賞目的の画質を維持しつつ、データサイズの大幅削減が実現できる。前述画像表現と自然画像を比較すると表1に示すような特性の違いがある。これらの違いによって、画像表現を非可逆圧縮すると、高周波成分が効

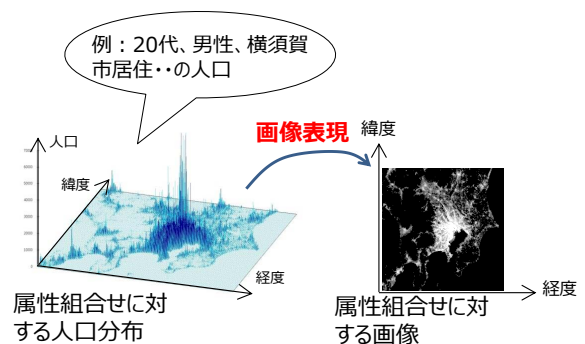


図1 画像表現:統計量の値を画素値に、位置情報を画像座標に、それぞれに対応付けると、統計データを画像表現にできる。

表1 画像表現と自然画像の特性比較

	自然画像	画像表現
画素間相関	高い	乏しい
値0の意味	輝度を表す量	統計値の有無

率的に圧縮できないため、符号化効率の低下が課題である。

課題解決のため、画素の再配置する手法を提案する。概要は図2に示す。再配置することで、元画像から自然画像に近い特性を持つ再配置画像と座標値画像が得られる。再配置画像は非可逆圧縮アルゴリズムにより、大幅なデータ削減が期待できる。デメリットとして、座標値画像というオーバーヘッドが発生する。

座標値画像は再配置画像と元画像の関連情報を保持しているため、表1により、統計値の無い、すなわち、元画像上の値

\* NTT メディアインテリジェンス研究所

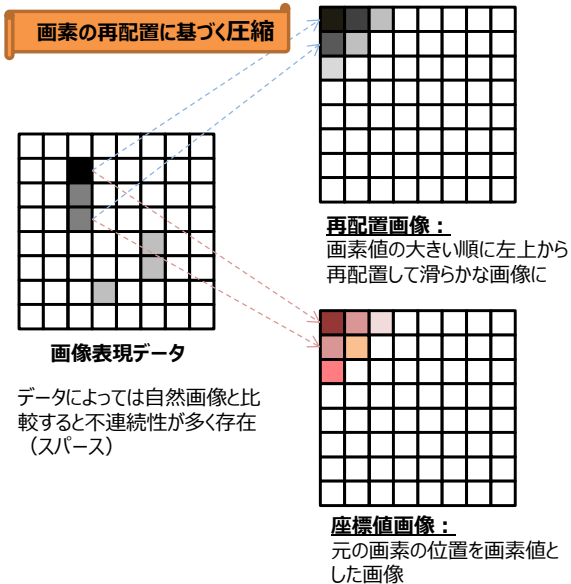


図2 画素の再配置: 疎らな画像表現の空間的連続性を高めるように画素の再配置を行う。

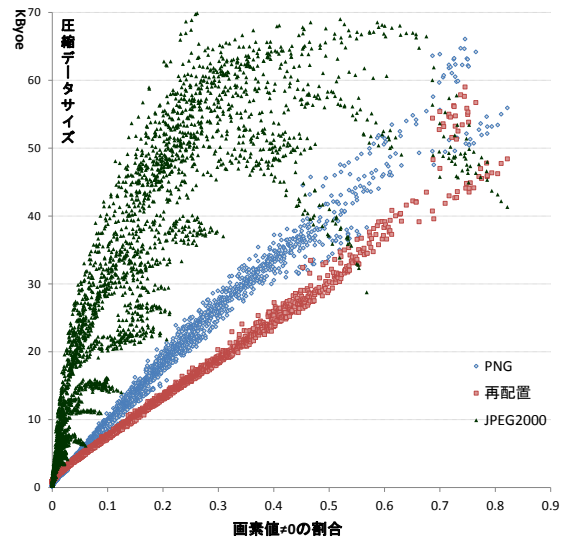


図3 評価実験: スパース度に応じてPNG, JPEG2000, 提案手法の圧縮データサイズの比較  
ただし, 横軸は $S$ の定義の補集合(値 $neq 0$ のロケーション)の割合で, 0が $S$ 最大で, 1が $S$ 最小を意味する

が0のロケーションについて, 元画像上の座標値を保持する必要はない. 従って, 上記オーバーヘッドは元画像上の値が0のロケーションの割合に依存すると言える. つまり, スパース度 $S =$  値が0のロケーションの割合と定義すれば, オーバーヘッドが $S$ に依存すると言える. 符号化効率を高める観点で, 再配置を行うか否かは,  $S$ に応じて判断すればよいと考えられる. 以下に,  $S$ に応じて適応的に画素再配置を行う提案圧縮手法のフローを示す. ただし,  $\theta$ はしきい値を表し, 経験的に得られるパラメータとする.

1. スパース度 $S$ を計算
2.  $\left\{ \begin{array}{l} \text{if } S \geq \theta, \text{ 画素再配置を行う} \\ \quad \left[ \begin{array}{l} \text{再配置画像を非可逆圧縮} \\ \text{座標値画像を可逆圧縮} \end{array} \right. \\ \text{if } S < \theta, \text{ 画素再配置を行わない} \\ \quad \left[ \begin{array}{l} \text{元画像を非可逆圧縮} \end{array} \right. \end{array} \right.$

### 3 評価実験

文献[1]の国勢調査の人口データを用いて実験を行った. 人口データから, 1次メッシュごとに年齢・性別属性の各組合せについて, それぞれ画像表現に変換し, 複数の疎らな画像を得た. これらの画像をPNG, JPEG2000(非可逆指定), 及び提案手法それぞれ用いて圧縮を行った. ただし, 画素再配置の効果を検証するため, しきい値 $\theta = 0$ とした. 図3に画像のスパース度に対し, それぞれの手法による圧縮データサイズのグラフを示す. 以下の傾向が見られる.

- 可逆圧縮(PNG)の場合, 横軸の値に比例してデータサイ

ズが単調増加する.

- 提案手法(再配置)の場合, PNGと同様, 横軸の値に比例してデータサイズが単調増加するが, 概ねPNGより圧縮率が高い. ただし, スパース度が小さい(横軸の値が1に近づく)場合, 圧縮率がPNGを下回るようになる.
- 非可逆圧縮(JPEG2000)の場合, スパース度が大きい場合, 前述課題のように他の圧縮方法よりデータサイズが大きく効率が悪い. ただし, スパース度が小さい(横軸の値が1に近づく)場合, 圧縮率がもっと大きくなる.

以上の実験結果から, 画素再配置の実施を切り替えることで, 圧縮率をを向上最大とすることができる. すなわち, 提案手法の有効性が確認できる.

### 4 まとめ

本稿では, 自然画像と異なり, 空間情報データの空間的不連続な特性を考慮し, 疎らの度合いに応じた適応型圧縮手法を提案した. 圧縮率の向上が確認できる一方, 今後の課題として, スパース度合いが小さくなる画像表現の方式の検討も欠かせないといえる.

### 参考文献

- [1] 統計局. URL: <http://www.e-stat.go.jp/SG1/estat/GL02100104.do?tocd=00200521>.
- [2] 件, 北原, 清水. 画像表現を用いた空間情報データの高速度検索と可視化手法. FIT2016, I-28.