

研究種目：基盤研究（C）

研究期間：2007～2009

課題番号：19500007

研究課題名（和文） マルチデータの無ひずみ圧縮とそのバリエーション

研究課題名（英文） Lossless Compression of Multi-Data Sets and its Variations

研究代表者

横尾 英俊（YOKOO HIDETOSHI）

群馬大学・大学院工学研究科・教授

研究者番号：70134153

研究成果の概要（和文）：日々生成される膨大な量のデジタルデータの管理の問題に対し、複数のデータを関連づけることによって解決する枠組みを設定し、その枠組みに属すいくつかの問題にデータ圧縮と情報埋め込みの手法を応用した。具体的には、データ圧縮と情報埋め込みの量的関連を明らかにし、同時に、情報埋め込みの効率的な手法の開発を行った。さらに、データの量のかかえる問題を解決する手段として、従来法の拡張に相当する新規なデータ圧縮法を提案した。

研究成果の概要（英文）：For the management of huge amount of data, we have developed a framework in which we address various issues by relating multiple data to each other. We have applied some techniques in data compression and data embedding to typical problems in the framework. We have extended a lossless data compression algorithm based on the BWT to yield a more general scheme, and also developed a syndrome coding algorithm as an efficient data embedding method with an emphasis on its application to gray-scale images.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,300,000	390,000	1,690,000
2008年度	1,100,000	330,000	1,430,000
2009年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：情報理論

科研費の分科・細目：情報学・情報学基礎

キーワード：情報数理，情報理論，データ圧縮，情報埋め込み，マルチデータ，情報源符号化

1. 研究開始当初の背景

(1) インターネット上の情報の爆発的増加に象徴されるように、日々生成されるデジタルデータは膨大な量にのぼる。これらは真に新規なデータである場合もあれば、既存のデータの改訂や発展など、相互に何らかの関連を有するデータの再生産である場合も少な

くない。むしろ、あらゆるデータは相互に深く関連しあっているとみるのがより適切である。その意味において、膨大な量のデータとは言え、その多くの部分は冗長であり、同時に、複数の冗長なデータの存在に伴う諸問題も顕著になっている。

(2) データの量的側面での問題を圧縮という技術によって解決しようとするデータ圧縮法は、情報技術の進歩に伴って飛躍的に発展し、データを個別に圧縮する手法だけでなく、分散情報源符号化など複数のデータを対象とする実用法の開発も進んでいる。また、分散情報源符号化の考えを情報の保護に応用しようとする研究も活発である。情報の保護にあたっては、保護すべきデータに別の情報を埋め込んで、認証や著作権保護に応用する手法が採られる。ただし、こうした情報埋め込みとデータ圧縮との直接の関連は十分に研究されるまでに至っていない。

2. 研究の目的

(1) データ圧縮とは、対象となるデジタルデータに対し、その見かけの大きさを縮小したデータを生成することである。このとき、圧縮が無ひずみであるとは、圧縮後のデータから元データが誤りなく完全に復元できることにほかならない。無ひずみデータ圧縮では、基本的には、二種類のデータ（圧縮前後のデータ）が主要な登場人物となる。これに対し、本研究ではこれを三種類以上に拡張したマルチデータの無ひずみ圧縮に着目した。そのようなマルチデータの無ひずみ圧縮に分類できる問題群に含まれる個々の問題については、従来から既に研究の対象となっているものが少なくない。本研究では、それらの一連の問題をひとつの問題群として統一的に扱うことによって、できるだけ見通しのよい解決を与えることを目的とした。

(2) 上記の問題群から特に基本的な問題として「データ圧縮とデータ埋め込み」との量的等価性の問題を選び、LZ 符号化と呼ばれるデータ圧縮法の枠組みでの等価性を示すことにした。また、データ埋め込みのためのシンδροーム符号化の手法を調査し、具体的な応用例の開発にも取り組んだ。さらに個別のデータの圧縮法の高性能化のために、無ひずみデータ圧縮の代表的手法であるブロックソート法の拡張を行って、その性能の解析を行うことを目的とした。

3. 研究の方法

(1) まず対象を抽象的なモデルによって表現することによって問題を定式化し、引き続き、モデルの解析、問題の解となるアルゴリズムの開発、さらに計算機によるアルゴリズムの実装を経て、提案法の理論的および実際の性能の評価を行った。

(2) データ圧縮とデータ埋め込みの量的等価性については、等価性の理論的証明を中核とし、成果の一部を国際会議において発表した。シンδροーム符号化によるデータ埋め込み

については、理論的な特徴づけに加え、より実際的な応用例を開発してその有用性を示した。また、ブロックソート法の拡張に際しては、新規な記号列変換法を提案し、その高速化をアルゴリズム論の視点で実践した。

4. 研究成果

(1) 情報保護や秘匿通信を目的とする情報埋め込みは、データ圧縮との直接の関連性を意識することなく研究されている。しかし、あるデータに別のデータを埋め込むという操作には、元のデータを圧縮して得られたすき間に別データを挿入するという原理が内在していると考えられる。本研究では、このような原理を浮き彫りにすることによって、データ埋め込みとデータ圧縮との量的な相補性について明らかにした。

データ圧縮の直接の目的は、通信の高速化、および記憶領域の有効利用である。後者の記憶領域の有効利用という観点からは、その目的を結果的に達成することさえできれば、利用しているデータ圧縮法の圧縮力の優劣にかかわらず、目的そのものは達成できていると言える。本研究では、データ圧縮の有効性を、他のデータが利用することのできる領域の確保という観点で評価した。具体的には、データ圧縮法の圧縮力をデータ埋め込みという手法によって補完する手法を提案し、その結果として確保できる領域を評価することで、純然たるデータ圧縮法との比較を行う。

データ圧縮法にとって、その冗長性はできるだけ排除するものというのが一般的な考え方である。しかし、データ圧縮法の冗長性は、除去が比較的容易なものがある一方、本質的な除去が困難な場合も少なくない。除去の困難なそのような冗長性に対しては、圧縮性能の向上のために冗長性を取り除こうとするだけでなく、冗長性の除去とは異なる視点で結果的に類似の効果が得られる手法を検討するという方向があってもよい。本研究では、無ひずみデータ圧縮法の代表的な手法である Lempel-Ziv 符号のうち、LZ77 と呼ばれる手法に着目した。LZ77 では、記号列の符号化対象部分から既に符号化済みの既出の記号列に対する最長一致を切出し、それを既出の記号列に対する参照によって符号化する。このとき、参照が一意的とは限らないという冗長性を生じることがある。このような冗長性は、除去が可能である限りデータ圧縮の観点からは除去もしくは削減すべきであるが、それが自明ではないため、参照多重性を逆に利用しようとする試みもなされている。

LZ77 の参照多重性の利用は、いくつかの側面からなされているが、ここでは、データ埋め込みという視点で定量的に評価した。データ埋め込み機能を有する無ひずみ圧縮器

に対し、圧縮によって確保できる空間の大きさと埋め込んだデータの長さの和を圧縮器が確保した領域と定義する。圧縮器が確保することのできる領域を定量的に評価することで、従来の純粋な圧縮力の評価にかわる新しい評価尺度を提案した。具体的には、LZ77の変種の例として LZSS と呼ばれる手法を選んで、圧縮に特化した手法に位置づける。一方、その対極にある手法として、圧縮をせずに、かわりに埋め込みだけを行う手法 (Size-Preserving LZ 法) を提案した。圧縮器が確保する領域長という点で、これら両極にある手法が漸近的に等しい性能を有することを示した上で、さらにこれらの中に位置する手法を構成した。すなわち、圧縮も埋め込みも同時に達成する手法である。提案法は、それに含まれるパラメータを調整することで、圧縮力と埋め込み力との混合比を変化させることができる。本研究では、このようなパラメータを適切に選ぶことによって、純然たる圧縮法や純然たる埋め込み法よりも、実際の性能に優れた方法が導出できることを実験によって示した。

以上の成果を国際会議において発表する (学会発表①) と同時に音声符号化への応用についても発表した (学会発表⑤)。

(2) 情報埋め込み, 別名, 情報ハイディングは、電子透かしとステガノグラフィとに大別される。ステガノグラフィは、興味の対象となっている主データの存在そのものを第三者に対し隠蔽するために、カバーデータと呼ばれるダミーのデータに主データを埋め込む技術である。その多くの手法は、検知されない程度にカバーデータを書き換えることによって実現される。「検知されない」ということの具体的な定義は、カバーデータの性質に依存するが、そのようなデータの種類に応じた手法ではなく、より抽象化した一般的なモデルも存在する。データをわずかに書き換えるということは、データ間に距離が定義されていることを意味する。ステガノグラフィとは、基本的には、書き換えの前後でデータの距離を一定範囲に制限するという条件のもとで、埋込みデータが再現可能なように書き換えの関数を設計する問題である。先行研究において、Galand と Kabatiansky は、カバーデータおよび埋め込みを行った後のデータ (これを「ステゴデータ」という) を全て 2 値系列であるとし、カバーデータとステゴデータとの距離をハミング距離によって測るモデルを定式化している。そのような定式化のもとで、彼らは問題の解が誤り訂正符号の 1 概念である被覆 (covering) 符号に帰着できることを示し、距離限界と埋込み可能容量との関係を導出している。

被覆符号のもっとも基本的な例は、ハミン

グ符号である。実際、ステガノグラフィをハミング符号によって実現する例が種々の文脈で紹介されている。ハミング符号などの誤り訂正符号とステガノグラフィとの関係が明確に意識されるようになってからは、このような手法はシンドローム符号化もしくはマトリックス埋込みと総称されるようになってきている。ハミング符号のステガノグラフィへの応用では、ハミング符号の被覆符号としての、また完全符号としての性質が重要な役割をはたしている。一方、これらの性質を共有する 2 元ゴレイ符号についても、同様の目的への応用可能性への言及が従来なされている。しかし、実際の情報ハイディングにゴレイ符号を応用した例は見当たらない。

本研究では、ゴレイ符号の情報ハイディングへの応用可能性を検証した。特に、2 値画像への情報埋め込み法として具体化を行い、ハミング符号を利用した手法との比較を行う。さらに 2 値画像への情報埋め込み法を濃淡画像に拡張する手法を提案する。濃淡画像はビットプレーン分解によって 2 値画像の重量と考えることができる。2 値画像用の情報埋め込み法を各ビットプレーンに適用することで濃淡画像への適用が可能となる。本研究では、ビットプレーンごとに独立に埋め込む場合よりも、埋込み量を減らすことなく画質の劣化を抑えることのできる手法を提案し、その手法にゴレイ符号を組み合わせた情報埋め込み法の性能を評価した。その結果、埋込み量を減らすことなく画質の劣化を抑えることのできる手法の存在を確認し、さらに、ハミング符号を用いた手法と比較して、埋込み率が低い領域ではほぼ同等の性能が得られ、埋込み率が高い領域では、ゴレイ符号を用いた提案法の性能の方が優れていることを確認した。(雑誌論文④, 学会発表④)。

(3) 無ひずみデータ圧縮の代表的手法であるブロックソート法の圧縮性能と速度性能の更なる高性能化を目的として、その要素技術の再検証を行った。まず、ブロックソートデータ圧縮法の Burrows-Wheeler 変換 (BWT) の拡張として、Generalized Radix Permutation 変換 (以下、「GRP 変換」と略す) と名づけた変換法を提案した。本変換は、ブロック長と次数という二つの整数パラメータを有し、これらの選び方によって、BWT を含む種々の変換法を記述することが可能な一般的な手法になっている。

本研究では、ブロックソートデータ圧縮法の漸近的性能と類似の漸近的な性能を示すと同時に、現実のデータを対象とする、パラメータに応じた実性能の評価を行った。特に後者を目的として、BWT の変換後に適用される Move-to-Front 変換を GRP 変換向け

に修正し、それを組み込んだ無ひずみ圧縮法を構成した。圧縮実験の結果、BWTを明確に改善しているとは言えないものの、圧縮対象によっては、GRP変換がBWTの意味のある拡張になっていることを示すことができた。(雑誌論文①, ②, ③, 学会発表②, ③)。

以上の研究を通じて、マルチデータの圧縮という概念を確立し、それに属す問題群のいくつかについて新規かつ有効な解決手段を示すことができた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

- ① 稲垣和将, 富澤義宏, 横尾英俊, 一般化 Radix Permute 変換による無ひずみデータ圧縮, 電子情報通信学会論文誌A, vol.J93-A, no.8, 掲載決定, 査読有, 2010.
- ② H. Yokoo, Extension and faster implementation of the GRP Transform for lossless compression, CPM 2010, Lecture Notes in Computer Science, vol. 6129, pp.338-347, 査読有, 2010.
- ③ K. Inagaki, Y. Tomizawa, and H. Yokoo, Novel and generalized sort-based transform for lossless data compression, SPIRE 2009, Lecture Notes in Computer Science, vol. 5721, 査読有, pp.102-113, 2009.
- ④ 笠原祥一, 横尾英俊, ゴーレイ符号による情報ハイディングと画像データへの応用, 電子情報通信学会論文誌 A, vol.J91-A, no.6, pp.685-694, 査読有, 2008.

[学会発表] (計5件)

- ① H. Yokoo, File-size preserving LZ encoding for reversible data embedding, DCC2010, Data Compression Conf., 2010.3.25, Snowbird, UT, USA.
- ② 富澤義宏, 稲垣和将, 横尾英俊, 無ひずみデータ圧縮のための一般化 Radix Permute 変換, 第32回情報理論とその応用シンポジウム, pp. 325-329, 2009.12.2, 湯田温泉, 山口県.
- ③ 稲垣和将, 富澤義宏, 横尾英俊, GRP 変換によるデータ圧縮の実験, 第32回情報理論とその応用シンポジウム, pp. 330-335, 2009.12.2, 湯田温泉, 山口県.
- ④ 笠原祥一, 横尾英俊, ゴーレイ符号による情報ハイディングの濃淡画像への応用,

第2回マルチメディア情報ハイディング研究会, pp.17-22, 2007.11.9, 宮城蔵王, 宮城県.

- ⑤ 側高英俊, 横尾英俊, MPEG-4 ALS マスクLZ符号化におけるデータ埋め込み, 第2回マルチメディア情報ハイディング研究会, pp.23-28, 2007.11.9, 宮城蔵王, 宮城県.

6. 研究組織

(1) 研究代表者

横尾 英俊 (YOKOO HIDETOSHI)
群馬大学・大学院工学研究科・教授
研究者番号: 70134153

(2) 研究分担者

()

研究者番号:

(3) 連携研究者

()

研究者番号: