

フラクタル圧縮の音声データへの応用と 音声データに有利なアフィン変換

謝花 蔵 河野 真治
Osamu JAHANA Shinji KONO

琉球大学大学院理工学研究科情報工学専攻
Specialty of Information Engineering, University of the Ryukyus.

概要

フラクタル画像圧縮では、画像の中の自己相似性を縮小アフィン変換を選ぶことにより求め、それを符号化することにより画像データの圧縮を行なっている。

本研究では、フラクタル圧縮を音声データに応用し、異なる3つの方法で圧縮を行ない、音声データに有利なアフィン変換を提案する。

1 フラクタル圧縮

ここで、フラクタル圧縮のアルゴリズムの説明を参考文献[1]のフラクタル画像圧縮に基づき行う。

1.1 レンジブロックへの分割

フラクタル圧縮は、データ内での自己相似性を見つけだし、それを符号化することでデータの圧縮を行う。しかし、一般のデータにおいて、自己相似性があったとしても見つけることは容易なことではない。そこで、データをある大きさのレンジブロックとよばれるブロックに分割する(図1)。その各レンジブロックに対して、最適なアフィン変換とドメインブロックを見つけだし、それを符号化する。

1.2 最適なアフィン変換とドメインブロックの決定

まず、分割されたレンジブロックに対して行われることは、クラス番号を見つけることである。クラス番号は、与えられたレンジブロックを4分割したときの輝度による順序付けによって決定される。レンジのクラス番号がわかれば、同じク

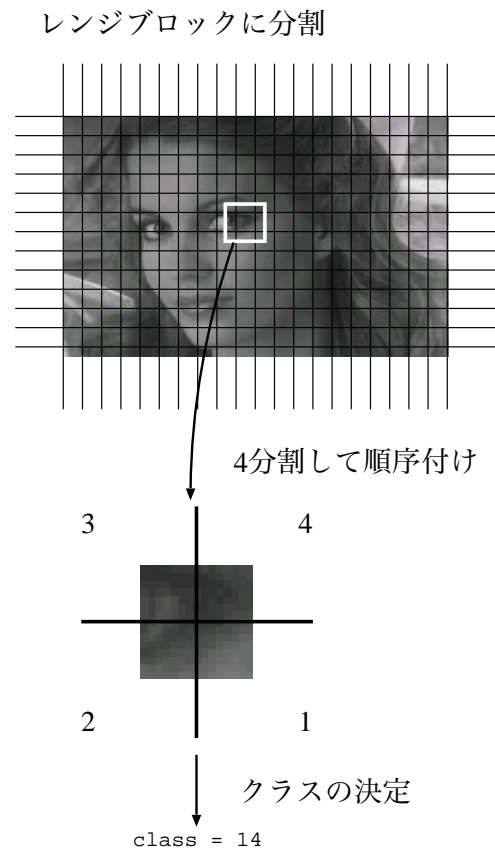


図1: レンジ分割からクラス番号の決定

ラスに属するドメインと比較することによって、うまく一致する可能性がほとんどないドメインブロックとレンジブロックの比較を避けることができる。レンジブロックと同じクラス番号を持つドメインブロックをとってきて、ドメインブロックからレンジブロックへの最適なアフィン変換を決定する。ここで用いるアフィン変換は、ドメインブロックに対して0.5倍の縮小と平行移動、コントラストと輝度の調節のみを許している。平行移動に関するパラメータは、レンジブロックとドメインブロックの位置から求められ、コントラストと輝度に関しては式(1)を最小にするものをとる。

$$E(c, b) = \sum_{i=1}^n (cd_i + b - r_i)^2 \quad (1)$$

c はコントラスト要素、 b は輝度のオフセット、 n はレンジ内のピクセルの個数、 d_i はドメイン内のピクセルの値、 r_i はレンジ内のピクセルの値である。 c と b の最適な値は、 $E(c, b)$ の c と b に対する偏微分が両方とも0になった時に得られる。その連立方程式を解けば、 c と b は容易に得られる。このとき得られる c と b は、最初にとってきたドメインに対しては、 $E(x, b)$ を最少にする値であるが、異なったドメインをとってきたときに、 $E(c, b)$ がさらに小さな c, b が存在するかもしれない。よって、 $E(c, b)$ がある一定の数値を下回るまで、同一のクラスに属するドメインを試していく。同一のクラスのドメインをすべて試しても、まだ $E(c, b)$ が条件を満たさない時は、レンジを4分割して、分割された各レンジに対して、上に述べた処理を同様に行なう。

以上に述べたことをすべてのレンジブロックに対して行ない、コントラスト要素、輝度のオフセット、そしてドメインの位置を符号化して、圧縮は完了する。

1.3 復号化

各レンジに対するコントラスト要素 c 、輝度のオフセット b そしてドメインの位置を読み込んで、ドメインの各ピクセルに c 倍したあと、 b をたす。それをすべてのレンジに対して行なう。この作業を繰り返すことで、画像は徐々に収束していき、8回ほどで完全に収束する。

2 フラクタル圧縮の音声への応用

ここでは、フラクタル圧縮を利用したいくつかの音声圧縮を説明する。

2.1 音声を画像として圧縮

1次元の音声データがある一定の長さで切って、並べて2次元データとする。それをグレースケール画像として、1節で説明したフラクタル画像圧縮で圧縮を行う。サンプリングレートが22,050Hzの約4秒の音声データを画像化したものを図2に示す。

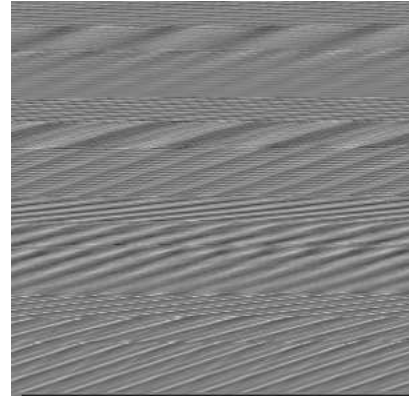


図 2: 画像化された音声データ

圧縮率は、 $\frac{14,949\text{byte}}{89,732\text{byte}} = 0.167$ で、復号後の音声データを聞いた感じは、全体的に white noise がかかっているように聞こえる。

2.2 周波数領域を圧縮

この方法は、音声データに対して、フーリエ変換を用いて周波数領域を求め、それに対して、フラクタル画像圧縮を行う。音声データの N 個の連続したサンプル $d_k(t)$ に対して、式(2)のような離散フーリエ変換を行ない、フーリエ係数 $c_{k_n}(n = 0, \dots, N - 1)$ を得る。

$$c_{k_n} = \sum_{t=0}^{N-1} \omega^{-nt} d_k(t) \quad (2)$$

(ただし ω は、1の原始 N 乗根)

実関数 $d_k(t)$ に対する、フーリエ係数には、

$$\begin{cases} c_{k_0} \in R \\ c_{k_n} = c_{k_{N-n}}^* \quad (n = 1, \dots, N - 1) \end{cases}$$

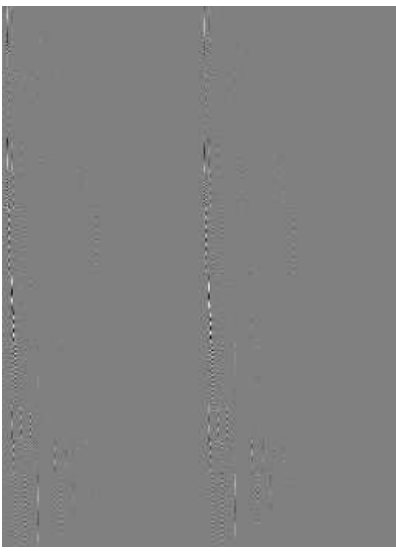


図 3: フーリエ変換後の音声データ

が、成り立つことが一般に知られている。これによって、必要なデータは、 $c_{k_0}, c_{k_{\frac{N}{2}}}$ の実部 (虚部は必ず 0 である) と c_{k_n} ($n = 1, \dots, \frac{N}{2} - 1$) の $\frac{N}{2} - 1$ 個の複素数である。つまり、フーリエ係数 c_{k_n} は、 N 個の実数で表される。この N 個の実数を量子化したのち、画像を作成する。画像 $f(x, y)$ は、

$$f(x, y) = Qu(Re(c_{y_x})) \quad (x = 0, \dots, \frac{N}{2})$$

$$f(x + \frac{N}{2}, y) = Qu(Im(c_{y_x})) \quad (x = 1, \dots, \frac{N}{2} - 1)$$

のように定義する。 $Re()$ は実部、 $Im()$ は虚部を得るための、 $Qu()$ は量子化のための関数とする。このようにして作られた画像を図 3 に示す。

この画像に対し、フラクタル画像圧縮を施す。圧縮率は、 $\frac{4,035byte}{89,732byte} = 0.044$ で、復元された音はひずんでいる。

2.3 生データを圧縮

この方法は、1 節で述べたフラクタル画像圧縮のアルゴリズムをそのまま音声データに適応したものである。使用するアフィン変換 A は、音声データを $D(t)$ としたとき、次のような形をとる。

$$A : D(t) \rightarrow aD(2t) + b$$

ここで用いるアフィン変換は、縮小写像でなくてはならないので、

$$0 \leq a < 1$$

となる。b については、1 サンプルあたりのビット数に依存するが、ここでは、8bit をもちいているので、

$$0 \leq aD(2t) + b \leq 255$$

を満たす値をとる。

このアフィン変換の例を図 4 に示す。アフィン変換やドメインブロックの探索には、1 節で述べた方法を用いている。

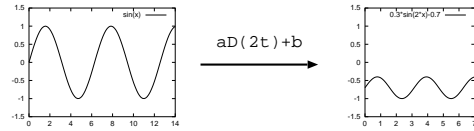


図 4: 音声データのアフィン変換

復号化時の様子を図 5 に示す。ドメインに対し、dynamic range、offset を求め、range を適用する。8 回ほどくり返すと収束する。

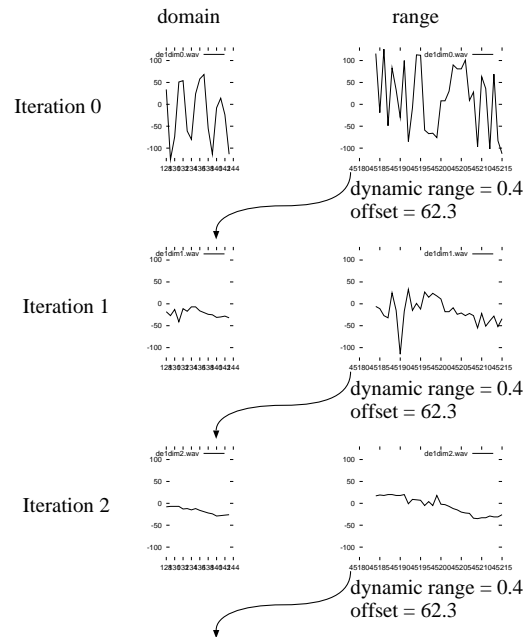


図 5: アフィン変換を適用していく

この方法での圧縮率は、 $\frac{16,127byte}{89,732byte} = 0.179$ で、復元された音にはノイズが含まれる。

3 結果と考察

まず、上で述べた3つの方法の圧縮率の一覧を表1に示す。どの圧縮方法にもノイズやひずみが

圧縮方法	圧縮率
音声を画像化	16.7%
周波数領域	4.4%
生データ	17.9%

表 1: 圧縮率一覧

存在し、その度合は圧縮率のよい順に悪くなっていく。つまり、音質の面から考えると、よい順に生データを圧縮 >> 音声を画像化して圧縮 >> 周波数領域を圧縮、という順になる。

音声を画像化して圧縮する方法と生データを圧縮する方法は、アルゴリズム的に似ていて圧縮率やノイズの入り方もほぼ同じである。ただ、生データを圧縮するときは、(画像としてでは無く)直接音声データとして扱うため画像化して圧縮する方法よりも波形が乱れにくく、ノイズが減るようである。

周波数領域に対してフラクタル圧縮を行ったときの圧縮率が抜きん出ていることについて注目したい。

この理由として、2つ考えられる。

一つは、フラクタル画像圧縮で用いたアルゴリズムと相性がよかったこと。一般に音声データを周波数解析すると低周波成分の方にデータが偏り、高周波成分は flat になる。ここで用いているフラクタル画像圧縮のアルゴリズムでは、flat なレンジブロックに対しては、レンジブロックを分割する必要はないため、可能な限り大きなレンジブロックを用いることができるため、符号化するアフィン変換の数が減り、圧縮率が上ることになる。

もう一つは、音声の特徴とフラクタル圧縮との関係によるものだと考えられる。画像の場合、自然風景の真中に巨大なビルがそびえている絵があったとしたら、自然風景がビルによって切られてしまうことになり、自然風景の持つ対称性を利用することができなくなってしまう。しかし、音声の場合、2つの音が混じりあっても周波数解析を行えば、大抵の場合分離することが可能で、データの対称性が失われることは少ない。つま

り、フーリエ変換することにより、音声データのもつ対称性を明確にし、それをうまく圧縮することができたためだと考えられる。

このことは、フラクタル圧縮にとって重要なことである。フラクタル圧縮で自己相似性を探す時、なんらかの邪魔によって、自己相似性が隠され、大幅な圧縮率の低下につながるものが予想される。しかし、音声データの場合では、データの持つ自己相似性は失われにくく、フラクタル圧縮はデータの持つ対称性を圧縮することができる。このことからフラクタル圧縮は、画像圧縮よりも音声圧縮に適しているといえよう。

4 結論と今後の課題

本稿では、フラクタル圧縮の音声の応用の利点とフーリエ変換を用いたときのフラクタル圧縮の利点を述べた。現在、周波数領域を圧縮する場合、まずフーリエ変換をした後、フラクタル圧縮を行っているが、今後この過程を一つのアフィン変換で行ないたい。このことによって、画像化するときの量子化の過程でのロスや、今のフラクタル画像圧縮での影響(コントラストが小さくなる、ぼける)によるロスをなくし、音声データのフラクタル圧縮の利点をそのままに、高圧縮、高音質な圧縮が期待される。

参考文献

- [1] M. ネルソン, J.L. ギャリー 著/萩原 剛志 山口 英 訳, “データ圧縮ハンドブック”, トッパン,
- [2] M.F. バーンズレイ/ 蔡 東生・江守 正多 訳/, “マルチメディア・フラクタル画像圧縮”, トッパン, 1995
- [3] マリオ ペルギア=著/ 柳原 二郎 川瀬 眞 陳 啓浩=訳, “離散的反復関数系の理論”, トッパン, 1995
- [4] 大石 進一 著, ”フーリエ解析”, 岩波書店