

ビデオ圧縮符号化技術の昨日・今日・明日

和田 正裕

(株) KDD 研究所

● 昔々

テレビが発明されたとき、人々が次に期待したものはテレビ電話であったという。真贋の程は定かではないが、なるほどテレビ電話はテレビの素直な発展形であるといえる。ただし、音声とテレビ(ビデオ)では信号の帯域が3桁ほど異なるので、既存の電話線でビデオ信号を送ることは不可能である。そのため、さまざまな「圧縮」、つまり信号の帯域を制限する(情報レートを削減する)処理(これを高能率符号化という)が研究された。事実、1960年代には日米の電話会社がテレビ電話サービスの試作を行っており、1970年の大阪万博ではそのデモも行われた。

当時は、モノクロ動画像をアナログ的に圧縮伝送していたが、アナログ処理には限界がありうまくいかなかった。その後、だんだんとビデオ信号のデジタル処理が可能となるにつれ、はるかに複雑なアルゴリズムを試せるようになった。そして現在ではLSIの大きな進歩により、複雑な圧縮アルゴリズムの経済的な実現が可能となってきた。

ではテレビ電話は完成したのか、と問われると、その答えは曖昧である。一般的な通信インフラの速度(～64kbps?)で放送テレビの画質を得ることは実現されていないが、表情が分かる程度の画質は得られている。将来はネットワーク側の高速化により、テレビと同じ画質が得られるかもしれないが、圧縮アルゴリズムの改善だけでは非常に困難である、というのが一般的な理解であろう。

ビデオの圧縮符号化は、むしろ衛星によるデジタル多チャンネル放送、SNG(衛星によるニュース素材伝送)、DVD、ビデオCD(カラオケでおなじみ)などを実現した点で、我々の生活に深くかかわっている。道路や銀行のビデオ監視システムなど、目

につかないところにも圧縮伝送技術が使われており、すでに社会のインフラ技術となっている。

● ビデオ圧縮符号化の原理¹⁾

ビデオ圧縮符号化の原理は、他の情報圧縮技術と同様、信号の性質を上手に使う、何らかの操作によりその統計的性質を偏らせ、これをエントロピー符号化することである。後述するが、たとえばビデオ信号の振幅(画面の明るさに相当し、デジタルでは標本点(画素と呼ぶ)の値である)自体は図-1(a)のようにランダムに分布するが、隣接画素との差分をとると、図-1(b)のようにゼロを中心としたシャープな分布を示す。これに可変長符号を割り当てる、つまり発生頻度の高いゼロ付近の信号には短い符号語を、頻度の低い大きな値の信号には長い符号語を、それぞれ割り振れば、画素値をそのまま固定長符号で伝送するよりも情報量を減らすことができる。

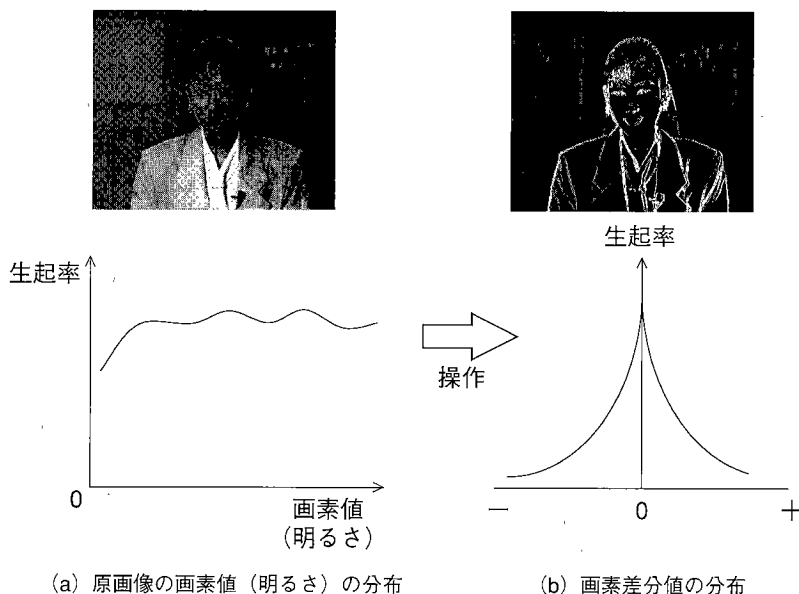


図-1 圧縮符号化の原理(例)

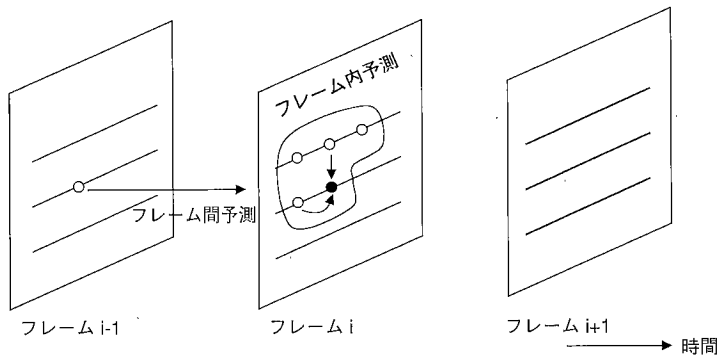


図-2 予測符号化の原理

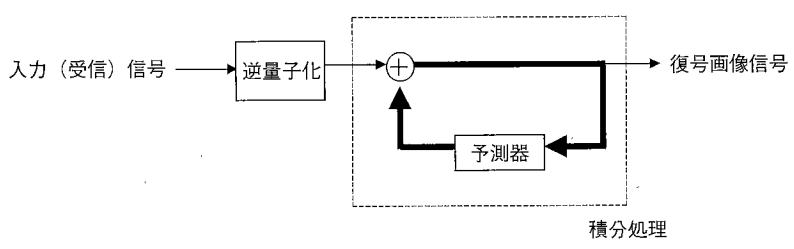


図-3 予測符号化の受信処理

ビデオ圧縮符号化の特徴は、医療用画像など特殊なものを除いて非可逆符号化（符号化処理により情報が失われるような符号化）が用いられることである。これにより信号に歪みが発生し、復号しても完全には元に戻らないが、ビデオ信号は「目で見る」ものであるから、視覚的に問題がなければかまわないと考える。これが汎用のファイル圧縮など他の圧縮技術と大きく異なる点である。画像符号化では、可逆圧縮の場合にはわざわざ「ロスレス符号化」などという。ロスレス符号化では、一般的に高い圧縮効率は見られない。

この非可逆符号化における近似操作を量子化といい、近似の度合い（たとえば量子化幅）により発生する情報レート（符号化レート）を制御する。また、この近似方法に工夫をして人間の視覚特性を上手に利用することにより、数値（SN比など）的に同じでも、劣化が目立たない符号化を実現することができる。たとえば、人間の視覚は斜め方向の精細さに対して鈍感であるので、斜め方向の解像度（信号的には高周波成分）を落としても画質の劣化は目立ちにくい。また、動物体のエッジなど画像の変化や動きが激しい部分では歪みが目立ちにくいので、荒く量子化してもかまわない。さらに、シーンチェンジ直後の数フレームは、目が追いつかないので劣化があっても許容される。

●ビデオ符号化の要素技術²⁾

ビデオ圧縮符号化では、以下のような「要素技術」が（通常はいくつか組み合わせられて）使用される。

予測符号化 (DPCM)

ビデオ信号では隣接する画素の相関が高いため、図-2のように周辺の画素から予測を行い、予測値からの誤差を計算すると、図-1で示したようなゼロを中心に指数関数的な分布をとる偏った信号となる。したがって、画素値自体ではなく予測値からの誤差を伝送すると情報レートを削減できる。これを予測符号化という。予測値として同じフレームの画素を用いるものをフレーム内符号化、隣接フレームの画素を用いるものをフレーム間符号化と呼ぶ。空間予測、時間予測などという場合もある。

直感的に考えて、動きの少ないビデオ画像ではフレーム間予測がよく当たり、動きが激しくぼけたような画像ではフレーム内予測がよく当たる。このように画像の性質を考慮して予測値を切り替えるものを適応予測と呼ぶ。

予測符号化では、図-3のように受信した誤差情報を過去の信号（予測値）に順次加算することにより再生画像を作り出す。このように、予測符号化は基本的に積分処理であるため、伝送路に誤りが起こると再生画像が長時間劣化することがある。

変換符号化

変換符号化では、信号を時間空間領域（画素の値で表現される）から位相周波数領域（変換係数の値で表現される）へ変換し（変換係数の個数は、変換前の画素の個数と同じである）、信号のパワーが直流付近に集中することを利用して情報レートを削減する（図-4）。つまり、小さな値をとる確率の高い高次の変換係数に可変長符号を割り当てる、または値の小さな高次係数は伝送しないことで、情報圧縮を行う。変換には、係数が0, 1のみで演算が簡単なアダマール変換、パワーの集中効率が高い離散コサイン変換 (DCT) などがあり、最近ではLSIの普及によりDCTがよく用いられる。

変換符号化は原理的に画像の空間的冗長性を利用するものであるから、時間的冗長性を利用するフレーム間予測符号化と組み合わせられる（フレーム間予測誤差信号を変換符号化する）場合が多く、これをハイブリッド符号化と呼ぶ。MPEGなど現在主流となっている符号化方式である。また変換符号化では、処理の都合上、画像を複

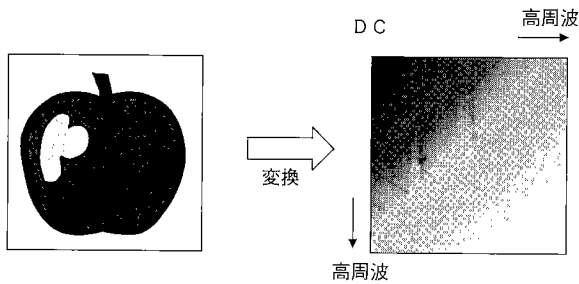


図-4 変換符号化の原理

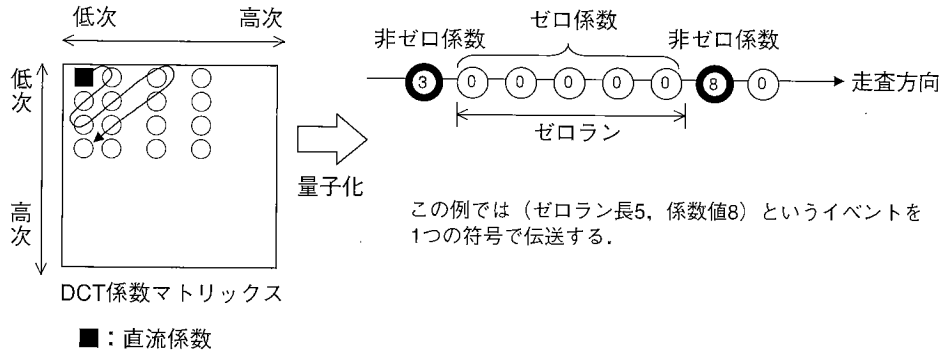


図-5 DCT係数のジグザグ走査

数画素の集まりである（通常矩形の）ブロックに区切って処理する。これをブロック符号化と呼ぶが、低レートでの符号化ではタイル状の目立つ歪みを引き起こすため、視覚特性上は好ましくない。

補助的技術

(1) ベクトル量子化

非可逆符号化における量子化では、通常は画素ごとに量子化代表値への近似操作（スカラ量の近似）が行われるが、これを複数画素まとめて1つの量子化代表値（この場合はまとめる画素数の次元を持つベクトル）に近似する操作（ベクトル量の近似）をベクトル量子化と呼ぶ。独立に量子化するよりも効率は高いが、代表ベクトルの集合（コードブック）の設定方法や処理が難しい。ただし量子化技術であるので、他の符号化技術と容易に組み合わせることができる。

(2) 動き補償

フレーム間予測符号化では、画像に動きがあると予測効率が落ちる。この時フレーム間の動きを検出して対応する画素の位置を補正すれば、予測効率は向上する。これを動き補償と呼び、動き検出LSIの実用化とともに広く使われるようになった。動き補償は、通常ブロック符号化におけるブロック（またはその整数倍）単位で行われる。

(3) 可変長符号化

ビデオ符号化では、エントロピー符号化のための可変長符号としてハフマン符号が多く用いられる。より適応性の高い算術符号は、処理量の点でビデオの実時間符号化に用いることは難しい。

●ビデオ符号化アルゴリズムの標準化

標準化の必要性

ビデオの圧縮符号化は、前述のように画像通信を目的として始まったため、相互接続性の点からアルゴリズムの標準化に対して強い要望があった。また、符号化アルゴリズムは非常に高速で複雑な処理を必要とするため、かなりの規模の専用ハードウェアが必要となり、これをLSI化するためにも経済的な意味で標準化は不可欠であった。

符号化アルゴリズムは、対象とする画像や想定するビットレート（符号化レート）に依存する部分があるため、目的や用途に応じて標準化を行うことになるが、ここでは現在の代表的な符号化方式を紹介する目的で2種類の方式に触れる。

勧告H.261³⁾

現在主流のビデオ圧縮符号化は、動き補償フレーム間予測の誤差をDCT符号化するハイブリッド符号化であるが、その最初の枠組みを確立したのがITU-T (旧CCITT) 勧告H.261である。テレビ電話・会議用なのでカメラ固定の動きが少ない画像を対象とし、適用ビットレートは64kbps～2Mbpsである。

H.261の基本部分はハイブリッド符号化であるが、新たに確立された技術として、DCT係数（変換した値）をジグザグ走査し、連続するゼロ係数の数（ラン長）とそれ

に続く係数値をまとめて符号化する方法がある(図-5)。予測がよく当たったり量子化が荒い場合にはゼロになる係数の割合が多いため、効率的な符号化法である。

また、標準化はある意味で技術の凍結であるから、その進歩を妨げる弊害が指摘されるが、H.261では相互接続性に関係のある部分のみを規定し、その他の部分は「実装の問題」として規定していない。たとえば動ベクトルの検出方法、DCTの高速演算法などで、特に後者は相互接続性に直接関係するから、本来ならば演算方式まで一意に決めてしまいたいところであるが、特許などの問題とも絡むため、演算精度を規定するという間接的な方法を採用している。

MPEG⁴⁾, 5)

ご存じのように、DVD、デジタル衛星放送、そして2003年に予定されているデジタル地上波放送で採用されている符号化方式であり、ISO、ITU共同で標準化された。MPEGは一般画像を対象としており、ビデオCDなど1Mbps程度のレートを対象としたMPEG-1と、数Mbps～15Mbpsを対象としたMPEG-2(HDTVの場合には50Mbps程度まで)があるが、画像のフォーマットが違う以外、符号化の原理はほぼ同じである。

その基本部分はH.261のハイブリッド符号化を踏襲するが、MPEGでは図-6のように予測方式がフレームごとに異なり、符号化対象フレームを挟む両フレームからフレーム間予測を行うことが特徴である(H.261は片方向のみ)。このため、符号化処理(と伝送)の順番が通常とは異なり、比較的大きな遅延が生じるが、会話型通信ではなく放送やパッケージメディアを対象としているので支障はない。この両方向予測により、物体の陰から新たに現れる背景に対しても予測が有効となり、符号化効率が向上する。

●ビデオ圧縮符号化の実現技術⁶⁾

ビデオ圧縮符号化のコーデックは高速演算を必要とするため、専用ハードウェアが必要である。最も演算量の多い処理は動ベクトル検出であり、かなり早い時期からLSI化が行われてきた。それでも数年前までは、MPEG-2クラスのコーデックは大型の家庭用電子レンジ程度はあったが、現在では数社のメーカーで1チップエンコーダが開発されており、PCカードサイズのコーデックも発表されている。HDTV用はまだかなり大きいですが、LSI化の進展により小型化も時間の問題であろう。

一方、ここ数年、汎用CPUの高速化に伴い、ソフトウェアによる符号化処理が試みられており、1Mbps程度の低レートの符号化処理(MPEG-1やH.261など)は、ほぼ

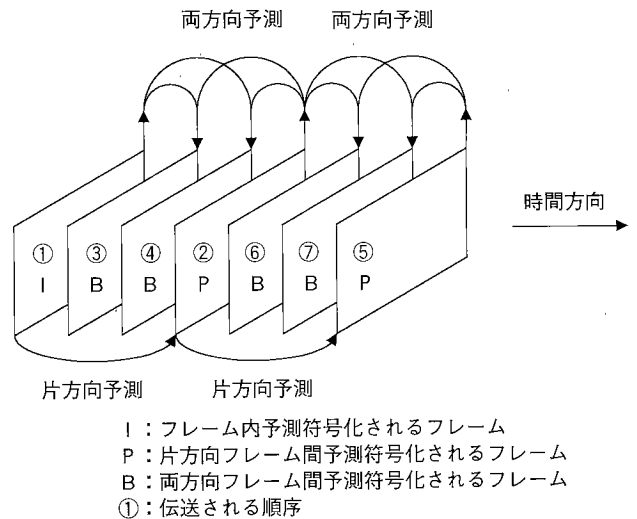


図-6 MPEGの予測方式

満足できる圧縮効率で実行可能となってきた。また復号は、MPEG-2クラスの処理がリアルタイムで行えるようになりつつある(いわゆるDVDプレーヤ)。このクラスの符号化処理は、現在はまだ困難であり(一部にリアルタイム処理を標榜するものもあるが、動き補償などの処理を簡易化したものが多く、符号化画質は不十分である)、オフライン処理を行うものがいくつか発表されている⁷⁾。しかしながら、現状のCPUの高速化のスピードからみて、MPEG-2符号化の完全なソフトウェア処理も近いうちに実現されるに違いない。

●ビデオ符号化アルゴリズムの現状

残念ながら、というべきか、実用上のビデオ圧縮符号化の効率は、近年大幅には向上しておらず、符号化レートと画質の関係はほぼ分かっている(もちろん、研究レベルではいろいろな試みがあり、また、ノウハウの蓄積などにより実用レベルの画質も毎年着実に向上しているが)。符号化レートと画質の関係をまとめたのが表-1である。符号化レートは慣習として伝送ハイアラーキに対応しているが、インターネット時代には無意味であろう。また、符号化レートに幅があるのは、許容される画質劣化が状況により異なるためである。

●昨日・今日、そして明日

ビデオ圧縮符号化技術の今までをまとめると図-7のようになろう。

最近の研究動向をみると、圧縮効率の向上を目指す従来型の研究に関しては、ハイブリッド符号化の延長線上で改良を試みるもの(たとえばブロック歪みをなくすため、DCT以外の非ブロック型変換を用いるものや、画面

画質レベル (アプリケーション)	代表的な符号化レート	代表的な符号化アルゴリズム	備考
テレビ電話	16kbps~384kbps	H.261, H.263	・カメラ固定, 1人の顔のバストアップ ・低解像度 (180×120画素以下) ・「なんとなく表情がわかればよい」
テレビ会議	128kbps~2Mbps	H.261, H.263	・カメラ半固定, 数人の顔のバストアップ ・中解像度 (360×240画素程度) ・「表情がはっきりわかる」
SNG (衛星ニュース素材伝送)	1Mbps~10Mbps	H.261, MPEG-1 MPEG-2	・カメラフリー, 一般画像 ・中~高解像度 (360×240~720×480画素程度) ・「劣化があまり気にならない」
SDTV放送 (2次分配)	4Mbps~10Mbps	MPEG-2	・一般画像 ・高解像度 (720×480画素) ・「一般人には劣化がわからない」
SDTV素材伝送 (1次分配)	10Mbps~50Mbps	MPEG-2	・一般画像 ・高解像度 (720×480画素) ・「専門家にもほとんど劣化がわからない」

表-1 符号化レートと画質の関係

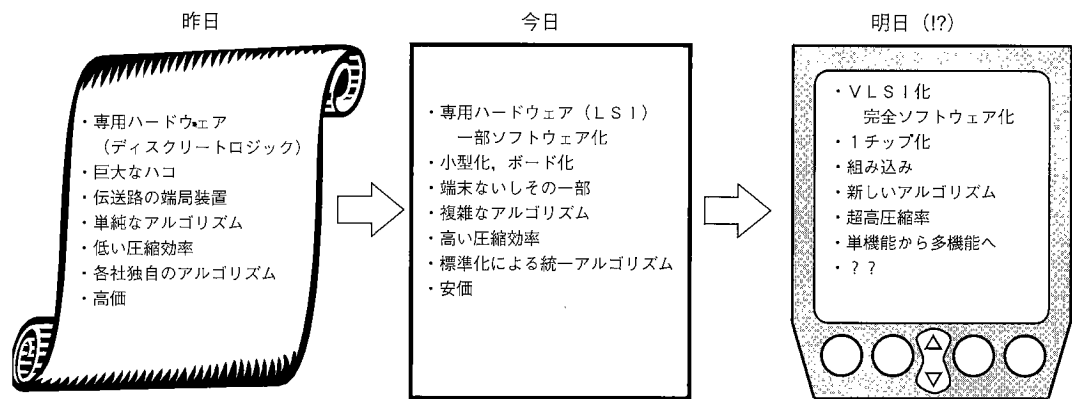


図-7 昨日・今日・明日

をいくつかの領域に分割してそれぞれに最適な符号化を行う領域分割符号化など)と、さらなる大幅な符号化効率の改善を目指す新しい符号化の試みがある。後者には、画像に内包されている自己相似性(画像の部分と全体の相似性。たとえば唐松の枝は唐松全体の形と相似である)を利用したフラクタル符号化や、送信側と受信側に画像に関するデータベースを用意し、送信側で画像認識を行ってデータベースを検索し、受信側で画像情報を合成する知的符号化(知識ベース符号化、認識合成符号化とも)などがある。これらの研究が成功すれば符号化効率は一気に2桁程度向上することになるが、現状では適用できる画像に限られているようである。

一方、別方向の研究として立体画像の符号化がある。ステレオ視であれば従来の符号化技術をほとんどそのまま適用できるが、3次元構造を持った立体画像を圧縮するには新しいアルゴリズムが必要である。CG画像に対する研究は多いが、一般画像に対する検討はこれからである。

今まで述べてきたように、ビデオ圧縮符号化の技術は現在、収穫期、実用期にあるため、最近ではその応用研究

が盛んになってきた。たとえば、ビデオ符号化に関する国内の代表的なワークショップであるPCSJと、併催される画像処理応用のワークショップであるIMPSは、1999年に初めて発表件数が逆転した。

また前述のMPEGでは、新しいMPEG-4(中低レート汎用符号化)規格において、圧縮効率の向上とともにインターネット応用やPC処理、アニメーションを考慮しており、現在検討が行われているMPEG-7では、画像圧縮ではなく画像検索のための検索キーとなるメタデータの規格を策定している。このように、ビデオ圧縮符号化は、単純に圧縮効率の向上だけを目指した時代が終わり、今後はビデオ圧縮に関連するさまざまなアプリケーションと融合して研究が進むものと予想される。新たな展開を期待したい。

参考文献

- (社)映像情報メディア学会編: 画像情報圧縮, オーム社.
- (社)映像情報メディア学会編: 目で見える画像圧縮, コロナ社.
- ITU勧告H.261: Video Coding for Audiovisual Services at px64kbps.
- マルチメディア通信研究会編: 最新MPEG教科書, アスキー出版局(1994).
- <http://www.cselt.it/ufv/leonardo/index.htm>
- 古閑: マルチメディア端末, (社)映像情報メディア学会編, 昭晃堂.
- <http://www.lab.kdd.co.jp/jpn/frame.html>

(平成11年12月2日受付)

