

記録系動画像符号化方式の検討

大町 隆夫 太田 睦

日本電気(株) C&C情報研究所

CD-ROMに代表されるデジタル記録媒体に長時間の動画像を記録するための動画像符号化方式について検討した。CD-ROMの最大転送レートは150Kbyte/secである。この符号化レートに対しては従来通信系用として開発されたフレーム間予測符号化方式と離散コサイン変換とのハイブリッド符号化方式が極めて有効であることが確認されているが、本報告ではこの方式が特殊再生という通信系とは異なる機能が要求される記録系に対しても有効であることを示した。さらに記録系における符号化効率改善手段のひとつとしてノンリアルタイム処理を用いた背景予測方式を提案し、効果を確認した。

A Study of Motion Video Coding
for Digital Storage Media

Takao Omachi Mutsumi Ohta

C&C Information Technology Research Laboratories, NEC Corporation

4-1-1, Miyazaki, Miyamae-ku, Kawasaki-shi, Kanagawa JAPAN

A coding method is discussed for video signals on digital storage media, especially for CD-ROM systems. The maximum transmission rate is 150kbyte/sec on CD-ROM interface. It is well known that Hybrid Coding method is very effective on this data rate in transmission systems. In this paper, effectiveness of this coding algorithm is discussed and confirmed also in digital storage media, though required functions have some differences from transmission systems. Moreover, non-realtime background prediction method is proposed, and it is clarified that this method improves coding efficiency.

1. はじめに

近年大容量、高速読出しが可能なデジタル記録媒体が開発され動画データが長時間デジタル記録することも不可能でなくなってきた。例えばCD-ROMは540Mbyteの記録容量を持ち、最大150Kbyte/secでデータが読み出せる[1]。またデジタルオーディオテープ(DAT)の読み出し速度は192Kbyte/secであり約2時間の記録が可能である[2]。このようなデジタル記録媒体に動画を長時間記録できれば、教育・娯楽・ビジネス等広範囲な応用分野が期待できる。既にこのデジタル記録媒体を対象とした記録用動画信号の標準符号化方式に対する検討がISO(国際標準化機構)の場でも開始されている。[3]。

動画データはそもそもの信号帯域が広く、音声信号の約1000倍とされている。CCIRのスタジオ用のデジタルテレビ信号規格[4]に従えば約20Mbyte/secのデータレートで情報が発生する。またテレビ会議・テレビ電話用の国際規格であるCCITTのCIFフォーマット[5]では約4Mbyte/secの情報量が発生する。従って現在使用可能な上記の記録媒体、或いはこれから数年間の間に使用可能な記録媒体に動画信号を記録するためにはCIFフォーマットからでもおよそ1/20-1/30の情報量圧縮を行う必要がある。

この規模の情報量圧縮において、現在のところ、最も効率的とされる符号化方式としてフレーム間予測符号化方式と直交変換符号化方式を組合せたハイブリッド符号化方式[6]がある。ただしこの符号化方式は通信系の符号化方式として研究開発が進められてきたため記録系符号化方式として適用する場合にはもう一度その最適性を見直す必要がある。

本報告ではこのハイブリッド符号化方式が記録系の符号化方式にも適応できることを示す。特に記録系符号化方式において必要とされる特殊再生機能をハイブリッド符号化方式で実現できることを示す。つぎに記録系符号化方式で許される符号

化効率改善項目を整理し、ノンリアルタイム処理による効率改善の一つとして背景画像の抽出およびそれを用いた予測方式を提案し効果を示す。

2. 記録系符号化に要求される機能

動画符号化方式は、従来テレビ会議等の通信系システムを対象として検討されてきた。これに対して新たに記録系システムを対象とした符号化方式を検討する場合に考慮しなければならない事項として特殊再生機能がある。特殊再生機能は対象とするシステムのアプリケーション分野によって異なるが、その主なものは以下の通りである。

(1) ランダムアクセス

任意の動画データ位置から再生を開始する機能。ディスク系の記録媒体では重要な機能であり、この機能によって教育、カタログ、データベース、娯楽用など多様なアプリケーションが期待できる。

(2) 高速サーチ

高速に動画を再生する。何処にあるかが正確に解らないデータを一連の動画シーケンスの中から探し出すために用いる。サーチは時間方向に対して順方向と逆方向が行わなければならない。これも広範囲なアプリケーションで求められる機能である。

(3) ストップモーション

動画の再生をストップしその時再生していた画像をフリーズして表示する。使用者がシステムとインタラクティブにやり取りをする場合に合ったほうが好ましい機能。

(4) スロー再生

低速動画再生。教育、娯楽用に用いられる。

(5) 逆転再生

逆時間方向の再生。娯楽、ゲーム用に用いられる。

この他にも様々な機能が考えられるが、符号化アルゴリズムに直接影響を与えない、もしくは機能仕様がまだ不明確である、等の理由から上記の5つを挙げた。この5つの内、ランダムアクセス、サーチ、ストップの機能は殆どのアプリケーション

ンで必要とされる機能と思われ、必ず実現する必要がある。

3. フレーム間符号化方式の記録系への適用

3. 1. ハイブリッド符号化方式

現在の所、ディジタル記録媒体として有望なものとしてCD-ROMとDATなどがある。この内、ランダムアクセスが可能のため広範な応用が期待できるCD-ROMについて以下の議論を進めることにする。CD-ROMの最大読み出し速度は1.2Mbit/sec (150kbyte/sec)である。この程度の符号化レートで現在最も効率的な符号化方式とされているものにフレーム間予測符号化方式と離散コサイン変換符号化方式(DCT)を組合せたハイブリッド符号化方式がある。この方式は通信系システムの標準化委員会であるCCITTでは採用が決定されているもので符号化レートとして64K-2Mbit/secを対象としている[6]。

このハイブリッド符号化方式はフレーム間予測符号化方式を基本としており、非常に高効率な情報圧縮を可能にする。そこで本章ではこのハイブリッド符号化方式が記録系でも有効かどうかについて検討を行う。

図1にハイブリッド符号化方式のブロック図を示す。この方式はフレーム間の差分信号を用いる

ことで動画信号の持つ時間方向の冗長性を、DCTを用いることで空間方向の冗長性を利用して情報量圧縮をおこなうものである。DCT変換されたフレーム間差分信号は量子化され可変長符号化され記録される。従って記録される情報はフレーム間差分信号がDCT変換され量子化された量子化インデックス、量子化特性、フレーム内/フレーム間モード信号、動きベクトル信号などであり、これらの信号から画像が再生される。

3. 2. 特殊再生機能の実現方法

このハイブリッド符号化方式はフレーム間符号化を基本的に用いているため特殊機能を実現する上で解決しなくてはならない問題が生じる。それらを各機能毎に検討していく。

(1) ランダムアクセス

フレーム間符号化方式では、フレーム間の差分信号を用いて符号化を行っているため任意の場所から復号が出来ない。しかしながら画像データは通常、数秒から数10秒の長さを持つシーン(ショット)に分解することが出来る。ここでシーンとはふたつのシーンチェンジに挟まれた区間を言う。またシーンチェンジとは隣接するフレームに連続性が無く完全に異なる点を指す(図2(a))。異なるシーンの画像間には相関が存在しないから各シーンの先頭画像をフレーム内モードで行い、

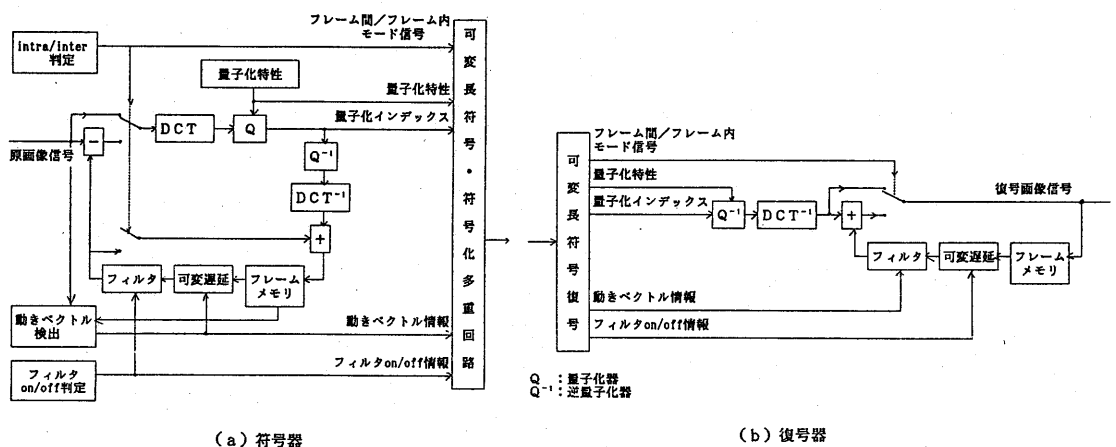


図1 ハイブリッド符号化方式ブロック図

そのデータアドレスをテーブル化していればランダムアクセスは可能になる。この場合ランダムアクセスはシーン単位でしか行われないが実用上なら問題は無いと思われる。

(2) 高速サーチ

ランダムアクセスの考え方を適用することで解決できる。すなわちシーンを更に分割し、例えば30フレーム毎あるいは100フレーム毎にフレーム内モードによる符号化を行う(図2(b))。これらのフレームのデータアドレスもテーブル化する。このアドレステーブルを参照しながら、これらのフレーム内モードのフレームだけを順次復号化することで高速サーチは実現される。時間に対して順方向にも逆方向にもサーチが可能となる。この方法は30フレームあるいは100フレームづつスキップして復号化するもので「見たい画像を素速く探索する」という高速サーチの目的を十分満足する。

(3) ストップモーション

ストップモーションを行うためには記録媒体からのデータ供給を停止し、同時に出力画面をそのまま維持するという二つの動作を行わなければならない。またストップモーションを解除し通常動作に戻る動作も円滑に行う必要がある。このストップモーションを行うための構成方法はいくつかあると思われるが、例えば読み込みバッファにおけるデータ制御でこれらの動作を全て行うことができる。すなわち記録媒体からの読み出しを停止し、復号器にはフレーム間モードを与え続け、量子化インデックス、動きベクトル信号をすべてゼ

ロとすればストップモーションが実現できる。この方式なら通常動作に戻ることも容易である。

(4) スロー再生

スロー再生はストップモーションを断続的に行うことで実現できる。すなわち1フレーム毎に1フレーム分のストップモーションを行えば1/2のスロー再生が可能になる。

(5) 逆転再生

ハイブリッド符号化方式はフレーム間符号化方式を基本としている。この方式では復号化にあたってフレーム間差分信号が順次加算されていくことで復号画像を得る。従ってフレーム間差分信号がフレーム単位に逆時間順に入力され、減算を行えば逆転再生されたことになる(図3)。ただ、ハイブリッド符号化方式ではフレーム間/フレーム内モードが適応的に用いられ、最終フレームの処理が必要なこと、動き補償とループ内フィルタが用いられていることから図3に示すブロック図よりは付加回路が必要になる。特にフレーム間/フレーム内適応符号化と動き補償フレーム間に関しては回路だけではなく逆転再生に当て必要な付加情報が必要となる。更にCD-ROMからの読み出しにおいてフレーム単位で逆転した状態にデータを並び替える仕組が必要となる。

以上により読み込みバッファに対してアドレステーブルを持つなどの改良を加えれば(5)を除いた4つの機能が実現できる。

4. 記録符号化による符号化効率の改善

4.1 改善項目

次に通信系の符号化方式では制約条件であった

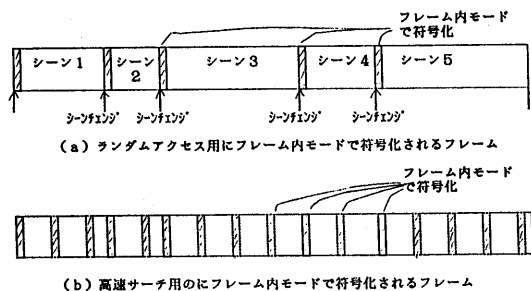


図2 符号化データ構成

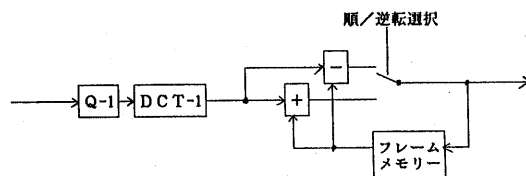


図3 逆転再生が可能な復号器

が記録系符号化方式では弛めることができ符号化効率改善につながる項目を整理する。

(1) 復号化遅延が許される

テレビ会議、テレビ電話などを対象とした従来の符号化・復号化方式では遅延は重大な品質劣化と考えられ、出来るだけ遅延を抑える必要があったがこの条件は記録系においては弛めることが出来る。

この事により復号器の読み込みバッファを拡大することが出来、インパルス的に発生する情報を吸収することが出来るため、急激な動きに対する画品質の劣化をある程度改善できる。

(2) 定レート符号化でなくて良い

従来の伝送系では回線交換系を前提としていたために符号化レートは一定にする必要があった。しかしながら記録系ではこの限りではない。すなわち読み込みバッファを用いて、データ読み込みを間欠的に行えば可変レートでデータが読み込める。もちろん、最大読み出し速度を越えることは出来ないがそれ以下の任意のレートで読み出すことが可能になる。

可変レート符号化方式の利点は記録媒体への効率的記録が可能になる点である。Judiceらの報告[7]によれば映画等で大きく情報が発生するのは大方10数%のシーンでしかなく、その他の部分でそれほど情報が発生する訳ではない。また最大発生情報量と平均発生情報量には大きく隔たりがある。もし最大発生情報量にあわせた定レート符号化を行うならば大量のダミーデータを用いねばならず効率的な記録の妨げになる。可変レート符号化を行えばそのようなダミーデータは必要無く効率的な符号化が可能である。

(3) 符号化をリアルタイムに行う必要が無い

これはCD-ROMなどの読み出し専用記録媒体についてのみ成立する。すなわち従来の動画符号化ではリアルタイムで処理されることが要求されていたが、これらの読み出し専用記録媒体においては必要条件ではなくなる。このため従来は出来なかった処理が可能になる。

リアルタイムの符号化という制約がはずされた

ことによって期待される符号化効率の改善としては二つ考えられる。ひとつは1フレーム単位の処理をより最適化(複雑化)すること。もうひとつは、画像データを先読みすることによって、複数フレームに対する符号化効率を改善することである。

以上の改善項目のなかから(3)のノンリアルタイム処理による改善を取りあげ、その一つとしての背景予測方式をつぎに提案する。

4. 2 ノンリアルタイム背景予測方式[8][9]

複数のフレームからなるシーン全体を符号化に先だって見渡したうえで変動の少ない画像部分から背景画像[10]を抽出し、この背景画像を予測符号化に用いることにより効率的な符号化が可能となる。

(1) 背景抽出方式

まず、フレームワークの固定されたシーンに対する背景画像抽出方式を述べる。画像は予め定められた大きさのブロックに分割しておき、このブロック単位で一連の処理を行う。

背景画像は、複数の画像フレームからなる1シーンにおいて変動の少ない部分から抽出する。まず各ブロック毎に複数フレームにわたって信号レベルの変動が抽出しきい値 Th より小さい期間(平坦期間と呼ぶ)を複数検出する(図4の①②③)。次に、他の平坦期間から信号レベルが大きくは

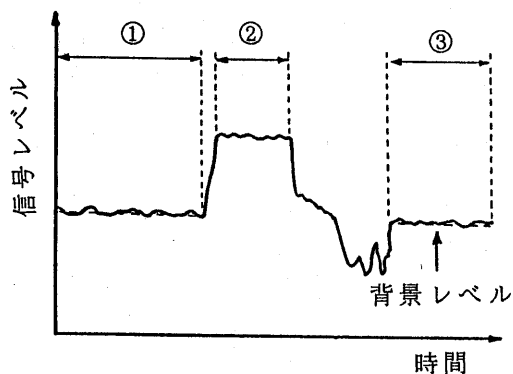


図4 信号レベルの平坦期間

ずれた平坦期間（図4では期間②）は排除して、残った平坦期間（背景抽出期間）での信号レベルの平均値として背景画像を生成する。

なお背景の抽出にあたり、しきい値 Th は小さいほど動領域を誤って背景部分として抽出するという誤抽出を防ぐことができる。しかし一般に、静止領域でもエッジ部では信号のフレーム間変動が大きいため、複雑な背景を抽出するためには、しきい値 Th もこれに応じて大きめに採る必要がある。

つぎに、背景全体が並行移動するパンニング画像における背景画像抽出方式を説明する。パンニング画像の場合にはそのパンニングの範囲をカバーする大きな背景画像を符号化に先だって抽出する。焦点距離の十分長いパンニングでは画像全体の動きをひとつの動きベクトルで表すことができる。そこで図5のように各フレーム毎にこの動きベクトルでパンニングの動きを補償することで、背景画像上の同一位置に相当する画素のフレーム内位置を決定することができる。このような補償を行うことにより、固定フレームワークと同様の方式で背景画像を抽出することができる。

なおパンニングの動きは画像各部分の動きから検出した。まず 8×8 画素からなるブロック毎に全探索ブロックマッチングで動きベクトルを検出する。そしてこのベクトルの最尤値をパンニングの動きベクトルとした。但し、マッチング誤差が大きいブロックは動領域と考え最尤値を求める際のカウントには含めない。またベクトルは抽出背景画像のボケ抑制のために小数精度 0.25pel で探索した。

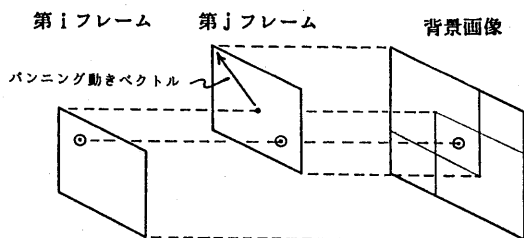


図5 パンニング動き補償

これにより動き補償位置が画素の間にくる場合には、その値を近傍4画素の値から内挿予測した。

(2) 背景予測方式

次に、上記の方法で抽出された背景画像を用いた予測符号化方式について説明する。予測方式として、動き補償フレーム間予測、フレーム内予測及び背景予測の3モードを適応的に選択する方式とした。ただし、背景予測を適用した場合にのみ予測差分値を全て零とし、背景画像に置き換える処理を行った。図6に符号化ブロック図を示す。前処理部においては、符号化に先だって背景画像を抽出し背景予測部に記憶させるとともに、符号化に同期させて現在符号化している画素が背景抽出期間内であるかどうかをモード情報として出力する。符号化においてはこのモード情報に従って、背景抽出期間内では背景予測が選択されるとともに、予測誤差信号は強制的にゼロレベルに抑えられる。また背景抽出期間外では動き補償フレーム間予測とフレーム内予測が適応的に選択されて用いられる。なおパンニング画像においては、背景予測ではパンニングの動きベクトルにより背景画像からの読みだし位置を操作して背景予測値を得る。

本方式では背景画像をシーン単位の初期情報として別途符号化する必要があるが、各フレームでの符号化効率の改善によりシーン全体の符号量は削減することができる。

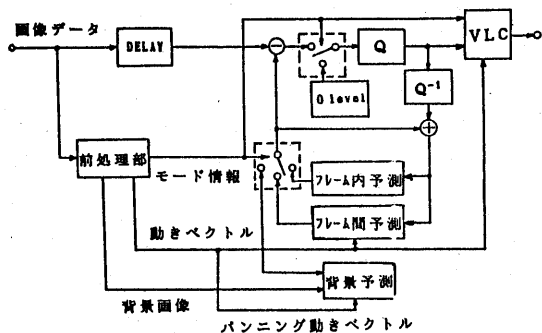


図6 背景予測符号化ブロック図

(3) シミュレーション結果

背景予測を用いた符号化方式の効果をシミュレーションにより確認した。固定フレームワーク画像として用いた画像はTV画像(会話シーン, 40フレーム)で、人物の動きが比較的小さくかつ背景の単純なものである。またパンニング画像として用いた画像は、ポスターなどの背景の前を人物が歩行するものである。符号化パラメータを表1に示す。本方式では背景予測の選択を符号化前処理により決定する予測モード検定型としたのでモード選択情報も符号化する必要がある。そこでモード情報の増加を防ぐためモード選択は8ライン×8画素のブロック単位で行った。また比較のため図6のブロック図から前処理及び背景予測を除いた従来方式(動き補償フレーム間、フレーム内の2モード適応予測)についても評価した。図7(a)に固定フレームワークにおける結果を、図7(b)にパンニングにおける結果を示す。符号化効率を表す指標として予測誤差信号のエントロピーを示している。この結果から背景予測により固定フレームワークで20~30%の効率改善が、パンニングにおいて30~40%の効率改善が可能であることがわかる。また背景予測の効果は復号画像上では背景部分のノイズ除去効果として現れ、張り付き等の問題は生じなかった。

以上の結果から固定フレームワークのみならずパンニング画像に対しても、ノンリアルタイム処理としての背景予測方式が符号化効率改善に効果を発揮することがわかった。

Image Size	384 pixels × 240 lines
One Sequence length	40 frames
Block Size for MC and Mode-select	8 pixels × 8 lines
Motion Detection Area	(±7, ±7) full search
Q dead zone threshold	2
Q step size	2 (linear quantization)

表1 符号化パラメータ

5. まとめ

記録系動画像符号化方式に関して検討し、従来通信系システムで用いられているフレーム間予測を基本としたハイブリッド符号化方式が記録系に対しても有効であることを示した。また通信系と比較した場合の記録系符号化における符号化効率改善の一つとしてノンリアルタイム背景予測符号化を提案し、その効果を示した。

参考文献

- [1] N. V. Philips and Sony, "Compact Disc Read Only Memory System Description," (Feb, 1985)
- [2] 日本電子工業会規格CP-340, 「デジタル・オーディオ・インタフェース」 (Nov. 1987)
- [3] ISO-IEC/JTC1/SC2/WG8 Coded Representation of Picture and Audio Information, Documents of Moving Picture Experts Group
- [4] CCIR Rec. 601
- [5] R. C. Nicol and N. Mukawa, "Motion Video Coding in CCITT SG XV - The Coded Picture Format," proc. of IEEE GLOBCOM'88, 31-1, (Nov. 1988)
- [6] R. Plompen, Y. Hatori, W. Geuen, J. Guichard, M. Guglielmo and H. Brusewitz, "Motion Video Coding in CCITT SG XV - The Video Source Coding," proc. of IEEE GLOBCOM'88, 31-2, (Nov. 1988)

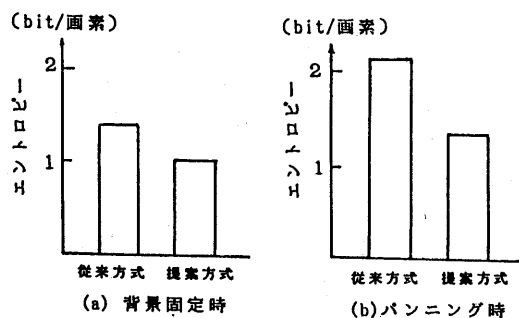


図7 符号化効率の改善

[7]C. N. Judice and M. J. Jaquez, "High Compression coding of Entertainment Video for Packet Networks,"proc. of IEEE COMSOC International Workshop, p. 4-5, (Nov. 1987)

[8]宮本他「背景予測を用いた適応符号化方式」画像符号化シンポジウム' 88、昭和63年9月, p. 93-94

[9]宮本他「パンニング画像における背景予測の検討」信学全大、1988年3月

[10]黒田他「動き補償・背景予測を用いたフレーム間予測方式」信学論誌'85/1, Vol. J68-B NO. 1, p. 77-84