

動画のリアルタイムフォーマット変換

ーモバイルマルチメディア通信の用途を広げる動画処理技術ー

上野山 努 松下電器産業 (株) マルチメディアシステム研究所 uenoyama@tri.mei.co.jp

行武 剛 松下通信工業 (株) ワイヤレスソリューション研究所 Takeshi.Yukitake@yrp.mci.mei.co.jp

現在、テレビ放送、DVD、テレビ電話など、多くの用途で身近にデジタル映像が利用されている。また、動画通信の可能なモバイル端末もまもなく実用化される予定である。一方ネットワークのシームレス化が進み、これらの映像情報を相互に利用したいという要求が高まっている。この実現のためには、動画の圧縮フォーマットを異なるフォーマットに変換する、フォーマット変換が必要になる。本稿では、モバイルテレビ電話端末向けの技術を中心に、動画の圧縮フォーマットと変換技術について概説する。

るため、さまざまな用途が考えられる。その中には、出張先からTV会議に参加する、といった用途もあるだろう。この場合、従来の狭帯域ISDNのTV会議端末とモバイルマルチメディア端末とで動画画像を含めた相互通信が必要となる。イメージを図-1に示す。

従来のTV会議端末と次世代のモバイルマルチメディア端末とでは、採用されている動画画像の圧縮フォーマットが異なる。既存のTV会議端末にはH.261が用いられているし、第3世代移動体通信ではMPEG4/H.263が標準フォーマットとなる。そのため、相互通信を可能にするためには、動画画像の圧縮フォーマットを、それぞれ互いの圧縮フォ

ーマットに変換する必要がある。そこで用いられるのが、トランスコーディング技術である。

本稿では、既存のTV会議端末とモバイルマルチメディア端末との相互接続に必要となる、H.261-MPEG4間のトランスコーディングを中心に、動画画像の圧縮フォーマットと、その変換技術について概説する。

動画画像の圧縮フォーマット

●圧縮の基本アルゴリズム

後のトランスコーディングの説明が必要となる、DCT係数、量子化特性値、動き補償を中心に、動画画像圧縮の基本アルゴリズムについて説明する。

「ケータイ」でも動画通信の時代に

2001年、日本で第3世代移動体通信のサービスが開始される。第3世代移動体通信では伝送帯域が拡大するため、その高速性を活かしたさまざまな利用が予定されている。音声に加え、動画画像もやりとりできる「携帯TV電話」、ビジュアルホンもその1つである。ビジュアルホンなどのモバイルマルチメディア端末は、どこでも動画画像通信が可能にな

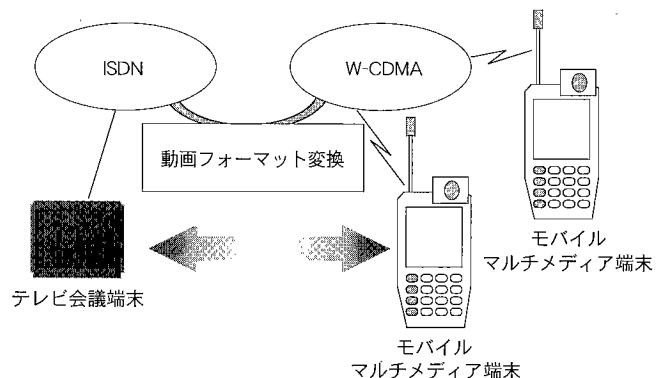


図-1 モバイル端末とISDN端末間のビジュアル通信

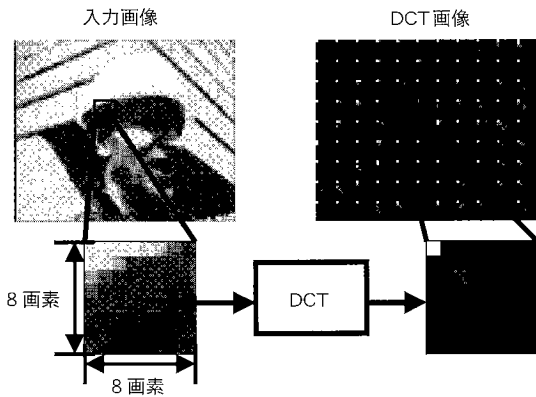


図-2 フレーム内DCT処理の例

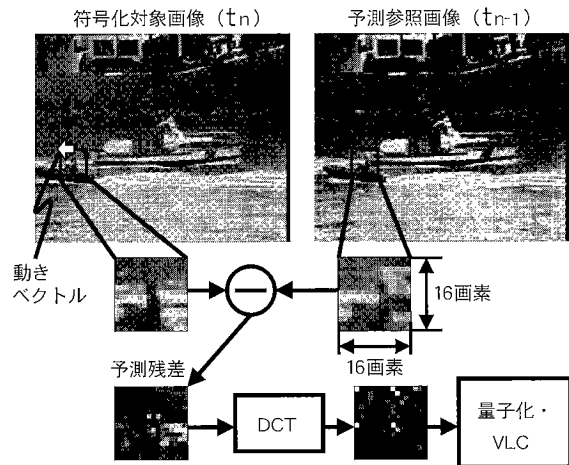


図-3 フレーム間圧縮の概念図

H.261, MPEG4を含む一般的な圧縮フォーマットの多くは、フレーム内圧縮とフレーム間圧縮の組合せで高能率な圧縮を実現している。

フレーム内圧縮は、動画像の空間方向の冗長性を利用して符号量を削減する。そのために、DCT (Discrete Cosine Transform : 離散コサイン変換)、量子化、可変長符号化の組合せを用いる。このフレーム内圧縮の過程について、図を用いて説明する。

最初に入力フレームに対し、DCTを施す。図-2は1枚のフレームに実際にDCTを施した例である。

フレーム内DCT処理を行うには、入力画像(図-2左上)を8画素×8画素のブロックに区切る(図-2左下)。そしてブロックごとに8×8の2次元DCTを施し、周波数領域に展開する。DCTの結果はDCT係数(図-2右下)になる。DCT係数は、左から右に向かって横方向の周波数の低い順に、上から下に向かって縦方向の周波数の低い順に並べられている。

一般に自然画像では、縦・横とも低い周波数成分(ブロック左上)が大きくなる。これは自然画像の画素値(輝度、色差)が空間上で滑らかに変化することが多いためである。そのため、DCTにより、低周波に情報を偏らせ、高周波の係数を0かそれに近い値にすることができる。

次に、この係数(図-2のDCT画像

の画素値)を量子化する。量子化は、DCT係数値を量子化ステップサイズで割る処理である。量子化ステップサイズは量子化の精度を示す値であり、その範囲などは圧縮フォーマットにより異なるが、2から64程度の整数値が一般的に用いられる。後で説明するH.261やMPEG4では、量子化ステップサイズは直接符号化されず、量子化特性値と呼ばれる値に変換して符号化され、より符号量が削減されている。量子化により係数値を小さくするとともに、値が0の係数を増やすことができる。

そして最後に、量子化された係数を可変長符号化する。この可変長符号化では、小さい値ほど少ないビット量の符号が割り当てられる。特に、値が0の係数は、何個連続しているかという個数(ランレングス)が符号化されるため、特に大きく圧縮される。

フレーム内、フレーム間を問わず、このような量子化を伴う圧縮では、量子化特性値(圧縮率)と画質の間にトレードオフの関係がある。量子化特性値の値を大きくすると、小さな係数値や値が0の係数が増えるため、符号量は少なくできるが、量子化誤差が増え、画質は悪くなるためである。

フレーム間圧縮は、動画像の時間方向の冗長性を利用して符号量を削減する。この方法は動き補償と呼ばれる。図-3に、時間的に連続するフ

レームをフレーム間圧縮する過程を示す。図-3のように、連続したフレームでは、内容が多少動いたり、変化したりしているが、似通った画像である場合が多い。図-3において時刻 t_n の画像をフレーム間圧縮する場合、1フレーム手前(時刻 t_{n-1})のフレームを参照画像とし、参照画像との「違い」を符号化する。この「違い」は、動きベクトルと、予測残差という2種類の情報として表される。

具体的な処理を説明する。対象フレームを縦、横16画素のマクロブロック(MB)の単位に分割し、各MBに近い画像を持つ領域を参照画像から見つける。対象MBに対する、見つけた領域の相対位置が動きベクトルとなる。図-3では、線で囲んだ領域ではボートが右から左に水平に6画素程度動いている。そのため参照画像の中でこのMBに近い画像領域は(-6, 0)の位置となり、この(-6, 0)が動きベクトルとなる。予測残差は、対象画像の画素値と参照画像の画素値の差分情報であり、MBごとに求められる。図-3では対象画像の線で囲んだMBと参照画像の線で囲んだ領域との差分が予測残差となり、図はこの絶対値を示している(見やすくするため、値は強調している)。

この予測残差に対し、フレーム内圧縮と同様に、8画素×8画素単位にDCT、量子化を施す。この結果、

フレーム間圧縮された情報は、動きベクトルと量子化・DCTされた予測残差の2つの情報として符号化される。また、フレーム間圧縮でも予測残差は可変長符号化されるため、予測残差が小さいほど符号量は少なくてすむ。別の言い方をすれば、同じ符号量で符号化する場合、元の係数値が小さいほど量子化特性値を小さくできるため、画質が良くなる。フレーム間圧縮において差分の係数値を小さくするためには、より対象マクロブロックに近い画像領域を、参照画像より見つける必要がある。

後述するH.261やMPEG4でもこのフレーム内圧縮、フレーム間圧縮の組合せが用いられる。そして、対象の画像をフレーム内圧縮するか、フレーム間圧縮するかはマクロブロックごとに選択することが可能である。本稿では、フレーム内圧縮されたマクロブロックをイントラMB、フレーム間圧縮されたマクロブロックをインターMBと呼ぶ。

●いろいろな動画圧縮フォーマット

前で触れたように、現在身近にあるデジタル映像には、いくつかの圧縮フォーマットが採用されている。これらの圧縮フォーマットは大別すると、国際規格として標準化されたフォーマットと、企業等が独自に開発したフォーマットに分けられる。

主要な動画圧縮フォーマットの国際規格の一覧を表-1に示す。

H.261/263は、主に有線系のTV電話やTV会議での利用を想定して規格化された。最初H.261が、主に狭帯域ISDN回線での動画通信のために、1990年に規格化された¹⁾。そして、その後の動画圧縮技術の進展を受けて、より高効率な圧縮と多様な映像への対応のため、1995年にH.263が規格化された。H.263はH.261の圧縮原理に基づいた動画圧縮方式

であるが、より多くの画像解像度に対応したり、高能率圧縮のためのオプションを多く備える点が、H.261と異なる。H.263はMPEG4の基本の動画圧縮方式と非常に近い圧縮方式である。

現在H.261は有線TV会議システムに採用されている。H.263は有線TV会議システムやPCのデスクトップTV会議システム採用されており、さらにMPEG4同様、第3世代移動体通信の動画通信にも使用することができる。

一方MPEGはMoving Picture Experts Groupの略で、標準化団体ISO/IEC JTC1/SC29の通称であり、その団体により標準化された動画などマルチメディア情報の符号化フォーマットの通称でもある。最初の規格であるMPEG1はCDに動画を蓄積することを目的として、1992年に規格化された。そのため、ビットレートはH.261/263より高い1.5Mbps程度をターゲットとしている。規格化にあたってはH.261が、圧縮方式のみならず標準化手法の面でも参考にされた。現在このMPEG1は、VideoCDやPCアプリケーション向けの動画フォーマットとして利用されている。

MPEG2はMPEG1を基に、より高画質な動画の蓄積や伝送を目的として、1994年に規格化された。MPEG2では現行のTV放送品質の動画や、高品位テレビ(HDTV)品質の動画を圧縮するため、現行TV放送で採用されているインタレース画像を効率よく圧縮するためのオプションなど、高画質圧縮のためのオプションを多く備えている。ビットレートはおおむね3Mbps以上をターゲットとしている。このMPEG2は、DVD Videoやデジタル衛星放送に採用されている。その後、低ビットレートでの圧縮を目的として、MPEG4が規格化された²⁾。

独自に開発したフォーマットには多くの種類があるが、RealSystemや

VDO Liveなど、PCを対象にしたインターネットストリーミングサービスに多く用いられている。これらは前者と同様のDCTと動き予測を用いたアルゴリズムによるものもあるが、フラクタル圧縮やウェーブレット変換など、それぞれの用途に適した圧縮アルゴリズムを採用しているフォーマットもある。

●有線TV電話の標準フォーマット、H.261

現在の狭帯域ISDN用TV会議端末などは、マルチメディア通信のシステムとしてITU-Tの勧告H.320に準拠している。H.320準拠の端末では、動画圧縮フォーマットとして必ずH.261に対応することが勧告されている。H.261は、INSネット64のような、 $p \times 64\text{Kbps}$ (p は整数)の伝送路で動画通信を行うことを目的としている。対象とする符号化ビットレートはおおよそ40Kbps~2Mbpsである。画像のサイズは、QCIF (Quarter Common Intermediate Format) と呼ばれる、横176画素、縦144画素の画像を扱えることが必須となっており、さらにオプションでCIF (Common Intermediate Format : 352 × 288画素)のサイズも利用できる。

H.261では、フレーム内圧縮、フレーム間圧縮の組合せで動画圧縮を行う。フレーム間圧縮では、予測残差、特に高周波成分のDCT係数を低減させるために、ループ内フィルタを用いることができる。ループ内フィルタは参照画像に対して施すローパスフィルタで、参照画像の高周波成分を低減させる。

●モバイル動画通信の標準フォーマット、MPEG4

前述したように第3世代移動体通信網では、マルチメディア通信が一般的になる。これらマルチメディア端末の動画圧縮フォーマットの標準として、MPEG4が採用される。

名称 (通称)	標準化時期	標準化団体	ビットレート	主な用途
H.261	1990	ITU-T	40K~2Mbps	有線テレビ会議
MPEG1	1992	ISO/IEC (WG29)	1~1.5Mbps	VideoCD
MPEG2	1994	ISO/IEC (WG29)	3~15Mbps	DVD, テレビ放送
H.263	1995	ITU-T	16K~1Mbps	有線/PCテレビ会議等
MPEG4	1999	ISO/IEC (WG29)	16K~1Mbps	次世代モバイル通信等

表-1 主な動画圧縮フォーマット

MPEG4は、マルチメディア高能率符号化のための幅広い規格である。MPEG4の基本部分（バージョン1と呼ばれる）は1999年3月に国際規格としての最終承認投票が行われ、1999年12月に国際規格（IS: International Standard）としてリリースされた。この中の画像符号化方式（MPEG4 Visual）は、従来の矩形自然動画像だけでなく、任意形状画像オブジェクト、顔アニメーションオブジェクト、メッシュオブジェクトなど、さまざまな画像符号化方式が規格化されている。さらに、従来の矩形自然動画像についても、多くの技術要素をツールとして用意しており、さまざまな目的に対する圧縮を可能にしている。

そのため、MPEG4では、必要となる技術要素と性能とを、プロファイルとレベルという指標で分類し、用途に合った適用と互換性の確保を実現している。プロファイルでは、シェイプ・コーディング（任意形状の符号化）や空間スケーラビリティ符号化（複数の解像度でのデコードが可能な符号化）など、使用可能な要素技術を定義している。レベルでは使用可能な画像解像度やフレームレートなど符号・復号に必要な処理能力を定義している。プロファイルとレベルによる分類は、MPEG2でも利用されていたが、MPEG4ではその幅広さから、さらに種類が多くなっている。

MPEG4の矩形自然動画像も、基本的にはH.261と同様にDCTによるフ

レーム内圧縮とフレーム間圧縮の組合せで圧縮されている。しかし、動き予測では、MPEG1/2と共通して半画素精度動き予測というアルゴリズムが採用されている。これは、参照画像の隣接する画素の平均値を画素と画素の中間の位置である「半画素」の画素値として、動き予測をこの精度で行うものである。これにより圧縮効率が向上するため、H.261で採用されたループ内フィルタは、MPEGでは採用されていない。その他にもMPEG4では、多様な高能率圧縮のための技術を利用することが可能である。

第3世代移動体通信のマルチメディア通信で用いる最も基本的な動画像圧縮フォーマットは、MPEG4 SP@L1（Simple Profile@Level 1）をさらに限定したものとなる。主な仕様を表-2に示す。

このフォーマットは3GPP（Third Generation Partnership Project）によって決定されている³⁾。3GPPは第3世代移動体通信に関連する標準化を行っている日欧中心のプロジェクトである^{☆1}。

H.261-MPEG4間のフォーマット変換

初めに述べたように、ISDNのTV会議端末とモバイルマルチメディア端末とをつなぐには、図-1のようにH.261-MPEG4間のフォーマット変換が必要となる。

具体例として、我々が試作したト

圧縮規格	MPEG4 Visual SP@L1
画像サイズ	縦144画素以下、横176画素以下の矩形（縦、横とも16画素単位）
フレームレート	15fps以下
ビットレート	64Kbps以下
その他	f_code=1（固定）、 intra_dc_thr=0（固定）など

表-2 モバイル通信向け動画フォーマットの主な仕様

項目	内容
画像サイズ	QCIF
フレームレート	15fps以下
入力圧縮フォーマット	H.261, MPEG4
出力圧縮フォーマット	MPEG4, H.261
入出力ビットレート	48Kbps

表-3 トランスコード基本仕様

ランスコードの基本仕様を表-3に示す。

このトランスコードは前述の用途への応用を目標にするため、基本仕様に加え次のような性能が必要となる。

- 双方向性：H.261→MPEG4, MPEG4→H.261両方向の同時変換が必要。
- リアルタイム性・低遅延性：リアルタイム、かつ低遅延での変換が望まれる。具体的には、会話を阻害しない200msec以内の遅延での変換が望まれる。
- 画質劣化が少ない：後述するように、H.261-MPEG4間の変換では、必ず画質の劣化が生じるが、劣化なるべく少ない変換が望まれる。

●簡単なフォーマット変換

最も簡単にフォーマットを変換するには、図-4のように入力ストリームをデコードし、変換先のフォーマットで再度符号化する。

この方法は、入出力の圧縮フォーマットによらず自由な変換ができる長所があるが、一方で再圧縮時の画質劣化が大きくなるという短所がある。

具体例としては、エンコーダのレ

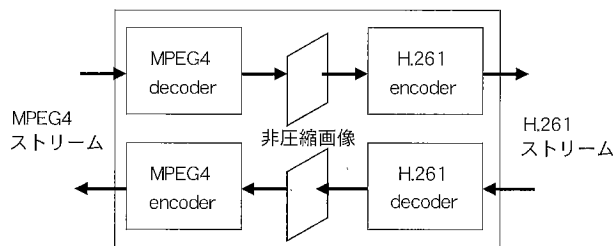


図-4 簡単なフォーマット変換器

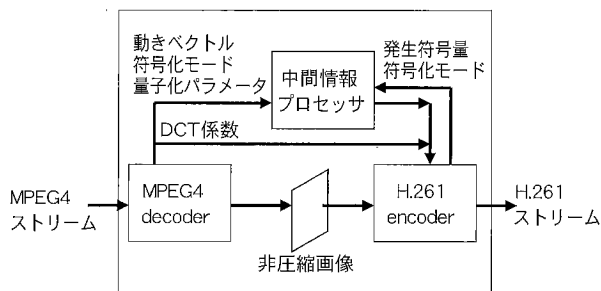


図-5 トランスコードのブロック図

復号時情報	利用方法
動きベクトル	動き検索の基点に使用
量子化パラメータ	量子化パラメータの下限値
符号化モード	符号化モードの候補
イントラDCT係数	直接再利用

表-4 再利用する符号化情報

項目	内容
主要コンポーネント	メインDSP, サブDSP (デュアルプロセッサ), コアDCT, 2次元アドレス発生器, 半画素演算器
アーキテクチャ	ベクトルパイプラインアーキテクチャ
インタフェース	画像I/F (NTSC), SDRAM, 16bitCPUバス
内部動作周波数	67.5MHz

表-5 実装対象のDSPの基本仕様

ート制御の問題が挙げられる。レート制御は、出力ストリームの目標ビットレートに合わせて、各MBにどれくらいのビットを割り当てて符号化するかを決定し、量子化ステップサイズを決定する処理である。ここで例として挙げているフォーマット変換の入力には、すでに高い圧縮率で圧縮された、つまり画質が劣化した映像が入力される。このような映像に含まれる成分は、その映像を見る人にとって2種類の意味を持つ。1つは圧縮前の映像に含まれる、必要な成分であり、もう1つは入力ストリームの圧縮時に発生したノイズ、つまり不要な成分である。図-4のフォーマット変換では入力映像を完全にデコードしてしまうため、再符号化時にはこの2種類の成分を区別せず、同様に符号化してしまう。結果、ノイズのような成分の符号化に無駄にビットが消費され、全体の画質劣化を増大させる。

●トランスコーディング

前記のような問題を解決するために、入力ストリームの符号化情報を使って出力ストリームの符号化を行う方法が考えられている。これをトランスコーディングと呼ぶ。「トランスコード」は直訳すれば「符号の

変換」といった意味になるが、動画のトランスコードは、入力ストリームの符号を直接変換するだけでなく、入力ストリームの符号化情報を利用した再符号化全般に、広く使われることが多い。たとえば、MPEG2ストリームを再圧縮してビットレートを変えるとといった、同一フォーマットによる再符号化処理も、トランスコードと呼ばれる^{4), 5)}。

H.261-MPEG4間のトランスコーディングは次のような固有の問題があり、直接出力ストリームの符号として利用できる符号化情報は少ない。

- 量子化方法の違い：H.261とMPEG4では量子化特性値から実際の量子化に用いる量子化ステップサイズに変換する方法が異なる。そのため、H.261とMPEG4の量子化特性値は、1対1で対応しない。
- 動き補償方法の違い：H.261は動きベクトルが1画素単位で、ループ内フィルタをかけることができるが、MPEG4は動きベクトルが半画素単位で、ループ内フィルタは使えない。そのため、動きベクトルなどの動き補償の情報が1対1で対応しない。
- 圧縮効率の違い：MPEG4では、半画素精度の動き予測や可変長符号の改良により、実際の圧縮効率が

H.261に比べて1割程度よい。そのため、仮にすべての情報を1対1に対応させて変換しても、入出力のビットレートが異なる。

このため試作したトランスコードでは、利用する符号化情報の多くは直接再利用せず、再符号化の制御情報として利用している。表-4にその内容を示す。

試作したトランスコードのブロック図を図-5に示す（簡単のため、MPEG4→H.261側のみ示す）。

このトランスコードの処理の一例として、これらの情報を使った画質向上のための処理について説明する。

この画質向上処理のポイントは次の2点である。

- 無駄のないビット割り当て（レート制御）
 - イントラMB情報の再利用
- 無駄のないビット割り当ては、前述したノイズに対し多くのビットを割り当ててしまう問題を解決するものである。この処理には入力ストリームの量子化特性値を利用する。トランスコード時のレート制御については、MPEGストリームを再圧縮する際の再量子化による画質劣化に関する研究がある⁴⁾。この研究報告では、入力に対する出力の量子化特性

値の倍率が1~2倍の間で画質が劣化することが指摘されている。これは出力ビットレートを入力よりも低くする場合により多く起こる問題である。今回のトランスコーダでは、入力と出力のビットレートが同じであり、さらにMPEG2に比べ圧縮率が高い設定のため、この問題に加えて、入力よりも小さい量子化特性値を割り当てるといった問題もある。

大きな量子化特性値で量子化された映像は、より多くノイズを含む。つまり、入力の量子化特性値が大きいマクロブロックに対して小さな量子化特性値(=多くのビット)を割り当てても無駄になる。そのため、エンコーダ内の通常のレート制御方法により求めた量子化特性値を入力量の量子化特性値と比較し、値の大きい方を実際の量子化特性値として採用する。これにより次のMBに使えるビット量をより多く確保でき、割り当てるべき領域により多くのビットを割り当てて画質の向上を図ることができる。

イントラMB情報の再利用では、イントラとして再符号化するMBを決定する方法が画質向上のポイントとなる。インターMBのDCT係数は動き補償の方式が異なるため再利用できないが、イントラMBのDCT係数はそのまま再利用することで、本来必要な逆DCT、DCTの繰り返しによる演算誤差の累積を防ぐことができる。そのため、入力ストリーム中の符号化モードをみて、入力ストリーム中でイントラ符号化されているMBは直接DCT係数を再利用して再度イントラ符号化すれば画質劣化を防ぐことができる。しかしながら、MPEG4→H.261のトランスコードを考えた場合、

- MPEG4ストリームは、無線伝送時のエラー耐性向上のため、イントラMBの割合が多い(H.261は符号化後のエラー対策があり、符号化時にイントラMBを増やす必要が

ない)。

- H.261はMPEG4に比べて圧縮効率が低い。

という特徴があり、入力のイントラMBすべてをイントラで再符号化することができない。そのため、各マクロブロックを過去どちらのモードで符号化したかの履歴を持っており、インターMBでの符号化がより長く続いたMBを優先的にイントラMBとして残すようにしている。

DSPへの実装

このトランスコーダの現実での利用を考えると、DSPへの実装によるハードウェア化は必須である。今回は表-5の仕様を持ったDSPへの実装を行った。このDSPはベクトルパイプラインアーキテクチャを採用した画像処理用のDSPで、動画像圧縮伸張用の専用ハードウェアを搭載している。

実装の手順としては、最初にエンコーダとデコーダの組合せによる単純なフォーマット変換器を実装して基本動作を確認し、順次復号時情報の再利用アルゴリズムを実装していく手順をとった。現在、単純なフォーマット変換器を実現し、入力の量子化特性値の再利用アルゴリズムを最初に実装した時点である。

実装対象のDSPは動画像処理用であるので、トランスコーダのアルゴリズムの実装は、基本的に効率よく実現できる。しかしながら、トランスコーディングではデコード時の情報を利用するため、実装に工夫を要する部分もある。たとえば、この後実装を計画している、入力ストリーム中のイントラMBのDCT係数を保存する方法である。メモリ管理の点から、この情報はデコードされた非圧縮画像に置き換わる形で画像メモリ内に保存する必要がある。しかし、DCT係数は値の偏りが大きく、1係数あたり16bitの記憶領域が必要であ

るが、画像メモリは8bit単位のアクセスに固定されている。そのため、トランスコーダでは入力ストリームの逆量子化前のDCT係数を画像メモリに保存し、再量子化が必要な部分の量子化特性値をデータメモリに保存することで、DCT係数の保存を可能にした。

これから広がるトランスコーディング

本稿ではH.261-MPEG4間のトランスコーディングを中心に説明した。このトランスコーディング技術はさらに改良を進めて、次世代のモバイルマルチメディア通信の用途を広げる技術として、実用化されることが期待されている。ただし、動画像のトランスコーディング技術は、当然ながらTV会議のためだけの技術ではない。最初に述べたように、現在、デジタル衛星放送、VideoCD、DVD Video、インターネット映像放送など、さまざま場面でデジタル動画像が利用されている。今後これらの機器のネットワーク化、IP化が進展すれば、機器間での映像のやりとりや機器のインフラをまたいだ映像の再利用などが増えていくことが予想される。このような場面では動画像のフォーマット変換は不可欠のものとなり、それぞれの用途やフォーマットに応じたトランスコーディング技術も、さらに開発されていくものと期待している。

参考文献

- 1) ITU-T Rec. H.261: Video Codec for Audiovisual Services at p×64 kbit/s, ITU-T (1990).
- 2) ISO/IEC 14496-2: Information Technology - Coding of Audio-visual Objects - Part2: Visual, ISO/IEC (1999).
- 3) Third Generation Partnership Project TR26.911: Codec for Circuit Switched Multimedia Telephony Service; Terminal Implementor's Guide V3.2.0, 3GPP (1999).
- 4) 角野真也, 柴藤 稔, 横矢直和: MPEG2再符号化における再量子化誤差の検討, 信学技報 IE99-32, pp.49-56 (July 1999).
- 5) Acharya, S. and Smith, B.: Compressed Domain Transcoding of MPEG, Proc. IEEE International Conference on Multimedia Computing and Systems, pp.295-304 (1998).

(平成12年3月3日受付)