

インターネット数理科学第10回

～ネットワークのこちら側を支える数理科学その3～

2006年12月14日

株式会社インターネット総合研究所代表取締役所長
東京大学大学院数理科学研究科客員教授

藤原 洋

1. ネットワークのこちら側を支える数理科学とは？
2. 動画像圧縮の原理のまとめ
3. 原型となった通信標準H. 261
4. 静止画標準JPEG
5. 蓄積動画標準MPEG1
6. 放送標準MPEG2
7. その他の標準方式

1. ネットワークこちら側を支える数理科学とは？

③(ネットワークの)あちら側

⇒「グラフ理論」「金融工学理論」に基づくデータベース、検索エンジン最適化、検索連動データベース、ネット金融サービス

①ネットワークそのもの

⇒「グラフ理論」による動的ルーティング、帯域制御、放送型ルーティング
「デジタル信号処理理論」に基づく変復調技術

②(ネットワークの)こちら側

⇒「デジタル信号処理理論」に基づくコンテンツ符号化技術
⇒「HTML」に基づくブラウザ技術

以下の3つの分野にわたって①②③⇒①②③⇒・・・順に

③ネットワークのあちら側を支える数理科学

⇒「グラフ理論」「金融工学理論」に基づくデータベース、検索エンジン最適化、検索連動データベース、ネット金融サービス

①ネットワークそのものを支える数理科学

⇒「グラフ理論」による動的ルーティング、帯域制御、放送型ルーティング
「デジタル信号処理理論」に基づく変復調技術

②ネットワークのこちら側を支える数理科学

⇒「デジタル信号処理理論」に基づくコンテンツ符号化技術

③(ネットワークの)あちら側

Web1.0(ポータル)⇒Web1.5(SNS)⇒Web2.0(ロングテール)

①ネットワークそのもの

ダイヤルアップ/2Gモバイル⇒ブロードバンド/3Gモバイル⇒IP放送/NGN/WiMAX

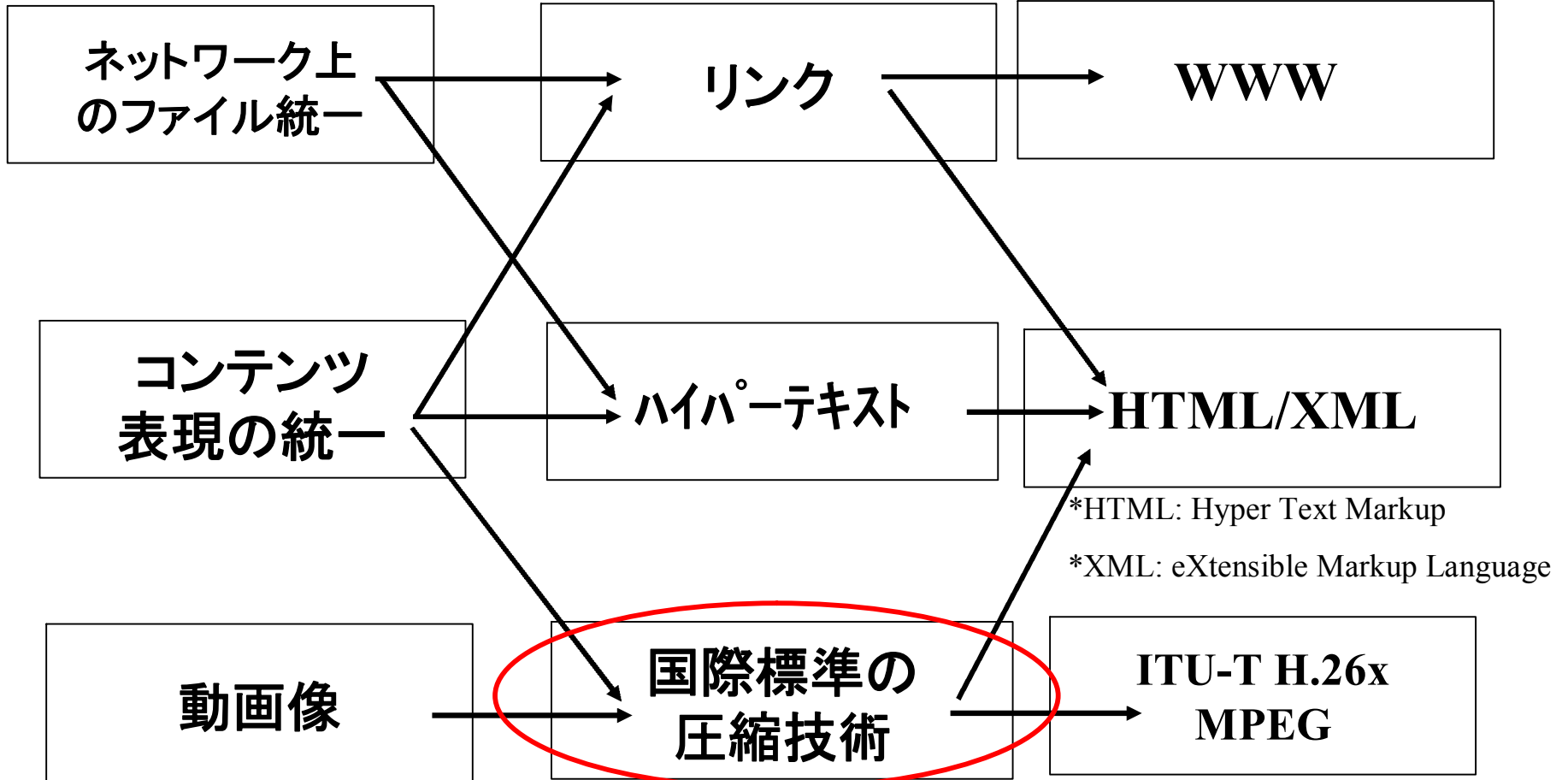
②(ネットワークの)こちら側

文字情報(Eメール)⇒HTML(ブラウザ)⇒動画(デジタル符号変換)

課題

着眼点

具体策



*HTML: Hyper Text Markup
*XML: eXtensible Markup Language

*ITU-T: International Telecommunications Union-Telecommunication Sector

*MPEG: Moving Picture Experts Group

「通信と放送の融合」というキーワードは、いつも通信業界から発せられてきた。旧電電公社時代の三鷹でのINS実験、狭帯域N-ISDNによる**TV電話の発展形としての動画像サービス**、ATM交換に基づくB-ISDNによるVoD、そして今話題となっているのが「ブロードバンド登場後のIP放送」。

通信ネットワークを用いた動画像サービスは、まさに4度目の挑戦となる。これまでの3回は、失敗だったが、今度はモノになるだろうか？この動向を展望する上で、現段階で最も積極的にサービスを展開している、「GyaO」と「TV-Bank」に目を向ける。

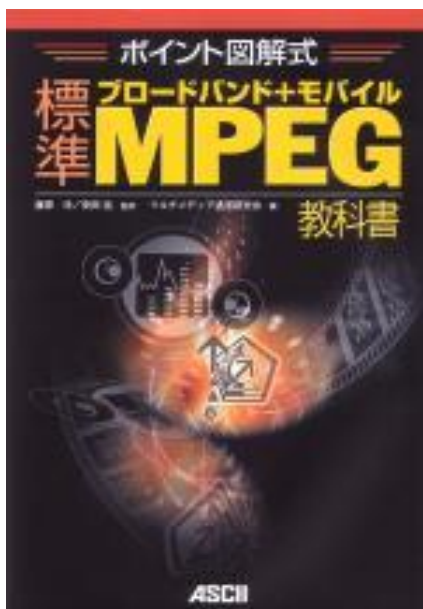
現行のTV放送を補完する2つのサービス

GyaOは、約1000タイトルのVoDとリアルタイム配信の組み合わせであり、TV-Bankは、他のサイトへのリンクを含めて約10万タイトルのVoDとリアルタイム配信の組み合わせです。GyaOが、編成と自主制作に力点を置いた擬似TV放送型であるのに対して、TV-Bankは、コンテンツ収集力に力点を置いたWeb1.0検索型であるといえる。また、ビジネスモデルとしては、GyaOが無料広告モデルであるのに対して、TV-Bankは、無料と有料の組み合わせモデルとなっています。どちらも現行のTV放送のテイストを残しつつも、多くの選択肢のあるVoDブロードバンド・インターネットならではの特長を活かしたサービスとなっている。

実際に、財団法人インターネット協会監修の2005年版『インターネット白書』のインターネット利用者の映像配信に対する利用動向調査によると、明らかに現行TV放送とは異なる、あるいはそれを補完するサービスのニーズが高いことがわかる。

以上のことから、GyaOとTV-BankというITベンチャーによる**インターネットを用いた動画像サービス**は、現行放送と競合するサービスではなく、補完するサービスであるといえる。また、両者合わせて1000万人程度のアクティブユーザーを獲得しているとみられ、事業的にも損益分岐点に到達するのは時間の問題という状況にあり、4度目の挑戦は、初めての成功が近いと思われるが、NHK、民放など、現行放送の約3兆円市場と比較すると、IP放送市場がその1%まで到達するには相当の時間がかかるだろう。

2. 動画像圧縮の原理のまとめ



ポイント図解式 ブロードバンド+モバイル 標準MPEG教科書

藤原洋/安田浩 監修、マルチメディア通信研究会 編

シリーズ:ポイント図解式

定価:本体3,800円+税

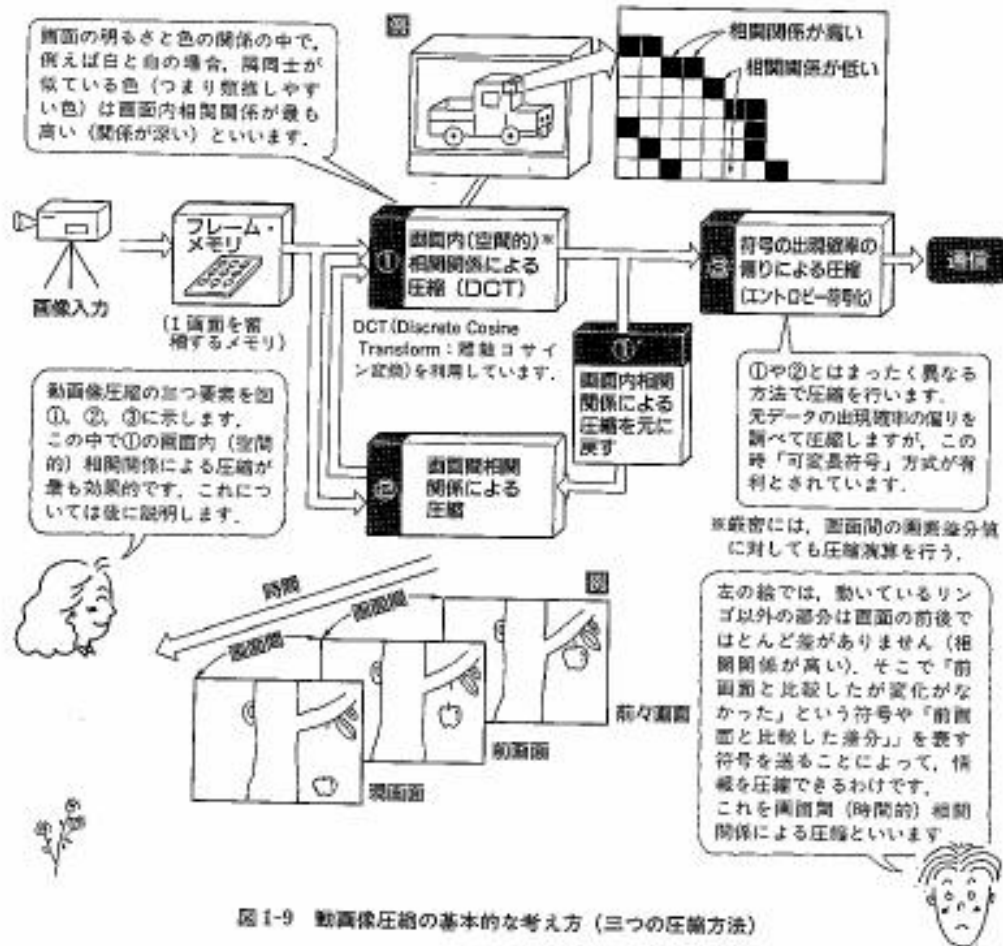
発売日:2003/02/10

形態:B5 (424ページ)

ISBN:4-7561-4241-9

ポイント

1 情報圧縮の三大要素 2 画像情報圧縮に最も効果的なのは DCT である



ポイント

- 1 画像にも精細さの波がある 2 自然画の絵柄は、画面上相関が少し高い

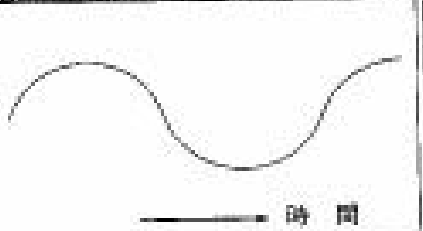
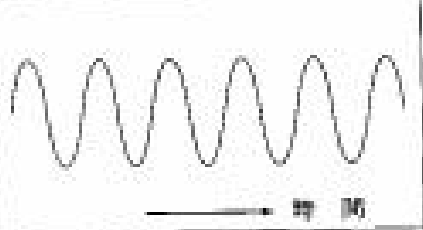
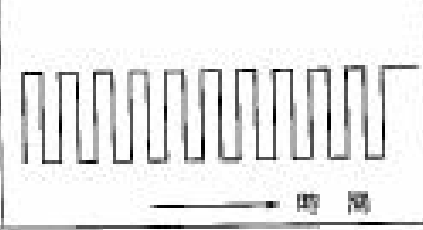
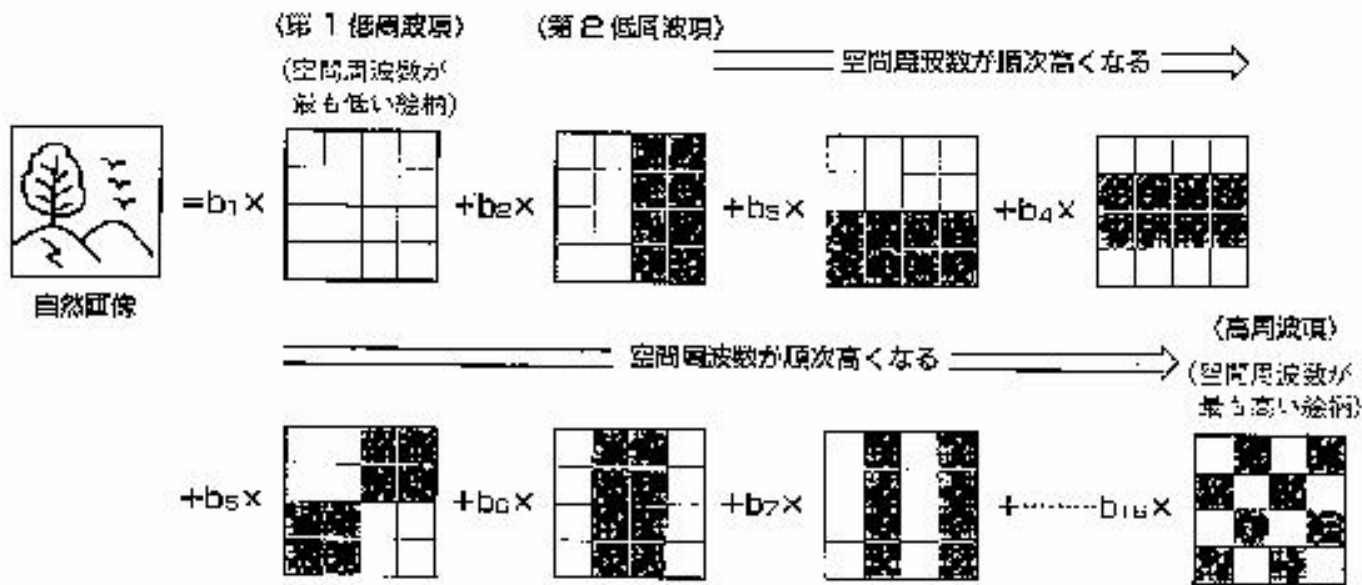
	水の波 (周波数)	画像 (デジタル) の空間周波数
(1) 周波数の低い波		(a) 空間周波数が最も低い絵柄 (無地に濃い) (画面上相関関係は最も高い)
(2) 周波数の少し高い波		(b) 空間周波数が少し高い絵柄 (自然画) (画面上相関関係は少し高い)
(3) 周波数の最高の波		(c) 空間周波数が最高の絵柄 (人工的, 幾何学的) (画面上相関関係は最も低い)

図 1-10 水の波と画像の空間周波数の関係

ポイント

- 1** 直交変換によって周波数分解可能 **2** DCTによって低周波項にエネルギーが集中



自然画像を、低い周波数成分から高い周波数成分に分解することができます。これを直交変換といいます。DCTでは変換後、低周波項に値が集中するメリットを生かして、高周波項を丸める（例えば図では b_7 以後）操作をして、情報圧縮を実現しています。



図1-11 一般的な直交変換（周波数成分の分解）の概念図



1 画像に対する直交変換 (DCT) は、8×8 正画面素ブロック単位で実行 2 直交変換と量子化の組み合わせが画像圧縮に有効

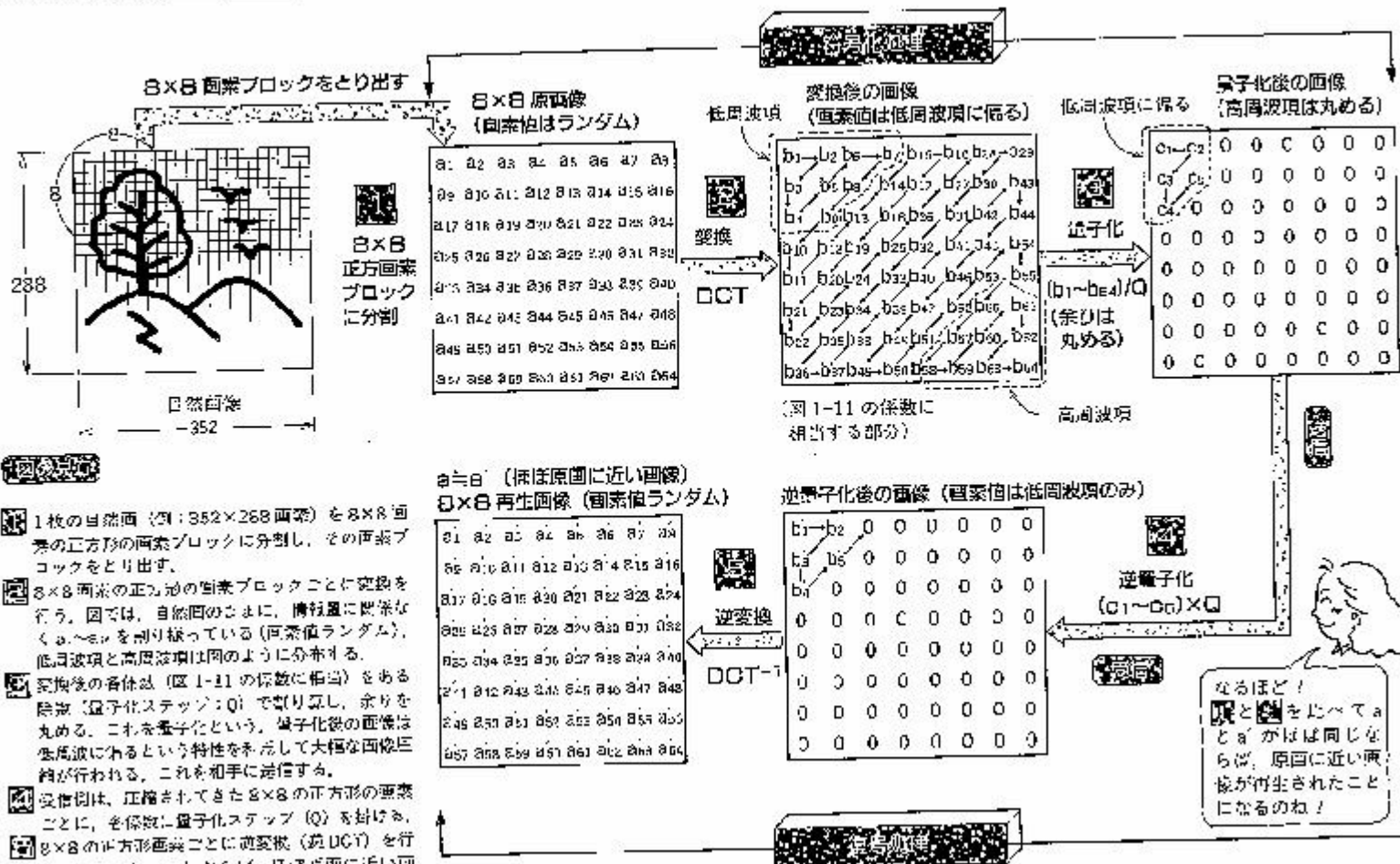


図 1-12 画素内 (空間的) 相関関係による情報圧縮の方法

ポイント

- 1 時間相関圧縮の基本は動き補償
- 2 動きベクトルを正方面素ブロック単位に抽出

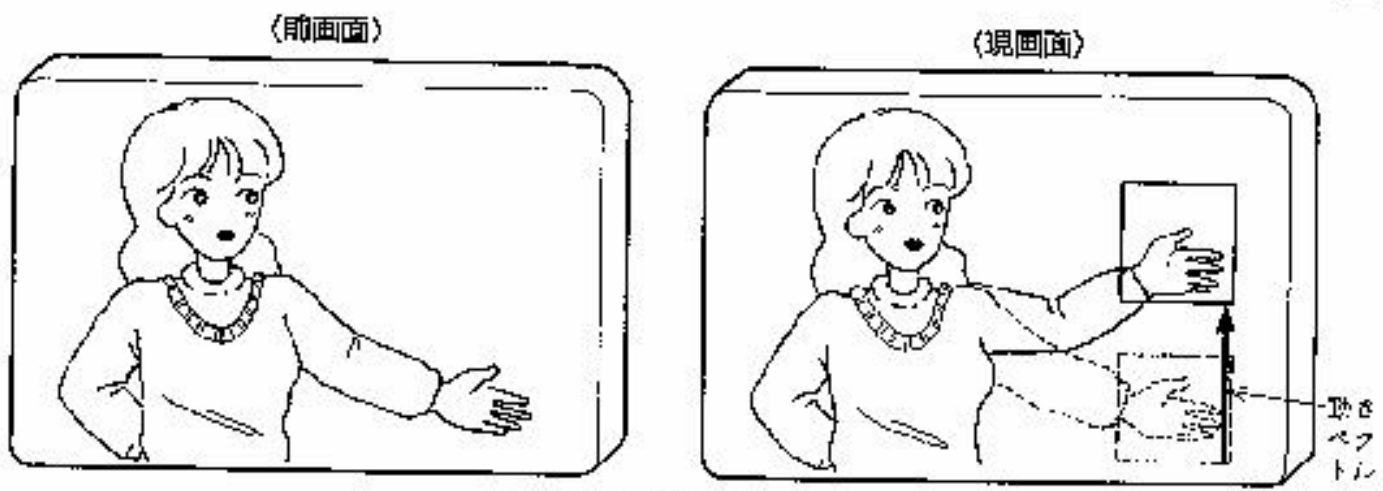


図 1-13 動き補償の考え方

表 1-3 可変長符号の例

元データ	出現確率	符号	符号長
0	0.6	0	1
1	0.2	10	2
2	0.1	110	3
3	0.1	111	3

ポイント

1 DCT・動き補償・可変長符号のハイブリッド符号化を採用

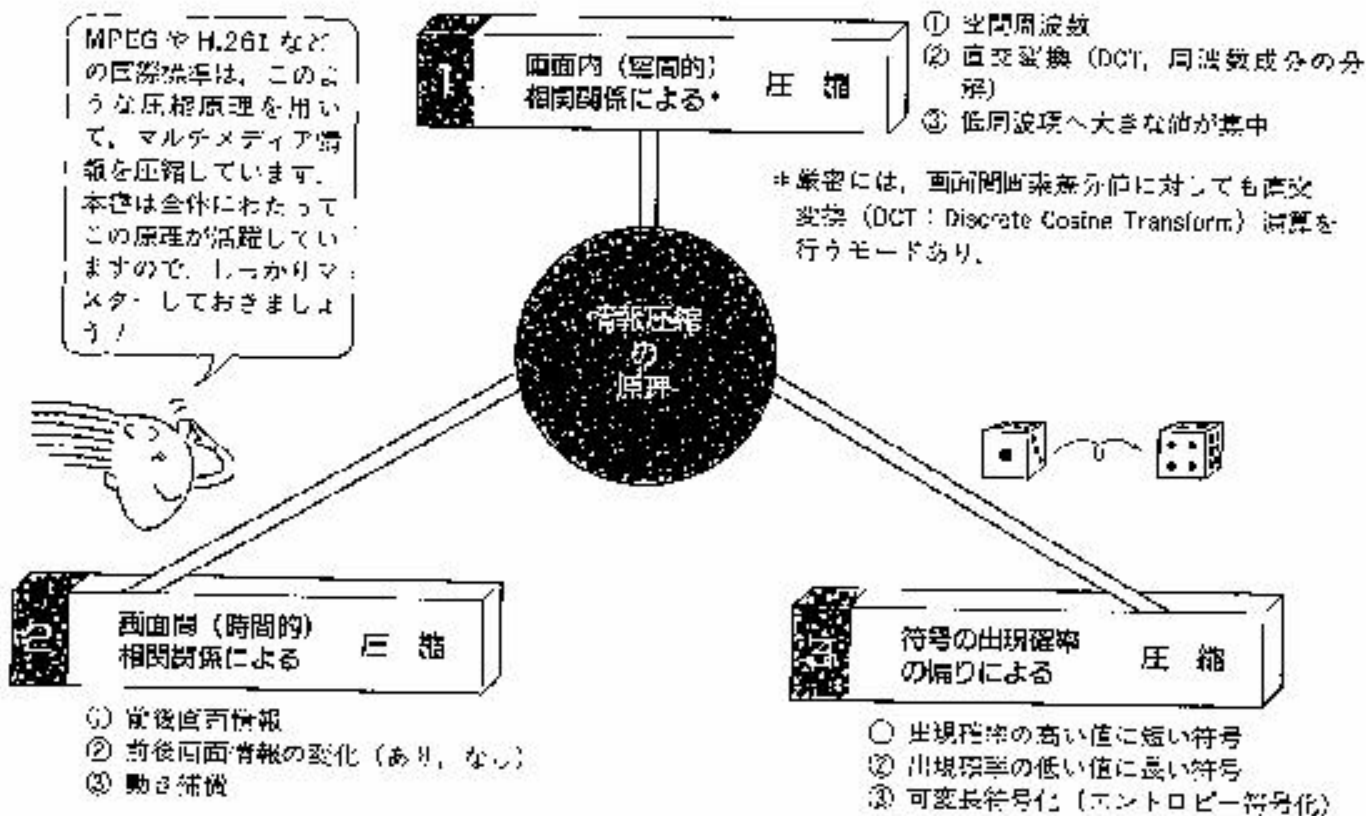


図 1-14 MPEG や H.261 などの国際標準の基本となっている情報圧縮技術

3. 原型となった通信標準H. 261

マルチメディア情報圧縮（インターネット時代の数学シリーズ）



藤原 洋編

税込価格：¥2,520（本体：¥2,400）

出版：共立出版

サイズ：A5判 / 146p

ISBN：4-320-01644-0

発行年月：2000.3

(1) ITU-T H.261

① 全体のデータ構造と処理系統

H.261 は、64 K から 2 Mbps 程度の伝送速度範囲をカバーする TV 電話/TV 会議用の標準で、通常の TV 品質よりも若干低品質であるが、384 Kbps 以上であれば TV 会議用に有効であるし、また 1 対 1 の TV 電話には 64 Kbps あれば実用に耐えるようになった最初の標準である。

H.261 コーデックのブロック図を図 2.1 に示すが、本図において情報源符号器と復号器は、各々、情報圧縮と伸張にとっての本質的要素である。

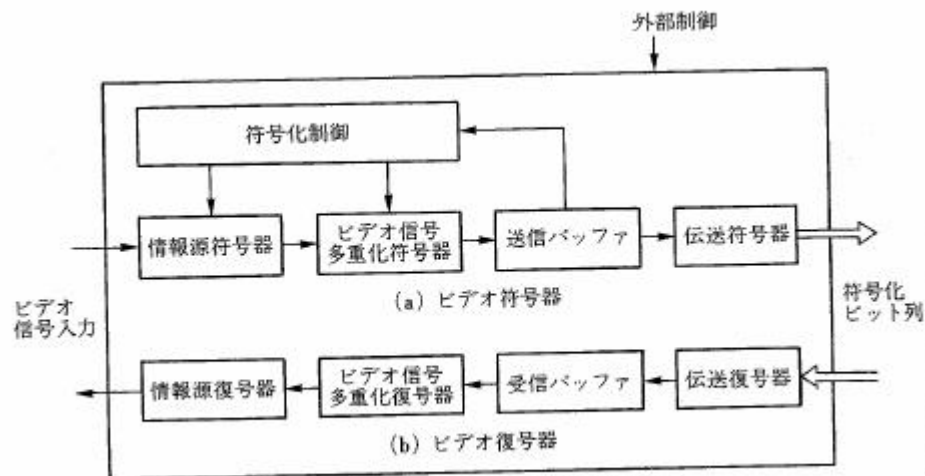


図 2.1 H.261 コーデックのブロック図

H.261の入力信号フォーマットは、欧米やアジアなどで異なるTV信号フォーマットに共通性をもたせるために、図2.2に示すCIF(Common Intermediate Format)と1/4のQCIFが規定されている。

次にCIF画面(ピクチャ)における階層型のデータ構造について図2.3に示す。この図にあるように1枚のCIF画面は、12個のGOB(Group of Blocks)によって構成される。次に1つのGOBは、33個のMB(Macro Block)で構成される。さらに各MBは、4つのYブロック(輝度信号)と2つのC(色差)ブロックとから成る。Cブロックが4つのYブロックに対して半分の2つしかない理由は、人間の視覚特性として輝度(明るさ)に対する解像度の方が色(色差)よりも高いことを利用している。各Y,Cブロックとも、8×8の画素ブロックとなっている。

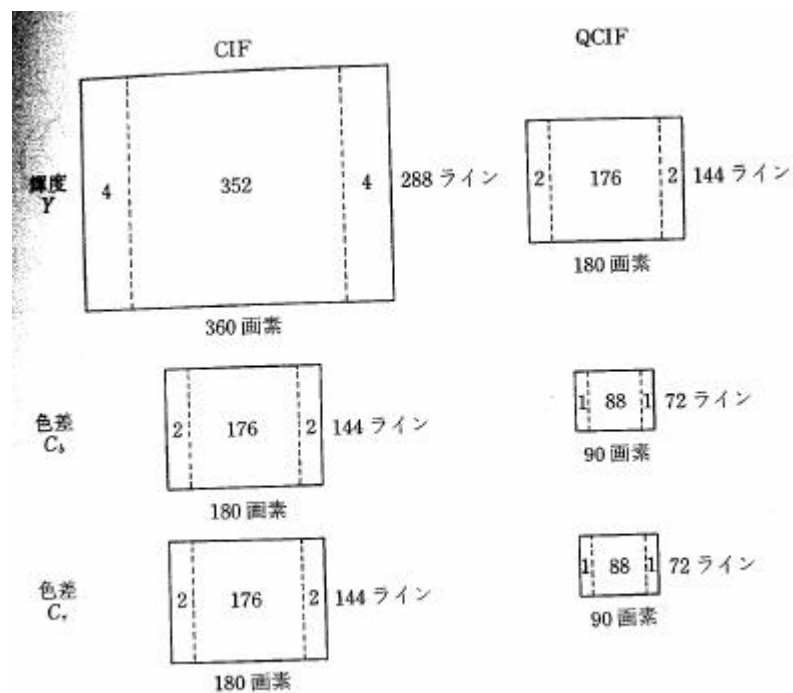


図 2.2 H.261 の入力信号フォーマット

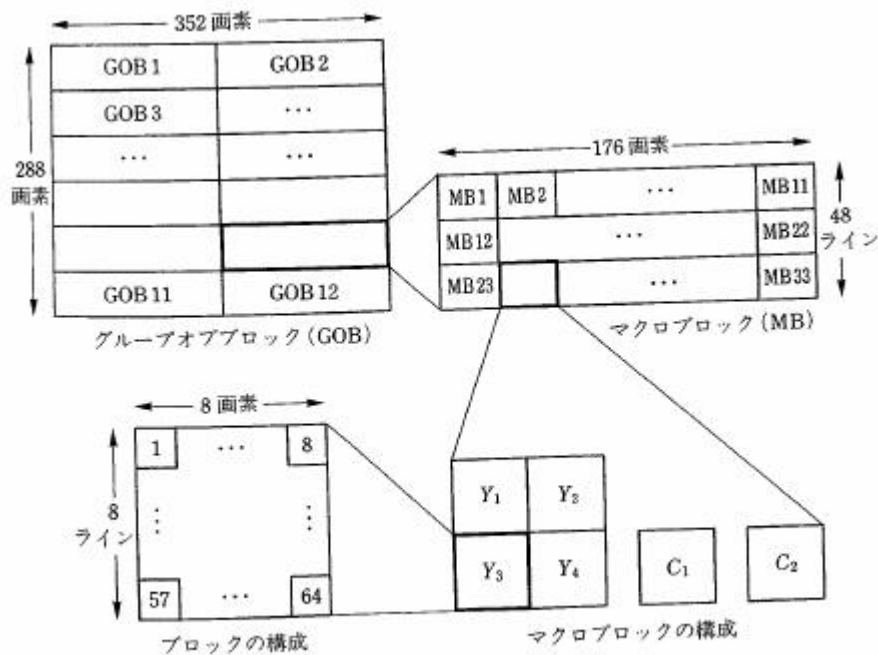


図 2.3 CIF 画面の階層型のデータ構造

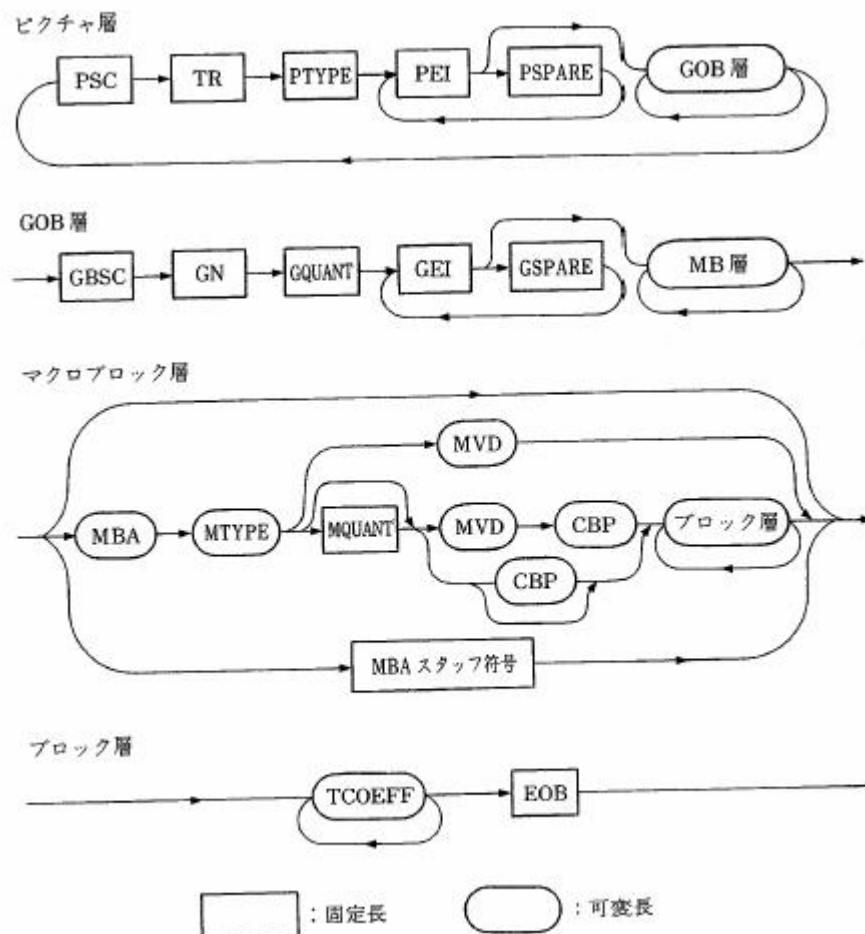


図 2.4 各階層型データ構造における処理系統図

以上のことからピクチャ、GOB、MB、およびBの四階層構造を形成する。この各四階層における処理系統図を図2.4に示す。また、この図中の各レイヤにおける符号種別一覧表を表2.1に示す。

② 情報源符号化方式

上記の処理系統の元となる最も重要なH.261の情報源符号化方式におけるエンコーダ部の制御ブロック図を図2.5に示す。

この図中のイントラとはIntra Frame Codingモードのことで、初期符号化画面やシーンチェンジが起こった時のように、時間相関をとらずに1枚の画面内に閉じた符号化を行うモードであることを示している。これに対して、

表 2.1 各レイヤにおける符号種別一覧表

PSC	ピクチャ開始符号 (20 ビット, 0000 0000 0000 0001 0000)
TR	テンポラルリファレンス (5 ビット FLC[固定長符号]) は元の CIF のピクチャ番号を表し, デコーダで使われる (この使用法は H.261 の規格外である). ピクチャレートを知らせるためのものではない.
PTYPE	ピクチャタイプ (6 ビット FLC). これはスプリットスクリーンの表示のオン・オフ, 書画カメラの表示のオン・オフ, フリーズピクチャの解除のオン・オフ, そしてフォーマットが CIF(1) か QCIF(0), を表す.
PEI	ピクチャ拡張情報
PSPARE	将来のための拡張ビット (0, 8, 16, ..., ビット)
GBSC	GOB 開始符号 (16 ビット, 0000 0000 0000 0001)
GN	12 の GOB を表すグループ番号 (4 ビット FLC)
GQUANT	グループ量子化情報 (5 ビット FLC). GOB の中で使われる 31 種の量子化器の 1 つを表し, 後から現れる MQUANT 情報で書き換えられるまで使われる.
MBA	MB アドレスで GOB 内の位置を示す (11 ビット以下の VLC). MBA は GOB 内でマクロブロックの絶対アドレスとその 1 つ前に送られたマクロブロックのアドレスの差で表される.
MTYPE	VLC (可変長符号) テーブルに示される MB タイプ情報.
MQUANT	MB 量子化情報 (5 ビット FLC)
MVD	動きベクトルデータ (11 ビット以下の 32 種の VLC)
CBP	有意ブロックパターン (9 ビット以下の 63 種の VLC)
TCOEFF	変換係数 (ジグザグスキャンされ, 8 ビットの FLC あるいは 13 ビット以下の 66 種の VLC)

インターとは Inter Frame Coding モードのことで、画面間の時間相関，すなわちフレーム間相関符号化モードであることを示している。ループフィルタは、動き補償を実行した時に発生するブロック歪み（動き補償単位の正方画素ブロック境界の不連続性）を平滑化処理するためのフィルタである。ここで、イントラ/インター判定処理は、実装は自由であるが、H.261 では、図 2.6 に示す判定曲線を推奨している。

なお、この図の判定曲線中の VAR(Variance) は、統計学上の分散であり、MSE(Mean Square Error) は、式 (2.1) で示される o:original (原画素)，mc:motion compensation (動き補償された画素) を意味する添え字である。

$$\text{MSE} = \frac{1}{256} \sum_{m=0}^{15} \sum_{n=0}^{15} (x_o(m, n) - x_{mc}(m, n))^2 \quad (2.1)$$

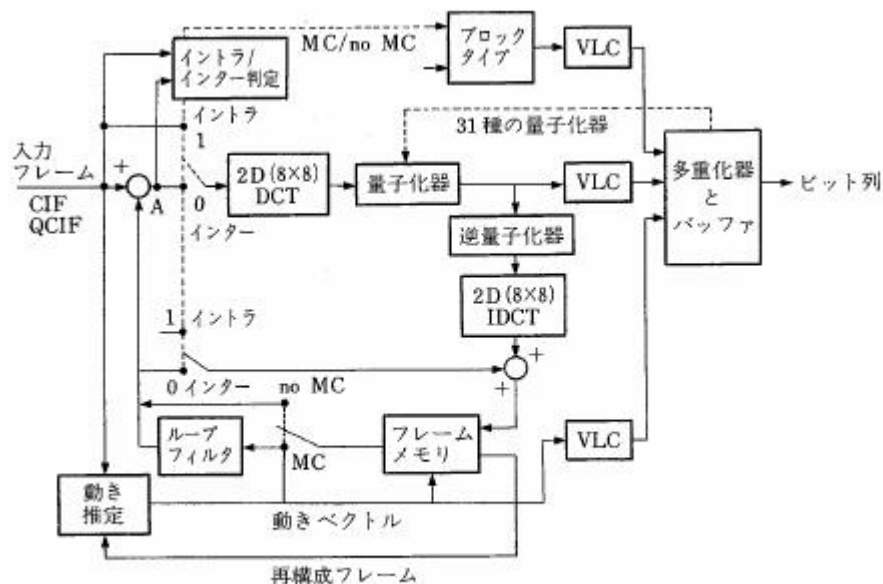


図 2.5 H.261 の情報源符号化方式におけるエンコーダ部の制御ブロック図

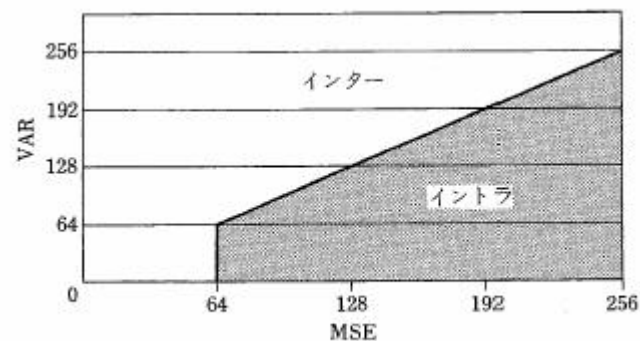


図 2.6 イントラ/インター判定処理の判定曲線

量子化については、図 2.7 に示すようにイントラ DC 係数（64 個の DCT 係数のうち左上隅の係数）とインター DC と全 AC 係数とで異なる量子化が実行される。

また図 2.5 に示したビデオエンコーダで用いるループフィルタは、式 (2.2) で規定される。

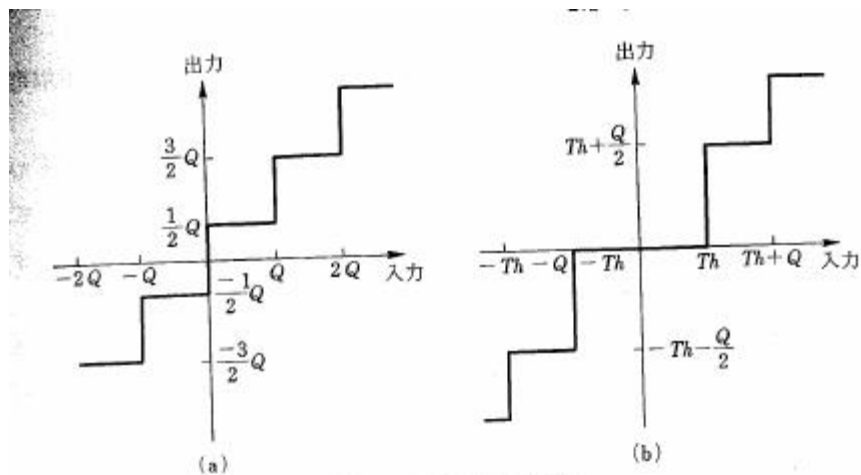


図 2.7 H.261 の量子化

1. ブロック内部の画素に対して

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

2. ブロックの縁の画素に対して

$$\frac{1}{16} \begin{bmatrix} 3 & 1 \\ 6 & 2 \\ 3 & 1 \end{bmatrix}$$

(2.2)

(2) ITU-T H.263

① 全体のデータ構造と処理系統

H.263 は、アナログ電話のモデムを前提としているため、32 Kbps 程度以下の伝送速度範囲を対象とする TV 電話用の標準で通常の TV 品質よりも低品質である。H.261 よりも約 5 年後で作られた標準であり、低ビットレート下でも高画質である。また、インターネット上で実用に耐えるようになった最初の標準である。

H.263 の入力信号フォーマットは、H.261 で欧米やアジアなどで異なる TV 信号フォーマットに共通性をもたせるために、導入された CIF の 1/4 サイズの QCIF (176 画素 × 144 ライン) を基本としている。しかし、sub-CIF (128 画素 × 96 ライン)、CIF (352 画素 × 288 ライン)、4CIF (CIF のタテ・ヨコ

とも 2 倍), 16CIF (CIF のタテ・ヨコとも 4 倍) のフォーマットが使用可能である。

動き予測は, H.261 のような整数画素単位のマクロブロック単位に 1 つの動きベクトルだけでなく, 半画素精度で中央値ベースの動き予測に加えて高度予測モードでは, マクロブロック単位に 4 つの動きベクトル, およびオーバーラップブロック動き補償を行うようになっている。

フレーム種別は, 前方向予測に加えて, オプションとして, MPEG と同様の両方向予測フレームが導入されている。図 2.8 に H.263 映像ストリームの各階層型データ構造における処理系統図を示す。

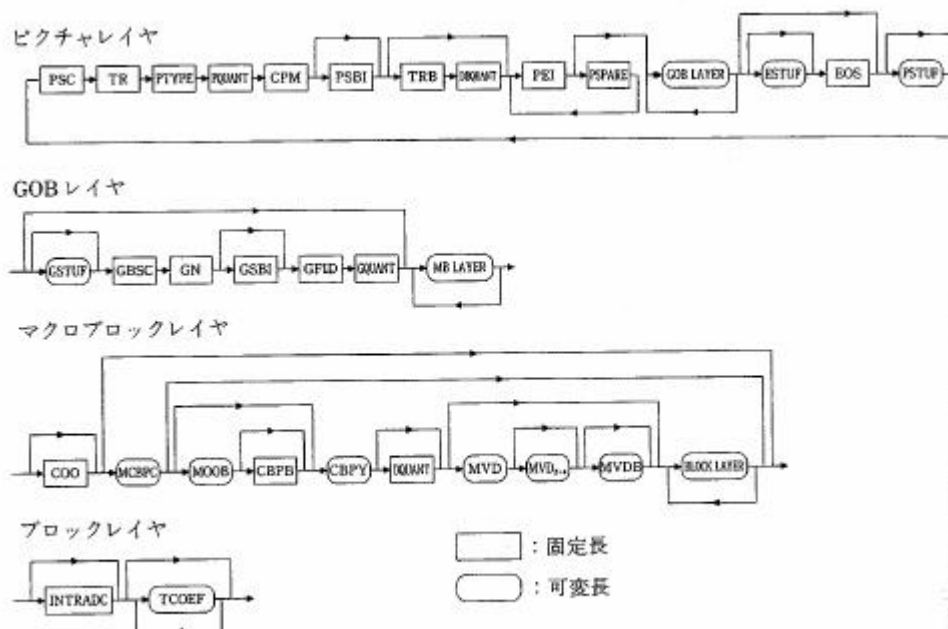


図 2.8 H.263 映像ストリームの各階層型データ構造における処理系統図

次に、画面（ピクチャ）における階層型のデータ構造については、H.261と同様にピクチャは、GOBによって構成される。次に1つのGOPは、33個のMBで構成される。さらに各MBは、4つのYブロック（輝度信号）と2つのC（色差）ブロックとから成る。

② 情報源符号化方式

上記の処理系統の元となる最も重要なH.263の情報源符号化方式におけるコーデックのブロック図を図2.9に示すが、図2.5に示したH.261と共通点が

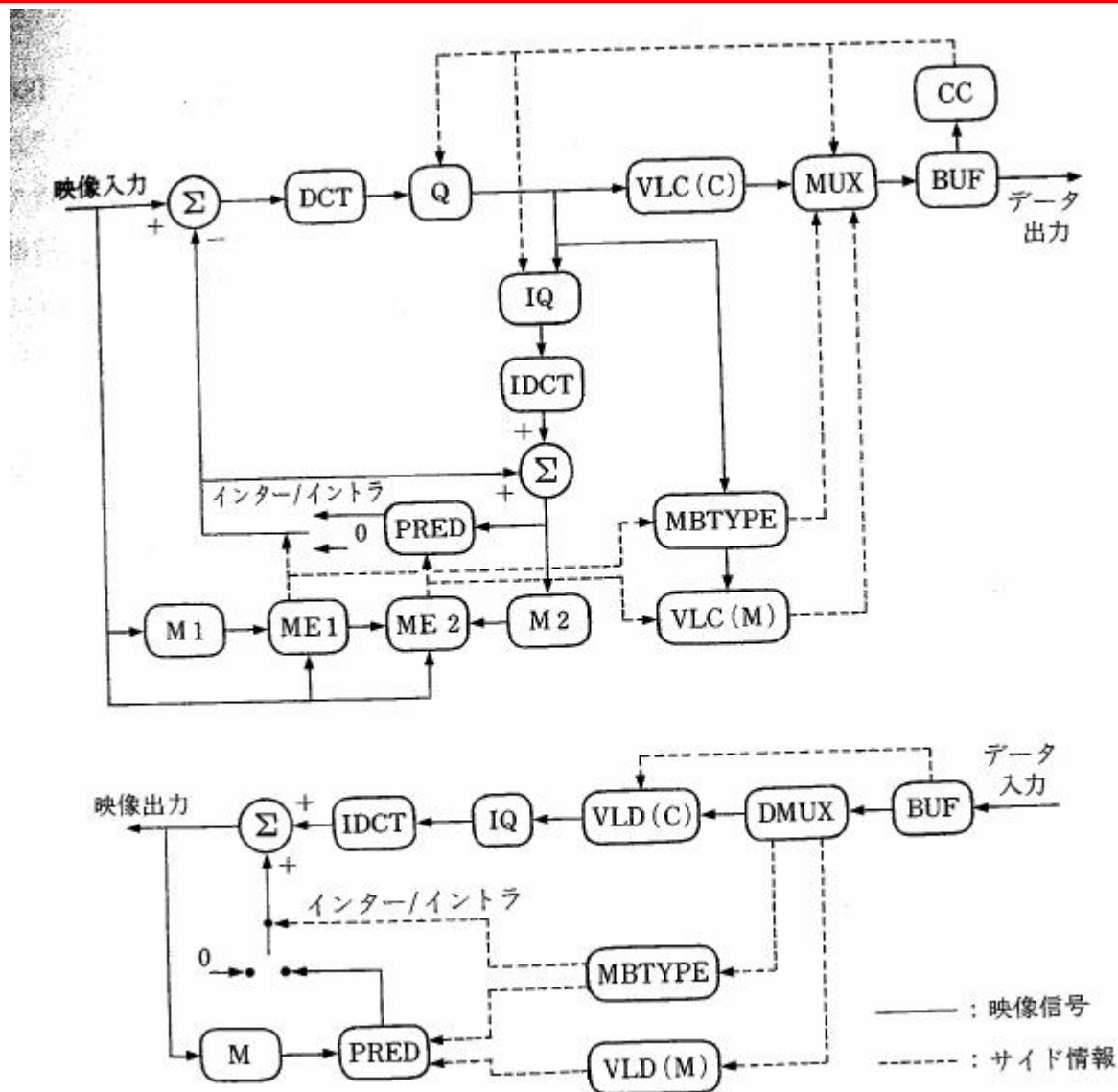


図 2.9 H.263 の情報源符号化方式におけるコーデックのブロック図

4. 静止画標準JPEG

2.2 JPEG

JPEG (Joint Photographic Experts Group, 合同静止画符号化専門家会合) は、もともと ISO (International Organization for Standardization, 国際標準化機関) と IEC (International Electro-technical Commission, 国際電気標準会議) の合同技術委員会である JTC1 (Joint Technical Committee) 内の WG8 (Working Group, 作業部会) の中にある 1 つの作業グループの名前であった。同様に次節で解説する MPEG があつたが², 両作業グループともに発展し, その後, WG1 と WG11 に昇格した。この標準化作業グループの名前がそのまま

国際標準のニックネームとなっている。正式規格名は, ISO10918 である。ISO と IEC は, とともに国家間の条約に基づいて制定される国際標準で, 日本では通産省の主管となっている。一方, 2.1 節で述べた ITU は, 郵政省の主管である。しかしながら, マルチメディアは, 通信, 放送, コンピュータ, 家電機器などの間の境界が不鮮明になってきており, この JPEG と後述の MPEG 2 は, ISO と ITU-T (勧告 T.82) の共通標準として規格化されている。これを共通テキストと呼んでいる。

JPEG は, 圧縮率 1/2 から 1/100 程度をカバーするモノクロとカラーの連続階調静止画に対する符号化規格であり, カラー FAX, デジタルカメラなど多くの製品分野で最も普及している標準である。JPEG には, 以下に示す

4つの符号化モードがある。

- ① シーケンシャル符号化モード
- ② プログレッシブ符号化モード
- ③ ロスレス符号化モード
- ④ 階層符号化モード

これらの技術は、主として Part1 で詳しく述べた二次元 DCT が基礎となっている。JPEG に DCT が採用された理由は、実装の容易さと圧縮性能であった。JPEG の性能指標となる画像品質と圧縮性能の目安としては、以下に示す 4 つの段階がある。

- ① 通常の品質 : 0.25 – 0.5 bpp(Bit Per Pixel)
- ② 良好な品質 : 0.50 – 0.75 bpp
- ③ 優れた品質 : 0.75 – 1.0 bpp
- ④ 原画と同等品質 : 1.50 – 2.0 bpp

では、次に、個々の符号化モードについて述べる。

(1) シーケンシャル符号化モード

シーケンシャル符号化モードは、画像表示の際に左上隅から右・下方向へ順次表示していくモードである。対象画像の精度と圧縮率の要求性能によって、同じシーケンシャル符号化モードにおいてもいくつかの実現手法がある。図 2.10 は、シーケンシャル符号化モードにおける実現手法の関連図である。この図からシーケンシャル符号化モードでも 5 種類の実現方法があることがわかる。というのは、画素サンプルを 1 つとっても 8 ビットサンプル精度と 12 ビット

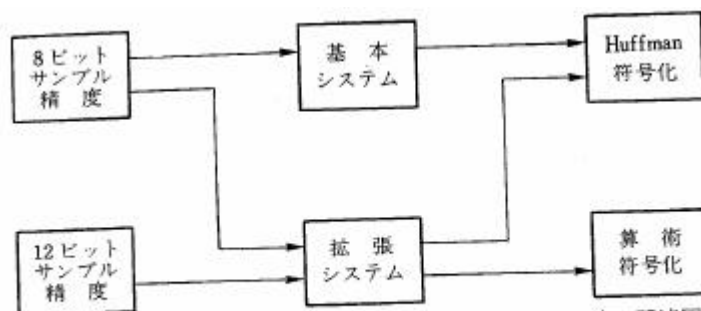


図 2.10 シーケンシャル符号化モードにおける実現手法の関連図

トサンプル精度がある。また、可変長符号化についても、標準的な Huffman 符号と、これより 10 % 程度有利な算術符号が用意されている。
 また、JPEG 基本エンコーダとデコーダの処理ブロック図を図 2.11 に示す。

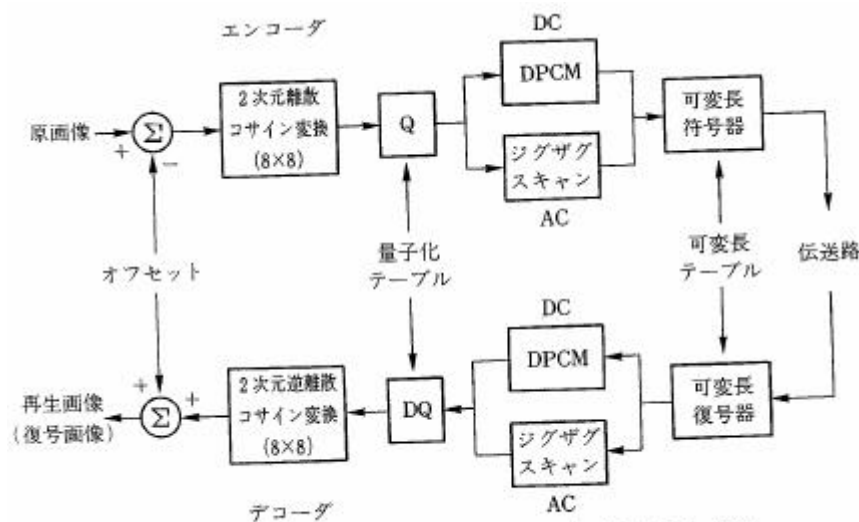


図 2.11 JPEG 基本エンコーダとデコーダの処理ブロック図

この図中における量子化テーブルを表 2.2 と表 2.3 に示す。量子化テーブルは、輝度と色差に各々利用される。

量子化された DCT 係数 S_{quv} は、式 (2.3) で表される。

$$S_{quv} = \text{Nearest integer} \left(\frac{S_{uv}}{Q_{uv}} \right) \quad (2.3)$$

このときの量子化換算図を図 2.12 に示す。

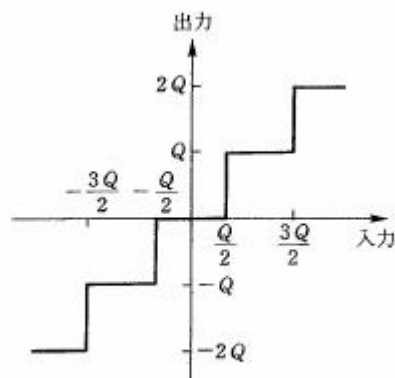


図 2.12 JPEG の量子化換算図

表 2.2 輝度量子化マトリクス Q_{uv}

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

表 2.3 色差量子化マトリクス Q_{uv}

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

また、DC 係数から生成されるブロックは、人間視覚システムを適用することによって高画質化される。具体的には、DC 係数は 63 個の AC 係数と分離して扱われ、第一次予測によって差分符号化される。これを式 (2.4) と図 2.13 に示す。この図に示したように空間周波数が、左上隅から右下隅に行く

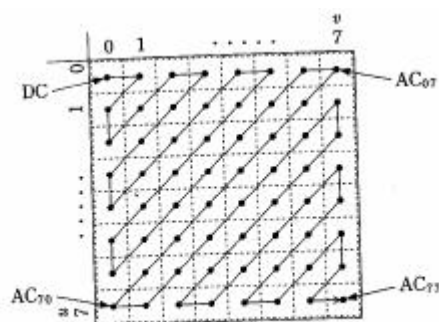


図 2.13 量子化後の DCT 係数のジグザグスキャン構造

に従って高くなっていく性質を利用すると便利なフォーマット化ができる。

$$\text{DIFF} = \text{DC}_i - \text{DC}_{i-1} \quad (2.4)$$

次にこのジグザグスキャンに従って DCT 係数をフォーマット化する際の DC 係数と AC 係数について割り当てられた分類を各々表 2.4 と表 2.5 に示す。

表 2.4 DC 係数に割り当てられた差分の分類

SSSS	DIFF 値
0	0
1	-1, 1
2	-3, -2, 2, 3
3	-7...-4, 4...7
4	-15...-8, 8...15
5	-31...-16, 16...31
6	-63...-32, 32...63
7	-127...-64, 64...127
8	-255...-128, 128...255
9	-511...-256, 256...511
10	-1023...-512, 512...1023
11	-2047...-1024, 1024...2047

出典：©1993 ITU-T.

(2) プログレッシブ符号化モード

プログレッシブ符号化モードは、通信回線を介して画像データベースの検索を行う際などに便利なように、最初の表示で画面全体を粗く表示し、2回

表 2.5 AC 係数に割り当てられた分類

SSSS	AC 係数
1	-1, 1
2	-3, -2, 2, 3
3	-7...-4, 4...7
4	-15...-8, 8...15
5	-31...-16, 16...31
6	-63...-32, 32...63
7	-127...-64, 64...127
8	-255...-128, 128...255
9	-511...-256, 256...511
10	-1023...-512, 512...1023

出典：©1993 ITU-T.

目, 3 回目と回を重ねるごとに, 精細度を順次上げていく符号化方式である. この機能をフレキシブルに実現するために図 2.14 に示すような選択肢が用意されている.

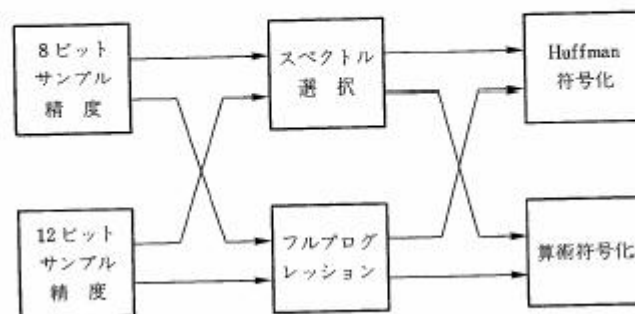
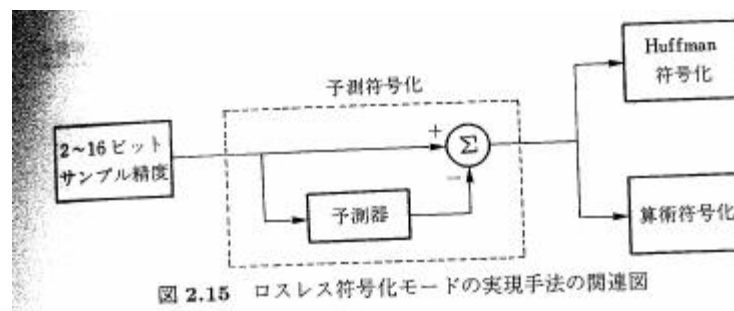


図 2.14 プログレッシブ符号化モードの実現手法の関連図

(3) ロスレス符号化モード

ロスレス符号化モードは、復号時に完全に情報が損なわれずに再生できる符号化モードを指す。このためDCTのような非可逆過程は適用できない。したがって、図 2.15 に示したような予測符号化とエントロピー符号化が中心となる。



5. 蓄積動画標準MPEG1

2.3 MPEG 1

MPEG (Moving Picture Experts Group, 動画像符号化専門家会合) は、もともと 2.2 節で述べた JPEG と同じく ISO と IEC の合同技術委員会である JTC1 内の WG8 の中にある 1 つの作業グループの名前であった。MPEG は JPEG とともに発展し、その後、WG11 と WG1 に昇格した。この標準化作業グループの名前がそのまま国際標準のニックネームとなっている。正式規格番号は、ISO11172 である。

MPEG 1 は、もともと CD-ROM を中心とする蓄積メディアを対象とする動画像符号化規格として標準化作業に入った。圧縮率 1/20 から 1/30 程度をカバーするカラー動画像用の符号化規格であり、ビデオ CD を中心としたカラオケシステムなどの分野で普及している標準である。MPEG 1 は、先に述べた ITU-T H.261 と JPEG の技術を継承した内容となっている。

(1) ビデオ符号化アルゴリズム

MPEG 1 は、ITU-T H.261 と JPEG の技術を継承しているが、CD-ROM を中心とする蓄積メディアを対象とするため、以下に列挙する新たに取り入れられたのが、ランダムアクセス機能である。このため、フレーム間相関符号化を取り入れるが、一定フレーム数内に必ずイントラフレームを 1 枚含むとすることを規定している。ここで MPEG 1 での規定は、デコーダ (復号器)

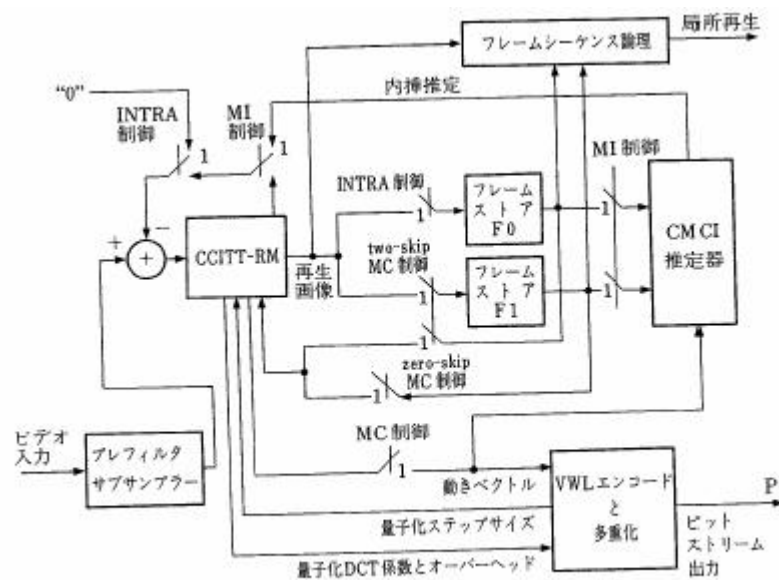


図 2.16 Puri らによって提案された MPEG 1 エンコーダのブロック図

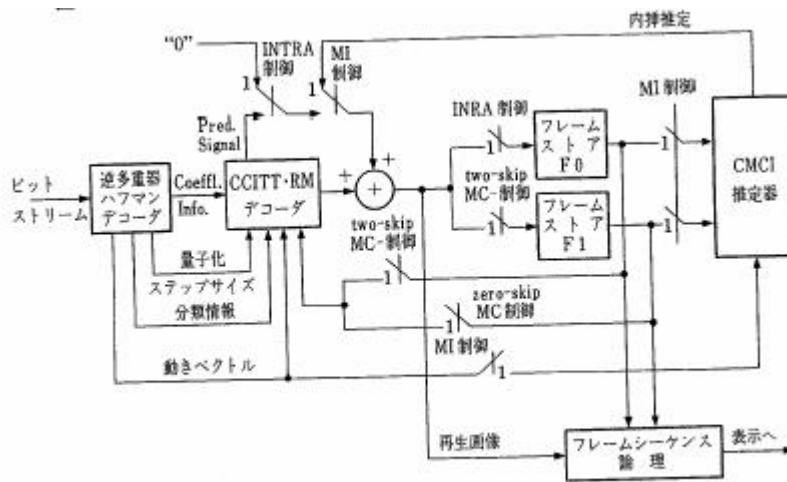


図 2.17 Puri らによって提案された MPEG 1 デコーダのブロック図

だけであり、エンコーダ（符号化器）の仕様は自由である。そこで、デコーダ規格の仕様を守るエンコーダとしていくつかのものが提案されているが、Puri らによって提案された MPEG 1 エンコーダとデコーダのブロック図を各々図 2.16 と図 2.17 に示す。ここで、CCITT-RM は、前述の H.261 を標準化

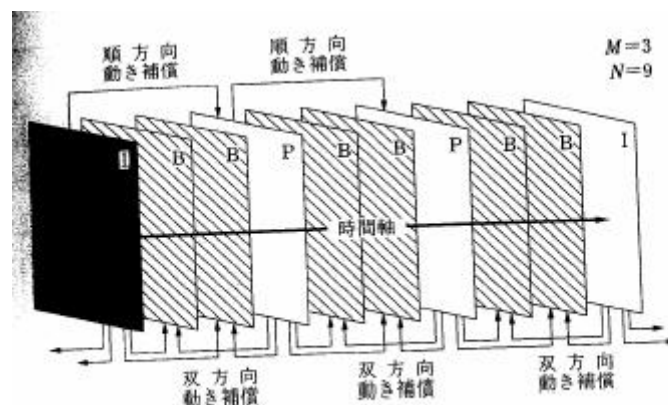


図 2.18 MPEG 1 ピクチャ間の時間相関図

表 2.6 MPEG 1 ビデオの解像度

		(NTSC)	(PAL)
		525/30	625/25
輝度 (Y) 画素/画像	ITU-R	720 × 480	720 × 576
	SIF	360 × 240	360 × 288
	SPA	352 × 240	352 × 288
色素 (C_b, C_r) 画素/画像	ITU-R	360 × 480	360 × 576
	SIF	180 × 120	180 × 144
	SPA	176 × 120	176 × 144
フレームレート (Hz)		30	25

する過程で採用された RM(Reference Model) のことであり、また CMCI は、条件付き動き補償内挿のことである。

次に重要なのは、H.261 とは異なり、MPEG で初めて導入された双方向動き補償の考え方である。図 2.18 に示すようにフレーム内予測画面 (Intra frame coding picture) を I ピクチャ、前方向フレーム間予測画面 (Predictive Inter frame coding picture) を P ピクチャ、および双方向フレーム間予測画面 (Bi-directional Inter frame coding picture) を B ピクチャと呼ぶことにしている。この双方向予測の考え方は、後述の MPEG 2 にも継承されている。

(2) 全体のデータ構造

MPEG 1 の画面構造を規定するフォーマットを SIF(Source Input Format) と呼んでいる。その仕様は、表 2.6 に示すようにテレビジョンフォーマットの各々 NTSC と PAL とに合わせている。この表から容易にわかるように SIF は、

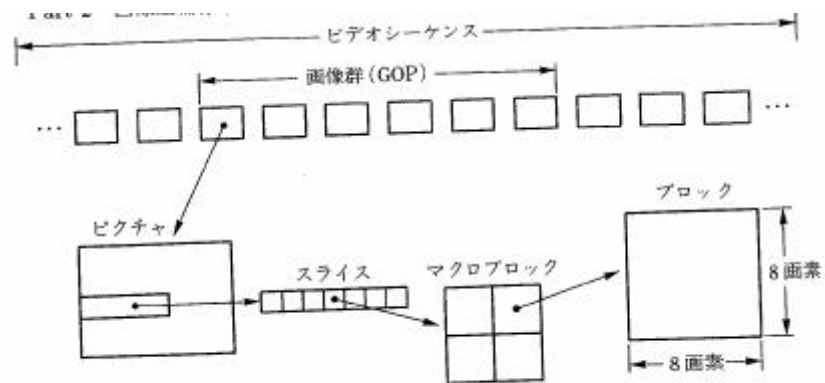


図 2.19 MPEG1 ビデオのデータ構造

表 2.7 MPEG 1 ビデオの各レイヤの機能比較

レイヤ	機 能
シーケンスレイヤ	1つ以上の画像群
画像群レイヤ	シーケンスへのランダムアクセス
ピクチャレイヤ	基本的符号化単位
スライスレイヤ	再同期単位
マクロブロックレイヤ	動き補償単位
ブロックレイヤ	DCT 単位

ITU-R (放送品質用の規格) と比較してタテ・ヨコともに半分の解像度となっている。

次に、MPEG 1 で用いられる SIF 上の各レイヤの機能比較とデータ構造を、各々表 2.7 と図 2.19 に示すが、6 つのレイヤ構造となっている。この表でピクチャレイヤまでは H.261 と本質的に同様である。ただし、H.261 で用いられていた GOB レイヤの代わりに、より自由度の高いスライスレイヤが定義されている。そして、画像群 (GOP: Group of Pictures, 少なくとも 1 つの I ピクチャを含むようにした画像群) レイヤとシーケンシャルレイヤとは、新しく導入された概念である。

6. 放送標準MPEG2

2.4 MPEG 2

MPEG 2 も MPEG 1 と同様に MPEG が規格案を作成してできたものである。しかし、これは、同じく ISO と IEC の合同技術委員会である JTC1 内の

WG 中にある 1 つの作業グループの最初の標準化作業 MPEG 1 が大成功を収めたため、ITU-T も同一規格とすることを決め、ATM を基本とした広帯域ネットワーク用の放送品質動画画像符号化標準である ITU-T H.262 と MPEG 2 とを共通テキストとすることとした。MPEG 2 は、MPEG 1 以上に発展し、これまでの通信、家電、コンピュータ業界に加えて放送業界から多くの参加を得て標準化作業が進行し、世界のデジタル放送の標準にも相次いで採用されるようになった。正式規格番号は、ISO13818 である。

MPEG 2 は、CD-ROM を中心とする蓄積メディアを対象とする動画画像符号化規格と違って、より高画質のデジタル放送や DVD(Digital Video Disc) への適用を目的としており、HDTV (高解像度テレビ) もその標準化対象となった。圧縮率 1/20 (現行 TV 品質) から 1/80 (HDTV 品質) 程度をカバーするカラー動画画像用の符号化規格であり、日本国内ではデジタル CS 放送や DVD ですでに利用されており、また BS4 後発機による HDTV 放送にも MPEG 2 の利用が計画されている。MPEG 2 は、先に述べた MPEG 1 の技術を継承・発展させた内容となっている。

(1) ビデオ符号化アルゴリズム

MPEG 2 は、MPEG 1 の技術を継承しているが、放送品質の高品質画像を対象とするため、表 2.8 に列挙する機能比較表に示したように、新たに取り入れられたのが、多様なビデオフォーマット、スケーラビリティ、および飛び越し走査（インタレーススキャン）の 3 つである。

ここで、ビデオフォーマットについては、4:2:0 だけでなく、4:2:2 と 4:4:4 のスタジオ品質ビデオフォーマットが導入されている。4:2:0 と 4:2:2 とのビデオフォーマットの違いは図 2.20 に示すように輝度と色解像度の相違に関するもので、4:2:2 とは、色についてはヨコ方向だけ輝度に比べて解像度が半分であることを示している。4:4:4 は、全く同一である。

スケーラビリティとは、符号化列の一部を取り出しても復号可能であることを示している。SNR(Signal To Noise Ratio) スケーラビリティとは、量子化ステップの粗い符号化列とさらに精細化した符号化列とが階層構成になっている構造を示している。空間解像度の粗精度符号化列と高解像度符号化列との階層構造を、空間解像度スケーラビリティと呼んでいる。また、時間スケーラビリティとは、時間解像度の粗精度符号化列と高精度符号化列とを階層構

表 2.8 MPEG 1 と MPEG 2 との機能比較表

	MPEG 1	MPEG 2
ビデオフォーマット	SIF 順次走査	SIF, 4:2:0, 4:2:2, 4:4:4, 順次走査/飛び越し走査
画像品質	VHS	放送品質/スタジオ品質
ビットレート	可変 (≤ 1.856 Mbps)	100 Mbps まで可変
低遅延モード	<150 ms	<150 ms (B ピクチャなしの場合)
アクセス性	ランダムアクセス	ランダムアクセス/チャンネルホッピング
スケーラビリティ		SNR, 空間, 時間 同時送信, データ分割
互換性		前方, 後方, 上方, 下方
伝送誤り	誤り保護	誤り耐性
ビットストリーム編集	あり	あり
DCT	飛び越し走査非対応	フィールド対応あるいはフレーム対応
動き補償	飛び越し走査非対応	フィールド, フレーム, デュアルプライム, 16x8MC (フィールド時のみ)
動きベクトル	P ピクチャ, B ピクチャに対する動きベクトルのみ	P ピクチャ, B ピクチャに対する動きベクトルだけでなく, 誤り隠しのためのコンシールメント動きベクトルを I ピクチャにも付与可能
DCT 係数スキャン	ジグザグスキャン	ジグザグスキャンと飛び越し走査ビデオのためのオルタネートスキャン

成にしたものである。図 2.21 に SNR スケーラビリティの例を示す。

MPEG 2 の大きな特徴の 1 つは、飛び越し走査への対応である。これは、DCT と動き補償のときに効果を発揮する。

さて、MPEG 2 エンコーダとデコーダのブロック図は、細部を除くと MPEG 1 の前節で示した各々図 2.16 と図 2.17 と原理的に同等である。フィールドモードを採用した時に動き補償と DCT 演算のブロック構成が変わるだけである。図 2.22 に 8×8 DCT 中におけるフレーム/フィールドブロック構成を示す。

次に重要なピクチャ構成は、MPEG 1 で初めて導入された双方向動き補償の考え方を踏襲している。前節の図 2.19 に示すように I ピクチャ、P ピクチャ、および B ピクチャとで構成されている。この双方向予測の考え方に基づいて

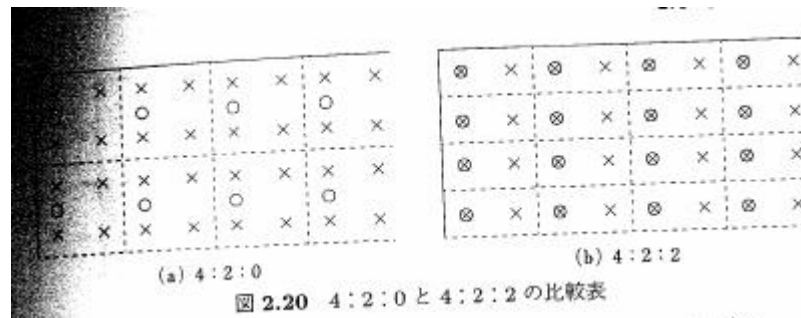


図 2.20 4:2:0 と 4:2:2 の比較表

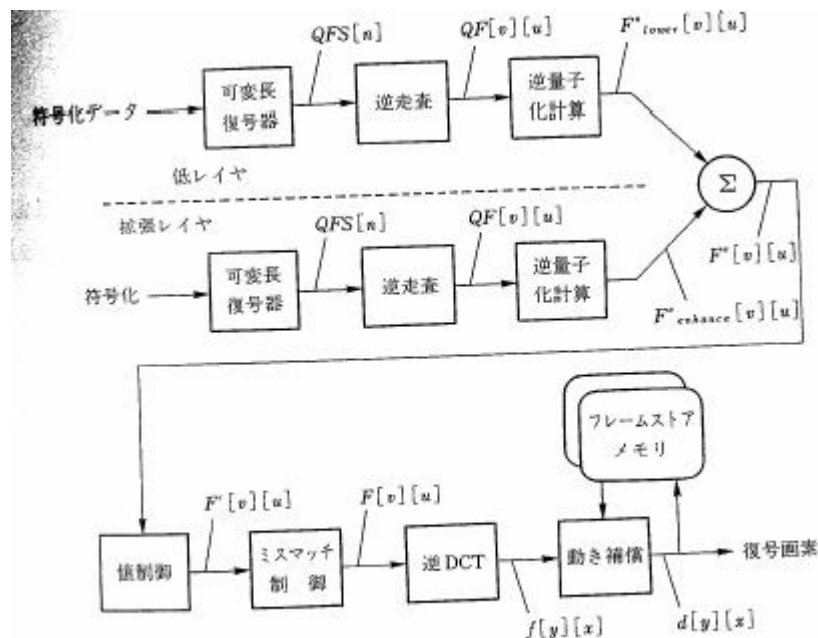


図 2.21 SNR スケーラビリティの例

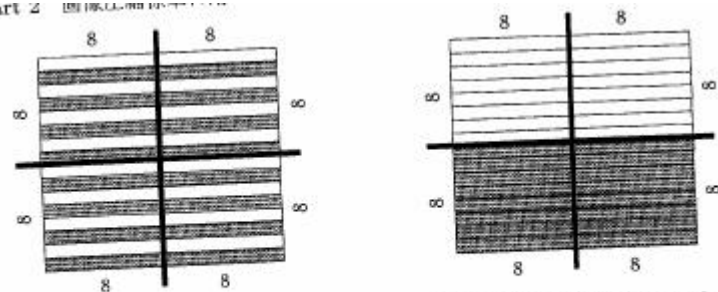
(2) 全体のデータ構造

MPEG 2 の画面構造を規定するフォーマットは、現行 TV のテレビジョンフォーマットである NTSC と PAL に準拠した Main Level が基本となっているが、MPEG 1 と同等の SIF をローレベルとして定義しているし、また、高解像度の HDTV をハイレベルとして規定している。このように、解像度に関してはレベルという名称で規定しているが、符号化ツールのセットはプロファイルとして規定している。このレベルとプロファイルの関係を表 2.9 に示す。本表から容易にわかるように MPEG 2 は、放送品質用の規格として全てをカ

バーした内容となっている。

次に MPEG 2 の各レイヤの機能であるが、基本的に前節の表 2.7 と図 2.18 に示す MPEG 1 で用いられる SIF 上の各レイヤの機能比較とデータ構造と同様の 6 つのレイヤ構造となっている。

58 Part 2



(a) マクロブロック中の4つのフレームブロック

(b) マクロブロック中の4つのフィールドブロック

図 2.22 8x8DCT 中におけるフレーム / フィールドブロック構成

(3) MPEG 2 の応用例 ATV

ATV(Advanced Television) プロジェクトは、アメリカ FCC (連邦通信委員会) が競争入札により策定した国家放送規格である。もともと日本のハイビジョンに対抗したものだだったが、MPEG 2 仕様が世界に認められたため路線変更し、ビデオ標準としては MPEG 2 を採用し、オーディオ標準としては Dolby 研究所の AC-3 を採用している。また、変調方式には、Zenith 社が提唱した 8VSB (残留側帯波) 方式が用いられることとなった。最終的に合意された方式を表 2.10 に示す。

本表の背景にある ATV の考え方としては、以下のような 4 層の階層構造がある。

① ピクチャレイヤ

HDTV については、720P (1280×720 順次走査) と 1080I (1920×1080 飛越し走査) の 2 つのピクチャフォーマットがあるが、このような画面仕様を規定するレイヤである。

② 圧縮レイヤ

所定の伝送速度 (正味のデータ速度 18.4 Mbps) 内において、MPEG 2 による画像圧縮を、また 384 Kbps で Dolby AC-3 によるオーディオの圧縮方式を

規定するレイヤである。

③ トランスポートレイヤ

MPEG 2-TS (トランスポート・ストリーム) による、画像、オーディオ、およびデータの同期と多重化の方式を規定するレイヤである。

④ 伝送レイヤ

帯域 6 MHz の地上波放送チャンネル内で、19.3 Mbps の伝送速度を実現するために、トレリス符号化 8-VSB 変復調方式について規定するレイヤである。

表 2.9 MPEG 2 ビデオのレベルとプロファイル

プロファイル	レベル	水平サイズ (pels)	垂直サイズ (pels)	フレーム レート (Hz)	ビット レート (Mbps)	VBV サイズ (Mbits)	動き ベクトル範囲 (pels)
シンプル	メイン	720	576	30	15	1.835	-128~ 127.5
	ロー	352	288	30	4	0.489	-64~ 63.5
メイン	メイン	720	576	30	15	1.835	-128~ 127.5
	ハイ 1440	1440	1152	60	60	7.340	-128~ 127.5
	ハイ	1920	1152	60	80	9.787	-128~ 127.5
SNR スケーラブル	ロー	352	288	30	3 (4)	0.367 (0.489)	-64~ 63.5
	メイン	720	576	30	10 (15)	1.223 (1.835)	-128~ 127.5
空間 スケーラ ビリティ	ハイ 1440	720 (1440)	576 (1152)	30 (60)	15 (40) (60)	1.835 (4.893) (7.340)	-128~ 127.5
	メイン	352 (720)	288 (576)	30 (30)	4 (15) (20)	0.489 (1.835) (2.447)	-128~ 127.5
ハイ	ハイ 1440	720 (1440)	576 (1152)	30 (60)	20 (60) (80)	2.447 (7.340) (9.786)	-128~ 127.5
	ハイ	960 (1920)	576 (1152)	30 (60)	25 (80) (100)	3.036 (9.787) (12.233)	-128~ 127.5

注：括弧内の数値は拡張レイヤを参照している。

表 2.10 アメリカの次世代 ATV の仕様

画像パラメタ	フォーマット 1	フォーマット 2
有効画素	1280H×720V	1920H × 1080V
全サンプル数	1600H×787.5V	2200H × 1125V
フレームレート	60 Hz 順次走査	60 Hz インタレース 30 Hz 順次走査 24 Hz 順次走査
色差サンプリング		4 : 2 : 0
アスペクト比		16 : 9
データレート		10~45Mbps または可変
カラーリメトリ		SMPTE240M
画像符号化タイプ		フレーム内符号化 (I) 予測符号化 (P) 双方向予測符号化 (B) I ビクチャ/順次走査
画像リフレッシュ		
ビクチャ構造	フレーム	フレーム/フィールド (60 Hz のみ)
係数走査	ジグザグ	ジグザグ, オルタネート
DCT モード	フレーム	フレーム/フィールド (60 Hz のみ)
動き補償	フレーム	フレーム/フィールド (60 Hz のみ) デュアルプライム (60 Hz のみ)
動きベクトル範囲		水平: シンタックス上無制限 垂直: -128, +127.5
MV 精度		1/2 画素
DC 係数精度		8 ビット, 9 ビット, 10 ビット
レート制御		前方分析器を用いた
フィルムモード処理		自動 3 : 2 ブルダウン修正 TMS
最大 VBV バッファサイズ		8 M ビット
量子化マトリクス		ダウンロード可能 (シーン依存)
VLC 符号化		フレーム内とフレーム間のランレングス/振幅 VLC
エラーコンシールメント		MC フレーム保存 (スライスレベル)
オーディオパラメタ		
チャンネル数		5.1
オーディオ帯域幅		20 kHz
サンプリング周波数		48 kHz
ダイナミックレンジ		100 dB
圧縮データレート		384 Kbps
トランスポートパラメタ		
多重化技術		MPEG 2 システムレイヤ
パケットサイズ		188 バイト
パケットヘッダ		4 同期を含むバイト
サービス数		サービススペースのスクランブルされたペイロード
条件付アクセス		4 ビット連続カウンタ
エラー処理		1 ビット/パケット
優先順位付け		多重プログラム能力
システム多重化		
伝送パラメタ	地上モード	高データレートケーブルモード
チャンネル帯域幅	6 MHz	6 MHz
超過帯域幅	11.5%	11.5%
シンボルレート	10.76 Mbps	10.76Mbps
シンボル当たりビット数	3	4
トレリス FEC	2/3 レート	なし
リードソロモン FEC	(208, 188) T = 10	(208, 188) T = 10
セグメント長	836 シンボル	836 シンボル
セグメント同期	4 シンボル/seg	4 シンボル/seg
フレーム同期	1/313seg	1/313seg
ペイロードデータレート	19.3 Mbps	38.6 Mbps
NTSC 共用チャンネル拒否	NTSC 拒否	N/A
パイロットパワーコントリビューション	受信器内フィルタ	0.3 dB
C/N スレッシュホールド	0.3 dB	28.3 dB
	14.9 dB	

7. その他の標準方式

2.5 MPEG 4

MPEG 4 は、2.3 節、2.4 節で述べた MPEG 1 と MPEG 2 との後継国際標準のニックネームであり、MPEG 3 は欠番となっている。正式規格番号は、ISO14496 である。

MPEG 1 は、もともと CD-ROM を中心とする蓄積メディアを対象とする動画画像符号化規格であり、MPEG 2 は、放送にも適用可能な高品質動画画像符号化規格として標準化作業を行った。当初、現行 TV 品質を MPEG 2、また HDTV 品質を MPEG 3 として考えていたが、MPEG 2 と MPEG 3 とは同一方式による圧縮が可能ということとなり当初の MPEG 3 の役割は、MPEG 2 が果たすこととなった。そして、1993 年に次の挑戦としてスタートした MPEG 4 の目的は、携帯電話やアナログ電話などの低ビットレート回線を前提としたものとなった。また、インターネットの普及は、この MPEG 4 の標準化作業を促進することとなった。そこで、MPEG 4 とは、コンピュータのための動画画像符号化規格として位置づけることができる。MPEG 4 は 1998 年 5 月に最終仕様が固まったが、圧縮ツールは既存の MPEG 1、MPEG 2 および ITU-T H.263 に使われているものが基本になっている。ただ、これまでの規格と大きく異なっているのは、画像の表示単位をオブジェクトとして定義し、これらを重ね合わせ表示するところである。これらの複数のオブジェクトを重ね合わせ表示することで、1 つのシーンが構成可能である。したがって、各オブジェクトに関するデータは、形状と透明度の情報をもっている。また、従来の自然画像や自然オーディオだけでなくコンピュータグラフィックスの利用が可能となっている。

(1) 要求仕様とアプリケーション

自然ビデオと合成映像との両方を扱えるのが MPEG 4 の大きな特徴であるが、考え方を明確にするために各々の要求仕様について、表 2.11 と表 2.12 に簡単にまとめる。要求仕様は、画像圧縮アルゴリズムを標準化する上での枠組みを規定したものである。このような要求仕様をまとめる背景には、MPEG 4 で想定されているアプリケーションがある。標準化会合で話題に上ったアプリケーションとしては、アナログ電話回線を用いたテレビ電話、携帯端末用テレビ電話・動画情報配信、インターネット、放送、電子図書館などがあつ

た。また、これらのアプリケーションを実現するために MPEG 2 でも用いられた、後で述べるプロファイルという概念を導入している。

表 2.11 MPEG 4 自然ビデオの要求仕様

要求項目	内 容
オブジェクト機能	任意形状オブジェクトによるシーン構成機能
ランダムアクセス	オブジェクト単位
自由度	オブジェクト数, 時間・空間解像度, 形状精度, などが可変
映像フォーマット	各種映像フォーマット, 8×8~2048×2048 まで自由に対応
伝送速度の目安	低: < 64 Kbps, 中: 64~384 Kbps, 高: 384 K~4 Mbps

表 2.12 MPEG 4 合成映像の要求仕様

要求項目	内 容
オブジェクトタイプ	任意オブジェクトに対応したシーン構成, テキスト, 2D および 3D, 静止, アニメ, など
メッシュ圧縮	2D/3D 頂点位置, 法線, 座標, トポロジーの符号化 とダウンロード機能
アニメパラメータ	face, body の定義およびアニメパラメータの圧縮機能
テキスト重ね合わせ	単独, 独立, 階層化, ビットマップ, 国際標準文字セットなど
テキストチャマッピング	2D/3D メッシュへの貼り付け機能
映像/グラフィックス	自然画像, グラフィックス単独・独立, アニメ, 階層化重ね合わせ, オブジェクトマニピュレーション

(2) プロファイル

プロファイルは、想定されるアプリケーションを実現するツールセットの規定である。また、レベルはその品質である。ここでもやはり自然画像と合成映像用に2つのプロファイルとレベルとが用意されているが、これらを表2.13と表2.14に示す。また、これらの表には、各々、最大参照メモリサイズ、および量子化モード、最大画素数、最大デコード時間についての目安が記述されている。

(3) ビデオ符号化アルゴリズム

MPEG4の画像符号化に関する部分は、ビジュアルパートと呼ばれており、符号化する対象を映像オブジェクトと呼ぶ。そして、この映像オブジェクト

表 2.13 MPEG 4 自然ビデオのプロファイルとレベル

プロファイル	レベル	画面サイズ	最大オブジェクト数	最大マクロブロック数	最高ビットレート
N ビット	L2	CIF	16	23760	2Mbps
	L4	1920×1088	32	489600	38.4 Mbps
メイン	L3	ITU BT.601	32	97200	15 Mbps
	L2	CIF	16	23760	2 Mbps
	L1	QCIF	未定	未定	未定
コア	L2	CIF	16	23760	2 Mbps
	L1	QCIF	4	5940	384 Kbps
シンプル・スケーラブル	L2	CIF	4	23760	256 Kbps
	L1	CIF	4	7425	128 Kbps
シンプル	L3	CIF	4	11880	384 Kbps
	L2	CIF	4	5940	128 Kbps
	L1	CIF	4	1485	64 Kbps

表 2.14 MPEG 4 合成映像プロフィールとレベル

プロフィール	レベル	最大オブジェクト数	最大ノード数	最高ビットレート
階層テキストチャ	L1	未定		
シンプル顔アニメ	L1	1		16 Kbps
	L2	4		32 Kbps
基本アニメ	L2	4mesh	480/mesh	64 Kbps
	L1	8mesh	1748/mesh	128 Kbps
二次元テキストチャ シンプル・スケーラブル	L2	4mesh	480/mesh	64 Kbps
	L1	8mesh	1748/mesh	128 Kbps

について、符号化・復号方式についての仕様を規定するものである。映像オブジェクトには、先に述べたように自然画像とコンピュータグラフィックス画像とがあるが、自然画像では矩形画像と任意形状画像とに対応可能である。

MPEG 4 では、自然画像を VO (Video Object) と呼称しており、VO ごとに各時刻に撮影した VOP (Video Object Plane) から構成される。なお VO は、形状とテキストチャとに分けて符号化する。

矩形形状 VO 符号化は従来型符号化であり、先述 H.263 と MPEG 2 との要素技術を取り入れている。特に、H.263 とは後方互換性を備えている。MPEG 4

04 動画の符号化と復号化

における H.263 と比較した特徴は、画像タイプ、画面内予測、画面間予測、動き補償、動きベクトル、量子化方法、インタレース対応、係数範囲について改良が加えられていることである。中でも、画像タイプに IPBS (S は、スプライト) の 4 種類あることと、動き補償を 16×16 だけではなく、 8×8 のブロックサイズにも対応しているところが大きな違いである。

ここで、 8×8 ブロックによる OBMC (Overlap Block Motion Compensation, オーバーラップブロック動き補償) は、図 2.23 に示すように、隣接ブロックの動きベクトルによる補償画素値との間で重み付け平均を求めることで予測する方式である。本図中において、 Z は対象ブロック、 R_1 と R_2 は垂直隣接ブロックの動きベクトルによる補償であり、 S_1 と S_2 は垂直隣接ブロックの動きベクトルによる補償である。これらの動き補償によって予測ブロックは、 P, Z, R, S, H の各成分について式 (2.5) から求められる。

$$p(i, j) = [z(i, j) \cdot h_0(i, j) + r(i, j) \cdot h_1(i, j) + s(i, j) \cdot h_2(i, j) + 4] / 8 \quad (2.5)$$

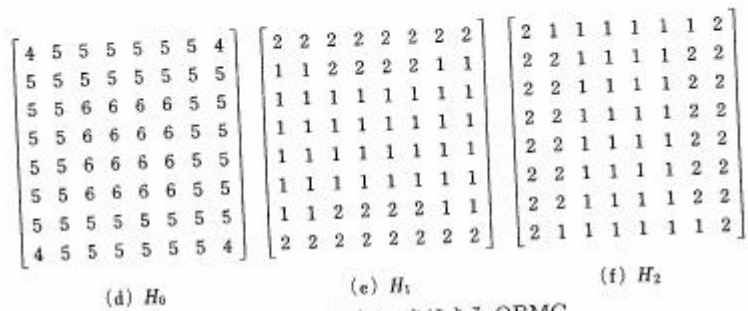
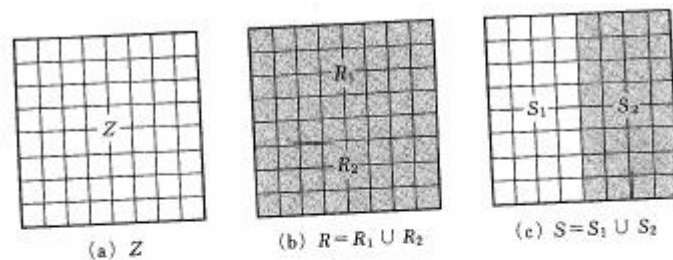


図 2.23 8×8 ブロックによる OBMC

また、MPEG 4 VO 符号化の特徴の1つであるスプライト符号化は、ビデオゲームやインターネット上のコンテンツ表現に有用である。これは、背景分離を行って抽出された VO を幾何変換して VOP ごとに符号化するもので

幾何変換には、静止、平行移動、アフィン変換、および平面透視変換の種類が用意されている。

任意形状 VO 符号化は、対象物体単位の符号化で、画素値、形状、および深度 (CG 分野ではアルファデータと呼ぶ) 情報に分けて符号化を行う。任意形状 VO の生成法としては、自然画像からの背景分離、自然画像からの切り出し、CG、およびアニメとがある。また生成後自然画との重ね合わせ時に VOP 境界との不連続性、特に高周波成分除去のために VOP 境界マクロブロック処理として、ベクトルパディング処理を行う。

コンピュータ・グラフィックス画像との親和性を確保するために MPEG 4 では、以上に述べた VO 符号化に加えて以下のような3つの手法を取りいれている。

第1にテキストチャ符号化に Wavelett 符号化を採用し、スケーラビリティを確保。

第2に3角形要素によるメッシュ符号化を行い、頂点座標の移動量表現法を採用。

第3に顔モデルを表す FDP (Facial Definition Parameter) と動きを表す FAP (Facial Animation Parameter) による符号化法を採用。

以上に述べたように MPEG 4 は、原理的には MPEG 1, 2, H.263 などと自然画処理の数学的基礎は共通しているが、CG 的な要素が入り、よりソフトウェアによる処理を想定しているところに新しさがある。

フレーム間予測

複数参照フレーム

従来技術では、フレーム間予測において参照フレームとして指定できるフレームは、Pフレームについては直前のI、Pフレーム、Bフレームについては直前および直後のI、Pフレームに固定。

複数の参照フレームを持つことによって、例えばシーンチェンジや移動物体を考慮してより前のフレームを参照フレームとして指定することが可能。

また、Bフレームについては未来方向のフレームを使わずに過去の2フレームを参照フレームとして指定したり、別のBフレームを参照フレームとして指定することが可能となっている。

可変ブロックサイズ

従来技術では、動き補償の単位は 16×16 画素のマクロブロックが基本で、H.263およびMPEG-4においては 8×8 画素ブロック単位の動き補償も利用。

H.264ではさらに単位ブロックサイズを追加し、 16×16 , 16×8 , 8×16 , 8×8 の4種類から選択可能。さらに、 8×8 画素ブロックについては、 8×8 , 8×4 , 4×8 , 4×4 の4種類のサブブロック分割も指定可能。

多数のブロックサイズを利用することで、形状や動きに適したブロックから予測が可能。

重み付け予測

H.264では、従来方式では画質向上が困難であった、フェードやディゾルブなどの特殊効果が用いられている動画の画質向上のため、参照フレームの予測誤差に重み付け係数を掛けてデコードする、重み付け予測(Weighted Prediction)を採用。

フェードやディゾルブは、前フレームと現フレームで一定のオフセットがかかったような画像であるため、そのことで予測差分に大きな値が生じることとなり、MPEG4などでは画質劣化の原因として問題となっていた。

1/4画素精度動き補償

動き補償の精度としては、MPEG-4 ASPで導入された1/4画素精度動き補償を使用。ゆっくり動くパンなどで特に効果的である。

1/2画素精度動き補償では6tapフィルターを用いて高周波まで再現を行っており、MPEG4で使用された線形補間よりも再現性が向上。1/4画素の生成は、再現性の高い1/2画素を用いてその線形補間で作成。

ご清聴ありがとうございました