

MPEG-4 の任意形状オブジェクト符号化技術

Video Coding Technology for Arbitrarily Shaped Objects in MPEG-4 Standards

山口 昇

YAMAGUCHI Noboru

任意形状オブジェクト符号化技術は、MPEG-4(Moving Picture Experts Group-phase 4)で初めて採用された従来の規格にはない機能を実現する技術である。矩形(くけい)の動画像を対象としていた従来の技術とは異なり、任意形状オブジェクト符号化技術では、動画像中の被写体(例えば、人物)のように任意形状の物体単位で圧縮が可能となる。

当社は、MPEG-4の規格化作業に参画して任意形状オブジェクト符号化の規格化に大きく寄与するとともに、応用アプリケーションの検討も併せて行っている。

Video coding technology for arbitrarily shaped objects is a technology for realizing this function, which is not in the conventional standard, adopted for the first time in the standardization of MPEG-4. Unlike the conventional technology for rectangular video, this technology makes compressibility possible for an object of arbitrary shape such as the subject in a video (for example, a person).

Toshiba has been taking part in the standardization of MPEG-4, and has significantly contributed to the standardization of video coding technology for arbitrarily shaped objects while studying its applications.

1 まえがき

MPEG-4における基本機能の大きな柱の一つとして、任意形状オブジェクト符号化が規格化された。当社は、規格化作業の初期段階から参画しており、当社提案も数多く規格に採用されている。

ここでは、任意形状オブジェクト符号化技術の概要及び当社における応用例を紹介するとともに、MPEG-4のビデオパートにおける任意形状オブジェクト符号化方式の詳細を、当社の提案を基に規格化された2値形状符号化方式を中心に述べる。

2 任意形状オブジェクト符号化の概要

MPEG-4における任意形状オブジェクト符号化のイメージを図1に示す。MPEG-4では、MPEG-1、MPEG-2と同様な矩形のビデオシーケンス(矩形オブジェクト)だけでなく、ビデオから抽出された人物など任意形状のビデオシーケンス(任意形状オブジェクト)や、物体の形状や手書き文字といったテクスチャ情報を持たない2値画像のビデオシーケンス(2値形状オブジェクト)を圧縮することができる。

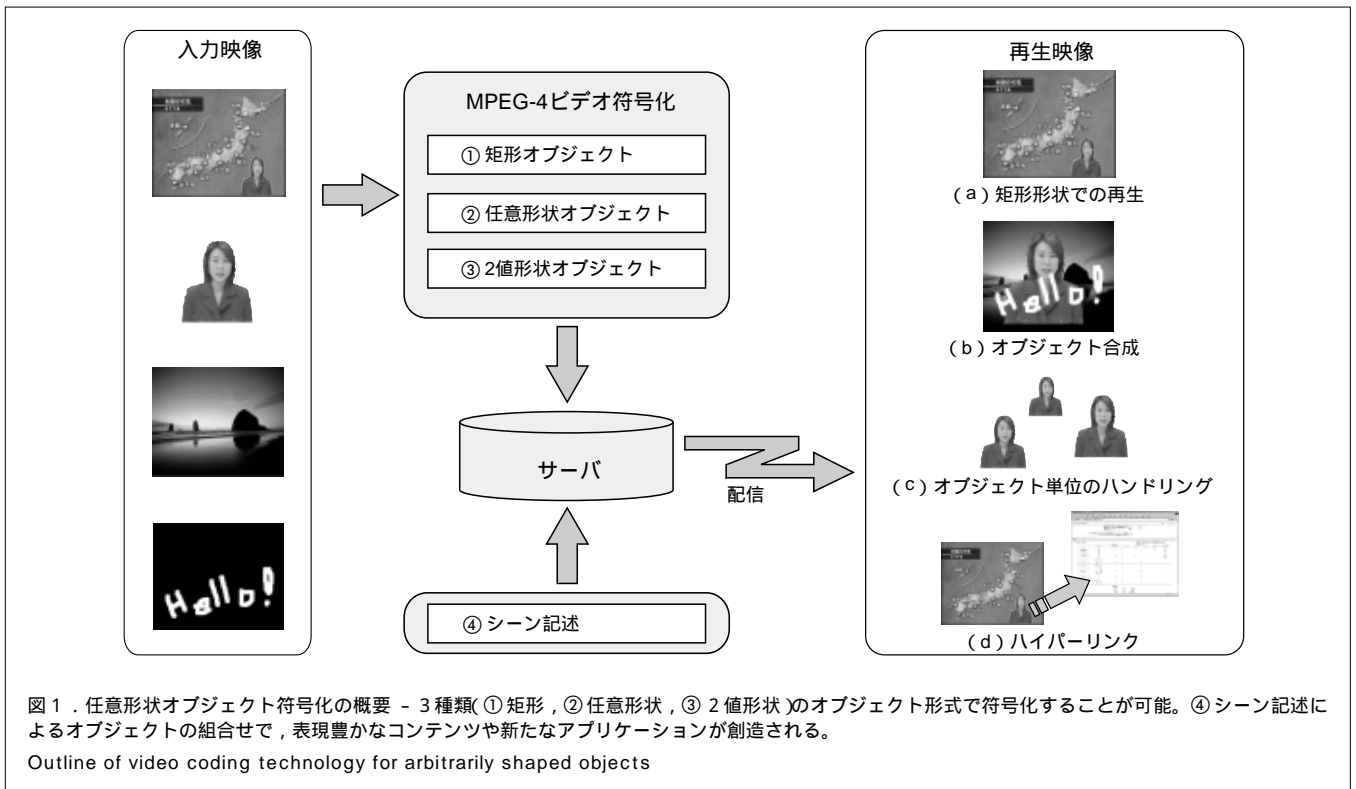
このようなオブジェクト符号化の機能を持つコンテンツは、“矩形オブジェクトの再生”はもとより、背景を入れ替えたり複数オブジェクトを“合成して再生”することができる。また、

オブジェクトごとにデータが分かれているため、特定のオブジェクトを拡大・縮小表示したり、一つのビットストリームから複数のオブジェクトを表示するなどの“オブジェクト単位のハンドリング”ができる。更に、あるオブジェクトに対して関連情報をリンクするなどの“ハイパーリンク機能”も容易に実現できる。

このような機能を実現するために、オブジェクト符号化のコンテンツには、オブジェクトの合成規則やオブジェクトと付帯情報の関係などを記した、シーン記述のデータが必要となる。このシーン記述に関しては、MPEG-4のシステムパートにおいてBIFS(Binary Format for Scene description)なるシーン記述手法が規格化されている。ただし、シーン記述にBIFSを使用することは必須ではなく、応用形態に応じて他のシーン記述手法を使うこともできる。

当社は、規格化作業と並行して、任意形状オブジェクトの符号化を実現するソフトウェアの開発を行った。そして、1999年5月に、世界に先駆けてMPEG-4ビデオ規格に準拠した任意形状オブジェクト符号化機能を備える、“Mobile Motion_{TM}V3”⁽¹⁾を発売している。

その後も、ウェブプレゼンテーションツール⁽²⁾や、この特集の別の論文(p.26参照)で紹介する顔チャット_{TM}システムなどの応用アプリケーションに、任意形状オブジェクト符号化技術を適用している。



3 任意形状オブジェクト符号化方式の詳細

3.1 形状情報の種類
任意形状オブジェクト符号化は, 形状符号化とテクスチャ (=オブジェクト内の画素値)符号化に分けて行われる。

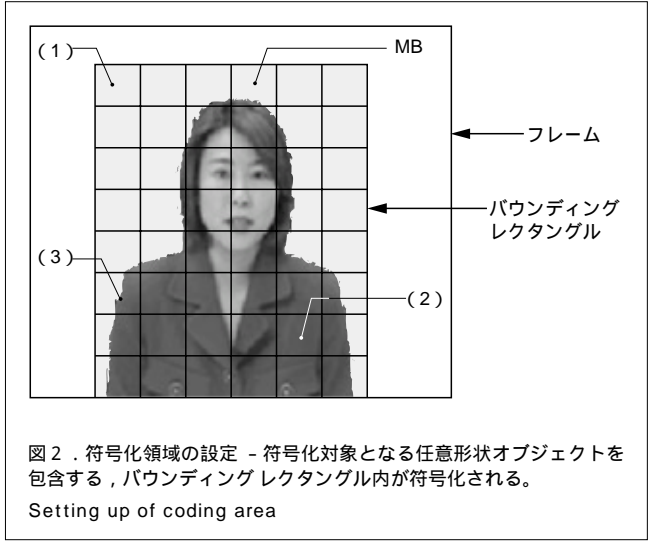
形状符号化には, オブジェクトを合成する際に利用されるアルファ値(値)の違いにより, 2値形状符号化(Binary Shape Coding)と多値形状符号化(Gray-scale Shape Coding)の二つのタイプに分類される。

3.1.1 2値形状符号化 値が, “0”か“ (=0 ~ 255)”の2値, つまり画素ごとにオブジェクト内に属するか否かのみを表すため, 値がオブジェクト内で一定となる。オブジェクトの形状情報が形状符号化の対象となる。

3.1.2 多値形状符号化 オブジェクト内の画素ごとに, 任意の値を取りうる。したがって, 画素ごとに値を変化させることが可能なため, 2値形状符号化と比較して自然な合成が可能となる。多値形状の符号化手順は, まず, オブジェクトの形状を2値形状符号化で符号化し, その後, 画素ごとに値をテクスチャ符号化と同様に符号化する。

このように, 形状符号化の本質は2値形状符号化にあるため, ここでは2値形状符号化にフォーカスを当てて述べる。

3.2 バウンディング レクタングル
まず図2に示すように, 符号化すべきオブジェクト(図2の例では, 女性の上半身)を包含する矩形領域(バウンディング



レクタングル)を設定する。バウンディング レクタングルの大きさ及び位置は別途符号化される。

次に, バウンディング レクタングル内を, 16 × 16画素のマクロブロック(MB)単位に分割して符号化処理を行う。このようにして分割されたMBごとに, 形状符号化とテクスチャ符号化に分けて符号化が行われる。

ここでは詳しく述べないが, テクスチャの符号化には, 動き補償予測(MC)+ 離散コサイン変換(DCT)符号化が適用されている。

図2を見るとわかるように, MBは次の3種類のタイプに分

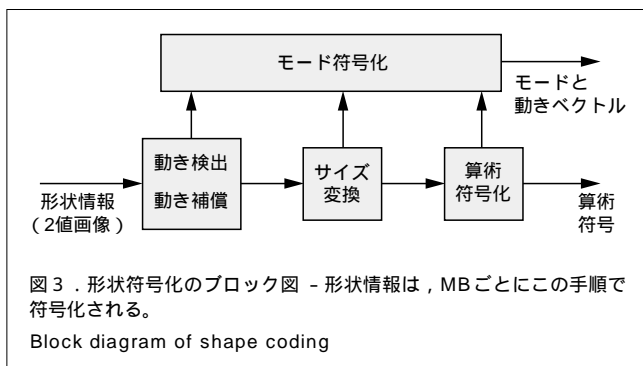
類される。

- (1) MB内のすべての画素がオブジェクトに属さない。
 - (2) MB内のすべての画素がオブジェクトに属する。
 - (3) MB内の一部の画素がオブジェクトに属する。
- (1)のMBには、テクスチャの情報が含まれないため、この種類のMBに対してテクスチャ符号化は行われない。

3.3 形状符号化⁽³⁾⁽⁴⁾

図3は、形状符号化のブロック図である。形状符号化は、2値画像として符号化処理を行う。

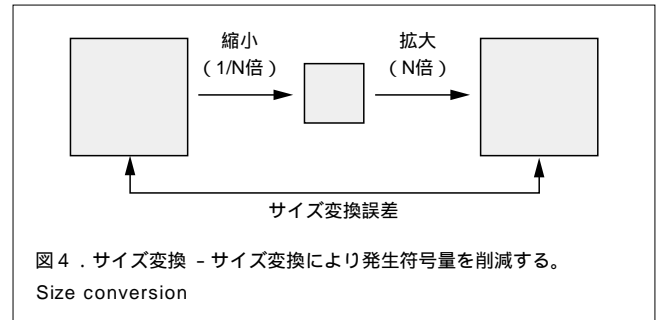
まず最初に、各MBが図2の(1)~(3)のいずれに属するかを判定し、形状情報が存在する(2)(3)のタイプのMBに対して“動き検出/動き補償”；“サイズ変換”；“算術符号化”が行われる。



3.3.1 動き検出/動き補償 動き検出は、ブロックマッチングにおける誤差(2値画像のため、排他的論理和で誤差を評価する)の累積値が最小となるように動きベクトル(MV)を検出する。ここで、ブロックマッチングの範囲は予測MVの周囲±16画素としている。この予測MVは、周囲のMBにある形状符号化のMVとテクスチャ符号化のMVの中から適切なものを選択する。ただし、予測MVを用いてブロックマッチングを行い、十分小さな動き補償誤差が得られれば、その予測MVをそのまま該当MBの形状符号化MVとする。

3.3.2 サイズ変換 後述する算術符号化は可逆符号化であるため、符号化すべき画素数が少ないほど発生符号量も少なくて済む。しかし、符号化すべきMBサイズを小さくすることは、元の形状情報が劣化して受信側に伝送されることを意味している。これは、テクスチャ符号化のDCT係数のように量子化を行うことで発生符号量を低減することに相当する。

サイズ変換比率には原寸大(1倍)、縦横1/2倍、縦横1/4倍の3種類があり、図4に示すように、サイズ変換誤差が、あらかじめ設定されたしきい値以下となる最小のサイズを選択する。ここで、サイズ変換誤差は、ブロックマッチング誤差



と同様に排他的論理和で評価する。

3.3.3 算術符号化 算術符号化は、図2の(3)タイプのMBに対して行われる。ここでは、JBIG(Joint Bi-level Image coding experts Group)に採用されている算術符号を基本とした手法が用いられている。サイズ変換されたMB内の各画素ごとに値の発生確率を決定していき、MB単位で閉じた算術符号化を行う。なお、各画素の値の発生確率を求める際には、値の空間・時間的相関を考慮して、以下のように確率を決定する。

- (1) フレーム内符号化モードの場合 現オブジェクト内の10画素を参照する。
- (2) フレーム間符号化モードの場合 現オブジェクト内の4画素と、動き補償を行った前オブジェクト内の5画素とを参照する。

また、算術符号化処理は、水平方向に順次行うモードと垂直方向に行うモードとがあり、発生符号量の少ないモードを選択できる。

3.3.4 モード符号化 各種モード情報の中には、フレーム内やフレーム間の相関が利用できるものが多い。例えば、あるMBが図2の(1)タイプであった場合、時間的に次のオブジェクトでも同じ位置にあるMBは、同じタイプのMBである確率が高い。つまり、現フレームの各モードを符号化する際に、前フレームのモードに応じて適切な可変長符号を選択することで、発生符号量を削減している。

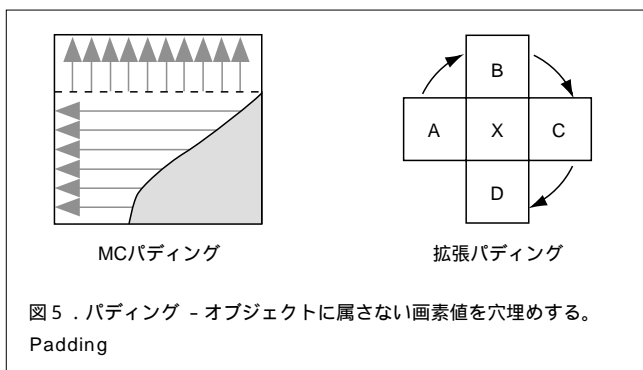
3.4 テクスチャ符号化

図2の(2)(3)タイプのMBに対してテクスチャ符号化を行う。

3.4.1 パディング あるMBを動き補償予測する場合に、MVの指し示す先がオブジェクトの外側になる場合がある。そのため、動き補償のリファレンスとなるフレームのバウンディングレクタングル内における、オブジェクトの外側を埋める処理が採用されている。この処理は、パディングと呼ばれている。

図5を用いてパディングの説明をする。図2の(2)タイプのMBには“MCパディング”を行う。

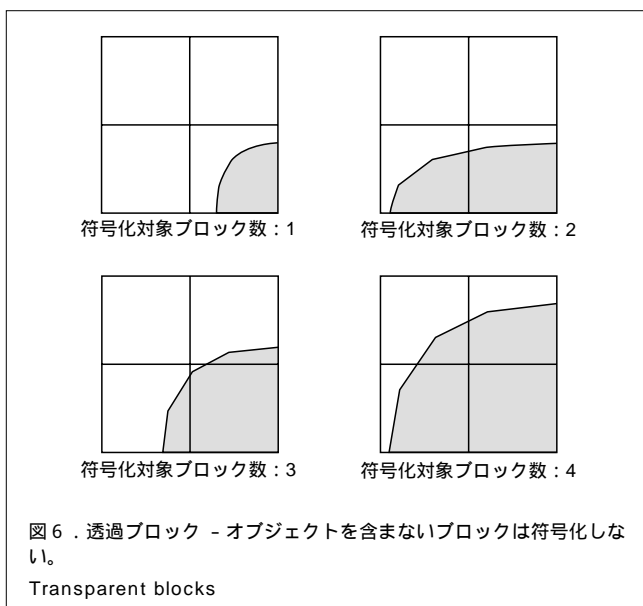
- (1) オブジェクトのもっとも外側にある画素値を、左右にコピーして穴埋めする。



- (2) 重なった拡張部分はその平均値で置き換える。
- (3) オブジェクトのもっとも外側にある画素値を、上下にコピーして穴埋めする。

図2の(1)タイプのMBには拡張パディングを行う。Xをパディング対象のMBとすると、次のようになる。

- (1) Xの4辺に接するMBであるA,B,C,Dに着目し、A,B,C,Dの順番でオブジェクトを含むMBか否かを調べる。
- (2) 最初に見つかったMBの画素値のうち、Xと接する画素値をX内にコピーすることで穴埋めする。



- (3) A,B,C,Dのすべてにオブジェクトが含まれていなければ、X内を値128で穴埋めする。

3.4.2 透過ブロック テクスチャ符号化では、動き補償予測誤差を8×8画素ブロックごとにDCT符号化する。また、MPEG-4には、8×8画素ブロックごとにMVを送るモードがある。しかし、図2の(3)タイプのMBには、図6に示すように、オブジェクトを含まないブロック(透過ブロック)が存在する。このようなブロックに対しては、DCT符号化やMVの符号を送らないことで発生符号量を削減している。なお、透過なブロックの情報は、形状符号化の結果から一意に決まる。

4 あとがき

MPEG-4の特長の一つであり、当社が規格化に寄与した任意形状オブジェクト符号化技術について述べた。この技術を用いることで、新たなマルチメディアアプリケーションの世界が広がることを期待する。

文献

- (1) (株)東芝 . Mobile Motion™ (オンライン), <http://www.toshiba.co.jp/mmotion/>, (参照 2002-04-02)
- (2) (株)東芝 . ウェブプレゼンテーションツール . (オンライン), <http://www2.toshiba.co.jp/mmlab/object/> (参照 2002-04-02)
- (3) 山口 昇,ほか . MMRを用いた物体形状情報符号化法の検討 . 電子情報通信学会技報 . IE96 - 142 ,1997 ,p.9 - 14 .
- (4) 渡邊敏明 . マルチメディアを目指す MPEG-4 標準化動向 : 形状符号化 . 映像情報メディア学会誌 . 51 ,12 ,1997 ,p.1984 - 1986 .



山口 昇 YAMAGUCHI Noboru

研究開発センター マルチメディアラボラトリー研究主務。
 動画符号化技術の研究・開発に従事。電子情報通信学会、
 映像情報メディア学会会員。
 Multimedia Lab.