

## MPEG-4 ビデオ符号化におけるグローバル MC の効果について

如澤 裕尚 上倉 一人 渡辺 裕 小寺 博

NTT ヒューマンインタフェース研究所

〒 238-03 横須賀市武 1-2356

TEL: 0468-59-2975, FAX: 0468-59-2829

Email: jozawa@nttvdt.hil.ntt.jp

本報告では、1996年1月に発行された MPEG-4 Video Verification Model Version 1.0 (VM) をベースに、グローバル動き補償 (MC) とローカル MC からなる二段階 MC を加えた方式を提案する。本方式では、第一段のグローバル MC で画面全体の大域的な動きを補償し、さらに第二段のローカル MC でマクロブロックごとの局所的な動きを補償する。グローバル MC は、パン、チルト、ズームといったカメラの動きを補償するが、画面中にはグローバル MC で予測できない部分もあるため、マクロブロックごとにグローバル MC を適応的に ON/OFF する。計算機シミュレーションの結果、グローバル MC とその適応 ON/OFF 制御を備えた提案方式は、特に高速のパンニングシーンに有効であり、オリジナルの MPEG-4 VM に比べ復号画像の SN 比が最高 3 dB 程度改善されることを確認した。

## Global Motion Compensated Prediction for MPEG-4 Video Coding

Hirohisa Jozawa, Kazuto Kamikura, Hiroshi Watanabe, and Hiroshi Kotera

NTT Human Interface Laboratories

1-2356 Take, Yokosuka, Kanagawa 238-03, JAPAN

TEL: +81-468-59-2975, FAX: +81-468-59-2829

Email: jozawa@nttvdt.hil.ntt.jp

This paper describes a two-stage motion compensation (MC) method based on MPEG-4 Video Verification Model Version 1.0 (VM) issued on Jan. 1996. The proposed two-stage MC consists of global MC and local MC. First, in the proposed MC method, global motion of a frame such as panning, tilting or zooming is detected and compensated. Then, local MC of each macroblock is carried out for the global motion compensated reference frame. Since the global MC cannot predict all regions in a frame, it is adaptively turned on or off for each macroblock. Simulation results show that the proposed video coding method using the two-stage MC performs effectively for fast panning sequences. Maximum PSNR improvements are about 3 dB in comparison with MPEG-4 VM without two-stage MC.

## 1 はじめに

ISO/IEC JTC1/SC29/WG11では、MPEG-2に続く次世代オーディオ・ビジュアル符号化規格としてMPEG-4 (ISO/IEC 14496)の標準化作業を進めている。MPEG-4のねらいは、以下の三つである。

- (a) 映像コンテンツ単位のインタラクティブ性の実現
- (b) 高圧縮化
- (c) 誤り耐性、および誤り回復力の強化

すなわち、従来標準にない新機能を実現すること、同じ符号化レートで比較した際に従来標準を凌駕する画質を実現することが目標である。

(a)のコンテンツ単位のアクセスを実現するためには、現在、フレーム、Group Of Blocks (GOB)、マクロブロックにより記述されているビデオ情報を、人物や背景といったオブジェクト単位に記述することが必要となる。このため、1996年1月に発行されたVideo Verification Model [1] (MPEG-2のTest Modelに相当)では、ビデオ・オブジェクト・プレーン (VOP) という概念が導入された。VOPは各オブジェクトのYUV信号と、その形状と位置を表すセグメンテーション・マスク (または多値の $\alpha$ プレーン) からなる。VOPの数、すなわちオブジェクト数は任意に設定することができ、オブジェクトの切り出し方法は問わない。各VOPの符号化はITU-T H. 263 [2]をベースとした方法で符号化される。

一方、(b)の高圧縮化のための検討も精力的に進められている。伝送レートが低いモバイル通信はもろんのこと、インターネット上での快適な画像通信を実現する上でも必須である。MPEG-4ビデオは、当初64 kb/s未満を想定していたが、現在は10 kb/s～1 Mb/sの幅広いレートがターゲットとなっている。このため、筆者等は64 kb/s未満の超低レートのみならず、64 kb/s以上のレートでも高い性能を発揮するH. 263をベースとし、グローバルMCとローカルアフィンMCを加えた方式を提案した[3]。グローバルMCは、カメラの動きによって生じるパンニングやズームを補償し、ローカルアフィンMCは、物体の平行移動、回転、拡大/縮小を補償するものである。グローバルMCはMPEG-4コア実験の一つとして選ばれ、VMに対して有意な性能差を出すべく複数機関で検討が進められている。

本稿では、グローバルMCとローカルMCの二段構成からなる提案方式をMPEG-4 VM [1]に組み込み、VM + グローバルMCの性能をオリジナルのVMと比較したので報告する。また、前回の

提案[3]におけるグローバル動き検出方法にも改良を施し、動き検出精度の改善も図った。

## 2 グローバルMC

カメラ操作によって生じるパン、チルト、ズームは画面全体の動きである。パン、チルトは平行移動量であるので、従来のマクロブロックごとの動き補償でも予測できる。しかし、各マクロブロックに同じベクトル値を持たせるよりも、画面全体を一組の動きパラメータで記述する方が効率的である。また、従来の平行移動モデルによる動き補償では、ズーム補償は不可能である。そこで、パン、チルト、ズーム等の画面全体の動きを大域的に補償するため、グローバルMC (GMC) [4]を導入する。

提案方式では、まずグローバル動き検出により三種類のカメラの動き、パンニング量 $H$ 、チルト量 $V$ 、ズーム率 $Z$ を求める。次に、これら三つのパラメータと参照画像からグローバル動き補償予測画像を生成する。以下、これらの手法について説明する。

### 2.1 グローバル動き検出 (GME)

現在の符号化フレームと参照フレームとの間で、三つのグローバル動きパラメータ $H, V, Z$ を求める。グローバル動き検出 (GME) は以下の三つのステップで行われる。

#### 2.1.1 ステップ1

符号化対象画像と参照画像 (局部復号画像ではなく原画像) を縦横1/2のサイズに縮小する。縮小にあたっては以下の二次元ローパスフィルタを用いる。

$$\begin{matrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{matrix} / 16$$

次に、 $8 \times 8$ ブロックごとに、ブロックマッチング法により平行移動ベクトルを求める。探索範囲は $\pm 56$ 画素で、輝度信号に対してのみ行う。

#### 2.1.2 ステップ2

符号化対象画像と参照画像 (局部復号画像ではなく原画像) の間で、 $16 \times 16$ ブロックごとに、ブロックマッチング法により平行移動ベクトルを求める。ステップ2では、ステップ1で求められた動きベクトルを初期ベクトルとし、その周囲の $\pm 15$ 画素を探索する。このため、最終的に求められる動きベクトルのレンジは $\pm 127$ となる。

### 2.1.3 ステップ3

原点,  $X$  軸, または  $Y$  軸に関し対称な二つのマクロブロック  $MB_1, MB_2$  (ペアブロックと呼ぶ) の平行移動ベクトル  $(V_{1x}, V_{1y}), (V_{2x}, V_{2y})$  からグローバル動きパラメータ  $Z, H, V$  を求める. 各パラメータは以下の式により計算される.

$$H = \frac{-2I_1W_1 + \{2(i_1^2 + i_2^2) + J_2^2\}W_2 + I_1J_1W_3}{2\{I_2^2 + J_2^2\}} \quad (1)$$

$$V = \frac{-2J_1W_1 + I_1J_1W_2 + \{I_2^2 + 2(j_1^2 + j_2^2)\}W_3}{2\{I_2^2 + J_2^2\}} \quad (2)$$

$$Z = \frac{2W_1 - I_1W_2 - J_1W_3}{I_2^2 + J_2^2} \quad (3)$$

$$I_1 = i_1 + i_2 \quad (4)$$

$$I_2 = i_1 - i_2 \quad (5)$$

$$J_1 = j_1 + j_2 \quad (6)$$

$$J_2 = j_1 - j_2 \quad (7)$$

$$W_1 = i_1V_{1x} + i_2V_{2x} + j_1V_{1y} + j_2V_{2y} \quad (8)$$

$$W_2 = V_{1x} + V_{2x} \quad (9)$$

$$W_3 = V_{1y} + V_{2y} \quad (10)$$

ここで,  $(i_1, j_1), (i_2, j_2)$  は, それぞれ  $MB_1, MB_2$  の中心座標である. なお, 座標原点は画面左上ではなく, 画面中心である.

表1に示すように, パラメータ  $H$  と  $V$  は偶整数に量子化され,  $\pm 126$  の範囲にクリップされる. また, ズームパラメータ  $Z$  は  $1/128$  の倍数となるよう量子化され,  $\pm 31/128$  の範囲にクリップされる. したがって, パラメータ  $H, V$  は各7ビット,  $Z$  は6ビットの等長符号で表現される.

Table 1 Global Motion Range

Parameter	Motion range	Quant. step
H	$\pm 126$	2
V	$\pm 126$	2
Z	$\pm 31/128$	$1/128$

ペアブロックの選び方は次の三通りのパターンである. まず, 始めに, 図1に示すように, ペアブロックを原点に関し対称となるよう選ぶ(パターン1). すなわち,  $i_2 = -i_1, j_2 = -j_1$  である. 原点に関し対称となる全てのペアブロックに対して(2)~(3)式の計算を実行し,  $H, V, Z$  それぞれの発生頻度を求める.

次に, 図2に示すように, ペアブロックを  $X$  軸に関し対称となるよう選ぶ(パターン2). すなわち,  $i_2 = i_1, j_2 = -j_1$  である.  $X$  軸に関し対称となる全てのペアブロックに対して(2)~(3)式の計算を実行し,  $H, V, Z$  それぞれの発生頻度を求める.

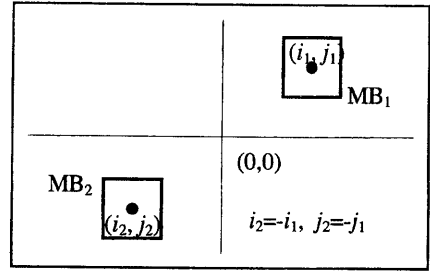


Figure 1 Block pair for GME (Pattern 1)

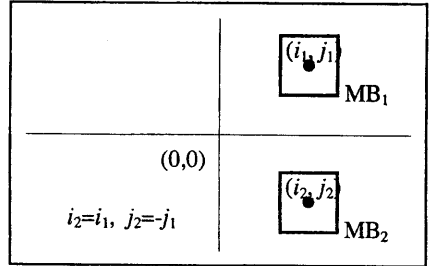


Figure 2 Block pair for GME (Pattern 2)

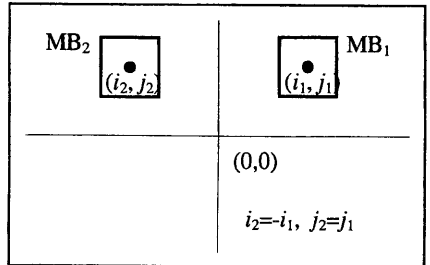


Figure 3 Block pair for GME (Pattern 3)

三番目に, 図3に示すように, ペアブロックを  $Y$  軸に関し対称となるよう選ぶ(パターン3). すなわち,  $i_2 = -i_1, j_2 = j_1$  である.  $Y$  軸に関し対称となる全てのペアブロックに対して(2)~(3)式の計算を実行し,  $H, V, Z$  それぞれの発生頻度を求める.

最後に,  $H, V, Z$  それぞれについて, パターン1~3の三つの度数分布をマージし, 最も発生頻度の高い値を最終的なグローバル動きパラメータとする. これは, 画面の中で最も支配的なパラメータを選ぶと同時に, 雑音や一部の局所的な動きによる誤検出を避けるためである.

前回の提案[3]ではパターン1のみから  $H, V, Z$  を求めていたが, 今回パターン2と3を加えることにより, バニング量  $H$ , チルト量  $V$  の検出精度が向上する.

## 2.2 グローバルMC予測

次段のローカルMEおよびMCにおける参照画像として、グローバルMC画像を生成する。グローバルMC予測は、画素単位に行なわれ、画素 $(i, j)$ におけるベクトル $v_x(i, j)$ ,  $v_y(i, j)$ は以下の式により表現される。

$$\begin{pmatrix} v_x(i, j) \\ v_y(i, j) \end{pmatrix} = \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix} \begin{pmatrix} i \\ j \end{pmatrix} + \begin{pmatrix} H \\ V \end{pmatrix} \quad (11)$$

但し、 $i - j$ 座標系は画面の中心が原点であるので、画面左上を原点とすると、(11)式は以下のよう書き換えられる。

$$\begin{pmatrix} v_x(x, y) \\ v_y(x, y) \end{pmatrix} = \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} H - L_x Z/2 \\ V - L_y Z/2 \end{pmatrix} \quad (12)$$

ここで、 $L_x$ ,  $L_y$ は、それぞれ画面の横方向および縦方向のサイズである。(12)式に示されるように、本提案における動きモデルは、アフィン変換のサブセットとなっており、パラメータ数を3から6に増やすことにより、容易にアフィン変換に拡張できる。

画素 $(x, y)$ の予測値は、参照画像中の座標 $(x + v_x(x, y)$ ,  $y + v_y(x, y))$ における値として求められる。 $v_x(x, y)$ ,  $v_y(x, y)$ は小数値を持つが、小数点以下については0.5刻みに量子化する。すなわち、各画素のグローバル動きベクトルは、ハーフピクセル精度となる。

また、動きベクトルが画面外を参照する場合は、H. 263 [2]のUnrestricted Motion Vector Modeと同様に、存在しない画素については最近傍のエッジ画素を予測値とする(図4)。

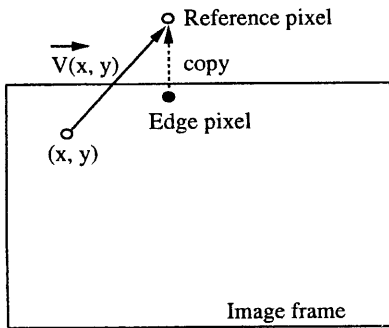


Figure 4 Unrestricted motion vector

## 3 予測モード選択

### 3.1 グローバルMC ON/OFF 判定

グローバルMCは画面全体の大域的な動きを補償するものであり、局所的な動きが支配的な領域で

は、グローバルMCは有効に作用しない。また、カメラが静止している場合、グローバルな動きが正しく検出されなかった場合等、グローバルMCにより逆に予測効率が低下する場合がある。したがって、全ての領域にグローバルMCを用いることは得策ではない。

以上の問題を解決するため、グローバルMCをマクロブロックごとに適応的にON/OFFする。図5に示すように、グローバルMCを施した参照画像と、グローバルMCをかけない通常の参照画像の両者に対してローカルMCを行う。グローバルMCとローカルMCの双方を用いた場合の誤差絶対値和 $SAD_{GMC_{on}}$ と、グローバルMCを用いずにローカルMCのみで予測した際の誤差絶対値 $SAD_{GMC_{off}}$ を以下の式により計算する。

$$SAD_{GMC_{on}} = \sum_{MB} |P - \hat{P}_{GMC_{on}}| \quad (13)$$

$$SAD_{GMC_{off}} = \sum_{MB} |P - \hat{P}_{GMC_{off}}| \quad (14)$$

ここで、 $P$ は現在フレームの画素値、 $\hat{P}_{GMC_{on}}$ はグローバルMCをONにした際の予測値を示す。また、 $\hat{P}_{GMC_{off}}$ はグローバルMCをOFFにした際の予測値である。両者を比較し、

$$SAD_{GMC_{on}} < SAD_{GMC_{off}}$$

の際にグローバルMCを用いる。

グローバルMCのON/OFFは1ビットの符号で表現する。このフラグはマクロブロックヘッダに挿入される。

### 3.2 INTRA/INTER 判定

図5に示すように、グローバルMCのON/OFF判定後、各マクロブロックごとにINTRA/INTER判定を行なう。INTRAモードは、

$$\sum_{MB} |P - \bar{P}| < SAD_{INTER} - 500 \quad (15)$$

の際に選択される。ここで、 $\bar{P}$ ,  $SAD_{INTER}$ は以下の通りである。

$$\bar{P} = \left( \sum_{MB} P \right) / 256 \quad (16)$$

$$SAD_{INTER} = \min(SAD_{GMC_{on}}, SAD_{GMC_{off}}) \quad (17)$$

## 4 計算機シミュレーション

### 4.1 シミュレーション条件

グローバルMCと、その適応ON/OFFの効果を見るため、以下の三方式の性能を比較した。

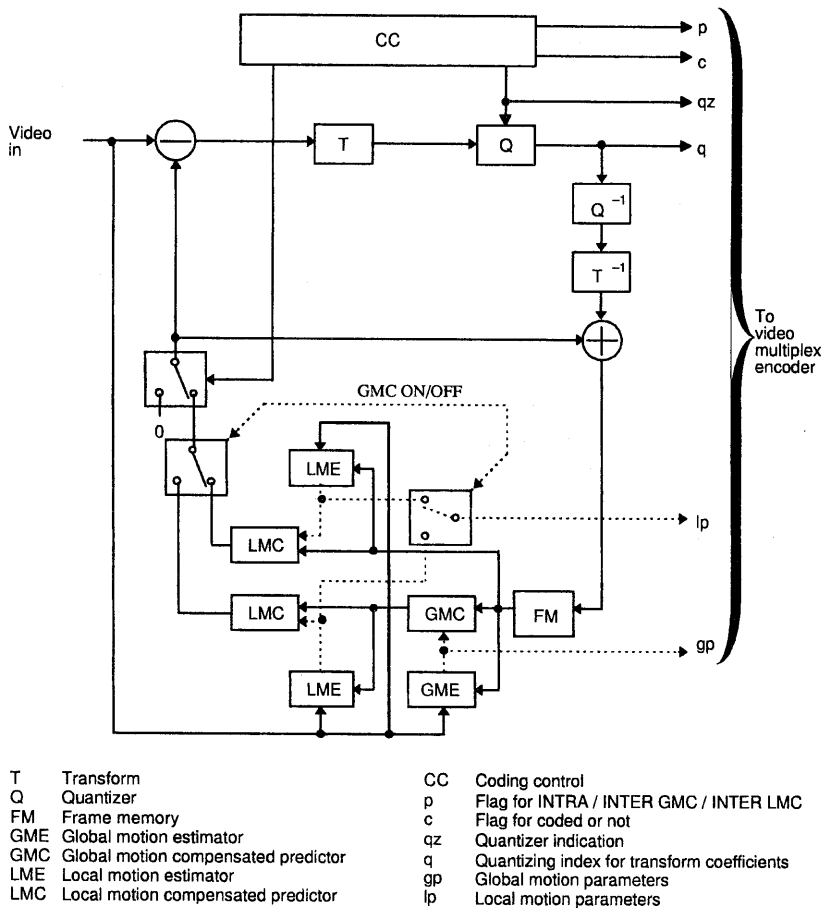


Figure 5 Blockdiagram of the proposed encoder

- (a) MPEG-4 VM Version 1.0 [1]
- (b) VM + グローバル MC (常時 ON)
- (c) VM + グローバル MC (適応 ON/OFF)

ここで、VMにおけるビデオ・オブジェクト・プレーン(VOP)の数は1である。したがって、実験におけるVMはH. 263とほぼ同じ仕様である。

表2に示すように、いずれの方式においても Unrestricted Motion Vector Mode (画面外参照動きベクトル), Advanced Prediction Mode (オーバーラップMC + 8 × 8ブロックMC)をONにしている。また、マクロブロックごとのMC(ローカルMC)の動きモデルは、今回はアフィン変換[3]ではなく、VMやH. 263と同じ平行移動モデルである。なお、ローカルMCの探索範囲は±32画素である。レートコントロールも、方式(a), (b), (c)全てに同じ方式[5]を用いている。各方式間の違いはグローバルMCの有無、あるいはグローバルMCをマクロブロックごとに適応的にON/OFF

する否かの違いのみである。なお、算術符号化とPB-frames ModeはVM 1.0では定義されておらず、今回の実験では使用していない。

テストシーケンスには、MPEG-4で用いられている“Foreman”, “Coast guard”, “Stefan”, “Table Tennis”の四つを用いた。“Foreman”, “Coast guard”はClass B画像と呼ばれ、中程度の空間ディテール、中程度の動き量を持つが、いずれも途中カメラが大きく振れる部分がある。MPEG-4では24, 48, 112 kb/sの符号化レートで実験されている。一方、“Stefan”と“Table Tennis”はClass C画像と呼ばれ、空間ディテールが高く、動きが大きいため、MPEG-4では320, 512, 1024 kb/sの符号化レートで実験されている。

#### 4.2 シミュレーション結果

“Foreman”, “Coast guard”を112 kb/sで符号化した際の復号画像SNRを図6, 図7に示す。処

Table 2 Simulation condition

	Method (a) (VM)	Method (b)	Method (c)
Global MC	OFF	All ON	Adaptive ON/OFF
Unrestricted Motion Vector	ON	ON	ON
Advanced Prediction (Overlapped MC, 8 × 8 MC)	ON	ON	ON
Syntax-based Arithmetic Coding*	OFF	OFF	OFF
PB-frames*	OFF	OFF	OFF
Local MC search area	±32	±32	±32

\*) Not defined in MPEG-4 Video VM Version 1.0 [1]

理フォーマットは CIF (Y: 352 × 288, C: 176 × 144), フレームレートは 10 frame/sec である。また, “Stefan”, “Table Tennis” を 320 kb/s で符号化した際の復号画像 SNR を図 8, 図 9 に示す。処理フォーマットは SIF (Y: 352 × 240, C: 176 × 120), フレームレートは 10 frame/sec である。

図 6 に示すように, “Foreman” においては, カメラが大きく振れる 200 フレーム以降からグローバル MC の効果が見られる。マクロブロックごとに適応的にグローバル MC を ON/OFF する方式 (c) (実線) は, オリジナルの VM (破線) に比べ最大 2.5 dB のゲインを得ている。また, カメラの動きが少ない 200 フレーム以前では, VM とほぼ同等の性能となっている。これに対し, 全マクロブロックに対してグローバル MC を用いる方式 (b) (点線) は VM, 方式 (c) のいずれよりも劣っており, 特にカメラが動かない前半では 1 dB 程度低い性能となっている。前半部分は人物の顔の動き (ローカルな動き) が支配的である上, カメラが若干ロール (回転) しており, グローバル動き検出がうまく行なわれていないためと考えられる。

“Coast guard” は, ほぼ全フレームに渡って背景が水平に等速度で動くシーン (但し, 80 フレームあたりでカメラが垂直に動く) であるが, 図 7 に示すように, グローバル MC の効果はほとんど見られない。“Coast guard” のように単純な動きは, ローカル動きベクトルの差分符号化により, 少ないベクトル量で十分な性能を出せるものと考えられる。

“Stefan” はカメラが左右に大きく振れる上, ズームも含まれているため, 図 8 に示すように, グローバル MC を用いた方式 (b), (c) は “Stefan” に対し非常に有効である。提案方式 (c) は VM を常時上回る性能を示し, 最大 3 dB のゲインを得ている。この効果は視覚的にもはっきり検知できるレベルであり, グローバル MC の効果が最もよく現れるシーケンスである。

最後に, “Table Tennis” では, 図 9 に示すように, 提案方式 (c) と VM はほぼ同等の性能である。

しかし, ズームを伴う 30 ~ 60 フレームにおいて, 提案方式 (c) は VM を 0.5 dB 程度上回っていることがわかる。

### 4.3 考察

以上に示したように, グローバル MC はカメラが高速に動くシーンやズームシーンに有効である。VM では, 動きベクトルのレンジは H. 263 と同じく [-32, 31.5] に制限されており, “Stefan” のように非常に速いパンには追従不可能である。これに対し, 提案方式ではグローバル MC で ±126 画素 (表 1), 第二段のローカル MC で [-32, 31.5] 画素の動きに対応できる。したがって, トータルでは [-158, 157.5] のレンジをカバーできる。VM の動きベクトルのレンジを, たとえば [-128, 127.5] に広げたとしても, 水平・垂直各コンポーネントごとに 2 ビットずつの付加が必要である。したがって, 1 マクロブロックあたり 4 ビットの付加情報となり, 1 フレームあたり 20 ビットのオーバーヘッドで済むグローバル MC より性能が劣ることは容易に推測できる。

グローバル MC において, マクロブロックごとの適応 ON/OFF は, 図 6 ~ 9 の結果より必須であると言える。しかし, ON/OFF のためのフラグがマクロブロックごとに 1 ビット必要であるため, 超低レートでは無視できないオーバーヘッドとなる。カメラが静止しているシーンでは, グローバル MC 自体の効果がないため, ON/OFF フラグが性能低下の原因となる。ON/OFF フラグの効率的な符号化方法が必要である。

## 5 むすび

カメラのパン, チルト, ズームを補償するグローバル MC と, その適応 ON/OFF 手法を提案し, MPEG-4 VM Version 1.0 に組み込んで VM に対する優位性を検証した。計算機シミュレーションにより, グローバル MC は特に “Stefan” のようにカメラが高速に動くシーンに有効であることを確認

した。グローバルMCにより、MPEG-4 VM に比べ復号画像SNRが最高3 dB程度改善できる。また、ズームシーンにも有効である。

グローバルMCの概念は、本稿のように画面単位の補償ばかりでなく、画面中の任意のオブジェクトの予測に対しても応用できる。また、グローバルMCの性能を一層向上させるためには、オーバーヘッド情報の効率的な符号化、回転等の高度な動きへの対応が必要となる。今後は、

- オブジェクトごとのグローバルMC
- グローバルMC ON/OFF フラグの効率的な記述方法
- より高度な動きモデル(アフィン等)

について検討する。さらに、1996年3月に発行されたVM Version 2.0 [6]に対する有効性を検証する。

#### 参考文献

- [1] MPEG Video Group, "MPEG-4 Video Verification Model Version 1.0," ISO/IEC JTC1/SC29/WG11 N1172, Jan. 1996.
- [2] Draft ITU-T Recommendation H. 263, "Video coding for low bitrate communication," Dec. 1995.
- [3] 如澤, 上倉, "適応グローバルMCとローカルアフィンMCを用いた低レート画像符号化", 情処学会研究報告, 96-AVM-12, pp. 21-28, 1996-02.
- [4] 上倉, 渡辺, "動画像符号化におけるグローバル動き補償法", 信学論 B-I, Vol. J76-B-1, No. 12, pp. 944-952, 1993-10.
- [5] Karl O. Lillevold, "Telenor H.263 encoder: tmn Version 1.6c," <ftp://bonde.nta.no/pub/tmn/software/>, Mar. 1996.
- [6] MPEG Video Group, "MPEG-4 Video Verification Model Version 2.0," ISO/IEC JTC1/SC29/WG11 N1260, Mar. 1996.

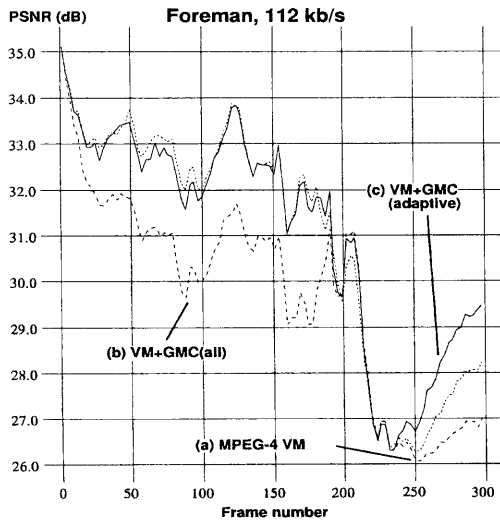


Figure 6 PSNR of decoded images.

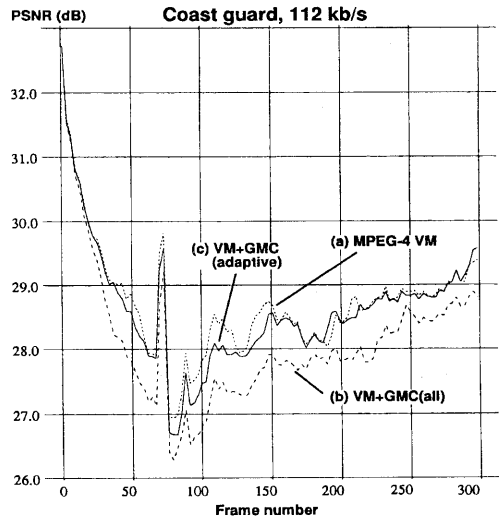


Figure 7 PSNR of decoded images.

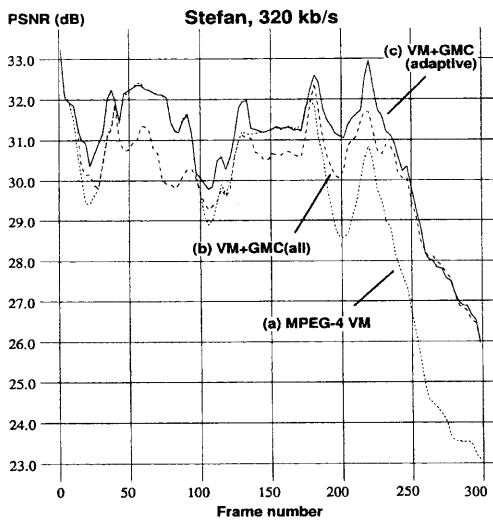


Figure 8 PSNR of decoded images.

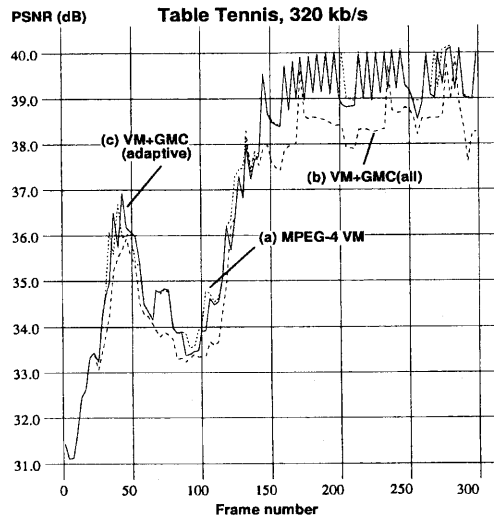


Figure 9 PSNR of decoded images.