

## 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

如澤 裕尚      上倉 一人

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、超低レートビデオ符号化勧告 ITU-T H. 263 をベースに、グローバル MC とローカルアフィン MC による二段構成の動き補償方法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 で物体ごとの局所的な動きを補償する。グローバル MC は、パニング、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 の ON/OFF を切り替える。一方、ローカル MC は H. 263 のオーバーラップブロック MC を基本としているが、動きモデルにアフィン変換を用いることにより、物体の平行移動ばかりでなく、回転、拡大、縮小等も補償する。計算機シミュレーションの結果、グローバル MC とローカルアフィン MC を加えた方式は動きの大きいシーンに有効であり、これらを用いない場合に比べ、復号画像の SN 比が 0.5~1 dB 程度向上することを確認した。

## Low Bit Rate Video Coding Using Adaptive Global MC and Local Affine MC

Hirohisa Jozawa and Kazuto Kamikura

NTT Human Interface Laboratories  
1-2356 Take, Yokosuka, Kanagawa, 238-03 JAPAN  
Phone: +81 468 59 2975 FAX: +81 468 59 2829  
Email: jozawa@nttvdt.hil.ntt.jp

This paper describes an H. 263-based video coding using a two-stage motion compensation. The proposed two-stage MC consists of global MC and local affine MC. In the first stage of the method, global motion of a picture such as panning, tilting or zooming is detected and compensated. Then, in the second stage, local motion compensation of each macroblock is carried out. Since the global MC cannot predict all the regions in a picture, it is adaptively turned on or off for each macroblock. The local MC employs an affine motion model in the context of the H. 263's overlapped block MC. The overlapped block affine MC compensates rotation and scaling of moving objects in addition to translational motion. Simulation results show that the proposed video coding using the two-stage MC performs effectively for the sequences with fast motion. PSNR improvements are about 0.5-1.0 dB in comparison with the conventional coding method without two-stage MC.

## 1. はじめに

現在、ISO/IEC JTC1/SC29/WG11 (MPEG)では MPEG-2に続く次世代オーディオ・ビジュアル符号化標準として MPEG-4の標準化作業を進めている。MPEG-4は、伝送レートが低く、誤り率の高い無線伝送路での通信や、対話型の AV データベース検索を主要なアプリケーションと位置付けている。MPEG-4への要求条件は、大きく分けて以下の三つである。

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

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

高圧縮符号化/低レート符号化の検討は MPEG-4 以前より行われているが、注目すべき国際標準として超低レート符号化勧告 ITU-T H. 263<sup>(1)</sup>がある (1996年3月勧告予定)。H. 263は、超低レート AV 端末勧告 H. 324 に付帯するビデオ符号化勧告で、当初は PSTN (Public Switched Telephone Network) 上での映像通信を目的に 64 kb/s 未満をターゲットとしていた。現在では、N-ISDN での使用も考慮されている。H. 263は、H. 261をベースとしながらも、MPEG-1, 2で用いられているハーフペル動き補償(MC)を導入し、さらにオプションとしてオーバーラップ MC, 8×8 MC, 画面外参照動きベクトル, 算術符号化, PB-frame 等の新技术を加えている。これにより、64 kb/s 未満の超低レートでも十分な画質を確保すると共に、64 kb/s 以上では H. 261 より大幅に優れた画質を実現している。

MPEG-4における高圧縮化の検討は、この H. 263 との差別化を図ることが重要である。MPEG-4 ビデオも、当初 64 kb/s 未満を想定していたが、現在は 10 kb/s~1 Mb/s がターゲットとなっており、幅広いレートで安定して高画質を達成することが要求される。そこで筆者等は、64 kb/s 未満の超低レートばかりでなく、64 kb/s 以上のレートにおいても高い性能を発揮する H. 263 をベースとし、これにグローバル MC とローカルアフィン MC を加えた方式について検討した。グローバル MC は、カメラの動きによって生じるパ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。本稿では、提案方式について解説すると共に、その性能をオリジナルの H. 263 と比較した結果を報告する。

## 2 提案方式の概要

提案するビデオ符号化方式の諸元を表 1に、エンコーダのブロックダイアグラムを図 1に示す。提案方式と H. 263 (あるいはそのテストモデル TMNS<sup>(2)</sup>) との主な相違点は以下の二点である。

- (1) 動き補償を、グローバル MC とローカル MC の二段構成とした。
- (2) ローカル MC の動きモデルにアフィン変換を導入した。

提案方式では、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 で物体ごとの局所的な動きを補償する。グローバル MC は、パニング、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 の ON/OFF を切り替える。一方、ローカル MC は H. 263 のオーバーラップブロック MC を基本としているが、動きモデルにアフィン変換を用いることにより、物体の平行移動ばかりでなく、回転、拡大、縮小等も補償する。

表 1 提案ビデオ符号化方式の諸元

|                                  |                                  |                                 |
|----------------------------------|----------------------------------|---------------------------------|
| Input picture format             | SIF-525 / CIF (SIF-625) / QCIF   |                                 |
| Global Motion Compensation (GMC) | Prediction unit:                 | macroblock (16 x 16)            |
|                                  | Number of motion parameter:      | 3 (panning, scrolling, zooming) |
| Local Motion Compensation (LMC)  | Prediction unit:                 | overlapped macroblock           |
|                                  | Motion model:                    | Affine transformation           |
|                                  | Number of motion parameter:      | 5                               |
|                                  | Translational motion parameters: |                                 |

|                                      |  |
|--------------------------------------|--|
|                                      | Scaling parameters: $C_x$ (H), $C_y$ (V)<br>Rotation parameter: $\theta$   |
| Prediction Mode                      | INTRA / INTER LMC / INTER GMC+LMC  |
| Transform                            | 8 x 8 DCT  |
| Quantization                         | 1 quantizer for INTRA DC coefficients (step size = 8)<br>31 quantizers for non-"INTRA DC" coefficients                                     |
| Coding of quantized DCT coefficients | 3D (Last, Run, Level) Variable Length Coding   |
| Coding of motion parameters          | DPCM + VLC for translational motion parameters $t_x$ and $t_y$<br>PCM + VLC for scaling and rotation parameters $C_x$ , $C_y$ and $\theta$ |
| Rate Control                         | TMN5-based rate control<br>Adaptive quantization using macroblock activity   |

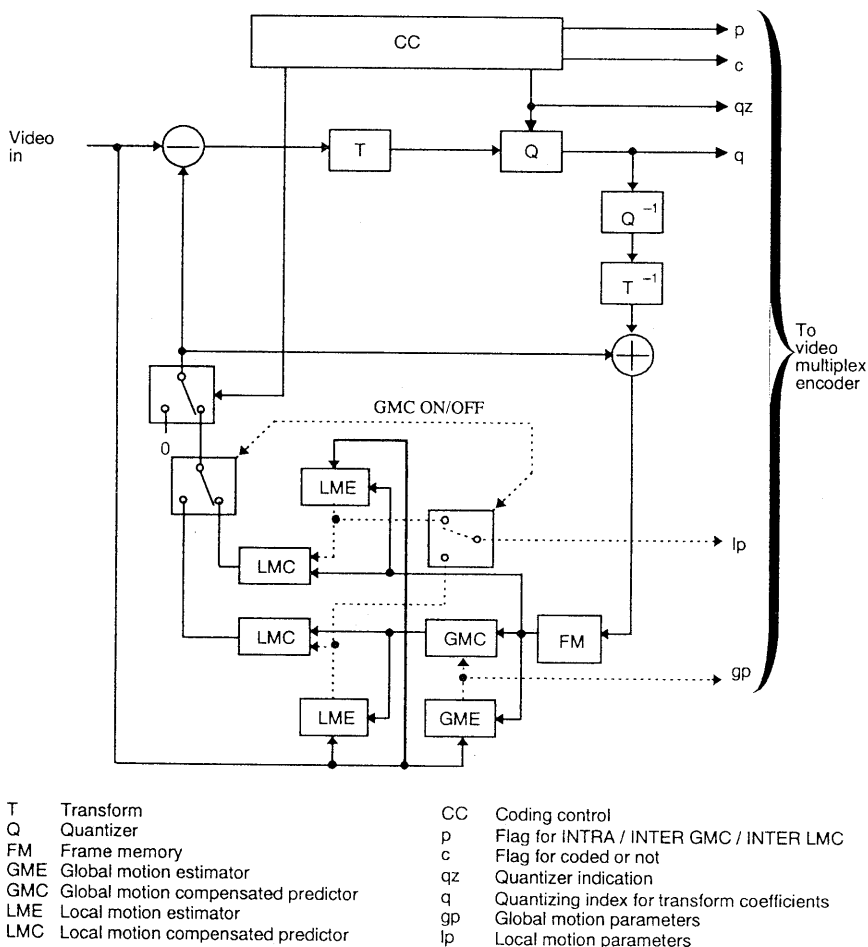


図 1 符号器のブロックダイアグラム

### 3 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。この種の動きは，マクロブロックの動きベクトルで局部的に記述するよりも，画面全体を一組の動きパラメータで

記述する方が効率が良い。また、従来の平行移動のみの動き補償では、ズームは予測不可能である。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<sup>(3)</sup>を用いる。提案方式では、 $Z, H, V$ の三つのパラメータによりカメラの動きを表現する。 $Z$ はズーム率、 $H$ はパニング量、 $V$ はチルト量を表す。提案で用いるグローバル動き補償は、

- (1) グローバル動き検出
  - (2) 予測参照画像のグローバル動き補償
- の二つに大別される。以下、これらの手法について説明する。

### 3. 1 グローバル動き検出

現在の符号化フレームと予測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 ( $Z, H, V$ ) 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。

#### 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。

$$\begin{array}{c} 1\ 2\ 1 \\ 2\ 4\ 2 \\ 1\ 2\ 1 \end{array} / 16$$

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 $\pm 56$ 画素である。また、輝度信号に対してのみ行う。

#### 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。

#### ステップ3

座標の原点を画面中央とし、図2に示す通り、原点に対し点対称の二つのマクロブロックの動きベクトル  $(V1_x, V1_y), (V2_x, V2_y)$  から、以下の式により  $Z, H, V$  を求める。

$$Z = \frac{i(V1_x - V2_x) + j(V1_y - V2_y)}{2(i^2 + j^2)}$$

$$H = \frac{(V1_x + V2_x)}{2}$$

$$V = \frac{(V1_y + V2_y)}{2}$$

ここで、 $(i, j)$ と $(-i, -j)$ はペアブロック  $MB_1$  と  $MB_2$ の座標である。ズームパラメータ  $Z$ は  $1/128$  刻みに量子化され、かつ  $\pm 31/128$ の範囲に納まるようクリップされる。パラメータ  $H, V$ も整数に量子化され、 $\pm 126$ の範囲に納まるようクリップされる。したがって、パラメータ  $Z$ は6ビット、 $H$ と  $V$ は各7ビットの等長符号で表現される。

この操作を画面内の全ペアブロックに対して実行し、 $Z, H, V$ それぞれの発生度数分布を求める。各パラメータで、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、雑音による誤検出を避けるためである。

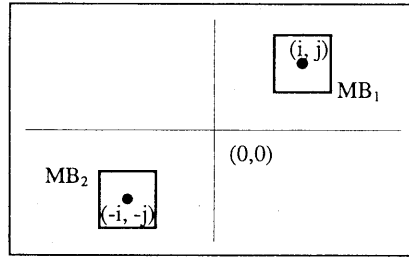


図 2 ペアブロックの位置関係

### 3. 2 予測参照画像の生成

3. 1により求められた三つのグローバル動きパラメータ ( $Z, H, V$ ) を用いて, もとの予測参照画像から, グローバル動き補償後の予測参照画像を生成する. 画素値  $S'_{ij}$  は, 図 3に示すように参照画像における座標  $(x, y)$  の値となる. 但し,  $x, y$  は小数値を取りうるため, 以下の式に示すように, 周囲の四画素  $S_{[x],[y]}, S_{[x]+1,[y]}, S_{[x],[y]+1}, S_{[x]+1,[y]+1}$ からの双一次内挿補間により求める. ここで,  $[x]$  は実数  $x$  の整数部を意味する.

$$S'_{i,j} = ([y]+1-y) * S_1 + (y-[y]) * S_2$$

$$S_1 = ([x]+1-x) * S_{[x],[y]} + (x-[x]) * S_{[x]+1,[y]}$$

$$S_2 = ([x]+1-x) * S_{[x],[y]+1} + (x-[x]) * S_{[x]+1,[y]+1}$$

$$x = \begin{cases} Z * i + H & (\text{luminance}) \\ Z * i + H / 2 & (\text{chrominance}) \end{cases}$$

$$y = \begin{cases} Z * j + V & (\text{luminance}) \\ Z * j + V / 2 & (\text{chrominance}) \end{cases}$$

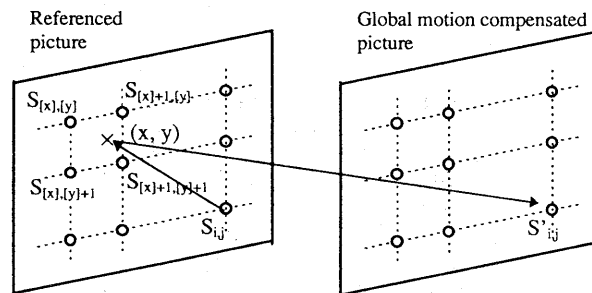


図 3 予測参照画像の作成

## 4. ローカルアフィンMC

### 4. 1 アフィン動きモデルを用いたローカル動き検出

従来の動き補償は物体の平行移動のみを補償するものであるが, 提案方式では動きモデルにアフィン変換を導入し, 平行移動に加えて物体の回転, 拡大/縮小を補償する. 動きは, 水平および

垂直方向の平行移動量  $(t_x, t_y)$ ，水平および垂直方向の伸縮率  $(C_x, C_y)$ ，回転角  $\theta$  の五つのパラメータにより表現される。予測の単位は，文献(4)のように任意形状の小領域ではなく， $16 \times 16$  のマクロブロックとした。座標  $(x, y)$  における画素の予測値は，予測参照画像における座標  $(x', y')$  の画素値として求められる。

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} C_x & 0 \\ 0 & C_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

ここで， $x'$  と  $y'$  は実数値を取りうるため，予測値は 3. 2 のグローバル動き補償画像生成と同様の双一次内挿補間により求める。

パラメータ探索はブロックマッチング法により行うが，計算量を削減するため，表 2 に示すような二段階サーチを行う。まず第一段で，符号化対象画像と予測参照画像（局部復号画像ではなく原画像）との間で，平行移動パラメータ  $(t_x, t_y)$  を  $\pm 15$  画素の範囲でフルサーチする。探索精度は 1 画素単位である。ここで最も誤差絶対値和が小さくなるものを選ぶ。次に第二段では，第一段で選ばれた  $(t_x, t_y)$  とその周囲 0.5 画素，合計 9 つの点に対して，伸縮パラメータ  $(C_x, C_y)$  と  $\theta$  を変化させ，最も誤差絶対値和の小さくなる組み合わせを最終的なローカルアフィンパラメータとする。なお，H. 263 のオプションである Unrestricted Motion Vector Mode も導入し，画面外を参照するようなアフィンパラメータも許している。

表 2 二段階サーチによるアフィンパラメータ探索

|              | Translation $(t_x, t_y)$ | Scaling $(C_x, C_y)$ | Rotation $(\theta)$                      |
|--------------|--------------------------|----------------------|--|
| First layer  | $\pm 15$ , step 1        | -                    | -  |
| Second layer | $\pm 0.5$ , step 0.5     | 0.8-1.2, step 0.1    | $-0.04\pi$ - $0.04\pi$ , step 0.02 $\pi$ |

表 3 伸縮率  $C_x, C_y$  用 VLC テーブル

| $C_x, C_y$ | Bit number | VLC code |
|------------|------------|----------|
| 0.8        | 4          | 1010     |
| 0.9        | 3          | 100      |
| 1.0        | 1          | 0        |
| 1.1        | 4          | 1011     |
| 1.2        | 5          | 11010    |

表 4 回転角  $\theta$  用 VLC テーブル

| $\theta$   | Bit number | VLC code |
|------------|------------|----------|
| $-0.04\pi$ | 4          | 1101     |
| $-0.02\pi$ | 3          | 101      |
| 0          | 1          | 0        |
| $0.02\pi$  | 3          | 100      |
| $0.04\pi$  | 4          | 1100     |

#### 4. 2 オーバーラップブロック・アフィン動き補償 (OBAMC)

動き補償画像を作成するにあたっては，H. 263 の Advanced Prediction Mode に採用されているオーバーラップブロック MC (OBMC) を用いる。但し，OBMC を行うのは輝度信号のみで，色差信号は通常のブロック MC である。

輝度信号ブロック内の各画素の予測値は，

- (1) 現マクロブロックのアフィンパラメータから予測される値，
- (2) 現マクロブロックの左または右のうち近い方のマクロブロックのアフィンパラメータから

予測される値,  
(3) 現マクロブロックの上または下のうち近い方のマクロブロックのアフィンパラメータから  
予測される値,  
の三つの値の重み付け加算により求められる。なお、重み係数は H. 263 と全く同じものを用いた。

## 5 予測モード判定

画面中にはグローバル MC で予測できない部分があるため、グローバル MC をかけることによりかえって予測効率が低下する場合がある。そこで、マクロブロックごとにグローバル MC の ON/OFF を切り替える。このため、図 1 に示すように、グローバル MC を行なった予測参照画像と、グローバル MC を行なわない通常の予測参照画像の双方にローカルアフィン MC を行う。その結果、誤差絶対値和の小さい方を選択する。グローバル MC の ON/OFF は 1 ビットの符号で表現し、他の情報と多重化して伝送する。なお、INTRA/INTER 判定は TMN5<sup>(2)</sup> と同じ方法とした。

## 6 シミュレーション実験

提案方式の効果を検証するため、MPEG-4 の標準画像のうち、“Table Tennis”と“Funfair”を用いて符号化実験を行なった。これらのテスト画像は 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 320～1024 kb/s の符号化レートで実験されている。

そこで、本実験でも、“Table Tennis”と“Funfair”をそれぞれ 320 kb/s と 1024 kb/s で符号化した。処理フォーマットは、“Table Tennis”の場合が SIF-525 (Y: 352×240, U, V: 176×120), “Fun fair”の場合が SIF-625 (Y: 352×288, U, V: 176×144) である。また、提案のグローバル MC とローカルアフィン MC そのものの効果を見るため、これらを ON にした場合と、OFF にした場合の復号画像の PSNR を比較した。その結果を図 4～図 7 に示す。なお、グローバル MC とローカルアフィン MC を OFF にしたものは、ほぼ H. 263 に近い仕様であるが、8×8 MC、算術符号化、PB-frame が入っていない。これは ON の場合も同じ条件である。また、双方とも同じレートコントロールを用いており、両者の差は、グローバル MC とローカルアフィン MC の ON/OFF の違いのみである。

図 4～図 7 より、グローバル MC+ローカルアフィン MC の効果が確認できる。“Table Tennis”, “Fun fair”とも、320 kb/s の場合で 0.5 dB 前後、1024 kb/s で 1 dB 程度の改善が見られる。レートが低い際に差が縮まるのは、主にローカルアフィン MC によるオーバーヘッド情報にあると考えられる。表 3、表 4 に示されるように、伸縮パラメータ  $C_x$ ,  $C_y$  は 1～5 ビット、回転パラメータ  $\theta$  は 1～4 ビットの語長を持っている。したがって、平行移動のみの MC に比べ、1 マクロブロック当たりの動きパラメータの符号量は最低 3 ビットから最高 14 ビット増加する。ローカルアフィン MC 自身の予測効率は高いため、レートが高い際には OFF の場合に比べて差がつくものの、レートが低くなると全体の符号量に占めるアフィンパラメータの割合が高くなり、差が縮まるものと考えられる。 $C_x$ ,  $C_y = 1$  でかつ  $\theta = 0$  のマクロブロックも多く、これらに 3 ビットもの符号を割り当てるのは不経済である。アフィンの ON/OFF 切り替えと、アフィンパラメータの効率的な符号化方法が必要とされる。

## 7 まとめ

H. 263 をベースに、グローバル MC とローカルアフィン MC を加えた方式について述べた。従来のローカル平行移動補償のみの場合に比べ、320 kb/s の場合で 0.5 dB 前後、1024 kb/s で 1 dB 程度 SN 比が改善されることを示した。また、符号化レートが低くなるにつれて、アフィンパラメータのオーバーヘッド分が大きくなり、性能の改善度が小さくなることを確認した。今後は、アフィンの ON/OFF 切り替えと、アフィンパラメータの効率的な符号化方法について検討する。

## 参考文献

(1) ITU-T SG15: “Draft Recommendation H. 263,” LBC95-251, Oct. 1995.

- (2) ITU-T SG15: "Video Codec Test Model, TMN5", LBC95-XXX, Jan. 1995.
- (3) 上倉, 渡辺: "動画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.
- (4) H. Jozawa: "Segment-based video coding using an affine motion model," SPIE VCIP '94, Vol. 2308, pp. 1605-1614, Sept. 1994.

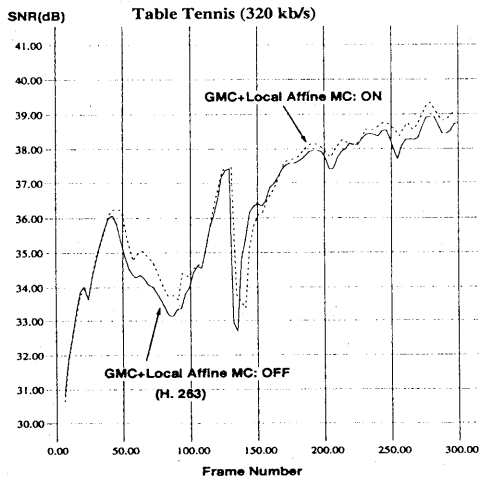


図 4 復号画像の PSNR (Table Tennis, 320 kb/s)

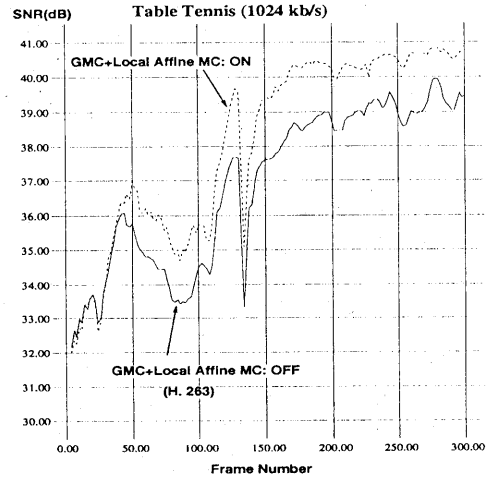


図 5 復号画像の PSNR (Table Tennis, 1024 kb/s)

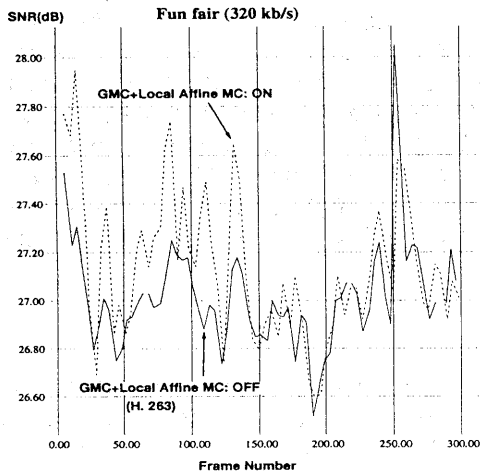


図 6 復号画像の PSNR (Fun fair, 320 kb/s)

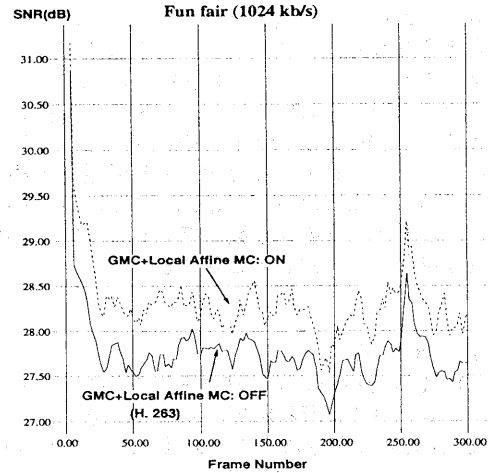


図 7 復号画像の PSNR (Fun fair, 1024 kb/s)