

# マクロブロックタイプを用いた MPEG2 圧縮動画像のカット点検出

青木 真吾<sup>†</sup> 森田 啓義<sup>††</sup>  
荒俣 吉壮<sup>†††</sup> 西 新幹 彦<sup>††</sup>

画像信号全体を復号することなく MPEG2 データから直接、瞬時ならびにディゾルブショットを検出する方法を提案する。提案方法は、MPEG データの I フレーム情報を利用した GOP 単位での粗い検出と、マクロブロックタイプを用いたフレーム単位での正確な検出という 2 つの手法を階層的に行うことにより、従来法に比べて、高い検出率と短い検出時間を可能にする。いくつかのビデオデータを用いた検出実験結果より、瞬時カット点検出では、通常の再生時間の 4 分の 1 程度の処理時間で、固定したしきい値を用いて 90~95% 近い検出率を達成し、誤検出率も 5% 程度と低く抑えられた。さらに、ディゾルブカット点検出に関しては、提案法は実験に用いたすべてのデータに対して 80% 以上の検出率を得、また検出性能は従来法より平均 20% 程度向上した。

## Cut Detection in MPEG2 Compressed Data Using Macro Block Types

SHINGO AOKI,<sup>†</sup> HIROYOSHI MORITA,<sup>††</sup> YOSHITAKE ARAMATA<sup>†††</sup>  
and MIKIHICO NISHIARA<sup>††</sup>

We propose a method to detect both of instantaneous and dissolved shots from MPEG2 data without decoding the entire volume of video signals. The proposed method consists of two detection algorithms in a hierarchical fashion. One algorithm coarsely detects shots in GOP unit by extracting I-frame information from the MPEG data stream. The other identifies the exact boundary between shots in frame unit using detailed information on types of macro blocks contained in the data stream. The proposed method perform faster and more efficient than traditional ones. Experimental results on some video data show that our method achieved 90~95% detection rate with about quarter of normal playback time for detecting instantaneous shots. Moreover, it suppressed the erroneous detection rate within 5%. As for dissolved shots, the proposed method attained more than 80% detection rate over all the video used in the experiment and the mean of detection ability is about 20% higher than the traditional ones.

### 1. はじめに

近年、CPU の高速化、ハードディスクの大容量化が急速に進んで、高性能パーソナルコンピュータが廉価化するにつれ、一般ユーザが映像データを利用するハードやソフトが広く普及してきている。しかし、さまざまな映像データが大量に蓄積されてくると、映像データの必要な部分だけを取り出してつなぎ合わせるといった編集作業や、莫大な映像情報の中から必要なデータを効率良く選択し、その情報に素早くアクセス

するビデオ検索技術が重要になる。

本論文では、ショット単位やシーン単位で映像にアクセスしたり、シーンの要約などのより高度な機能を実現したりすることを目的に、通常の映像データに現れるカット点の大半を占める、瞬間的に映像の切り替わる瞬時カット点と、スポーツ映像で頻繁に用いられる、2 つのショットが合成されながら切り替わるディゾルブカット点を、MPEG2 圧縮データから自動的に検出する手法について考察する。

MPEG2 (あるいは MPEG1) で圧縮された映像データからカット点を検出する方法はすでにさまざまな方法が提案されている<sup>3)~10)</sup>。それらの中でも、MPEG 圧縮データ中の I フレームに含まれる輝度 DC 成分を 2 次元に配列した、いわゆる、DC 画像を利用する方法<sup>3)~5),8)</sup> は、完全に映像データを復元するための処理時間を省くことができ、しかも瞬時カット点検出

<sup>†</sup> オリンパス株式会社  
Olympus Corporation

<sup>††</sup> 電気通信大学大学院情報システム学研究所  
Graduate School of Information Systems, The University of Electro-Communications

<sup>†††</sup> 株式会社菱友システムズ  
Ryoyu Systems Co., Ltd.

の場合、いったん復元してからカット点を検出する方法<sup>2)</sup>とほとんど変わらない精度で検出できるという点で注目に値する。反面、Iフレーム情報だけではカット点の正確なフレーム位置を同定するには不十分であること、映像に応じて検出に用いるパラメータをよく調整しないと、誤検出や未検出といった誤りが生じやすいという問題がある。

ディゾルブカット点検出に関しては、フレームごとの画素値の分散がディゾルブ部分においてU字形に変化するという特徴を利用して検出を行う方法<sup>8),9)</sup>や、輝度値の差分を利用する方法<sup>10)</sup>が知られているが、文献5)で指摘されているとおり、実際の映像では動きのあるシーンや局所的な変動によって理想的なU字型になることが少ないため、未検出や誤検出が多く発生する恐れがある。

これらの問題を解決するためには、MPEG符号で圧縮されたデータから必要な情報だけを抽出・加工する方法が、多くの映像データにおいて、効果的である。一般に、映像データにカット点が生じると、MPEG圧縮データに埋め込まれたマクロブロックタイプ情報には、MPEG符号器の動作に起因する特徴的なパターンが現れる。この特徴パターンを統計的に処理することにより、カット点の検出は画像再生にかかる時間を大幅に短縮することができるだけでなく、高い検出率でカット点を検出できることが報告されている<sup>4),6)</sup>。しかし一方で、瞬時カット点検出では、映像に対して検出のためのしきい値が最適値からずれると、物体やカメラワークが急激に変化するシーンにおける誤検出や暗いシーンにおける未検出といった誤りが生じやすい。またディゾルブカット点検出では、誤検出率は15~20%前後であることが報告されている<sup>6)</sup>。

本論文では、従来法のようにマクロブロック情報をBフレーム1枚ごとに処理するのではなく、隣接する2枚のBフレームのマクロブロック情報とカット点の関係に着目し、誤検出・未検出といった誤りの改善を図る。さらに、ディゾルブカット点検出に関しては、動きベクトル情報をマクロブロック情報に付加することによって、検出誤りが改善できることを示す。

以下、次章では、本論文の議論に最低限必要なMPEG2技術について述べる。3章では、マクロブロック・タイプ情報を用いたカット点検出法を提案する。続いて4章では、従来法と提案法の性能比較実験を行う。最後に、まとめを5章で行う。

## 2. カット点とMBTとの対応関係について

MPEG2は、主にフレーム内符号化、予測符号化、

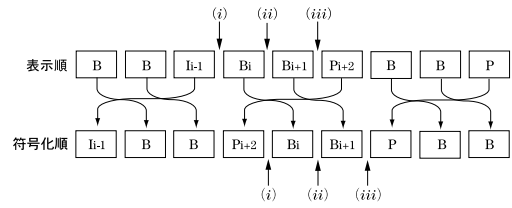


図1 フレームの表示順と符号化順の違い

Fig. 1 Difference between the order of display and encoding of frames.

双方向予測符号化という3種類の符号化手法から構成されており<sup>1)</sup>、ここでは、便宜上、順に、I符号、P符号、B符号と呼ぶ。

これらの符号に対応して、映像データの各フレーム(1画面)には、Iフレーム、Pフレーム、Bフレームという3種類がある。どのフレームがどの種類になるかは周期的に決まる。1周期分のフレームはGOP(Group of Picture)と呼ばれ、市販のエンコーダなどでは、1GOPは12~15個のフレームから構成されることが多い。

I符号はフレーム内の情報のみを使用して符号化を行うが、P符号では、直前のIフレームもしくはPフレームを予測値として参照することによって、あらかじめ冗長な部分を取り除き、高い圧縮効果を得ている。B符号は、P符号と同様の予測を行うが、直前と直後のIフレームもしくはPフレームを参照することができるので、P符号よりさらに圧縮効果が期待できる。

ただし、B符号を用いる場合、表示順のままフレームを符号化すると、復号側で直後フレームの参照が実現できなくなるので、符号化順は図1に示すように並べ換えられている。

いずれの符号を用いる場合においても、基本的にはフレームをさらに細分化した $8 \times 8$ のブロック単位に離散コサイン変換が施され、変換係数の量子化・エントロピー符号化によってデータが圧縮される。それぞれのブロックにどの符号を適用するかは、縦・横 $2 \times 2$ のブロックを1つにまとめたマクロブロック(MBと略す)を単位にして(図2参照)、符号化の際に最も圧縮効果が高くなるものが選ばれる。

なお、4つのブロックは輝度情報のみを表しており、MBにはそれらの色情報を表すためにほかに2つのブロックが含まれている。

符号化について注意すべきことは、フレームと符号が1対1に対応しているのではなく、表1に示すように、フレームの種類によってマクロブロックごとに選択可能な符号が異なっている点である。したがって、たとえば、Bフレームに含まれるブロックは、MPEG

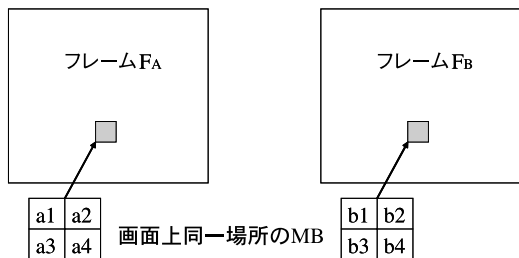


図2 マクロブロックとブロックの関係  
Fig.2 Macro blocks and blocks.

表1 各フレームと適用可能な符号の関係  
Table 1 The available codes for each frame type.

フレーム種類	I 符号	P 符号	B 符号
I フレーム			
P フレーム			
B フレーム			

(印は適用可能)

符号器において、I 符号、P 符号、B 符号で同時に符号化される。正確にいうと、P 符号を用いる際には、直前の I フレームもしくは P フレームを参照する順方向予測と、直後のそれを参照する逆方向予測がありうるので、合わせて、これら 4 種類の符号の中から最も圧縮効果が高いものが選ばれる。同様に、P フレームのブロックに関しては、I 符号と順方向予測の P 符号から選択される。

各ブロックをどの符号を用いて符号化したかという情報は MB 単位に符号化の際に圧縮データに埋め込まれている。この情報をマクロブロックタイプ (MBT と略す) と呼ぶ。

映像データにカット点が発生すると、符号器では予測がうまく働かず、P フレームや B フレームに対しても、I 符号を選択する傾向が強くなる。圧縮データから MBT を抽出するのは比較的容易なので、MBT の分布をフレームごとに調べればカット点の発生を検出できると思われる。これが、本提案方式を含め MBT を用いるカット点検出法に共通する基本的な考え方である。

実際、瞬時カット点とディゾルブカット点が発生する場合に分けて、MBT の分布がどのようになるかについて以下で考察する。なお、本研究では、GOP の構成は、I フレーム 1 枚、P フレーム 4 枚、B フレーム 10 枚、計 15 枚が IBBPBBPBBPBBPBB の順で並んだ標準的なものに固定している。

2.1 瞬時カット点が発生する場合

瞬時カット点の発生する位置は 1 対の連続した B フレーム  $B_i, B_{i+1}$  に注目すると、以下の 3 つの場合に

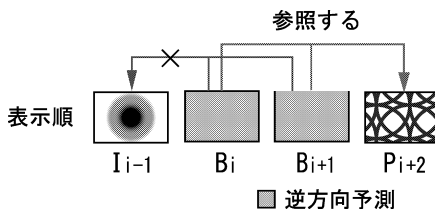


図3 (i) の場合  
Fig.3 Case (i).

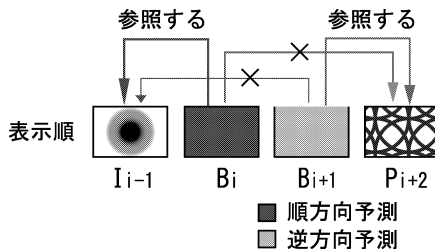


図4 (ii) の場合  
Fig.4 Case (ii).

分類できる。

- (i)  $B_i$  の直前にカット点が存在する場合
- (ii)  $B_i$  と  $B_{i+1}$  の間にカット点が存在する場合
- (iii)  $B_{i+1}$  の直後にカット点が存在する場合

ここで、表示順、符号化順のいずれにおいても、 $B_i$  の左隣のフレームと  $B_{i+1}$  の右隣のフレームは I もしくは P フレームであるが、どの場合でも扱いは同じなので、以下では、図 1 に示した表示順にフレームが、 $I_{i-1}, B_i, B_{i+1}, P_{i+2}$  と並んだ場合を例にとって説明する。

(i) では、B 符号を用いて  $B_i$  と  $B_{i+1}$  の各ブロックを符号化する際には、 $I_{i-1}$  と  $P_{i+2}$  が参照される。この場合、映像は  $I_{i-1}$  と  $B_i$  の間で切り替わるので、 $I_{i-1}$  を参照しても、予測がうまく働かず、 $P_{i+2}$  のみへの参照を符号器は採択する傾向にある。すなわち、 $B_i, B_{i+1}$  ともに逆方向予測を行うマクロブロック (BMB と略す) の数が増える (図 3)。(ii) の場合は、 $B_i$  と  $B_{i+1}$  の間で映像が切り替わるため、符号器では、 $B_i$  に対しては  $I_{i-1}$  への順方向予測を行うマクロブロック (FMB と略す)、 $B_{i+1}$  に対しては  $P_{i+2}$  への BMB の数が増えると考えられる (図 4)。(iii) の場合は、 $B_i, B_{i+1}$  はともに、 $I_{i+2}$  への FMB の数が大半を占める (図 5)。

以上の考察より、B フレーム対の FMB と BMB 数の大小から、GOP 中のどこで瞬時カットが発生したかを検出できる、

2.2 ディゾルブカット点が発生する場合

画像が徐々に変化していくディゾルブカット部分では、MBT の分布は瞬時カットの場合とは異なる特徴

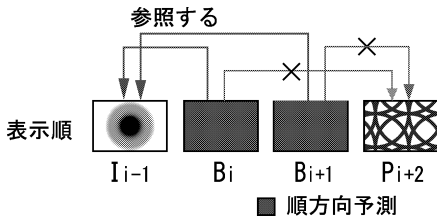


図 5 (iii) の場合  
Fig. 5 Case (iii).

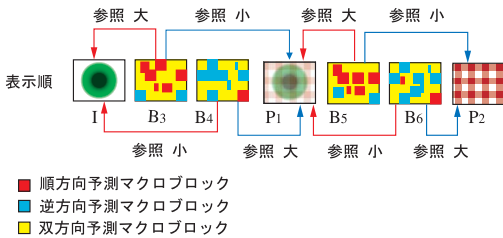


図 6 ディゾルブカットと MBT の関係  
Fig. 6 Dissolved cut and MBT.

を持つ変化をする。図 6 はディゾルブカット部分での MBT 分布の変化を例示したものである。図中の  $B_3$  に注目すると、参照できるフレームは、左隣の  $I$  と 2 つ先にある  $P_1$  であるが、 $I$  のほうが  $P_1$  より時間的に近いので  $B_3$  との相関が強く、予測結果が  $P_1$  を用いた場合に比べより正確に行えると考えられる。その結果、BMB に比べ、FMB の数が増える傾向にある。

一方、 $B_4$  では、 $I$  より  $P_1$  との相関が強いので、 $I$  への順方向予測より、 $P_1$  への逆方向予測のほうが頻繁に行われるため、BMB が多く発生すると考えられる。

同様に、対 ( $B_5, B_6$ ) については、 $B_5$  では  $P_1$  への FMB が増え、また、 $B_6$  では相関の強い  $P_2$  への BMT が  $P_1$  への FMT に比べてより多く現れる。

以上をまとめると、ディゾルブ部分では  $B_3, B_5$  で FMB が多く、 $B_4, B_6$  で BMB が多くなると考えられる。すなわち、B フレームごとに FMB と BMB の頻度を比べると、その大小関係は、ディゾルブ変化が生じている部分において、フレーム単位で交互に反転するという現象（ジグザグ現象と呼ぶ）が生じうる。この現象は、文献 6) で初めて報告された。

### 3. 提案カット点検出法

提案方式では、最初に、I フレームの DC 画像情報を用いて GOP 単位でカット点を粗く検出し、続いて、MBT を利用してフレーム単位で正確なカット点位置を検出するという階層的な手法を用いることによって、検出処理を効率良く高速化している。最初の GOP 単位の検出と、続くフレーム単位での検出には、瞬時カッ

表 2 瞬時カット点の判定ルール  
Table 2 Decision rule for instantaneous cut.

判定条件	判定結果
$\min\{b_i, b_{i+1}\} \geq T_\mu$	(i)
$f_i \geq T_\mu \wedge b_{i+1} \geq T_\mu$	(ii)
$\min\{f_i, f_{i+1}\} \geq T_\mu$	(iii)

ト点とディゾルブカット点の場合と異なる手法を用いているので、以下では 2 つに分けて説明する。

#### 3.1 瞬時カット点検出

##### 3.1.1 I 法<sup>3),5)</sup>

隣接する GOP それぞれの先頭に位置する 2 つの I フレームに含まれる輝度 DC 画像が大きく変化すればそれらの間でカット点が生じていると考えられる。輝度 DC 画像の変化は次のようにして定量的に評価できる。

まず、あるしきい値  $T_\delta$  と  $T_\nu$  をあらかじめ固定しておき、図 2 中のフレーム  $F_A$  と  $F_B$  を、隣接する GOP にそれぞれ含まれる I フレーム  $I_A, I_B$  として考える。

さらに、それぞれのフレームにおいて画面上で同じ位置を占める MB 中のブロック  $a_1 \sim a_4$  と  $b_1 \sim b_4$  に対して離散コサイン変換を施した後の輝度 DC 成分をそれぞれ  $\alpha_i, \beta_i$  ( $i = 1, \dots, 4$ ) とおく。これらの差の絶対値の和

$$\delta_I = |\alpha_1 - \beta_1| + |\alpha_2 - \beta_2| + |\alpha_3 - \beta_3| + |\alpha_4 - \beta_4|$$

に対し、不等式

$$\delta_I > T_\delta$$

が成り立つ MB の数を  $\nu_I$  とする。もし

$$\nu_I > T_\nu$$

ならば、両 I フレーム間のいずれかのフレームでカット点が発生したと判断する。

##### 3.1.2 MBT 法

I 法でカット点を含むと判断された GOP に対して、MBT の分布を用いてフレーム単位での検出を行う。2.1 節で考察した (i) ~ (iii) の場合に対応して、B フレーム対 ( $B_i, B_{i+1}$ ) の MBT から FMB の数  $f_j$  と BMB の数  $b_j$  ( $j = i, i+1$ ) があるしきい値  $T_\mu$  を超えるかどうかを表 2 の判定条件と照合し、3 つの場合のうちのどれが生じたかを判断する。

しきい値  $T_\mu$  の決め方は、画面の明るさと映像の動きの激しさに関係する。画面が暗いと、カット点が生じているフレームでも双方向予測を行う MB が半数近く残る場合があり、その結果、表 2 のどの条件にも該当しないという未検出誤りが発生しやすい。

提案法では、I 法で算出したブロックに対する  $\delta_I$  の値をフレームごとに平均をとり、あらかじめ定めた一

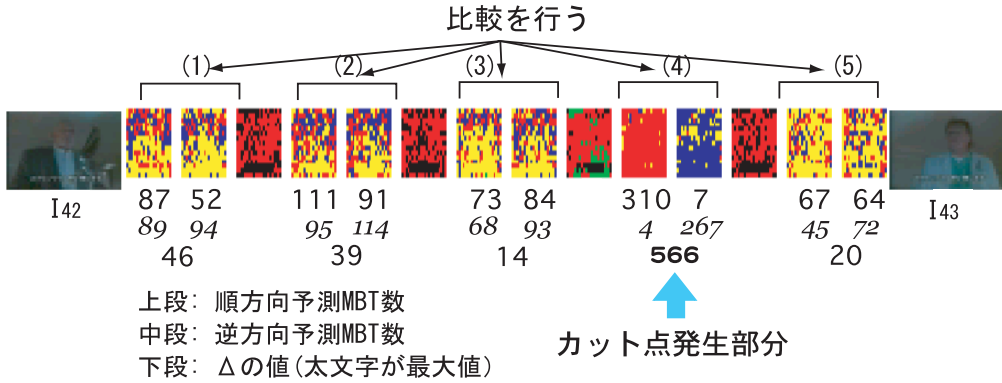


図 7 カット点発生部分の例  
Fig. 7 An example of instantaneous cut.

定値に乗じることによって、 $T_\mu$  の値を適応的に変化させ、未検出を減らしている。

しかし、この方法だけでは、画面が暗くかつ動きが激しいシーンでは、同じ GOP 内に判定条件を満たす B フレーム対がいくつも現れ、誤検出が発生する可能性がある。そこで、GOP 中の各 B フレーム対  $(B_i, B_{i+1})$  に対して、

$$\Delta = |f_i - b_i| + |f_{i+1} - b_{i+1}| \quad (1)$$

を計算し、最大値  $\Delta_{\max}$  をとる B フレーム対を求める。他のフレームのとり値と  $\Delta_{\max}$  の比が一定のしきい値  $T_\Delta$  以下であるとき、カット点が発生したと判断する (図 7 参照)。

### 3.2 ディゾルブカット点検出

ディゾルブカット点ではジグザグ現象が発生するが、この現象は、動きの激しいシーンにおいても生じうる。ただし、ディゾルブカット点の大半は継続時間が 1GOP 程度なので、以下で述べる  $kI$  法で、数 GOP に及ぶ動きの激しいシーン部分を除外し、次の修正 JYL 法で短時間のものを省くことで誤検出を抑える。

#### 3.2.1 $kI$ 法

ディゾルブカット点の検出を行うために、ディゾルブが継続する間隔を覆う  $k$  枚の I フレームの輝度 ( $Y$ )、色差 ( $Cb$ )、色相 ( $Cr$ ) それぞれについて、フレーム間のヒストグラムの変化量 (変動距離、ダイバージェンス) を求め、変化量の大きいところをディゾルブと判断する。詳しくは次の手順に従う。

まず、成分  $x = Y, Cb, Cr$  に対し、 $j$  番目の I フレーム ( $j = 1, \dots, k$ ) の各 DC 成分値を均等に  $q$  値に量子化する。このとき  $i$  番目の量子化区間に値が落ちる相対頻度を  $p_j^x(i)$  とおき、 $p_j^x$  から  $p_{j+1}^x$  への変化量を表すために変動距離  $V_j^x$  とダイバージェンス  $D_j^x$  を以下の式で計算する。

$$V_j^x = \sum_{i=1}^q |p_{j+1}^x(i) - p_j^x(i)|$$

$$D_j^x = \sum_{i=1}^q p_j^x(i) \log \frac{p_{j+1}^x(i)}{p_j^x(i)}$$

これらを 3 成分全体で平均をとった値をそれぞれ  $V_j, D_j$  とおき、

$$r_j = V_j D_j$$

とする。 $V_j^x$  と  $D_j^x$  に共通する特徴として、 $p_j^x$  と  $p_{j+1}^x$  が一致するときに 0 であり、違いが大きいくほど値が大きくなるという性質がある。したがって、 $k$  枚の I フレームにわたってディゾルブが発生しているとすると、中間部分では 2 つのシーンが重なるため  $r_i$  の値が際立って大きくなり、一方、シーンの変化が少ない両端部では、 $r_i$  の値は極端に小さくなると考えられる。

ところで、 $k$  のとるべき値はディゾルブの継続する時間によって変化するが、今回対象にした映像データではほとんどが 1 秒未満で切り替わっているため、 $k = 6$  として考え、

$$r_3 \geq T_1$$

が成り立ち、さらに、

$$(r_1 \leq T_2 \wedge r_2 \leq T_2) \vee (r_4 \leq T_2 \wedge r_5 \leq T_2)$$

が成り立つならば、中心位置を占める 3 番目の I フレームから始まる GOP 部分でディゾルブが発生したと判断する。しきい値  $T_1, T_2$  の最適な値は動きの激しさに依存して決まると考えられる。

#### 3.2.2 修正 JYL 法

$kI$  法でディゾルブがあると判断されたならば、さらにフレーム単位の検出を行いディゾルブがどこにあるかを、2 章で述べたようにジグザグ現象が起きるかどうかで判断する。

JYL<sup>(6)</sup> では B フレーム内の FMB 数  $f$  と BMB

数  $b$  から順方向マクロブロック比  $\rho$  を

$$\rho = \frac{f}{f+b}$$

として定義し、 $\rho$  の値の変化からディゾルブを検出している。

しかし、ディゾルブ変化は上下左右に物体が動くような変化ではなく、画像が浮き出てくるような変化なので、動きベクトルの大きさは比較的小さいと考えられる。

そこで本論文では、圧縮データから各 MB の動きベクトル情報を取り出し、ベクトルの大きさがしきい値  $T_m$  より大きな MB はあらかじめ  $f$  と  $b$  の計算から除外したうえで、順方向マクロブロック比  $\rho'$  を求める修正 JYL 法を提案する。

さらに修正 JYL 法では、B フレームの対 ( $B_i, B_{i+1}$ ) ごとに、 $\rho'_i, \rho'_{i+1}$  を計算したら、B フレームの対ごとに

$$\Delta\rho = \rho'_{i+1} - \rho'_i$$

を計算し、しきい値  $T_\rho$  より大きい  $\Delta\rho$  の値が連続するならば、ディゾルブが発生したと判断する。

しきい値  $T_m, T_\rho$  の最適な値はエンコーダによって検知された映像の動きに依存する。

#### 4. カット点検出実験

瞬時カット点検出、ディゾルブカット点検出それぞれについて、提案法の有効性を調べるために、従来法との比較実験を行った。従来法としては、瞬時カット点検出については、提案法と同様に MBT を利用している MJC 法<sup>4)</sup>、ディゾルブカット点検出に関しては、MBT 比のジグザグ現象の利用する JYL 法<sup>6)</sup> を用い、各方式のプログラムはすべて JAVA で書いた。

実験に使用した計算機は、CPU が Athlon (1.4 GHz)、搭載メモリは 512 MB の Windows 2000 Pro が稼働する PC で、市販の MPEG2 エンコーダ基板を PCI 接続している。MPEG2 圧縮データはすべて市販ビデオから作成し、解像度は  $352 \times 240$ 、ビットレート

は平均 2 Mbps、最大 4 Mbps の VBR (Variable Bit Rate) である。

カット点検出の結果を評価するため、検出率 ( $RCD$ )、誤検出率 ( $RFD$ )、検出性能 ( $DP$ ) の 3 つの基準を用いた。これらの値は、既知正解カット点数 ( $CSC$ )、正しく検出されたカット点数 ( $CD$ )、誤検出カット点数 ( $FD$ ) から、

$$RCD = \frac{CD}{CSC} \times 100$$

$$RFD = \frac{FD}{CD + FD} \times 100$$

$$DP = \frac{CD - FD}{CSC} \times 100$$

として求められる。

実験結果については、瞬時カット点とディゾルブカット点で異なる映像データを用いているため以下で分けて述べる。

##### 4.1 瞬時カット点検出実験結果

瞬時カット点検出では実験データとして表 3 に掲げるものを用いた。それぞれのデータのカット点位置は、あらかじめ目視により正確な位置を確かめた。しきい値は提案法ではデータ全体の検出率が最大値に近くなる条件のもとで誤検出が最小になるように選んだ値を固定して用いたが、MJC 法では各データごとに検出性能が最も良くなる値を選んだ。提案法の  $q$  の値に関しては相対頻度分布が疎とならないように  $q = 16$  とした。 $q$  の値を多少変えても分布は大きく変化しないので結果にはほとんど影響しないと考えられる。

検出率、誤検出率、検出性能の結果を表 4 に示す。表中の太文字はより良い方の値を示している。

提案法と MJC 法との比較すると、検出率では双方ともに 90% 以上の良好な結果が得られた。

全体的に暗い映像の (13) や部分的に暗い映像の (15) ~ (17) において検出率が向上している。しかし、映像 (14), (18) に対しては、提案法は従来法と比べて検出率が低減している。この原因は、2 つの映像では、

表 3 瞬時カット点検出に用いた映像データ

Table 3 The video data for experiments of detection of instantaneous cuts.

映像の種類	時間	カット点数	動き	明るさ
(11) ドラマ	10 分	204 個	少ない	明るい
(12) サッカー	10 分	78 個	激しい	明るい
(13) ホラー	10 分	113 個	少ない	暗い
(14) アニメ 1	60 分	866 個	激しい	明暗あり
(15) アクション映画	105 分	2020 個	激しい	明暗あり
(16) サスペンス映画	100 分	746 個	少ない	明暗あり
(17) ラブストーリー	113 分	1229 個	少ない	明暗あり
(18) アニメ映画	133 分	1662 個	激しい	明暗あり

表 4 瞬時カット点検出結果の比較 (上段: MJC 法/下段: 提案法) (百分率表示)

Table 4 The experimental results of detection of instantaneous cuts.

映像データ	(I1)	(I2)	(I3)	(I4)	(I5)	(I6)	(I7)	(I8)
検出率 (RCD)	100.	92.31	95.58	<b>93.76</b>	93.37	95.58	96.50	<b>93.86</b>
	<b>100.</b>	<b>96.15</b>	<b>99.12</b>	91.68	<b>94.01</b>	<b>99.06</b>	<b>97.97</b>	89.71
誤検出率 (RFD)	0.48	10.00	3.57	10.96	9.24	3.52	5.49	11.86
	<b>0.00</b>	<b>3.36</b>	<b>0.88</b>	<b>5.47</b>	<b>3.36</b>	<b>0.81</b>	<b>0.41</b>	<b>6.87</b>
検出性能 (DP)	99.51	82.05	92.04	82.22	83.86	92.01	90.90	81.22
	<b>100.</b>	<b>92.31</b>	<b>98.23</b>	<b>86.37</b>	<b>90.74</b>	<b>98.26</b>	<b>97.55</b>	<b>83.37</b>

表 5 ディゾルブカット点検出に用いた映像データ

Table 5 The video data for experiments of detection of dissolved cuts.

映像の種類	時間	カット点数	動き
(1) サッカー 1	10 分	18 個	全体的
(2) サッカー 2	10 分	9 個	全体的
(3) サッカー 3	10 分	57 個	全体的
(4) 野球	10 分	15 個	部分的
(5) ドキュメント	10 分	12 個	部分的

動きの激しいシーン中に瞬時カット点が発生している場面があり, 提案した MBT 法では, GOP 内に複数の B フレーム対で  $\Delta/\Delta_{\max}$  が  $T_{\Delta}$  を超えるため, カット点とは判断されないことによる.

誤検出率は従来法に比べ提案法は大幅に改善されている. 映像 (I6), (I7) はいずれも動きの少ないシーンで構成されているが, 一般に, 動きの少ない映像に対しては, MBT 法は誤検出率を低く抑えるのに有効である. 動きの激しい映像である (I2), (I4), (I5), (I8) に対しても, MJC 法に比べ誤検出率は 1/2 から 1/3 程度に減少している.

検出性能に関してはすべての映像データで提案法は MJC 法を上回り, ほとんどのデータで 90% を超える結果を得ている. 動きの激しい (I4), (I8) では検出率が低いため, 検出性能が 80% 台にとどまっている. 動きの激しい映像でのカット点検出率をより向上させることが今後の課題である.

検出時間に関しては提案法, MJC 法ともにデータの再生時間の約 4 分の 1 で検出することができた. 今回の実験ではプログラム作成に JAVA を使用したが, C 言語などを使用すれば, 検出時間はさらに短縮できると考えられる.

#### 4.2 ディゾルブカット点検出実験結果

ディゾルブカット点検出実験に用いた映像データを表 5 に掲げる. 瞬時カット点に比べるとディゾルブカット点は使用される頻度が低く, データ数も必然的に限られてくる. そのため, 動きの激しい部分にディゾルブカットが比較的多く発生しているスポーツ映像を中心に TV 放送から採取した. 50 分間の全データ中, 111 個のディゾルブカット点が含まれている.

表 6 ディゾルブカット点検出結果の比較 (上段: JYL 法/下段: 提案法) (百分率表示)

Table 6 The experimental results of detection of dissolved cuts.

映像データ	(D1)	(D2)	(D3)	(D4)	(D5)
検出率 (RCD)	78.78	66.67	75.44	<b>80.00</b>	66.67
	<b>100.</b>	<b>100.</b>	<b>84.21</b>	<b>80.00</b>	<b>83.33</b>
誤検出率 (RFD)	22.22	<b>0.00</b>	10.41	7.69	38.46
	<b>10.00</b>	10.00	<b>9.43</b>	<b>0.00</b>	<b>23.08</b>
検出性能 (DP)	56.56	66.67	66.67	73.33	25.00
	<b>88.89</b>	<b>88.89</b>	<b>75.44</b>	<b>80.00</b>	<b>58.33</b>

しきい値に関しては, JYL 法ではデータごとに検出性能の値が最も良くなるものを選び, 提案法ではすべてのデータを通して  $T_1 = 100$ ,  $T_2 = r_3/7$ ,  $T_m = 40$  に固定し,  $T_p$  はデータごとに検出性能の値が最も良くなるを選んだ. 検出率, 誤検出率, 検出性能の結果を表 6 に示す. 検出率では提案法は 80% 以上の結果を得ることができ, JYL 法より良い結果を得ている. 未検出となったものには, 類似シーン間でのディゾルブが多く見られた. しかし異なるシーン間での場合などは正確な位置で検出されていることからディゾルブカット点検出に有効である.

誤検出率に関しては, 大きな動きベクトルを持つ MB を数えないという修正 JYL 法の考え方が有効に働いている. しかし, (D2) では逆に誤検出率が JYL 法に比べて増加している. 実際に誤検出が生じたのはスロー再生のシーンであり, 動きベクトルの大きさは小さいため, 新たな工夫が必要と思われる. この点に関しては今後の課題とする. なお, 検出性能では, ほとんどの映像で 80% 近い結果を得ており, 従来法より平均で 20% 程度向上している.

## 5. ま と め

MPEG2 圧縮データから映像の瞬時カット点ならびにディゾルブカット点検出において, 動きの激しいシーンにおける誤検出, 暗いシーンにおける未検出という 2 つの誤りを改善するために, B フレーム対のマクロブロックタイプを利用する提案法の有効性が確か

められた。

今後の課題としては、動きのあるシーンなどでの誤検出を防ぐとともに、同系色でのディゾルブディゾルブカット点の検出精度を上げること、さらに、他の漸次カット点が生じた場合と MBT の関係について考察することがあげられる。

最後に、MPEG2 符号化方式には、圧縮率を上げるためにさまざまな画像処理技術が用いられている。これらを圧縮データに含まれている情報からの確に引き出すことによって、今後、カット点検出を含め、さまざまな応用の可能性が期待できる。

謝辞 本研究の一部は、電気通信大学と船井電機(株)の情報家電に関する共同研究(FUN-X プロジェクト)の援助を受けて行われた。

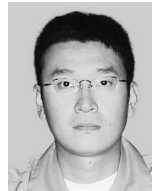
### 参 考 文 献

- 1) ISO/IEC 13818-2 Committee Draft (MPEG-2).
- 2) 大辻清太, 外村圭伸, 大庭有二: 突発的フィルタを用いた映像カット点検出法, 信学論 D-II, Vol.J77-D-II, No.3, pp.519-528 (1994).
- 3) Nakajima, Y.: A Video Browsing Using fast scene change detection for an efficient networked video database access, *IEICE Trans. Inf. & Sys*, Vol.E77-D, No.12, pp.1355-1364 (1994).
- 4) Meng, J., Juan, Y. and Chang, S.-F.: Scene Change Detection in a MPEG Compressed Video Sequence, *Digital video Compression Algorithms and Technologies*, SPIE, Vol.2419, pp.14-25 (1995).
- 5) 中島康之, 氏原清乃, 米山暁夫: 部分複合を用いた MPEG データからのカット点検出, 信学論 D-II, Vol.J81-D-II, No.7, pp.1564-1575 (1998).
- 6) Jun, S.-B., Yoon, K. and Lee, H.-Y.: Dissolve Transition Detection Algorithm Using Spatio-Temporal Distribution of MPEG Macro-Block Types, *Proc. 8th ACM international conference on Multimedia*, California, United States, pp.391-394 (2000).
- 7) 新倉康巨, 阿久津明人, 谷口行信, 佐藤 隆, 外村佳伸, 浜田 洋: MPEG 符号化映像ショットチェンジ検出のための動き補償解析ハイブリッド法の提案, 信学論 D-II, Vol.J81-D-II, No.8, pp.1838-1846 (1998).
- 8) Yeo, B.L. and Liu, B.: Rapid scene analysis on compressed video, *IEEE Trans. Circuits and Systems for Video Technology* (Dec. 1995).
- 9) Alattar, A.M.: Detecting and compressing dissolve regions in video sequences with a DVI multimedia imaging compression algorithm, *Proc. IEEE Int. Sym. on Circuits and Systems*, pp.13-16 (May 1993).
- 10) Gu, L., Tsui, K. and Keightley, D.: Dissolve Detection in MPEG Compressed Video, *IEEE Int. Conf. on Intelligent Processing Systems*, Beijing, China (Oct. 1997).

(平成 16 年 12 月 24 日受付)

(平成 17 年 7 月 18 日採録)

(担当編集委員 金子 俊一)



青木 真吾

2003 年電気通信大学大学院情報システム学研究科博士前期課程修了。同年オリパス(株)入社。現在に至る。



森田 啓義(正会員)

1983 年大阪大学大学院後期課程修了。同年豊橋技術科学大学助手。1990 年電気通信大学電気通信学部講師を経て、現在同大学大学院情報システム学研究科教授。3D 画像処理, 情報源符号化, フレーム同期に関する研究に従事。工学博士。電子情報通信学会, 情報理論とその応用学会, IEEE 各会員。



荒俣 吉壮

2002 年電気通信大学大学院情報システム学研究科博士前期課程修了。現在,(株)菱友システムズにて航空機ソフトウェア関連業務に従事。



西新 幹彦

1998 年電気通信大学大学院情報システム学研究科博士前期課程修了。2001 年同大学院同研究科博士後期課程修了。博士(工学)。同年から同大学同研究科助手。情報理論, データ圧縮, 実時間通信等の研究に従事。IEEE, 電子情報通信学会, 情報理論とその応用学会各会員。