

MPEG 動画像の高速カット検出方式

5N-2

金子 敏充 堀 修

(株)東芝 研究開発センター 情報・通信システム研究所

1 はじめに

カット検出は動画像の意味的な最小単位に分割する方法として注目され、様々な手法が報告されている[1]-[5]。その手法のほとんどは、カット以外のところでは連続する画像は類似しているが、カットの前後で画像は大きく異なっていることを利用したもので、連続するフレーム間の類似度を定義しておき、類似度がしきい値以下となるときにカットであると判定するというものである。ほとんどの場合、類似度は画像の画素値から算出されるものとして定義されている。従って、汎用性の高い類似度であるといえるが、一方では、圧縮符号化された動画像に対しては必ず復号処理を行った後にカット検出を行う必要がある。一般にサイズの大きな圧縮動画像をソフトウェアで復号するには多くの計算量を必要とするため、カット検出を高速化することが難しい。

ここで提案するカット検出手法は、MPEGにより圧縮符号化された動画像に特化した手法であり、完全に復号を行うことなくカット検出を行うため、高速に処理ができることが特徴である。画像間の類似度の定義としては、MPEGにより圧縮符号化された動画像データ中に含まれる動きベクトルデータの符号量を利用している。

2 MPEGと動きベクトル

MPEGは動き補償付き画像間予測とDCTのハイブリッド符号化方式である。符号化の際は、画像をマクロブロックと呼ばれる 16×16 画素サイズに分割し、このマクロブロックごとに画像間予測を行う。動き補償付きであるため、符号化の際はマクロブロックごとに参照画像との位置のずれを表す動きベクトルが情報として付与される。動きベクトルは、直前のマクロブロックの動きベクトルとの差分をとり、差分値だけを可変長符号化する。差分値に割り当てられる符号長は出現頻度を考慮して、小さな差分値に短い符号長が割り当てられている。また、予測がうまく当てはまらない場合には画像間予測

を行わずに符号化するため、動きベクトルは符号化されない。

動きベクトルの符号量とカットとの関係を考察すると、1. カットがない場合には画像間の予測符号化が使われ、しかも隣り合うマクロブロックは類似した動きベクトルを有するため、動きベクトルの符号量は小さくなる、2. カットにより全く異なる画像に変わると、カットを越えた画像間の予測が使われず、動きベクトルが符号化されなくなる、3. カット後も同じ様な色調であるためにカットを越えて画像間予測が行われる場合には、動きベクトルは不揃いになるため、動きベクトルの符号量は多くなる、という関係がある。従って動きベクトルの符号量と画像間の類似度とは反比例の関係にある。

3 カット検出方法

次式により計算される特徴量を時刻 $t-1$ の画像と時刻 t の画像間の類似度 $f(t)$ として定義する。

$$f(t) = \frac{1}{|MB(t)|} \sum_{i \in MB(t)} g(i),$$

$$g(i) = \begin{cases} 1/l(i), & \text{画像間予測符号化} \\ 0, & \text{その他} \end{cases}$$

ただし、 $l(i)$ は動きベクトル i の符号量、 $MB(t)$ は時刻 $t-1$ と時刻 t の画像をまたいで予測符号化を行うことが許されているマクロブロックの集合である。この定義では、画像間予測を行わずに符号化されている場合に対して、ペナルティとして符号長を無限大としていることになる。

MPEG符号化された動画像に対し、類似度を求めた例を図1に示す。図1のケースでは、単純なしきい値処理を行えばカットが検出できるが、動きの激しいシーンではカットのピークがはっきりと現れてこないため、ノイズの中からカットで生じるピークだけを検出する処理が必要となる。

そこで、まず類似度の両側の突出度 $f_r(t), f_l(t)$ を

$$f_r(t) = f(t) - \min_{0 \leq i \leq w_2} f(t + w_1 + i)$$

$$f_l(t) = f(t) - \min_{0 \leq i \leq w_2} f(t - w_1 - i)$$

"A Fast Detecting Method of Scene Breaks from MPEG Encoded Movies",
Toshimitsu Kaneko and Osamu Hori,
R&D Center, Toshiba Corporation,
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan.

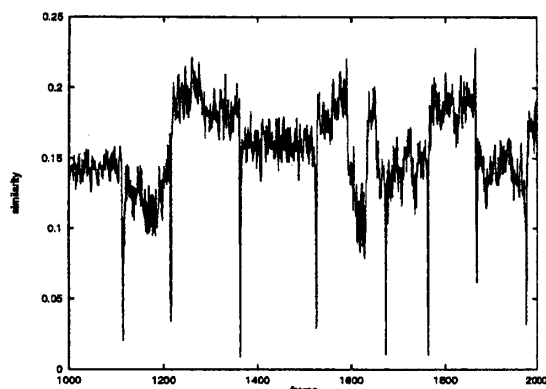


図1: 類似度

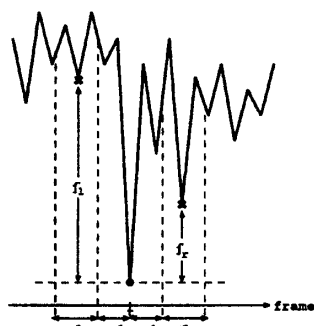


図2: フィルタリング処理

として求める(図2参照)。そして、3つのしきい値 T_b, T_i, T_s に対して、(1) $f_r(t) < T_b$ かつ $f_i(t) < T_b$ 、または、(2) $f_r(t) < T_i$ かつ $f_i(t) < T_s$ 、または、(3) $f_r(t) < T_s$ かつ $f_i(t) < T_i$ のいずれかを満たすときにカットとする。ただし、 $T_s \leq T_b \leq T_i$ とする。条件(1)は類似度のピークが両側ともはっきりと現われているとき、条件(2),(3)は片側の突出度は小さいが、逆側の突出度が大きい場合に相当する。 w_1 は大きく設定するほど鈍ったピークも検出されるようになるが、同時に動きの激しい場面での誤検出が多くなる。また、 w_2 は大きくしすぎると連続するカットが検出できなくなり、小さくすると誤検出が増える。

カット検出処理をより高速に行うため、前処理としてカット検出処理を行う区間を限定する処理を加えた。まず予測符号化が行われていないイントラピクチャ(Iピクチャ)のマクロブロックDC成分のみを復号する。次に得られた時間解像度の低い画像に対して輝度ヒストグラムを用いたカット検出処理を施す。Iピクチャ I_i と I_{i+1} との間にカットありと判定された場合、 I_i と I_{i+1} 間のフレームをカット検出処理区間とする。このとき、Iピクチャに対するカット検出処理はカットの見逃しが無いようにしきい値を設定した。

表1: 実験結果

総フレーム数	9000frames
総カット数	521
未検出	7/521 (1.3%)
過剰検出	11/521 (2.1%)
処理時間	1547.48sec (58.2 frame/sec)

4 実験結果

以上のカット検出処理を実際にMPEG2で符号化された50分の映画に対してEWS上で処理した結果を表1に示す。中にはディゾルブやワイプといった特殊なカットは含まれていないが、それを考慮しても、高い検出率が得られていることがわかる。未検出であったカットは暗いシーンでのカット、動きの大きいシーンをつなぐカットであった。一方、過剰に検出したのは速いカメラパンや画面一杯に映っている物体の高速な動きを含むシーンであった。

処理速度に関しては、30(frame/sec)以上の処理速度が達成されている。計算時間のほとんどはMPEG2の文法理解に要した時間であり、カット検出のために要している時間は僅かに0.41(sec)であった。ソフトウェアでのMPEG2の復号処理速度は4.7(frame/sec)程度であった。逆DCT等の処理を省略したためにかなりの高速化が達成されている。

本稿では瞬時のカットだけを検出対象としていた。しかしながら、映像によっては特殊なカットが多用されていること、特殊なカットにはその手法により映像製作者の意図が反映されていることがある、といった理由から、特殊なカットを検出、分類することは大変重要である。今後はこのようなカットの検出手法の開発を行っていく予定である。

参考文献

- [1] 大辻, 外村, 大庭, “動画データベースハンドリングに関する検討,” 信学技報 IE90-103, 1991.
- [2] 大辻, 外村, 大庭, “動画カット検出,” 信学技報 IE91-116, 1992.
- [3] 外村, 安部, “動画データベースハンドリングに関する検討,” 信学技報 IE89-33, 1989.
- [4] 長坂, 田中, “カラービデオ映像における自動索引付け法と物体探索法,” 情報処論, pp.543-550, 1992.
- [5] H.J.Zhang, A.Kankanhalli and S.W.Smoilar, “Automatic Partitioning of Full-Motion Video,” Multimedia Systems, pp.10-28, 1993.