

DCT成分を用いた動画シーンのクラスタリング*

— カット検出の一手法 —

6L-7

岩成 英一

有木 康雄

龍谷大学 理工学部

1 はじめに

情報のマルチメディア化に伴い、映像・音声・テキストといった情報が大量に伝送され、蓄積されつつある。マルチメディアにおいて動画データベースを構築しようとする場合、効率良く動画を圧縮できるとともに、効果的に検索できることが必要不可欠である。

この圧縮と検索を行なうためには、動画を同一シーンと思われるいくつかのクラスタに分割しておき、各クラスタ毎に内容を解析(シーン解析)することが妥当であろう。すなわち、同一シーン中にある、隣合うフレーム間では画像が類似しているという性質を使ってシーンの変わり目(シーンカット)を検出し、動画のクラスタを形成する。

本研究では、シーン解析の前処理として、シーンカットを検出し、動画をいくつかのシーンに分割することを目的としている。手法としては、DCTの第1(直流成分)、2、3成分(水平、垂直周波数成分)を用い、これらを各フレーム画像の特徴を表すパラメータとして扱う。以下にDCTによるカット検出の方法、および実験について述べる。

2 DCTによるカットの検出法

2.1 DCT成分の抽出

各フレーム画像をいくつかのブロックに分割し、そのブロックを次式に示す2次元DCTにより周波数成分に変換する。(式はブロックサイズが $N \times N$ の場合を表す)

$g(u, v) =$

$$\frac{2}{N} C_u C_v \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N}$$

ただし、

$$C_u, C_v = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } u, v = 0 \\ 1 & \text{otherwise} \end{cases}$$

このように変換された成分は図1のように、左上を直流成分とし、右下に向かって周波数が上がっていくよ

*Scene clustering in moving images by DCT components -one method of scene cut detection. Eiichi Iwanari and Yasuo Ariki (Ryukoku University)

うに成分が並ぶ。このうちの、 DC, AC_1, AC_2 の3つ

DC	AC_1	AC_5	...
AC_2	AC_4		
AC_3			
⋮			

図1: DCT成分の並び方

の成分を用いてシーンカットを検出する。

2.2 カットの検出法

処理の対象としている動画はテレビで放映されているニュース番組である。このような動画では、1秒という短い間に何度もシーンが変わることはない。そこで、一つのシーンカットから次のシーンカットまでの間には少なくとも30フレーム以上の同一シーンが存在すると仮定できる。

各フレームは 320×240 画素でモノクロ256階調の画像である。この画像を 40×40 画素の正方形のブロックで分割すると、 8×6 個のブロックが得られる。それぞれのブロックに対して上で述べたようなDCTを施し、直流成分(DC)、交流成分(AC_1, AC_2)の3つを取り出す。したがって、1フレームからは、 $8 \times 6 \times 3 = 144$ 個の成分が得られる。

この144次元空間の中で同一シーンは一つのクラスタを形成すると考える。今、問題を簡単にするために、144個の成分には相関がないと仮定する。144個の成分それぞれに対して、過去1秒間分の30フレームより平均 μ_i と分散 $\sigma_i^2 (1 \leq i \leq 144)$ を求めておく。新しく入力されたフレームのDCT成分 $x_i (1 \leq i \leq 144)$ が、

$$\left| \frac{x_i - \mu_i}{\sigma_i} \right| > 3$$

となるような成分の個数を求め、その数(over数)が総数の半分に達した時シーンカットを検出する。なお、1フレーム入力されるたびに、144個のDCT成分の平均 μ_i と分散 σ_i^2 は過去30フレームから再計算している。

従来法[1][2]に対するこの方法の利点は、動画にシーンに対応するクラスタに分割することが出来るため、

シーンの類似性を計算することが出来る点である。これはニュース番組においては、スタジオ → 現場 → スタジオ といった動画の大きな流れをとらえるのには都合が良い。

3 実験結果

実験 1 ニュース番組から約 2 分 (3753 フレーム) の連続画像を取り出して実験を行なった。図 2 に交流成分だけを用いてシーンカットを検出した結果を示す。パラメータの数は $96(8 \times 6 \times 2)$ 個なので、その半数である 48 を閾値とした。実際のシーンカットが 11 回起っ

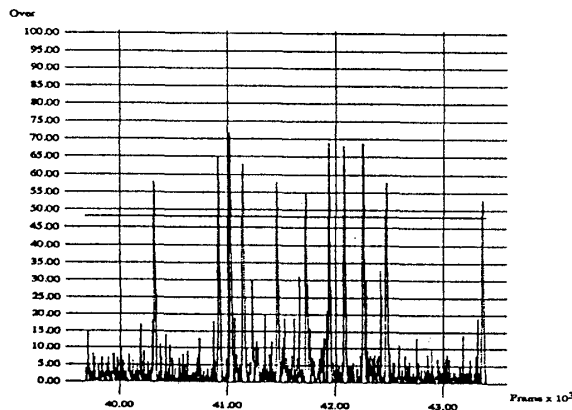


図 2: 各フレームに対する over 数

ているのに対し、交流成分のみを用いた場合には、図 2 のように 11 個全てを正しく検出することができた。次に、直流成分のみで実験した場合には、10 個のシーンカットは正しく検出したが、一つのシーンカットは検出することが出来なかった。検出に失敗した状況は、暗いシーンから暗いシーンへ 2 フレームにわたって重なりながら変わる場合であった。なお、直流成分のみの場合は、パラメータの数は、 $48(8 \times 6)$ 個なので、その半数である 24 を閾値としている。

実験 2 次に交流成分のみと直流成分のみを用いた場合、over 数がどのように変わるかを調べるために、ランプの灯りが暗くなるような画像系列に対して実験を行なった。結果を図 3 と図 4 に示す。図より、直流成分のみの場合と比べて交流成分のみの方が over 数の全体に対する比が少なく、安定であることがわかる。この理由としては、直流成分のみの時は平均値での比較になるので、画像の明るさのみが変化した場合にも反応してしまうためである。

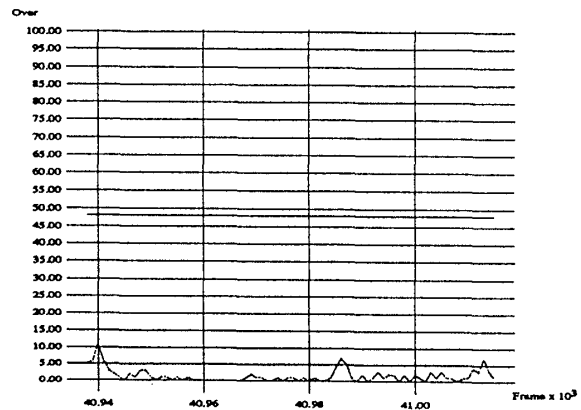


図 3: AC 成分のみの場合

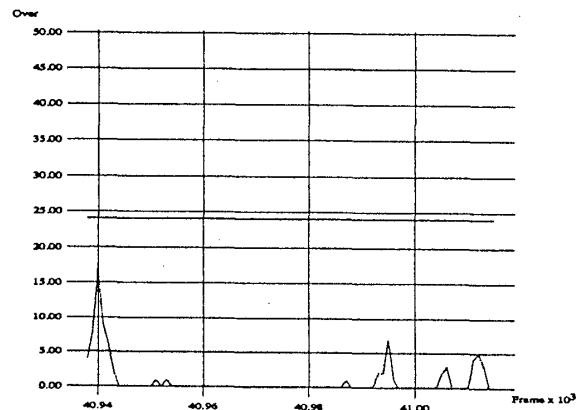


図 4: DC 成分のみの場合

4 まとめ

DCT の成分を用いて動画におけるシーンカットの検出を行なった。明るさが変化するような場合には交流成分のみを用いるのが有効であることがわかった。今後は、カメラワークや動物体によるシーンカット検出の影響を調べる予定である。

本研究は、科学技術庁の平成 5 年度科学技術振興調整費による「創造的研究支援のための自己組織型情報ベースシステムの構築に関する研究」の一環として行なわれた。

参考文献

- [1] 大辻 清太, 外村 佳伸, 大庭 有二: "動画カット検出", 信学技報, IE91-116, 1991.
- [2] 大辻 清太, 外村 佳伸, 大庭 有二: "輝度情報を使った動画ブラウジング", 信学技報, IE90-103, 1990.