

動画のシーンクラスタリングにおける同一シーンの決定法*

3M-4

岩成 英一

有木 康雄

龍谷大学 理工学部

1 はじめに

ニュース番組を計算機に取り込み、データベースを構築する研究を行なっている。これまでに、DCT成分を用いて動画像中のフレーム画像をクラスタリングし、形成されたクラスタ間でシーンをカットする方法（シーンクラスタリング）を研究してきた[1]。その結果、DCTの周波数成分を用いることによって、輝度値を用いていた従来法[2][3]よりも良好な結果が得られることがわかった。

今回の報告では、動画をシーンクラスタリングした後で、クラスタの類似性や同一性を判定する実験を行なったので、その方法と結果について報告する。この研究の利用価値は、同じような映像を見つけることにあるが、ニュース番組では次のような利用価値がある。ニュース番組における一つの記事では、ニューススタジオでキャスターが概要を述べた後、いくつかの現場のシーンを経て再びニューススタジオに戻る、という構成をとることが一般的である。従って、常に戻ってくるホームポジションのようなクラスタを見つけることによって、ニュース記事毎の切り出しが可能になる。

2 同一シーンの検出法

2.1 DCTによるクラスタの表現

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、左上を直交成分とし、右下に向かって周波数が上がっていくよ

うに配置される。このうちの、A1成分を用いて特徴空間を形成する。例えば、 $N = 40$ のときはフレーム画像の大きさを 320×240 画素とすると、 $96(8 \times 6 \times 2)$ 次元の特徴空間となる。今、シーンクラスタリングによってカット検出が終了し、既に動画がいくつかのクラスタに分割されているものとする。先ほどの特徴空間で、クラスタの正規分布を仮定し、各クラスタにおいて最初の30フレームより、DCT成分の平均ベクトル μ と分散共分散行列（対角行列） Σ を求め、これをクラスタのDCT表現とする。各クラスタで最初の30フレームを用いる理由は、カメラワークによってシーンのはじめと終りでは別のクラスタに遷移することも考えられるためである。

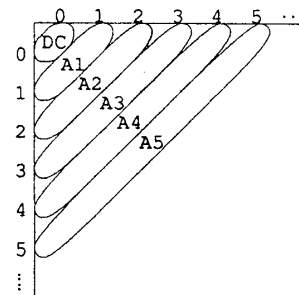


図1: クラスタ形成のためのDCT成分

2.2 シーンの同定

類似したシーンのクラスタは、特徴空間において互いに近いところに形成されると考えられる。従って、特徴空間でクラスタ間の距離を求めることによってシーンが類似しているか否かを判定することができる（シーンの同定）。クラスタ間の距離、すなわち二つの分布間の距離尺度についてはいくつかの方法が考えられるが、ここでは、クラスタの平均ベクトル間のユークリッド距離尺度を用いている。

3 実験結果

3.1 実験1

まず予備実験として、フレーム画像を特徴空間で主成分分析し、二つの直行する主軸を求めた。この2次

*A Method for Identifying Scene by Scene Clustering in Moving Picture
Eiichi Iwanari and Yasuo Ariki (Ryukoku University)

元平面にフレーム画像を図2のようにプロットし、類似したシーンのクラスタが近いところに形成されることを確かめた。

実験に用いたニュース番組では、図4に示すようにシーン0、6、17がニューススタジオで、同じシーンとなっている。また、シーン7と16も同じシーンである。これらの同一シーンの関係が図2のクラスタの位置にも反映されている。興味あることに、シーン11、12、13、14が近いところにクラスタを形成している。これらのシーンは、別なシーンであるにもかかわらず、画像の上半分が空であるという共通点を持っている。同一シーンを検出をするという点で誤りであるが、類似したシーンを探することができるという可能性を示しているといえる。

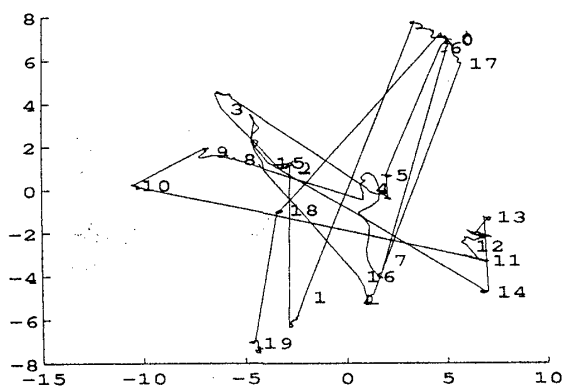


図2: 主成分分析によるクラスタ位置の表現

3.2 実験2

2.2で述べた方法により、各クラスタの最初の30フレームより求めた平均ベクトルを用いて、クラスタ間のユークリッド距離を求めた。結果を図3に示す。全てのクラスタ間のユークリッド距離求めたが、ここでは、シーン17と他の全てのシーンとの間で計算した距離を示す。

同じシーンである0と6では、距離が近い値を示しているが、その差があまり大きくないので、しきい値設定に問題を残している。

4 まとめ

DCT成分を用いてシーンのクラスタリングを行ない、そのクラスタから同一シーンを決定する方法について述べた。

同一シーンは特徴空間中で近い位置にクラスタを形成するので、クラスタ間のユークリッド距離がしきい値より小さいものを同一シーンと判定した。今後は、

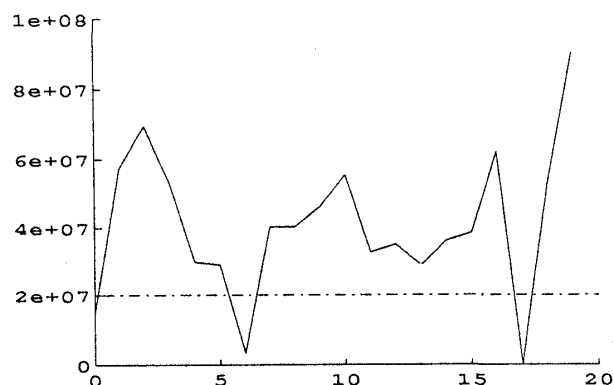


図3: シーン17と他のシーンとの距離

動画の記述について研究を進めていく予定である。



図4: 実験に用いた動画

本研究の一部は、科学技術庁の平成5年度科学技術振興調整費による「創造的研究開発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] 岩成 英一, 有木 康雄: "DCT成分を用いた動画シーンのクラスタリングとカット検出", 信学技報, PRU93-119, 1994.
- [2] 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.
- [3] 上田 博唯: "インタラクティブな動画画像編集方式の提案", 信学技報, IE90-6, 1990.