

マルチメディアデータベースシステムの 為の映像の自動抄録に関する研究

5V-2

西川博行 得丸孝行 打浪清一
九州工業大学

1 背景

放送局を始め、博物館でも、個人でも大量の動画像記録を保持する時代となってきた。しかし、十分な内容索引が出来ていないためマルチメディアデータが有効に活用されていない。

そこで、本研究では、大量の動画像データがある時、これからあまり人手をかけずにできるだけ自動的に実時間で抄録を作成する方法について検討し、その方法論を確立することを目的とする。

2 概要

抄録とは動画像（テープ、光ディスクなどに記録されたもの）が与えられた時、何が何処に記録されているか、その内容と位置を求めることをいう。つまり、シーンが変わるたびにそこへインデックスを付ける事と、そのシーンでは何が写っているのか、また、そのシーンの状態（ズーム、パン等）を求めることをいう。なお、シーンとは区切り目のない一連の映像をいう。また、区切り目とは、カメラが一旦止められ別の対象物を写し始めた時、及び、シーンの状態が変わる時をいう。

シーンの状態を、FIX、PAN、ZOOMの3つに分けた。今回の場合、カメラが一旦止められ別の対象物を写し始めるChangeを含める。なお、FIXはカメラが固定されている状況、PANはカメラがパンしている状況、ZOOMはカメラがズームインもしくはズームアウトしている状況をいう。

ここで、動画像を再生しながら、静止画として順次取り込みつつ、時間的に隣接する静止画を比較し、シーンを区切ってゆくシステムを考えた。

微小時間において隣接する2枚の静止画を動画像より抜きだし、その間の種々の変化を検出するために12種の

評価値を設定した。

これらの評価値が、シーンの抽出にどの程度有効か実験し、調査した。その結果、12種中4種の評価値が、動画像の特徴を顕著に捉えていることが判明した。そこで、この4種の評価値Diff、Col、HDiff、VDiffを用いた、後で述べる評価関数を設定した。

Diffは2枚の静止画の各々の対応する点で色情報の差をとり、これを画面全体で集計したものであり、漠然とした画像の違いを捉えられる。Colは1枚ずつ静止画の色情報の総和をとり、それぞれの差をとったものである。Diffと違い位置関係によらないため、ズームなどにより位置がずれても同じ画像か否か認識が出来る。HDiffは画像を水平にずらしながら相関をとったものがある。その対象となる2枚の静止画の違いは同じものが水平にずれることにより生じたものかどうか検出できる。VDiffも同様で、垂直方向についての演算である。

今回試作したシステムは、これらの評価値の特性を効果的に利用して、区切り目を検出するための評価関数を求めるための測定を行なうものである。

本稿では、与えられた動画像について時間軸について微小時間、今回は1秒程度離れたフレームの2枚の静止画について考案した評価値の演算を行ない、得られた値から、2枚の静止画像の違いを認識することの出来る評価関数を提案した。これにより、シーンの区切り目の位置だけでなくそのシーンの状態の種類も含めて抽出できる。

また、その評価関数を一般の動画像に適用し、その有意性を確認した。

3 評価関数

ここで、試作システムの出力結果より評価関数を求める。試作システムでは12個の評価値を計算したが、計測、評価の結果、すべての評価値を考慮しても、飛躍的に評価関数の信頼度が向上するわけではないことが確認できた。このため、そのうちの特に有効なものを用いて

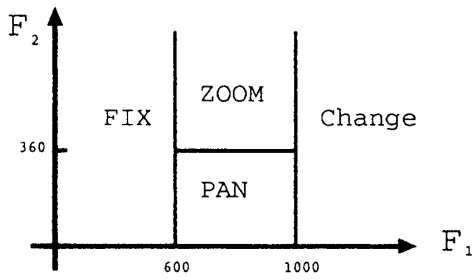


図 1: 評価関数

以下の2つの式をカット抽出評価関数とした。

$$F_1 = 0.748Diff + 0.595Col$$

$$F_2 = HDiff + VDiff$$

これらの評価関数の係数は、4つのシーンの状態の例について計測した結果により求めたものである。

F_1 は漠然とした画面の違いを抽出し、 F_2 はパンしているかどうかを抽出する。

F_1 の出力が600以下でFIX、1000以上でChangeと判別する。それ以外だった場合、 F_2 の出力を求め、360以上でZOOM、以下でPANと判別する。これを図1に示す。

4 実験結果

一般の動画画像を試作システムで処理し、提案した判別関数を適用した。FIX、PAN、ZOOM、Changeの各々の場合毎に20の例をとって判別した結果を表1に示す。

80種の場合について計測を行ない、そのうち誤認識が5つあった。ただし、そのうち2つはFIXをPANと誤認識したもののなのでX軸方向の移動量、Y軸方向の移動量が0であることから、誤りを訂正することができる。したがって、実質の誤認識率は、平均約4%である。

表 1: 評価関数による判別結果

状態	正認識	誤認識	FIX	PAN	ZOOM	Change
FIX	18	2	18	2	0	0
PAN	18	2	1	18	1	0
ZOOM	19	1	1	0	19	0
Change	20	0	0	0	0	20

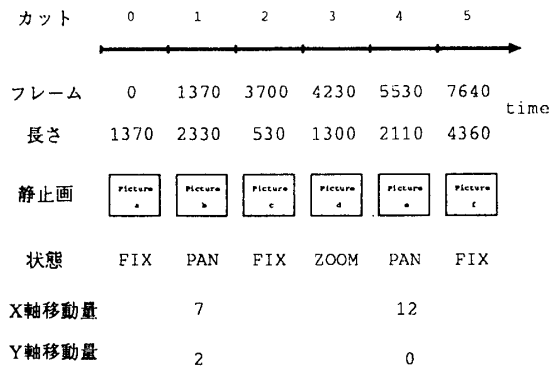


図 2: 動画画像の抄録

5 まとめ

提案方式を評価することにより、平均約4%の誤認識率でカットを区切り、判定することができることを確認した。これにより、抄録の自動化が可能であることを確かめることができた。

本稿での抄録例を図2に示す。

ここで、本方式の抄録のための計算時間は、 F_1 、 F_2 合わせて数秒程度であった。このため、今回の目的の一つであった実時間処理は、実現できなかった。

それゆえ、現段階では、動画画像をQuickTimeのMovieとして、一旦取り込んだ後、そのファイルから連続して画像をとりだしながら、本方式で提案した関数を適用していった。

使用した計算機はMacintosh Quadra840AVである。本研究で提案した評価関数をより高速のPowerPCで用いれば実時間で内容抄録を行なうシステムが実現できると考える。

6 謝辞

本研究の一部は、大川情報通信基金の研究助成を受けた。ここに記し、深く感謝します。

参考文献

- [1] 塩、スクランスキー：「動画画像からの歩行者検出法」、情処研報、CV75-5(1991)
- [2] 嶺、八木、谷内田：「時系列差分画像を用いた複数移動体の抽出及び追跡」、情処研報、CV81-7(1993)