

内容に基づいた動画像の構造分析*

2 P-4

村野井 亮治 趙 継英 太田 浩二 早坂 里奈 松下 温†
慶應義塾大学理工学部‡

1 はじめに

動画像は静止画像と違って、全体を一目で見ることは不可能である。そこで、動画像全体の概要把握や、動画像中の希望するポイントへのアクセスを可能にする機能が必要となる。しかし、ビデオデッキのような早送り・巻戻しといったものでは、非常に時間がかかり、また見落としてしまうといった問題が生じる。そのような問題を解決するために、各カメラカットの間の部分(ショット)を単位として、ショットの代表フレームを表示するシステムが現在までに提案されている [1]。しかしこれでは、ショット数が膨大な場合に表示量が非常に多くなり扱いづらい。

そこで本研究では、ショットよりも大きなまとまりとしてシーンに注目し、映画を対象として動画像中からシーンを抽出して階層的に表現するシステムを提案する。

2 動画像の構造解析

まずショットを抽出し、次に、ショット間の類似度をを用いてシーンを抽出する。

なお、本研究は MPEG-1 により符号化された動画像を対象とする。

2.1 ショットの抽出

動画像データの最小単位として、ショットの抽出を行なう。MPEG-1 符号化画像における B、P ピクチャの参照方向の変化によりフレーム間の相関を調べ、カット点を検出する。この方法では、動画像を完全に復号する必要がないため、高速な処理が可能となる。

2.2 代表フレームの選出

ショットを表す代表フレームを選出する。

この代表フレームは、シーン抽出の際にショット間の類似度を求めるために用いる他、ユーザにショットの内容を提示するためにも用いられる。

長く、動きのあるショットの場合はひとつのフレームでショット全体を代表させるのは難しいため、ひとつのショットに対し複数の代表フレームを選出する。

処理量削減のため I ピクチャのみを候補とし、次の手順で行なう。

1. ショット中の動きが少ない部分を抽出し、これを初期クラスタとする。
2. ショット中の I ピクチャを群平均法によってクラスタリングする。
3. 各クラスタについて、I ピクチャを平均した平均画像との距離が最も小さい I ピクチャを代表フレームとする。

2.3 ショットの統合によるシーンの抽出

会話のシーンのような似ているショットの繰り返しになる部分は、同じシーンに属する可能性が高い。この部分をまとめることでシーンを抽出する。

$shot_n$ と $shot_{n+1}$ との間の結合度 $connect_{n,n+1}$ を、まわりのショット同士の類似度から次のように求める。

$$connect_{n,n+1} = 1 - \prod_{i=n-N+1}^n \prod_{j=n+1}^{i+N} d_{ij} \quad (1)$$

ここで、 N は比較するショットの範囲を表す。 d_{ij} は $shot_i$ と $shot_j$ の類似度であり、それぞれの代表フレーム間の距離(各画素値およびヒストグラムの差)などからファジイ推論によって求められ、 $[0, 1]$ の値を取る。

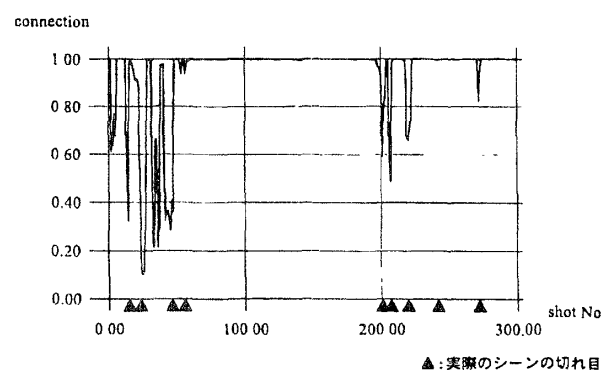


図 1: 結合度の変化

全ショットの結合度の変化は図1のようになる。この図からわかるように、結合度の変化が大きい点をシーンの切れ目とすることで、実際のシーンの切れ目を検出することができる。

*Content-based Analysis of Video Structure

†Ryoji Muranoi, Jiying Zhao, Koji Ohta, Rina Hayasaka, Yukaka Matsushita

‡Faculty of Science and Technology, Keio University



図 2: ビデオブラウザ

3 ビデオブラウザ

前章までに述べたアルゴリズムによって解析された結果を利用したビデオブラウザを図 2 に示す。従来のビデオデッキのような連続的なアクセス機能に、本研究で提案するランダムなアクセス機能を加えた。

図 2 の左上のウィンドウは、選択されたシーンおよびショットの再生を行なうモニターである。図 2 の右のウィンドウ(シーンウィンドウ)は、一目で概要の把握をするためのものである。各シーンを最高 3 枚の代表フレームで表示している。そのため、従来のショットの代表フレームを表示するものよりもはるかに表示量が減っており、それでいてシーンにおける概要がわかるようになっている。図 2 の左下のウィンドウ(ショットウィンドウ)は、さらに詳しい情報を知りたい時に、選択されたシーンに含まれるショットを表示するものである。このように 2 つのウィンドウは階層的に表示されるため、まずシーンウィンドウによって動画の

概要を把握してから必要に応じてショットウィンドウで詳細を把握したり、あるいは各シーンおよびショットの再生を行なうといったようなユーザのニーズに応じた操作が可能となる。

4 まとめ

以上のようにして、動画像の対象に映画を用いてシーン抽出を行ない、階層的に表示するシステムを提案した。表示量を削減しつつ情報伝達量は損なわないため、動画像の概要把握や検索を効率的に行うことができる。

参考文献

- [1] F.Arman, R.Depommier, A.Hsu, and M-Y.Chiu. Content-based browsing of video sequences. In *Proceedings ACM Multimedia '94*, pp. 97-103, 1994.