

内容把握のための動画像解析手法の一検討

3C-4

前田潤治 洪政国

日本アイ・ビー・エム(株)東京基礎研究所

1 はじめに

動画像は、静止画像の時系列の集合であるが、一枚一枚の静止画像(フレーム)を取り出しても意味をなさず、ある連続した時間を単位とするシーン(カット)の集合であると考えの方が自然である[1]。また、シーンの変わり目を検出する手法も数多く提案されており(例えば[2][3])、一連の動画像をシーンに分割して検索や編集のために利用する方法も提案されている。[4]

しかし、シーンの変化とは、フレーム間の物理的な変化であり、撮影者の目的を意味的にきめ細かに反映しているとは限らない。具体的には、一つの物理的シーンの中に、カメラの操作等によって複数の目的物を時間をずらせて写しこむことは頻繁にある。それを検索する立場からいえば、異なった時間に存在する異なった目的物を捜すときには、異なった検索結果が得られる方が自然である。一つのシーンにつき、その内容をよく表している一枚のフレームを代表フレームとして検索に利用する方法が提案されているが[4]、特にそのようなシステムの場合、同一シーン内であっても、異なる目的物に対しては異なるフレームを取り出して検索に利用できれば、その動画像の内容を把握する大きな助けになる。しかし、現在のところ、カメラ操作の推定は一般に高価なアルゴリズムを必要とする。(例えば[5])

そこで本稿では、実写の動画像を対象とし、シーン内の構造に着目し、安価なアルゴリズムで物理的なシーンを分割して、撮影者の意図をよりきめ細かに反映し、内容把握を容易にするような手法を提案する。

2 サブシーンの定義

サブシーンとは、直観的には、同一シーン内に存在するが互いに物理的な差異の大きなフレーム列であるとする。

"A Study on Moving-Picture Analysis for Supporting Content-Understanding"

Junji Maeda Hong Jung-kook
IBM Japan, Tokyo Research Laboratory,
1623-14, Shimotsuruma, Yamato, Kanagawa, 242, Japan

本研究の主眼は、このようなサブシーンという概念を導入し、それがどれくらい撮影者の意図を反映し、検索する者の内容把握を助けるかを調べることである。もちろん、シナリオやコンテといったレベルでの撮影者の意図を理解するのは極めて困難である。ここでは同一シーン内に複数の目的物が時間をずらせて存在するような場合を検知し、それぞれを別のサブシーンの目的物として分離できれば成功とする。

より具体的には、サブシーンとは、同一シーン内のフレームを時系列に沿っていくつか分割した集合で、自分より前後の集合とは内容の変化があるが、内部では(少なくとも時間的に前方では)比較的变化が一様であるようなものとする。「比較的变化が一様」という条件を導入したのは、変化の少ない部分は撮影者が興味を持った箇所であると考えられるからである。[6]

3 サブシーンの検出

ここではすでに、シーンの変わり目を検出する方法は与えられているとする。

サブシーンを以下のようにして検出する。

1. まず、シーンの先頭のフレームをそのシーンの先頭サブシーンの代表フレームとして登録する。
2. それに続く各フレームについて、
 - 代表フレームとの物理的变化
 - 直前のフレームとの物理的变化を求める。
3. これら、二種類の物理的变化について、以下の条件が同時に成り立った場合、そのフレームをサブシーンの変わり目と見なし、新たなサブシーンの代表フレームとして登録する。
 - 代表フレームとの差が大きくなり、適当な閾値を超えた
 - 直前のフレームとの差の絶対値の差分が負で、その絶対値が急に大きくなり、適当な閾値を超えた(これは、前フレームとの差が、大きなものから小さなものへ急に変化したことを意味する)

4 実験

データとして、国立民族学博物館の「アンデス民族学ビデオテーク」の番組の一部を使用した。

物理的変化量として、当該フレーム間のカラーヒストグラムの類似度 [3] を用いた。

検出されたサブシーンの代表フレームの例を図に示す。ここで、(b) が新たに検出されたサブシーンの代表フレーム、(a) はもともと (b) のフレームが属していたシーンの先頭フレーム、(c) は次のシーンの先頭フレームである。

(内容は、(a) は容器の中のいも、(b) はいもの皮を剥いているところ、(c) はこれらの情景を引いて撮ったものである。)

この例では、シーンの先頭の目的物と明らかに異なる目的物を取り出すことができた。

5 まとめ

本稿では、撮影者の意図をきめ細かに反映した構造としてサブシーンの概念を導入し、それを抽出する方法を提案した。動画像の内容によらない普遍的な閾値の決め方は一つの課題として残っている。

また、将来カメラ操作の推定が安価に行えるようになれば、撮影者の意図をさらに正確に反映させることが期待できる。

6 おわりに

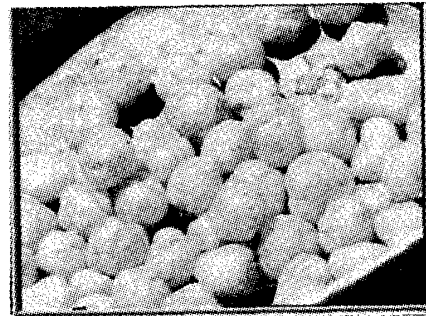
本研究は国立民族学博物館（民博）と日本アイ・ビー・エム（株）との共同研究「博物館におけるマルチメディアの有効利用」の一環であり、実験に使用されたデータは民博に所属する。ご指導いただいた民博の杉田繁治教授、データの準備などのご協力をいただいた民博の鈴木明、中川隆、宇治谷恵の諸氏に感謝いたします。

参考文献

- [1] Kawagoe, K., "Continuous Media Data Management," ACM SIGMOD, Records, Vol. 20, No. 3, pp.74-75, 1991.
- [2] Ioka, M., "Method of detecting scene change in moving picture," IBM Technical Disclosure Bulletin, Vol. 34, No. 10A, Mar. 1992.
- [3] 長坂, 田中, "ビデオ作品の場面変わりの自動検出法," 第40回情処全国大会, 1Q-5, 1990.
- [4] 草場, 高橋, 洪, "映像データベースにおける情報の入力と管理," 情処研報, 92-CH-15-2, 1992.

[5] 阿久津, 外村, 橋本, 大庭, "動画像インデクシングを目的としたカメラ操作の規定方法," 信学誌 D-II, Vol. J75-D-II, No. 2, pp.226-235, 1992.

[6] 内山, 米田, "内容把握を支援する動画インデックス," 1993年信学会春季大会 SD-9-4.



(a) 当該シーンの先頭フレーム



(b) 新たに抽出されたサブシーンの代表フレーム



(c) 次シーンの先頭フレーム

図. サブシーン抽出例