

## 大局的な視覚特徴追跡に基づく動画分類手法

石田 明久 †

吉田 謙一 ‡

高橋 成雄 ‡

東京大学 理学部 情報科学科 †

東京大学 大学院新領域創成科学研究科 複雑理工専攻 ‡

## 1 はじめに

近年の計算機性能の向上にともない、我々は大量の動画を扱う機会を容易に得ることができるようになってきた。従来そのような動画を検索・分類できる手法は、動画を信号ととらえ周波数解析を行うもの [1] や、動画からコンテンツをパターンマッチング等を用いて抽出するもの [2] など、自動で行う手法も開発されてきている。しかし、これら既存の手法では、人の視覚的な注意をひきつける特徴に関して、十分に配慮をしていないという問題があった。

本研究では、そのような視覚的注意の集中する特徴部分の時間変化に着目して、内容的に継ぎ目のない動画を自動的に分類する手法の開発を目指している。特に本論文では、そのような動画の分類を行うための、人の視覚特性に基づいた特徴抽出手法について示す。本論文の貢献は、時間変化を伴う注目度マップを用いることで、人の知覚に考慮にいたった動画の解析・分類手法を与えることである。この手法においては、まず時間変化する動画の各フレームから注目度マップを計算し、その注目度マップを3次元的に重ねることでボリュームデータを構成する。そしてそのボリュームデータから等値面を抽出することで、そこに内在する視覚的注意の集まる特徴領域を切り分けていく。この等値面の特徴を解析することにより、入力動画がもつ時空間内の大局的な特徴の分布を抽出していく。さらに、このような大局的な分布を符号化することで、人の注目する特徴に基づいて動画を効果的に分類するシステムを実現していく。

A Method for Categorizing Video Clips by Tracking Global Salient Features

†Haruhisa ISHIDA: Department of Information Science, Faculty of Science, The University of Tokyo.

‡Kenichi YOSHIDA and Shigeo TAKAHASHI: Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, The University of Tokyo

## 2 視覚的特徴の抽出

まず解析を行う対象である動画の特徴と、それを抽出する方法について述べる。今回解析対象とする動画の特徴とは、注目度マップを基にしたボリュームデータである。動画からこのボリュームデータを作成する方法は、以下のようなステップで構成される。

1. 動画の各フレームから輝度、色、方向性に基づいた注目度マップ [3] を別々に作成する
2. 動画をボリュームデータとみなし、その中からある閾値以上の値を持つサブボリュームを取得する

図1は、滝の流れの動画から注目度マップを抽出した例を示している。以後、このように抽出したサブボリュームから、解析・分類に必要な動画ごとの特徴抽出を行っていく。ここで、閾値の設定は全体のボリュームに対するサブボリュームの割合から適宜設定している。注目度マップの値の散らばりからの自動的な閾値の設定は、今後の課題である。

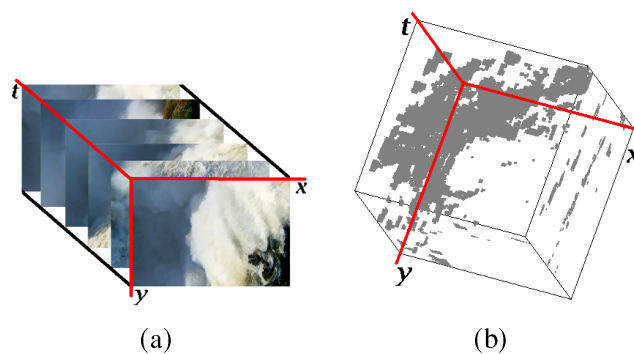


図1: (a) 入力動画である滝の流れの映像, (b) 抽出されたサブボリュームの例.

### 3 特徴の解析

#### 3.1 特徴量の取得

ここではポリウムデータから取得するパラメータと、その過程に関して述べる。先ほど取得した閾値以上の値を持つ部分は一般的に人の目を引く特徴ごとにいくつかの領域に分かれているため、まずそれを連結成分ごとに分離していく。さらに、それぞれの視覚的特徴の重要度を評価するため、その領域全体における注目度の積分を求める。ここで得られているサブポリウムのすべてに、人の視覚的注意が向けられるとは限らない。ここでは、すべてのサブポリウム連結成分ごとに注目度の積分値を計算し、その中の最大値を人の視覚の分解能を考慮したある定数で割り、その値を閾値として連結成分の取捨選択を行うことにする。

その後、残った連結成分の集合から、(a) ポリウムデータ内の連結成分の個数と個々の領域内の注目度の積分の値、(b) 量子化された各時間範囲内で、連結成分を時間軸に垂直な面で切断したときの得られる断面における重心位置の移動距離とその角度、そして(c) 断面積の時間変化量や断面の空間的位置を特徴量として抽出していく。これらの特徴量は、それぞれ(A) 人が動画を見る際に注目する対象の数、(B) 人の視覚的注意が局所的に集中するかあるいは全体を見るかといったこと、また(C) 注目する対象の動きに関する情報を取得し、そこから動画の特徴を取得するため定義されていることに注意する。

例えば、(a) において、個々の領域内の注目度の積分値を量子化して横軸にとり、ポリウムデータ内の連結成分の個数の縦軸にとると、図2のようなヒストグラムを得ることができる。これにより、積分値を量子化して得られる区間数を、次元としてもつ特徴ベクトルを定義することができる。(b) や(c) の特徴からも同様にして、特徴ベクトルを抽出していく。

#### 3.2 取得した特徴量の解析

ここでは得られた特徴量は、非常に大きな次元を持つベクトルとして定義されることになる。これを、クラスタリングすることにより動画の分類を行うことが今後の課題となる、具体的な作業としては、まずそのような特徴量が定義されている空間において、動画のサンプルの分布に対し主因子分析を行い、次元圧縮を行うことで特徴的空間を決定する。その後、

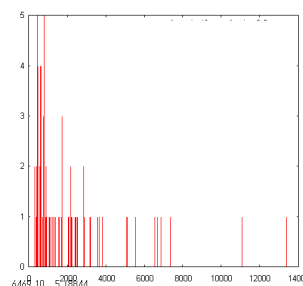


図 2: サブポリウムの積分値 (横軸) と連結成分の個数 (縦軸) のヒストグラム

動画を分類するために、特徴空間において動画を特徴づけるベクトルをクラスタリングする予定である。また次元圧縮に際しては、データ空間に内在する多様体を構築することで、より効果的に次元圧縮が行えるラプラシアン固有写像 [4] を用いる手法も検討していきたい。

### 4 終わりに

本稿では、動画内において人が注目する領域を取得し、その大局的な特徴を取得することで動画を分類する手法の導入部分に関して述べた。動画を自動で分類するために特徴を取得する際、人の視覚特性を考慮することでより効果的に特徴を取得することができると思われる。

#### 参考文献

- [1] Y. Sugiyama and Y. Araki. Classification of TV sports news by DCT features using multiple subspace method. In *IEEE Proceedings of the 14th International Conference on Pattern Recognition*, Vol. 2, pp. 1488–1492, 1998.
- [2] J. Assfalg, M. Bertini, C. Colombo, and A. D. Bimbo. Semantic annotation of sports videos. *IEEE MultiMedia*, Vol. 9, No. 2, pp. 52–60, 2002.
- [3] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254–1259, 1998.
- [4] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, Vol. 15, pp. 1373–1396, 2003.