

## 統計的な特徴量に基づく動画像検索手法

精廬幹人 橋本周司

早稲田大学理工学部 応用物理学科

{mit,shuji}@shalab.phys.waseda.ac.jp

## 1. はじめに

近年、MPEG/JPEGなどの規格化やネットワーク技術の発達や計算機の高速化に伴い大量の画像コンテンツが容易に取得できるようになってきた。それに伴い動画像を対象とした画像検索の重要性が高まっている<sup>[1]</sup>。

これまでに提案されている動画像検索はシーンチェンジの検出から得られたキーフレームの静止画像検索を基本とするものが多い<sup>[2]</sup>。これに対しここでは、動画像の持つ統計的な特徴量に基づいて、動画像をキーとして大量の動画像群から目的のシーンを検索する新しい動画像検索手法を提案する。

## 2. 手法

まずここでは動画像全体のことをストリームと呼び動画像中の一部分をクリップとして呼ぶことにする。

動画像の情報は非常に大きなものになるため、検索の際には情報量をいかに削減した特徴量を得るかが重要である。本手法は情報量の削減のために動画像が時系列に沿ったデータであることに着目しストリームの各フレームを粗く分類し、クリップ中のクラス間の遷移の統計的性質を特徴量としている。

具体的には、各フレーム間の画像の差分の総和を量子化によりクラス分けし、クリップの特徴量はこのクラス間の時間遷移を有向グラフにしたものを利用する。クラスの数  $C_{max}$  とすれば、この有向グラフは  $C_{max} \times C_{max}$  の隣接行列で表すことができる。そしてクリップ間の類似度を計算するために、この隣接行列をベクトル表記し、特徴ベクトルとする。クリップの類似度は検索対象の動画像のストリーム中のクリップの特徴ベクトルとキークリップの特徴ベクトルの内積をとって正規化したものを用いた。

## 2.1 ストリームのクラス列

まず、ストリームのクラス列を得るために動画像の各フレーム間の差分値を用いてクラス分けする。このとき時刻  $t$  のフレームが属するクラス  $C_i$  次のようにして求めるまず

$$C_i = \sum_{y=0}^{RGB} \sum_{x=0}^h \sum_{y=0}^w (I_{\{RGB\},t,x,y} - I_{\{RGB\},t-1,x,y}) \quad (1)$$

$h$ : 画像の高さ  $w$ : 画像の幅  
 $I_{\{RGB\},t,x,y}$ : 時刻  $t$  の  $(x,y)$  の画素値

The search method of moving image based on statistical feature.

Mikito Toguro and Shuji Hashimoto

Department of Applied Physics, Waseda University

55N4-10 3-4-1 Ohkubo Shinjuku-ku Tokyo 169-8555, Japan

として、 $C_i$  を線形量子化し、 $C_{max}$  個のクラスに分類する。このクラス分類を順次行いストリームをクラス列にする。

## 2.2 特徴ベクトルの作成

次にクラス列から遷移の有向グラフを表現する隣接行列  $M$  を作成する。

$$M = \begin{pmatrix} m_{00} & \cdots & m_{0j} \\ \vdots & \ddots & \vdots \\ m_{j0} & \cdots & m_{jj} \end{pmatrix} \quad (j=C_{max}) \quad (3)$$

ただし、 $m_{ij}$  はクリップ中でクラス  $C_i$  から  $C_j$  への遷移の回数である。

例えばクラス列が  $\{0, 1, 2, 1, 2, 1, 0\}$  ( $C_{max}=3$ ) の場合に  $M$  は次のようになる。

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} \quad (4)$$

この隣接行列をクリップ毎に作成しクリップ間の類似度を計算するためにベクトル表記する。これを特徴ベクトル  $V$  と呼ぶことにする。

$$V = \{m_{00}, \dots, m_{0j}, m_{j0}, \dots, m_{jj}, \dots, m_{j0}, \dots, m_{jj}\} \quad (5)$$

検索対象の動画像から、先頭を  $t$  フレームにして、キーストリーム長分切り出したサンプルクリップの特徴量ベクトルを  $V_{st(t)}$ 、キーストリームの特徴量ベクトルを  $V_k$  としたときのクリップ間の類似度  $V_i$  をベクトルの内積を正規化したもの(6)で定義する。

$$V_i = V_{st(t)} \cdot V_k / |V_{st(t)}| |V_k| \quad (6)$$

検索では、動画像長を  $L$ 、キーストリーム長を  $l$  としたときクリップ間の類似度  $V_i$  を  $t=0$  から  $t=T$  までシフトしながら計算する。ただし、 $T=L-l$  である。検索結果はこの類似度の中で最大の値をとる  $t$  から始まるクリップである。

## 3 実験

## 3.1 予備実験

まず最初に、以上のような手法が有効に動画像検索を行えることを確かめるためにランダムに生成した10000フレーム分のクラス点列から、5000フレーム目から30フレームのクラス列をキーとして

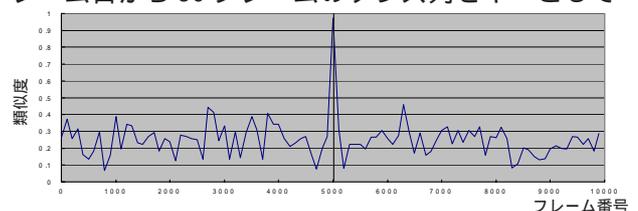


図1 予備実験結果

検索実験を行った。この結果を図1に示す。ただし  $C_{max}=10$ とした。この結果では切り出したクリップ部以外の部分では類似度が低いことがわかる。

### 3.2 実際の動画像を用いての検索実験

実際に放送されたニュース番組を利用して、検索実験を行った。ニュース番組は、ニュース素材として、さまざまな場面が映し出されているため、素材として適切であると考えられる。今回の実験では、NHKの7時のニュースを約20分録画しこれに対して処理を行った。処理の条件を表1に示す。

表1. 処理の条件

画像サイズ	352x240 (Video CD 規格)
フレーム数	38000[frame] (21分6秒)
クラス数	10

各処理では表2の環境を用いた。

表2. 処理環境

エンコーダ機	Sony VAIO S550
特徴量抽出処理	IBM AT互換機 Celeron 330MHz, メモリ128Mbyte OS: Linux
検索処理機	IBM Thinkpad X22 Pentium 700MHz, メモリ 386Mbyte OS: Windows

MPEG エンコード処理用ハードウェアが搭載されている Sony VAIO S550 で作成した MPEG ファイルを特徴量抽出処理機で特徴量抽出処理をした後、検索処理機においてニュースキャスタが画面に現れている 11317 フレーム目から 230 フレームの部分を選択して切り出し検索処理を行った。

このニュース番組中の類似度の変化を図2に示す。ただしキークリップの位置に縦線を引いてある。

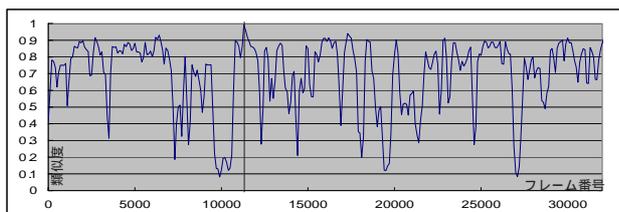


図2. ニュース番組中の類似度の変化

### 3.3 類似度の高いクリップ

次に、類似度の高いクリップがどのようなクリップであるかを調べた。

図3は図2のグラフのうち、類似度が0.9以上のピークを示している部分に関して、初めのフレームの画像を切り出して並べたものである。各ピークから線で示されている画像がそのクリップの先頭の画像である。

この実験では、ニュースキャスターの上半身が映し出されているクリップをキーにして検索したわけであるが、この結果を見る限り、人の上半身が映し出されているクリップが多く検出されていることがわかる。

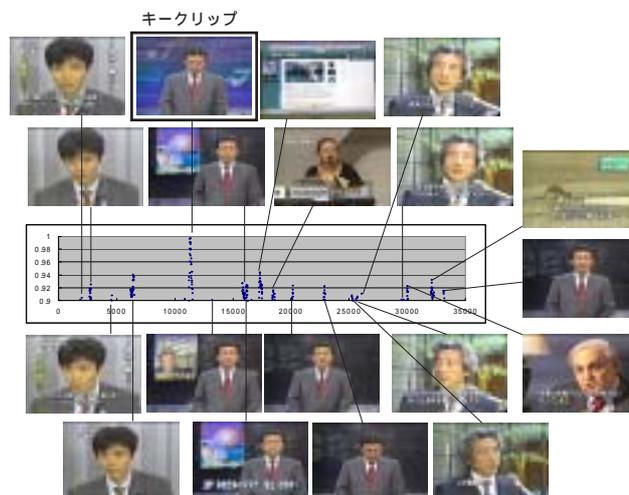


図3. 類似度の高いフレーム

今回の手法では、フレーム間の画素の差分値でクラスタリング処理を行い検索しているため、明示的に画像同士の相関を見ているわけではないが、画像的に類似したクリップで類似度が高いという結果が得られた。

### 3.4 処理時間

最後に処理時間を表3示す。検索のプログラムの処理の都合上、MPEG ファイルを一度 PNM フォーマットに変換をかけているが、処理時間の88%はこの処理にかかった時間である。

表3. 処理時間

処理時間	2時間1分
(内MPEGデコード処理)	1時間47分
(内特徴量抽出処理)	14分31秒
特徴量ファイルサイズ	75998[Byte]
検索時間	約4秒

以上から特徴量抽出処理および検索時間については実用的な時間内で行うことができることが判る。

## 4 まとめ

統計的な特徴量に基づいた動画像検索手法を提案し実際に放送されたニュース画像を用いて有効性を検討した。類似度の高いクリップの内容から、本手法の有効性が確認できた。今回は各フレームの分類にフレーム間の差分を利用したが、この分類を行う方法については議論の余地がある。たとえば MPEG のストリームでは、動き予測ベクトルが得られるが、これを利用して動的特徴の分類を行う、または MPEG ストリームのビットレートの変化を元に、目的とする動画を検索する手法などが考えられており、現在検討中である。

### 参考文献

- [1] “大規模自動装填ロボットによるペタバイト級映像アーカイブシステム”, 丹野 義和, 前原 文雄, 関谷 里美, 伊藤 学, 露峰 浩, 長谷川 文雄, 電子情報通信学会誌 Vol. J84-D2 No.6 pp.1102-1111, 2001
- [2] “動きベクトルと色情報を用いた階層的 MPEG 画像検索”, 大野 剛宏, 青木工太, 吉田 俊之, 酒井 善則, 2000 年度映像メディア処理シンポジウム講演論文集, 2000