

# Tiny LSH 法による異なる特徴量の動画検索性能の比較検討

村林 昇† 吉田 健一†

筑波大学大学院 ビジネス科学研究科†

## 1. はじめに

Web 上における動画の投稿や検索サービスが普及し、著作権を侵害した大規模な複製動画の高速検出技術の重要性が増している。我々は、先の研究 [1] で、Locality Sensitive Hashing (以下 LSH) [2] と同様の機能を持つ Tiny LSH と Direct-mapped cache を組み合わせた動画の高速検索手法である Tiny LSH 法を提案し、検索時間と検索正解率の有効性を確認した。本研究では、音声特徴量や画像特徴量など様々な特徴量に対して Tiny LSH 法に適用した場合に、検索性能の違いを比較検討する実験を行った。

## 2. Tiny LSH 法に基づく高速検索手法

### 2.1 RGB 色空間に基づいた画像特徴量

文献 [3] は、RGB 色空間に基づいた画像特徴量を用いたビデオ検索の研究で検索時間は長いながらも検索性能が良い結果を報告している。本研究ではその手法と同様のアプローチをとった。図 1 に RGB 色空間に基づいた画像特徴量の処理方法を示す。図 1(4) で生成した RGB 特徴量から R と G データを取り出し、RG 特徴量とする。

図 2 に RG 特徴量の Tiny LSH 処理を示す。文献 [1] で提案した Tiny LSH 法を RG 特徴量で用いるためにビット処理の仕方を変えている。

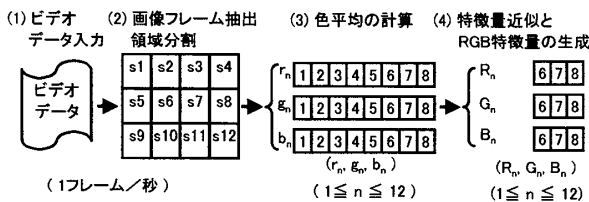


図 1 RGB 画像特徴量の処理方法

RG Tiny LSH関数  $F_v(R_n, G_n)$  ( $1 \leq n \leq 12$ )

```

1:  $C_n = (R_n + G_n \gg 1) \gg 1$ 
2: if ( $C_n == 0$ )  $C_n = (R_n + G_n)$ 
3: Return  $C_n$ 
    
```

図 2 RG 特徴量の Tiny LSH 処理

### 2.2 音声の振幅に基づいた音声特徴量

図 3 に音声特徴量の処理方法を示す。画像特徴量の抽出時間 (1 秒) と整合性を取るため、同じ時間単位で特徴量を生成する。ここでは、特徴抽出区間を  $a_1 \sim a_5$  の区間に分割し、音声信号の振幅平均値  $b_1 \sim b_5$  を音声特徴量とする。

図 4 に音声特徴量の Tiny LSH 処理を示す。文献 [1] で提案した Tiny LSH 法を音声特徴量で用いるためにビット処理の仕方を変えている。

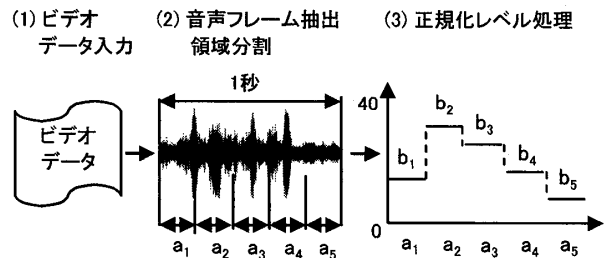


図 3 音声特徴量の処理方法

音声 Tiny LSH関数  $F_a(b_1, b_2, b_3, b_4, b_5)$

```

1:  $M = b_1 \gg 2 + b_2 \gg 2 + b_3 \gg 2 + b_4 \gg 2 + b_5 \gg 2$ 
2: if ( $M == 0$ )  $M = b_1 \gg 1 + b_2 \gg 1 + b_3 \gg 1 + b_4 \gg 1 + b_5 \gg 1$ 
3: if ( $M == 0$ )  $M = b_1 + b_2 + b_3 + b_4 + b_5$ 
4: Return  $M$ 
    
```

図 4 音声特徴量の Tiny LSH 処理

### 2.3 Direct-mapped cache に基づいた高速検索

提案手法は、RG 特徴量と音声特徴量のそれぞれに対して、Tiny LSH 法の概念と文献 [4] で用いた高速データ検出の手法を組み合わせることで検索処理を高速化した。

図 5 に、画像特徴量に対する Direct-mapped cache とビデオデータベースからなるハッシュテーブルのデータ構造を示す。Tiny LSH 処理は RG 特徴量を用いるが、検索のためのビデオ画像のフレーム間距離は図 1(4) で生成した RGB 特徴量を用いてユークリッド距離を計算する。

Comparison of the video retrieval performance by the several features based on the Tiny LSH method.

† Noboru Murabayashi (noboru@gssm.otsuka.tsukuba.ac.jp)

† Kenichi Yoshida (yoshida@gssm.otsuka.tsukuba.ac.jp)

Graduate School of Business Sciences, University of Tsukuba

(†) Otsuka 3--29--1, Bunkyo-ku, Tokyo, 112--0012, Japan

音声特徴量に対するハッシュテーブルのデータ構造は、図 5 のビデオデータベース部を図 3(3) で生成した音声特徴量を蓄積する音声データベースで置き換え、Direct-mapped cache 部を音声用にするだけで同様に考えることができる。

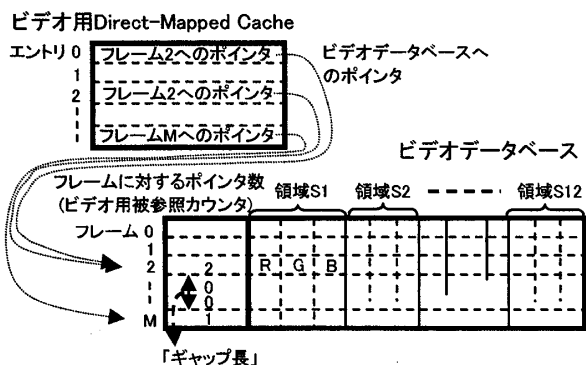


図 5 画像特徴量の Direct-mapped cache

### 3. 実験

#### 3.1 テストデータ

テストデータを用いて各特徴量に対する提案手法の性能評価を行った。テストデータは図 6 に示すように、連続テレビドラマの原ビデオに対して画像フレームの縦と横をそれぞれ 10% 切り取り、画像サイズをスケールして記録レートを変えた 88.5 時間の改変ビデオを用意した。

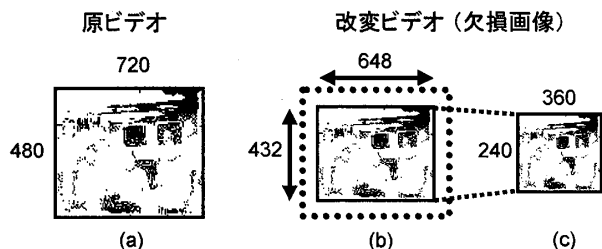


図 6 テストデータの生成方法

#### 3.2 実験結果

実験は、RGB 特徴量、RG 特徴量、音声特徴量の 3 種類の特徴量を用いて行った。原ビデオの各特徴量をビデオデータベース、または音声データベースに蓄積し、30 秒～300 秒の長さのビデオ断片にした改変ビデオが正しい位置で検索できるか実験した。検索の正解率は、(正しく検索できた断片数) / (全断片数) と定義する。

図 7 に検索の正解率を示す。音声特徴量の場合がもっとも正解率が良く、正解率の良さは RG 特徴量、RGB 特徴量の順になっている。

図 8 は検索時間の結果で、音声特徴量の場合がもっとも検索時間が短く、次に RG 特徴量、RGB 特徴量の順になっている。どの特徴量の場

合も 300 秒のビデオ断片に対して 0.04 秒以下の高速で検索できた。

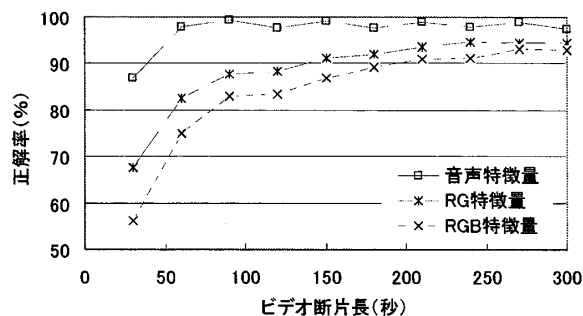


図 7 検索の正解率

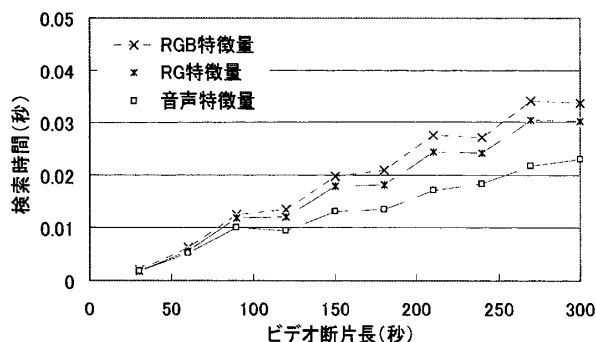


図 8 検索時間の結果

### 4. まとめと今後の課題

異なる特徴量を用いた Tiny LSH 法による動画検索性能の比較検討を行った。今回よりも大規模、改変の種類を多くしたビデオを用いた実験の再確認が今後の課題である。

#### 参考文献

- [1] K. Yoshida and N. Murabayashi, "Tiny LSH for Content-based Copied Video Detection", Proc. of SAINT2008, IEEE CS, pp. 89-95, 2008
- [2] M. Datar, P. Indyk, N. Immorlica and V. Mirrokni, "Locality-sensitive hashing using stable distributions", Nearest Neighbor Methods in Learning and Vision: Theory and Practice, MIT Press, 2006
- [3] 長坂 晃朗, 宮武 孝文, "時系列フレーム特徴の圧縮符号化に基づく映像シーンの高速分類手法", 電子情報通信学会論文誌 D-11, J81-D-11, No. 8, pp. 1831-1837, 1998
- [4] Kenichi Yoshida, Fuminori Adachi, Takashi Washio, Hiroshi Motoda, Teruaki Homma, Akihiro Nakashima, Hiromitsu Fujikawa and Katsuyuki Yamazaki, "Density-based spam detector", KDD2004, pp. 486-493, 2004