

時間特徴量による動画像検索手法の解析

高橋 秀和<sup>†</sup> 児玉 明<sup>†</sup> 金田 和文<sup>†</sup> 山下 英生<sup>†</sup>

<sup>†</sup> 広島大学

〒 739-0046 東広島市鏡山 3-10-31

TEL : (0824) 21-3646

FAX : (0824) 21-3639

E-mail : hide@eml.hiroshima-u.ac.jp

あらまし 本稿では、動画像検索として時間特徴量に着目した。輝度値の頻度分布、及び、フレーム前後間の分布相関値を用い、また、入力データに応じて自動決定したグループ化しきい値により相関値のグループ化を行うことで、マッチングに用いるデータ量を削減し、検索処理時間の短縮を実現した。また本提案方式について、入力データがシーンチェンジを持つ場合と、持たない場合についてシミュレーション実験を行い、入力データの持つ特徴によるグループの構造、及びマッチング結果について、解析を行った。

キーワード 動画像検索手法、相関値、グループ化処理、シーンチェンジ、高速化アルゴリズム

An Analysis of Video Searching Methods from Temporal Characteristics

Hidekazu TAKAHASHI<sup>†</sup> Mei KODAMA<sup>†</sup> Kazufumi KANEDA<sup>†</sup> Hideo YAMASHITA<sup>†</sup>

<sup>†</sup> HIROSHIMA University

3-10-31, Kagamiyama, Higashi-Hiroshima,

739-0046 JAPAN

TEL : +81-824-21-3646

FAX : +81-824-21-3639

E-mail : hide@eml.hiroshima-u.ac.jp

**Abstract** Today, we often use the video data in user personal equipment in multimedia communication services. Therefore, we consider the matching method of using the temporal characteristics in video searching system. In this paper, we pay attention to the luminance variances and the correlation value of the luminance variances between current frame and next frame. In addition, we propose matching algorithm the grouping correlation value by threshold and matching time can be short. Finally, we consider the structure of group and matching results in query video which have scene changes and not have ones.

**key words** Video Retrieval Method, Correlation Value, Grouping, Scene Change, Fast Algorithm

## 1. はじめに

近年の、高速インターネットや大容量データベースの整備、また、コンピュータの処理能力の進歩により、多くの情報を伝達することのできる動画データを取り扱うことが多くなってきている。それに伴い、膨大なコンテンツ情報の中からユーザが情報を手に入れるために、動画検索システムが要求されている。

従来の動画検索手法の1つとして、画像の明るさや色情報といった空間的な特徴量を利用した手法が挙げられるが、動画データは、テキストや静止画データと比べてデータ量が非常に大きいため、この手法ではハードウェアに負担がかかってしまい、検索時間が長くなるという問題点がある。そのため、動画データの時間方向の特徴量を利用した手法の開発が行われている [1][2]。

本稿では、輝度値の頻度分布に着目し、特にフレーム前後間の輝度分布の相関値を利用したマッチング手法について検討する。さらに、動画の相関値の偏った頻度特性を利用して、入力データに応じて自動決定したしきい値による相関値のグループ化、及びその時間的特徴をもつパラメータを利用したマッチングアルゴリズムを提案する。また、本手法を用いたシミュレーション実験を行い、検索時間、検索精度における有用性を示す。

## 2. 全画素・全フレーム比較によるマッチング処理

まず始めに、動画の全画素・全フレームの比較によるマッチング処理について述べる。

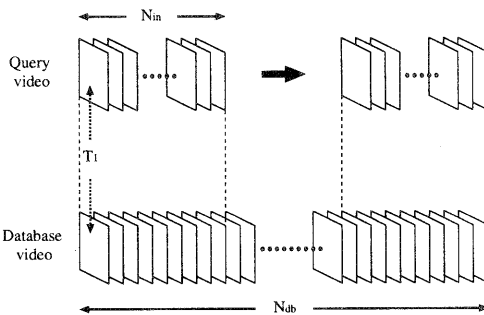


図 1: 全画素・全フレーム比較によるマッチング

入力動画のフレーム数を  $N_{in}$ 、データベース動画のフレーム数を  $N_{db}$  とし、図 1 に示すように、まず  $N_{in}$  フレームの比較を行ってから、入力動画を 1 フレームずつシフトさせて、 $N_{db}$  フレーム全てとの比較を行う。1 フレームの比較に要する処理時間を  $T_1$  とすると、データベース動画すべて

との比較に要する処理時間  $T_a$  は、式 (1) で表すことができる。

$$T_a = T_1 \times N_{in} \times (N_{db} - N_{in} + 1) \quad (1)$$

ここで、 $N_{in} = 90$ 、 $N_{db} = 90750$  としたときの処理時間は、 $T_1 = 0.143[s]$  より、 $T_a = 1166807.1[s]$  を要することになる。これは、ユーザがデータベースに対して検索を行うということから考えて、大きすぎる処理時間である。

## 3. 提案動画検索方式の処理手順

2 節で示したように、動画データにおいて、入力データとデータベースデータの全画素・全フレームの比較によるマッチングは、多大な処理時間を要するため、動画のデータ量を削減し、かつ動画の特徴を生かしたマッチング処理を行うため、我々は、動画の時間方向の特徴量に着目した。

その時間方向の特徴量を利用した検索処理の流れを図 2 に示す。本検索方式は、大きく特徴量テーブル作成処理とマッチング処理から構成される。次に、そのそれぞれの処理について述べる。

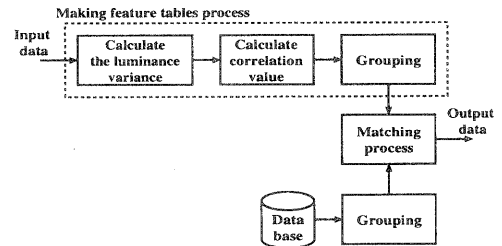


図 2: 提案動画検索処理の流れ

### 3.1 特徴量テーブル作成処理

本処理は、次の各処理器から構成される。

- 輝度分布算出器  
入力データに対して、輝度値  $Y$  についての頻度分布を求める。ここで、輝度値は 8bit 精度で扱う。
- 相関値算出器  
求めた頻度分布を用いて、式 (2) により、フレーム前後間の分布相関値を求める。

$$C = \frac{\sum_{j=0}^{255} (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=0}^{255} (x_j - \bar{x})^2 \sum_{j=0}^{255} (y_j - \bar{y})^2}} \quad (2)$$

ここで、 $\bar{x}, \bar{y}$  は平均値である。

- グループ化器  
求めた分布相関値を用いて、図3に示すように、相関値があるしきい値より高いか低いかで相関値のグループ化を行う。グループ化に用いるしきい値  $Th_{tbl}$  は、求めた入力データの分布相関値より、式(3)を用いて算出する。

$$Th_{tbl} = \frac{1}{N-1} \sum_{i=0}^{N-2} C_i \quad (N \geq 2) \quad (3)$$

ここで、 $N$  は入力フレーム数、 $C_i$  は上で求めた分布相関値である。

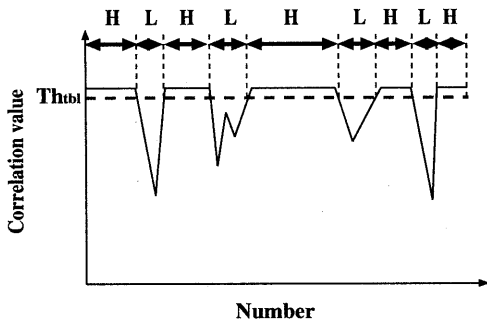


図 3: 相関値のグループ化

以上で求めた特徴量テーブルを比較することにより、マッチング処理を行う。

### 3.2 マッチング処理

グループ化によって得られた H, L の各パラメータを用いて、下記の処理手順によりマッチング処理を行う。

1. 入力データのグループ化テーブルにおいて、最初と最後のグループを省く。
2. 図4に示すように、入力データのグループ構造数である  $GN_{in}$  から最初と最後のグループ分の 2 を引いた、 $GN_{in} - 2$  個のグループについて、データベースデータのグループとの比較を行う。その際、式(4)により判定を行う。

$$|G_{in}[i] - G_{db}[i]| < Th_{grp} \quad (4)$$

ここで、 $G_{in}[i]$  は、入力データの  $i$  番目のグループの値、 $G_{db}[i]$  は、データベースデータの  $i$  番目のグループの値、 $Th_{grp}$  は、しきい値である。

3. 2の比較の処理の際、式(4)を満たすグループ数をカウントし、そのカウント数が式(5)を満たすものをマッチング結果として出力する。

$$count > Th_{count} \quad (5)$$

ここで、 $count$  は、式(4)を満たすグループ数のカウント数、 $Th_{count}$  は、しきい値である。

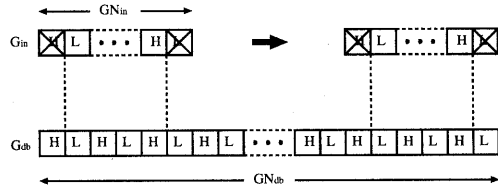


図 4: 相関値のグループ化によるマッチング処理

## 4. シミュレーション実験

### 4.1 実験方法

シミュレーション実験に用いたデータを表1に示す。表1の条件のもとで実験を行い、全検索処理の各処理器におけるマッチング処理時間とマッチング数、及びグループ構造数を測定する。なお使用計算機は、Fujitsu GP7000M25 400MHz UltraSPARC-2 である。

### 4.2 実験結果

実験結果を、全検索処理の各処理器における処理時間とマッチング数の観点から評価する。各処理器における処理時間とマッチング数を、入力データにシーンチェンジがある場合の結果を表2に、シーンチェンジがない場合の結果を表3にそれぞれ示す。また合わせて、入力データのグループ構造数 ( $GN_{in}$ )、及びデータベースデータのグループ構造数 ( $GN_{db}$ ) も表中に示す。ここで、本検索システムでは、データベース中から入力データと同一のシーンを検索することを目的としているため、マッチングに用いる式(4)、式(5)のしきい値は、最も精度がよくなるような、 $Th_{grp} = 0, Th_{count} = GN_{in} - 2$  とした。

### 4.3 考察

マッチング手法において、頻度分布およびフレーム前後間の分布相関値を用いることは、マッチングのための情報量を少なくでき、かつ動画像の時間特徴量を利用できる面で有効である。そのことは、

表 1: 画像データ

|                 |  |
|-----------------|--|
| シーケンス           | アニメ、スポーツ 1、スポーツ 2、音楽 1、音楽 2、映画、ニュース、スポーツ 3、スポーツ 4、スポーツ 5、Bus、Flower Garden、Mobile & Calendar、Popl、Table Tennis<br>計 15 シーケンス                             |
| データベース<br>フレーム数 | 9000 フレーム (アニメ、スポーツ 1、スポーツ 2、音楽 1、音楽 2、映画、ニュース、スポーツ 3、スポーツ 4、スポーツ 5)<br>150 フレーム (Bus、Flower Garden、Mobile & Calendar、Popl、Table Tennis)<br>計 90750 フレーム |
| 入力フレーム数         | 90 フレーム  |
| 画像<br>フォーマット    | ITU-R BT.601 4:2:0<br>704[pe]x480[line]  |

表 2: 各処理器における処理時間とマッチング数 (入力データにシーンチェンジがある場合)

| 入力      | 輝度分布<br>算出時間 [s] | 相関値<br>算出時間 [s] | グループ化<br>処理時間 [s] | マッチング<br>時間 [s] | マッチ<br>ング数 | $GN_{in}$ | $GN_{db}$ |
|---------|------------------|-----------------|-------------------|-----------------|------------|-----------|-----------|
| アニメーション | 7.210            | 0.538           | 0.392             | 0.016           | 133        | 3         | 3182      |
| スポーツ 1  | 7.224            | 0.554           | 0.404             | 0.024           | 176        | 3         | 3686      |
| スポーツ 2  | 7.330            | 0.548           | 0.400             | 0.008           | 3          | 5         | 3290      |
| 音楽 1    | 7.140            | 0.538           | 0.384             | 0.066           | 592        | 3         | 3296      |
| 音楽 2    | 7.180            | 0.564           | 0.386             | 0.012           | 1          | 21        | 2304      |
| 映画      | 7.424            | 0.516           | 0.404             | 0.010           | 1          | 5         | 3492      |
| ニュース    | 7.252            | 0.562           | 0.384             | 0.104           | 798        | 3         | 3814      |
| スポーツ 3  | 7.154            | 0.556           | 0.392             | 0.004           | 2          | 4         | 3464      |
| スポーツ 4  | 7.280            | 0.538           | 0.394             | 0.086           | 700        | 3         | 3398      |
| スポーツ 5  | 7.094            | 0.576           | 0.386             | 0.078           | 693        | 3         | 3372      |

表 3: 各処理器における処理時間とマッチング数 (入力データにシーンチェンジがない場合)

| 入力      | 輝度分布<br>算出時間 [s] | 相関値<br>算出時間 [s] | グループ化<br>処理時間 [s] | マッチング<br>時間 [s] | マッチ<br>ング数 | $GN_{in}$ | $GN_{db}$ |
|---------|------------------|-----------------|-------------------|-----------------|------------|-----------|-----------|
| アニメーション | 7.436            | 0.548           | 0.398             | 0.062           | 1          | 41        | 13106     |
| スポーツ 1  | 7.274            | 0.558           | 0.404             | 0.036           | 8          | 5         | 16900     |
| スポーツ 2  | 7.180            | 0.562           | 0.420             | 0.060           | 1          | 31        | 17194     |
| 音楽 1    | 7.178            | 0.560           | 0.392             | 0.086           | 147        | 3         | 11424     |
| 音楽 2    | 7.272            | 0.574           | 0.406             | 0.042           | 1          | 34        | 9614      |
| 映画      | 7.170            | 0.558           | 0.418             | 0.066           | 1          | 36        | 16996     |
| ニュース    | 7.308            | 0.578           | 0.404             | 0.070           | 1          | 39        | 14900     |
| スポーツ 3  | 7.282            | 0.528           | 0.420             | 0.074           | 1          | 38        | 17180     |
| スポーツ 4  | 7.380            | 0.572           | 0.410             | 0.040           | 1          | 27        | 10454     |
| スポーツ 5  | 7.256            | 0.624           | 0.414             | 0.032           | 1          | 35        | 8398      |

検索処理時間の結果によくあらわれており、表 2、表 3 に示すように、グループ化を行った場合、特徴量テーブル作成として相関値のグループ化を行っておけば、データベースの特徴量テーブルとのマッチング処理は、0.1[s] よりも短い時間で行うことができる。これは、2.節で示した全画素・全フレームを比較する手法と比較して、格段にマッチングの高速

化がなされており、また、グループ化を行わない場合に約 1.6[s] の時間を要する<sup>[2]</sup>ことを考えても、マッチング時間の高速化に関してグループ化処理が有効であることが分かる。

マッチング数の観点から見ると、入力データのグループの構造とマッチング数との関係としては、構造数が多いほど、マッチング数が少なくなっている

ことが分かる。この理由として、入力データの構造数が少なくなるということは、マッチングのための情報量が少なくなるということが挙げられる。特に、マッチング処理において、入力データグループの最初と最後のグループを省いてマッチングに使用しないため、構造数3の場合を例にとってみると、比較を行うためのグループは、1つしか残らないという現象が起こる。したがって、1つのグループの値をデータベースのグループの値と比較するため、入力データの特徴があらわれず、マッチング数が増加する原因となる。

次に、表2と表3を比較すると、入力データにシーンチェンジがない方が、シーンチェンジがある場合に比べて、マッチング精度が良いことが分かる。この現象は、グループ化しきい値を自動決定する処理に起因している。シーンチェンジがない、つまり、時間特徴量の変化が見られない入力データは、図5のような波形をしている。ここで、図中の直線は自動決定したグループ化しきい値を表している。この波形をグループ化する際に、グループ化しきい値を入力相関値の平均値と定めているため、波形は上下に分かれ、グループ化される。一方、シーンチェンジのある入力データの場合、図6にあるように、シーンチェンジの部分で減少した相関値がグループ化しきい値(相関値の平均値)を下げるため、細かく変化している部分のグループ化ができないため、大きな分類だけが行われ、入力データの構造数が少なくなるといった現象が起こる。

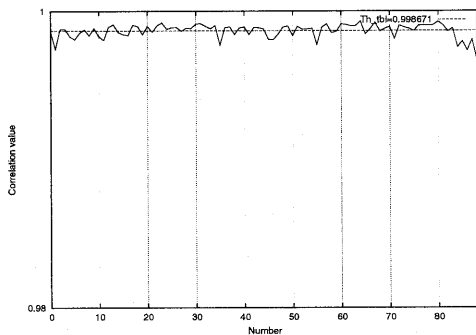


図5: シーンチェンジのない入力データにおける相関値の波形

データベース構造数を比較すると、シーンチェンジのない入力データの場合、図5のような波形においてグループ化しきい値を算出するため、しきい値は1に近い値となり、そのしきい値を用いてデータベースのグループ化を行うことから、データベース構造数が、シーンチェンジのある入力データの場合と比べて増加する。また、マッチング時間はデータ

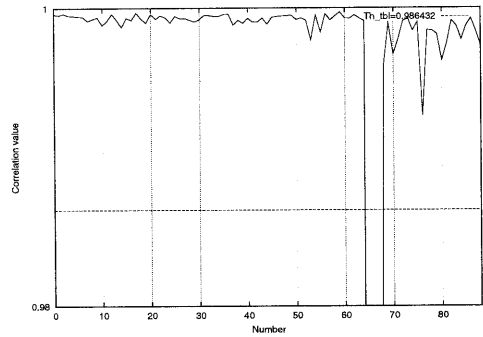


図6: シーンチェンジのある入力データにおける相関値の波形

ベース構造数とマッチング数に依存している。

以上のことから、本提案手法は入力データにシーンチェンジがない場合に特に有効であると考えられる。また、シーンチェンジがある入力データについては、本手法で用いたH,Lのパラメータだけでなく、シーンチェンジにおける相関値、つまり、シーンチェンジにおいて相関値が減少する幅をパラメータに加えると、より入力データの特徴を利用できると考えられる。

## 5. まとめ

フレーム前後間の分布相関値を用いたマッチング手法として、入力データに応じて自動決定したグループ化しきい値による相関値のグループ化を行うマッチング手法について述べた。また、入力データにシーンチェンジがある場合とない場合について、シミュレーション実験を行い、グループ構造数とマッチング数の関係についての解析、及び、グループ化を行う本手法の有用性の実証を行った。今後の課題としては、入力データにシーンチェンジがある場合についての本手法の拡張、及び、縮小画像や符号化画像などの誤差のある入力データに対してのマッチング手法の確立が挙げられる。

## 参考文献

- [1] 高橋 克直, 寺島 信義, 富永 英義: “画紋情報を用いた動画画像検索手法に関する検討”, 映像学技術報, BCS98-61, pp. 1-8 (1998).
- [2] H.Takahashi, M.Kodama, K.Kaneda and H.Yamashita: “Video Retrieval Methods from Temporal Characteristics”, IS2000, pp. 376-379 (2000).