

## 同一映像探索における高速な特徴抽出

荒川 豊<sup>†</sup> 青木 輝勝<sup>†</sup> 安田 浩<sup>†</sup>

<sup>†</sup> 東京大学先端科学技術研究センター 〒153-8904 東京都目黒区駒場 4-6-1

E-mail: † {arakawa, aoki, yasuda}@mpeg.rcast.u-tokyo.ac.jp

あらまし ここで言う探索とは、具体的な映像をクエリーとし、それと同一の動画像が大量の映像データのどこにあるかを探すものである。従来、探索に利用する映像の特徴として輝度分布が使われてきたが、映像データは通常圧縮された形で流通・保存されるため、特徴抽出には計算量の大きな復号処理が必要であった。そこで前論文では、特に MPEG 動画像に着目し、復号処理をほとんど必要としない、高速な抽出が可能な特徴を検討・提案した[1]。本稿では、それに改良を加え、様々な符号化パラメータを変化させて符号化した MPEG 動画像でも探索可能となることを示した。また、特徴抽出方法を決定するパラメータがいくつか存在するが、探索可能でありながら高速に探索を行なうためには、それぞれのパラメータ値がどれ程の範囲内であれば良いかを明らかにした。

キーワード 映像探索, 高速探索, MPEG, 動き情報, 動きベクトル, マクロブロックタイプ

## Fast Feature Extraction for the Same Video Retrieval

Yutaka ARAKAWA<sup>†</sup> Terumasa AOKI<sup>†</sup> and Hiroshi YASUDA<sup>†</sup>

<sup>†</sup> Research Center for Advanced Science & Technology, The University of Tokyo

4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904 Japan

E-mail: † {arakawa, aoki, yasuda}@mpeg.rcast.u-tokyo.ac.jp

**Abstract** *The Same Video Retrieval* means the method to seek out the specific video from long video data. This type of video retrieval is useful for copyright protection, automatic video clustering, commercial message detection from a television broadcast, etc. We proposed the fast feature extraction from the video data in the MPEG compressed domain for this video retrieval in previous paper. In this paper, at first, the fast feature extraction is improved upon. Secondly, the parameters defining the way of feature extraction are examined in the respects of retrieval-possibility and speed.

**Keyword** Video Retrieval, Fast Retrieval, MPEG, Motion Information, Motion Vector, Macro Block Type

### 1. はじめに

#### 1.1. 背景と目的

ネットワークと記憶装置の大容量化により、インターネット上でやりとりされるコンテンツの量は増大し続けている。こうした中で、大量のコンテンツの中から、目的のコンテンツを検索・探索する技術の重要性が高まってきている。本稿では特に、動画像コンテンツの探索技術に焦点を当てる。動画像コンテンツは、データ量が大きいため流通が困難であったが、近年それも容易となり、今後その流通量がますます増加することは必至である。

ここで検索とは、キーワードや色合いなど、なんらかの条件をクエリーとし、大量のコンテンツの中から適合するものを選び出す技術であり、様々な研究がなされてきている。一方、探索とは、具体的な動画像データをクエリーとし、それと同一の動画像が大量のコンテンツの中のどこにあるかを探すものである。図1に検索と探索をそれぞれ図示する。

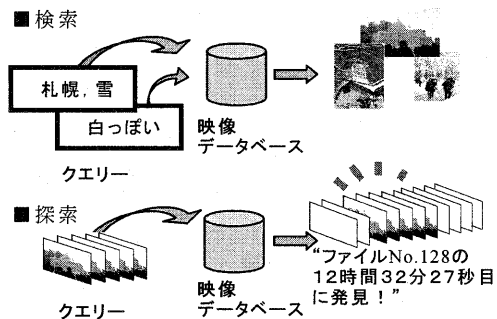


図1. 検索と探索

欲しいコンテンツを探すためには、一般的には検索技術が用いられることが多いが、著作権法違反の検出や、長時間放送データの中にCM映像が実際に流れているかの調査、あるいは連続ドラマの自動収集など、探索技術の応用範囲も広い。

従来の映像探索技術では、映像の特徴として輝度分布が用いられてきた[2]。しかしながら、通常映像は符号化された形で流通・保存されるため、そこから輝度分布を得るためには復号処理を施す必要がある。今後更に動画コンテンツの流通量増加が見込まれるのであるから、特徴抽出に要する時間は極力小さくしなければならぬ。そこで我々は、前論文において、特にMPEG形式の動画に焦点を当て、ほとんど復号せずに高速に抽出することが出来る特徴の提案をした[1]。[1]では提案に留まったが、本稿ではその改良と有効性の検証、そして抽出の仕方についてのより細かな検討を行なう。

## 1.2. 提案したMPEG動画の特徴

我々が提案した、MPEG動画ファイルから抽出する特徴は次のようなものであった。

- MPEG動画ファイル中のマクロブロックタイプ（以下、MBタイプと呼ぶ）と動きベクトル（以下、MVと呼ぶ）から特徴を抽出する。

- MBタイプのうち、Pピクチャ中のMV≠0であるMC（前方向予測動き補償予測されたMB）、Bピクチャ中のMV≠0であるFwd（前方向動き補償予測されたMB）、MV≠0であるBwd（後方向動き補償予測されたMB）、Interp（双方向動き補償予測されたMB）、P・Bピクチャ中のIntra（イントラ符号化されたMB）を動きのある部分（動MBと呼ぶ）と解釈する。

- MBタイプのうち、P・Bピクチャ中のSkip（一つ前に復号したフレーム中の、対応するMBと変化がないMB）、Pピクチャ中のMV=0であるMC、Bピクチャ中のMV=0であるFwd、MV=0であるBwdを動きのない部分（静MBと呼ぶ）と解釈する。ただし、Coded（予測誤差情報あり）の場合には、MV=0でも変化があると見なすことが出来るので動MBとする。

- フレーム内で映像をM×N個に分割し、それぞれのサブブロックに属するMBの平均をとることにより動きのある／ないサブブロックであるかを決定し、その分布をそのフレームの特徴とする。

- 時間軸方向にも平均をとる。このとき一定時間ごとに区切り、それぞれの区間に含まれるフレームの特徴を平均化する。

まず高速に抽出出来るという理由から、MBタイプとMVに着目した。さらに様々な符号化パラメータによりエンコードされたMPEG動画でも探索可能となる、すなわち同じ特徴となるよう、動MB/静MBへの分類と空間・時間方向の平均化を導入した。

## 2. 符号化動画特徴抽出方法の改良

[1]においては、実験用に作成したシンプルなMPEG動画を使い、符号化パラメータが異なっても変化しにくい特徴を探った。しかしながら、より複雑なMPEG

動画を用いて探索を行なってみると、思ったように探索出来ない場合があることが判明した。これを調査した結果、符号化パラメータにより、同じMBでも

- ① Skip⇔Interp, Coded, MV=0⇔Interp, Coded, Quant, MV=0⇔Interp, Not Coded, MV=0
  - ② Fwd, Coded, MV=0⇔Fwd, Coded, Quant, MV=0⇔Fwd, Not Coded, MV=0
  - ③ Bwd, Coded, MV=0⇔Bwd, Coded, Quant, MV=0⇔Bwd, Not Coded, MV=0
  - ④ Skip⇔No MC, Coded, Quant, MV=0
- と変化することが分かった。

MV=0のInterpは動MBと分類していたが、①ではSkip、すなわち静MBにも変化していることが分かる。MV=0のInterpが前後の参照MBから線形補間するものであるため、前後の参照MBが異なればどちらの参照MBとも異なり、動きMBとなるが、前後の参照MBが同じ場合にはどちらの参照MBとも同じとなり、静MBとなる。つまり、MV=0のInterpはそれだけでは動MBとも静MBとも判断できない、ということの意味していると考えられる。

また、CodedならばMVが0であろうとなかろうと、予測誤差があるため参照MBとは異なり、動MBと見なすことが出来ると考えた。しかしながら、①～④を見るとCodedのMBが静MBにも変化することが分かる。これはCodedかNot Codedかの違いはMVが0かどうかの違いよりも変化への影響が小さく、エンコード時には許容誤差範囲としてCodedがNot Codedにもなり得ることを示していると考えられる。そこでCodedでもMV=0ならば静MBに分類することとした。

最終的に、MBタイプ、MVにより、図2のように動/静MB、判定不可能の3つに分類することとした。

■Iピクチャ	
Intra	判定不可能（比較対象外）
Intra, Quant	判定不可能（比較対象外）
■Pピクチャ	
MC, Coded	if MV=0 then 静MB else 動MB
No MC, Coded	静MB
MC, Not Coded	if MV=0 then 静MB else 動MB
Intra	動MB
MC, Coded, Quant	if MV=0 then 静MB else 動MB
No MC, Coded, Quant	静MB
Intra, Quant	動MB
Skip	静MB
■Bピクチャ	
Interp, Not Coded	判定不可能（比較対象外）
Interp, Coded	判定不可能（比較対象外）
Bwd, Not Coded	if MV=0 then 静MB else 動MB
Bwd, Coded	if MV=0 then 静MB else 動MB
Fwd, Not Coded	if MV=0 then 静MB else 動MB
Fwd, Coded	if MV=0 then 静MB else 動MB
Intra	動MB
Interp, Coded, Quant	判定不可能（比較対象外）
Pwd, Coded, Quant	if MV=0 then 静MB else 動MB
Bwd, Coded, Quant	if MV=0 then 静MB else 動MB
Intra, Quant	動MB
Skip	静MB

図2. 各MBタイプの動/静MB判定方法

### 3. 改良した特徴抽出方法の検証

改良した特徴抽出方法により、実際に MPEG 動画像の探索を行なう実験を行なった。実験には MPEG-7 テストシーケンス ITEM V28 に含まれる ETRI\_od\_A.mpg を用いた。実験方法としては、ETRI\_od\_A.mpg を様々な符号化パラメータで再符号化した MPEG 動画像を用意し、その中で基準とした一つの MPEG 動画像の一部をクエリーとし、その他のものから探索を行なった。変化させた符号化パラメータは、動き探索範囲、GOP 構造、フレームレート、量子化行列、画像サイズである。実験には市販のソフトウェアエンコーダを用いた。基準とした MPEG 動画像は動き探索範囲を“標準”の設定に、また GOP 構造は IBBPBBPBBPBBPBBP、フレームレートは 29.97fps、量子化行列はソフトウェアエンコーダのデフォルトのもの、画像サイズは 240×352pixel である。クエリーは基準 MPEG 動画像の 50 秒目から 10 秒間とし、特徴抽出における平均化は空間方向にフレームを 15×22 個に分け、時間方向に 0.1 秒ごとに分割した単位で行なった。平均化の方法としては、その単位内に動 MB が一つでもあればその部分は動きのあるものと判断した。結果を図 3 に示す。

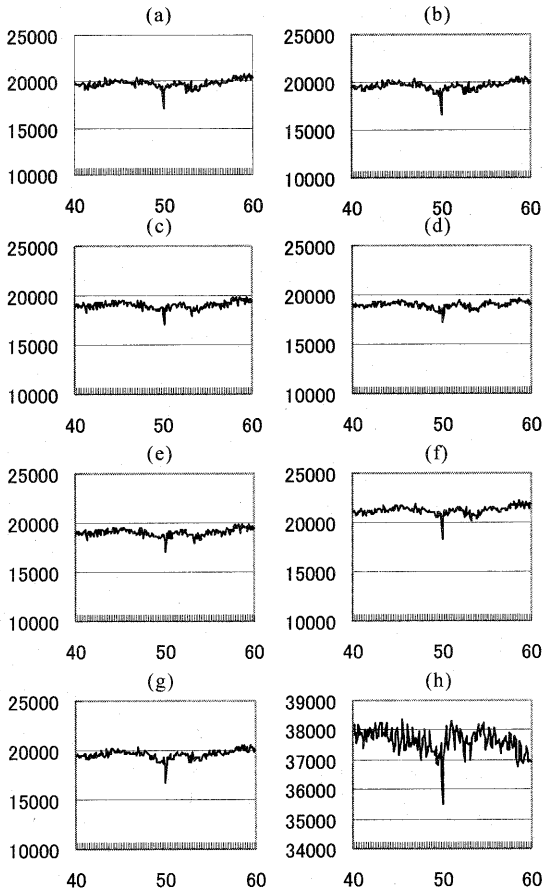


図3. 様々な符号化パラメータを変えて符号化した映像から、改良した特徴抽出方法により探索した結果

横軸は映像の時間軸(sec)、縦軸は一致しなかった特徴数を表す。また、それぞれ(a)動き探索範囲が“最小”の映像、(b)動き探索範囲が“最大”の映像、(c)GOP 構造が IBBPBBPBBPBBPBBPBBP の映像、(d)GOP 構造が IBBBPPBBBPPBBBPPBBBPPBBBPP の映像、(e)GOP 構造が IPPPPPPPPPPPP の映像、(f)フレームレートが 50fps の映像、(g)量子化行列が MPEG 標準の映像、(h)画像サイズ 480×704pixel の映像、から探索した結果を示す。

図 3 を見ると、期待通り 50 秒の時点で特徴が最も一致していることが分かる。すなわち、改良した特徴抽出方法により、どの符号化パラメータを変化させた MPEG 動画像からでも、探索可能であることが分かる。

### 4. 特徴抽出パラメータの検討

MPEG 動画像の特徴としては、MB タイプと MV から図 2 に従って抽出し、それを空間方向、時間軸方向それぞれについて平均化する。この特徴抽出の際、変化させ得るパラメータとして、以下のものがある。

- ・ M, N : フレームのサブブロックへの分割数(縦方向に M 個、横方向に N 個に分割)
- ・ AT : Average Time. 時間方向の分割区間
- ・ QL : Query Length. クエリー動画像の長さ

これらのうち、M, N, QL は小さい方が、また AT は大きい方が比較すべき特徴の数が減り、望ましい。しかしながら、あまり大きな単位で平均化し、特徴を均してしまうと他と区別できず、探索出来なくなると考えられる。つまり、それぞれの特徴抽出パラメータには、探索可能であるための限界値が存在する。また、どの符号化パラメータが異なった映像から探索するかにより、これらの特徴抽出パラメータの限界値も変わってくると考えられる。

そこでそれぞれの場合について、特徴抽出パラメータの限界値を調べるため、次のような実験を行なった。

- 3 章で使用した、符号化パラメータを 1 種類ずつ変えた MPEG 動画像を使い、様々な特徴抽出パラメータ値で探索を行なう。
- 特徴抽出パラメータは一つずつ変化させ、その際、他の特徴抽出パラメータは 3 章に記述した実験で用いた値と同じものを用いる。
- 特徴抽出パラメータはそれぞれ次の値を試す。

M : 8, 4, 2  
 N : 11, 6, 3  
 AT(sec) : 0.2, 0.4, 0.8, 1.6  
 QL(sec) : 10, 5, 2.5, 1.75

- それぞれの条件で、探索できるかどうかを調べる。ここで、「異なる特徴数」が探索した全領域の中で最小となるとき、探索できると判断した。

3 章における結果も加え、実験結果を表 1 に示す。

表 1. 様々な符号化パラメータが異なる MPEG 動画像間で探索を行う場合において、特徴抽出パラメータを変化させたときの探索可能性

図中、○は探索可能、×は探索不可能を示す。また、(a)~(h)は図3のものと同様の映像を示す。

M					N				
	15	8	4	2		22	11	6	3
(a)	○	○	○	○	(a)	○	○	○	○
(b)	○	○	○	○	(b)	○	○	○	○
(c)	○	○	○	○	(c)	○	○	○	○
(d)	○	○	○	○	(d)	○	○	○	○
(e)	○	○	○	○	(e)	○	○	○	○
(f)	○	○	○	○	(f)	○	○	○	○
(g)	○	○	○	○	(g)	○	○	○	○
(h)	○	×	×	×	(h)	○	×	×	×

AT(sec)

	0.1	0.2	0.4	0.8	1.6
(a)	○	○	○	×	×
(b)	○	○	○	×	×
(c)	○	×	×	×	×
(d)	○	×	×	×	×
(e)	○	○	○	×	×
(f)	○	○	×	×	×
(g)	○	○	○	×	×
(h)	○	×	×	×	×

QL(sec)

	20	10	5	2.5	1.75
(a)	○	○	○	○	×
(b)	○	○	○	×	×
(c)	○	○	×	×	×
(d)	○	○	×	×	×
(e)	○	○	○	×	×
(f)	○	×	×	×	×
(g)	○	○	○	×	×
(h)	○	×	×	×	×

表 1 より、様々な符号化パラメータが異なる MPEG 動画像間で探索を行う場合の、各特徴抽出パラメータの限界値を大まかに知ることが出来る。

M, N, AT, QL のどの場合を見ても、(h) (画像サイズが異なる MPEG 映像間) では、沢山の特征情報が必要であることが分かる。また、M, N の結果を見ると、(h) 以外では分割数を大きく減らしても大丈夫であることが分かる。図には記載していないが、特に N の場合では、小さい方が探索しやすい (一致した時間での特徴合致数とそうでない時間での特徴合致数の差が大きい) という結果が得られた。これは、ETRI\_od\_A.mpg が左右への動きが多い映像であるため、動きがあると

判断される部分が符号化パラメータの違いにより左右にずれることが多く、左右に広いサブブロックに分割した方が正しくマッチする確率が高くなるためと考えられる。しかしながら、一般に動画像には様々な方向への動きが存在するため、分割方式としては M, N ともにある程度小さくするのが適当であろう。AT を見ると、(h) の他に (c), (d), (f) でも AT を長くとれないことが分かる。これらには参照フレームまでの距離が異なることによる誤差が存在し、AT が長いほど多くその誤差が蓄積するためと考えられる。QL でも、あまり QL が短いと参照フレーム距離の違いによる誤差により他と区別がつかなくなり、探索できないことが分かる。しかしながら、CM 一本を 15 秒とすると、CM 一本分程の長さのクエリーがあれば、どの符号化パラメータが異なっても、探索が行なえるだろうと考えられる。

これら特徴抽出パラメータの限界値はどの符号化パラメータが異なるかにより変化するが、符号化パラメータがどれ程異なるかにも当然依存すると考えられる。特徴抽出パラメータの値を選ぶ際には、著作権違反である不正コンテンツを見つけるならば符号化パラメータが大きく異なるものでも見つけられるように、また CM 探索ならば画像サイズの違う映像の探索は考慮しなくて良いなど、用途に合わせた選び方をすべきであろう。

## 5. おわりに

本稿では、高速な同一映像探索をめざし、[1]において提案した、MPEG 動画像ファイルからほとんど復号することなく高速に抽出出来る特徴について改良を加え、その特徴により符号化パラメータが異なる映像でも探索が可能であることを示した。また、特徴抽出方法を決定するパラメータ群の速度・探索可能性から見た場合の、限界値の範囲を実験により示した。

今後は、高適合率を実現するために、より様々な種類の動画像でも探索実験を行い、適切な特徴抽出パラメータを探る必要がある。また画像サイズの異なる映像間で探索が難しいのは、平均の取り方として、単位内に一つでも動きのある MB があればその単位を動く部分としていることに問題があると考えており、検討を行なっていく。また、現在 MPEG1 形式の動画像を対象に実験を行なっているが、MPEG2 への対応も考慮する予定である。

## 文 献

- [1] 荒川 豊, 青木 輝勝, 安田 浩, “MPEG 動画像探索における高速な特徴抽出の検討,” IMPS'02, pp. 111-112, Nov. 2002.
- [2] 柏野 邦夫, 黒住 隆行, 村瀬 洋, “ヒストグラム特徴を用いた音や映像の高速 AND/OR 探索,” 信学論, vol. J83-D-II, no. 12, pp. 2735-2744, Dec. 2000.