

画紋情報を用いた動画画像検索システムの実装

高木 真一†* 小館 亮之* 大串亮平‡ 奥野洋平‡ 越井剛‡ 土橋健太郎‡
池上大介‡ 大崎佑紀‡ 奥村真澄‡ 金田瑞規‡ 田邊集‡ 中神央二‡
服部しのぶ‡ 湧井道子‡ 富永 英義††*

†早稲田大学大学院国際情報通信研究科

* 早稲田大学国際情報通信研究センター

〒169-0051 東京都新宿区西早稲田 1-3-10 早稲田大学 29-7 号館

Tel: (03)5286-3831

E-mail: takagi@tom.comm.waseda.ac.jp

‡早稲田大学理工学部電子・情報通信学科

〒169-8555 東京都新宿区大久保 3-4-1

Tel: (03)5286-3385

あらまし 我々の研究グループでは従来より動画像に含まれる普遍的識別情報として”画紋”という概念を提案しており、テレビ放送 CM 送出確認を一つの応用例として挙げ、画紋情報を用いた動画画像検索システムを開発してきた。本稿では、マルチメディアコンテンツ記述インタフェースの国際標準 MPEG7 の仕様がほぼ凍結したことをうけ、従来システムの操作性と実用性を大幅に改良し、さらに MPEG7 へ準拠したものとすることで、MPEG7 を利用したシステムの実現例を示す。

キーワード 動画画像検索 MPEG7 画紋情報 CM

Implementation of Video Searching System Using Videoprint

Shin'ichi TAKAGI†* Akihisa KODATE* Ryouhei OOGUSHI‡

Youhei OKUNO† Gou KOSHII‡ Kentarou DOBASHI†

Daisuke IKEGAMI‡ Yuki OOSAKI‡ Masumi OKUMURA‡ Mizuki KANADA‡

Shuu TANABE‡ Ouji NAKAGAMI‡ Sally HATTORI‡ Michiko WAKUI‡

Hideyoshi TOMINAGA††*

† GITS, WASEDA University

* GITI, WASEDA University

Bldg.29-7,1-3-10 Nishi Waseda, Shinjuku-ku, Tokyo, 169-0051 Japan

Tel: +81-3-5286-3831

E-mail: takagi@tom.comm.waseda.ac.jp

‡Dept. of Electronics Information and Communication Engineering, Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555 Japan

Tel: +81-3-5286-3385

Abstract We have developed a video-video searching system using what we had proposed and called "Videoprint," which had been adopted to MPEG7 as Visual Color Layout Descriptor. This system searches for TV commercials, by extracting and comparing it with the Videoprint of the commercials. Although we use TV commercials as an example here, video that this system can search for is not limited to commercials, but could be arbitrary.

key words Retrieval MPEG7 Videoprint CM

1. はじめに

映像や音声に代表されるマルチメディアコンテンツは、テキスト情報と比較すると一般の検索エンジンでは直接コンテンツを検索することができないなど、検索はその一例ではあるが、コンテンツに対する高度な利用技術が十分提供されていない点が挙げられる。この問題点の解決に向けた動きのひとつとして、マルチメディアコンテンツの高度利用のための記述インタフェースの標準化のひとつとして、XML(eXtensible Markup Language)技術を適用したMPEG7(ISO/IEC 15938: Multimedia Content Description Interface)の標準化が1996年頃より行われており、2001年7月のMPEGシドニー会合で、その第1版の技術仕様がほぼ固まるに至った。

我々の研究グループでも、従来よりマルチメディアコンテンツの高度利用を目的とした、情報の構造化とその応用に関する検討を行ってきており、(高橋さん文献)その第1ステップとして、マルチメディアコンテンツに対する高度な検索機構が重要であると考え、我々の提唱する「画紋」と呼ぶ、動画画像の特徴情報を用いた映像情報をキーとした映像の検索システムの研究開発を行ってきた。中でもテレビ放送におけるCMの送出確認を想定した映像検索システムの構築を行い、検討を行ってきた[1][2]。

本稿では、コンピュータ技術の進歩やXML関連技術の普及に伴い、MPEG7の第1版の技術仕様がほぼ凍結されたことを受け、我々が従来研究開発してきた動画画像検索システムの操作性や実用性を大幅に改良させ、MPEG7を利用した具体的なシステムを開発したので、その報告する。

本稿は、まず2章で画紋情報による動画画像検索方式について示す。次に3章で今回構築した動画画像検索システムの内容について示し、4章で考察を述べ、5章で結論を述べる。

2. 画紋による動画画像検索方式

本章では、我々の研究グループの提唱する「画紋」の概念とその特徴、作成方法、画紋による動画画像検索方式について簡単に述べる。詳しくは、文献[1][2]を参照されたい。

2.1 画紋について

我々は、動画画像にも人間の指紋と同様の性質、すなわち、少ない情報量で動画画像の識別が可能とする動画画像固有識別情報があると考えている。我々はその動画画像固有識別情報のことを画紋(Videoprint)と呼んでいる。

動画画像検索を行うことを考えると、動画画像同士を比較することは、処理量が大きくなるため、ハードウェアの性能が上がっているとは言えあまり現実的ではなく、動画画像から動画画像の

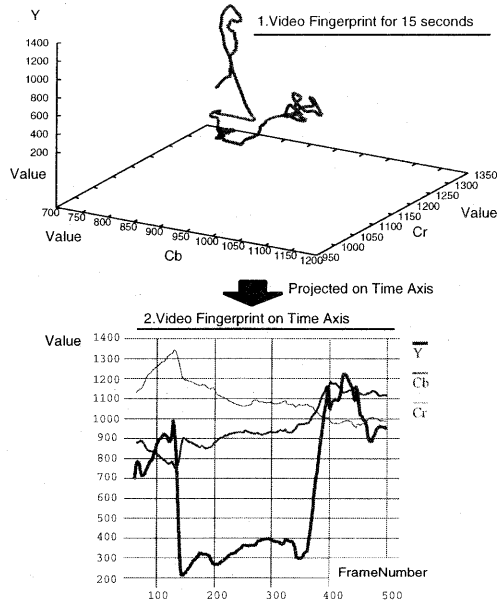


図1 15秒間の動画画像から生成された画紋の例 (高橋による)

特徴を抽出し、特徴情報を用いて動画画像検索を行うのが、より現実的かつ一般的と言える。

2.2 動画画像の画紋の作成方法

画紋値として、カラーヒストグラムや、オブジェクトの形状・座標、位置関係等を利用することが可能である。しかし、それらの情報を用いて画紋を作成すると、情報量が大きくなってしまふ。そして、特徴情報抽出のために複雑な処理が必要となり、また、特徴情報の精度が抽出アルゴリズムに依存し、元の動画画像品質に精度が左右され、画紋としての動画画像の識別情報としての普遍性が希薄になる。そこで、画紋値の抽出(作成)精度が、抽出アルゴリズムに非依存、かつ動画画像品質に特徴情報が変化することなく、しかも各フレームの内容に依存する少ない情報量の特徴情報を用いて画紋を作成することとする。その考えでは、各フレームの平均色情報を用いるのが有効な手段と考えられる。また、各フレームをまとめて行き、平均値化を行い、情報量を減らすのが有効な手段と考えられる。

2.3 MPEG ストリームからの画紋生成

動画画像を構成するフレームから、内容に依存した小さい情報量の特徴情報(画紋値)を求め、画紋にすると考えると、画面の平均値化を行い情報量を減らすのが有効な手段と考えられる。そのためには、元のフレームの細かいピ

クセル値の情報まで必要なく、DC成分で求められるブロック単位の平均画素値を用いても良い。例えば、画面の全平均値は、I-frameであれば、全ブロックのDC成分の全平均値を求めれば良く、また計算量の多いコサイン演算(DCT演算)もなく少ない演算量で容易に求められる。本稿では、主に動画像を構成するフレームのY,Cb,Cr値のそれぞれの全平均値(フレームを1点(pixel)にする)を画紋値とし、それらを時間軸に並べ画紋を作成(図1参照)し、画紋を用いたCM送出確認(CM検索)システムを構築した。

2.4 画紋情報による動画像検索方式

原理的に画紋は同じ動画像であれば、同じ画紋が得られるはずである。そのため、テレビ放送から抽出された画紋情報の時間変化波形に対して検索クエリであるCMから抽出された画紋情報の波形をフレーム単位でずらしながら、両者の波形を比較していけば、同じ内容の映像であれば波形が重なるはずである。本稿で構築したシステムでは、この原理に基づき、両画紋の波形の平均誤差を閾値処理することにより、動画像検索を行っている。詳しくは、文献[1][2]を参照されたい。

3. 動画像検索システムの実装

3.1 従来システム

我々が構築してきた従来の動画像検索システムでは、主に以下に示す手順で動作するようになっている[1]。

1. MPEG ストリームからの画紋作成
 - (a) MPEG2-PS → MPEG2-ES
 - (b) MPEG2-ES から DC 成分の抽出
 - (c) DC 成分の画紋化
2. 画紋比較による CM 検索
3. 結果表示画面生成 (HTML 生成)

これらの手順は、システム構築当時の技術的な背景もあり、ある工程が全て終了してからでないと次工程に進めない構造になっていた。そのため、MPEG映像をシステムに入力してから、CMの検索結果の表示までに人手による作業を多く要求するようになっており、実用的なシステムとしての完成度が十分でなかった。その後のコンピュータ技術の進展やXML関連技術の普及、MPEG7の仕様凍結といった背景の大きな変化があったため、今回システムに大きな改良を加えた。

3.2 システム設計方針

本システムの構築は、操作性や実用性の向上を主要目的とし、以下の方針でシステム設計を行った。

- 動画像検索の基本的なアルゴリズムには、大きな変更を加えない。
- MPEG ストリームからの画紋情報抽出、比較検索、結果表示までを1回のクリック操作のみで実現させる。
- システムの検索速度を、入力映像(TV放送)の再生速度より高速にする。(動画像のリアルタイム再生速度以上の検索速度の実現)
- システムで出力する各種インスタンス文書は、極力MPEG7(ISO/IEC 15938:2001年3月の主要パートFCD(Final Committee Draft)相当)に準拠させる[3]。
- 主に画紋情報抽出、比較検索、MPEG7インスタンス生成、表示画面作成等を実行するサーバと、主に結果表示、システム操作を行うクライアントにわけ、インターネット(HTTP,TCP/IP)を通じてクライアントからサーバにアクセス、操作できるようにする。
- クライアント側は、通常のWebブラウザのみを要求するように各種機能をできるだけサーバ側に構築する。
- MPEG7以外のXML技術も多く取入れ、画紋情報の表示方法の改良を行う。

3.3 システム構成

本システムは、2台のサーバと1台の表示クライアントから構成される。サーバは、検索サーバと表示サーバの2台のコンピュータから構成され、検索サーバでは主にMPEGストリームからの画紋情報抽出、検索機能を提供し、表示サーバでは、検索サーバでの検索結果データ等を元に、MPEG7を含む各種インスタンス生成、表示画面生成等の機能を提供する。表示クライアントは、簡単に操作できるよう、Webブラウザを用いてGUIによる操作と結果表示を行うためのものである。そのため、CUIによる操作、文字列のみによる結果表示だけで良ければ、表示用サーバと表示用クライアントは特に必要としない構成とした。

本システムにおける各サーバ、クライアントコンピュータのスペックを表1に示す。検索サーバと表示サーバについては、今回使用したコンピュータとソフトウェアを、表示クライアントについては、必要となるソフトウェアを示した。表1からわかるように、特別に特殊なソフトウェアは特に要求しない。

3.4 システムの動作の流れ

本システムでは、CMのMPEGストリームから画紋情報抽出をし、その表示とMPEG7

表1 サーバ、クライアントのスペック

検索サーバ	
ハードウェア	IBM AT 互換機
CPU	Pentium III 1GHz
Memory	512MB
OS	Linux(Vine Linux2.1)
Software	Java(JDK1.3.1), Perl apache
表示サーバ	
ハードウェア	IBM AT 互換機
CPU	Celeron 400MHz
Memory	128MB
OS	Windows2000
Software	Java(JDK1.3.1, JMF2.1.1) Perl, AnHTTPd
表示クライアント	
必要な Software	Internet Explorer5.5 SP2 以上 MSXML3.0, SVG Viewer

インスタンスの生成をする「検索クエリー作成」と、テレビ放送の MPEG ストリームから、画紋情報抽出をし、クエリーの検索と結果表示する「動画像検索」の2つの機能を提供できる。図2に示すように、本システムは、画紋生成モジュール、検索モジュール、結果表示作成モジュールの3モジュールから構成されており、それぞれの機能において以下の手順で処理が行われる。

3.4.1 検索クエリー作成機能

まず画紋情報抽出モジュールで、MPEG ストリームをデコードさせながら、各フレームの全マクロブロックの Y,Cb,Cr のそれぞれの DC 成分を抜き出し、その合計値を全マクロブロック数で割ることで、そのフレームの画紋値とする。DC 成分のみを用いることから計算量の多い IDCT は行っていない。つぎに、抽出した画紋情報を用いて、MPEG7 インスタンスの生成、画紋情報の時間変化のグラフ、画紋情報のカラーバーの作成を行う。

3.4.2 動画像検索機能

まず検索クエリー作成と同様に、MPEG ストリームをデコードさせながら、各フレームの全マクロブロックの Y,Cb,Cr のそれぞれの DC 成分を抜き出し、その合計値を全マクロブロック数で割ることで、そのフレームの画紋値とする。DC 成分のみを用いることから、計算量の多い IDCT は高速化のため行っていない。つぎに、抽出した画紋情報を検索モジュールへ渡し、検索クエリー (CM) の画紋情報 (MPEG7 インスタンス) を参照しながら、基本的に従来システムと同様のアルゴリズムで検索を行う。検索結果 (検出されたクエリー名と入力ストリーム中の存在位置 (Frame 番号) 等) を、結果表示作成モジュールへ渡す。

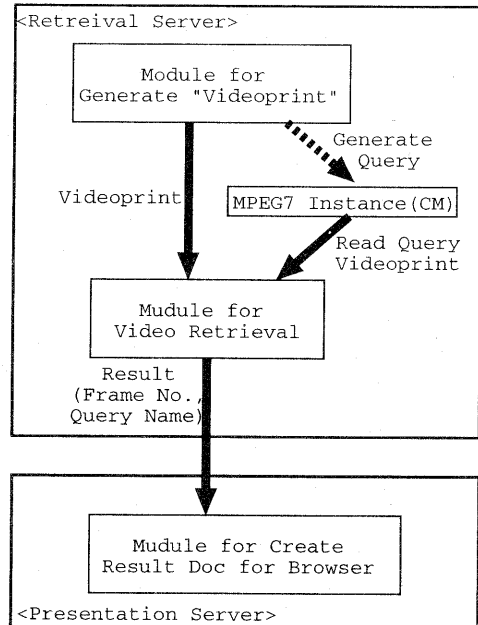


図2 処理の流れ

検索結果画面では、クエリー映像 (CM) の代表フレームの静止画 (JPEG 画像)、SVG で作成した画紋情報の時間変化のグラフ、検索された CM の代表画像と検索された位置 (Frame 番号) 等を表示させため、結果表示作成モジュールでは、XSL の作成等を行う。各モジュール間は、ソケットを用いて情報転送しており、結果が出力され次第、順次、次のモジュールへ転送される。

尚、CM の代表画像については、クエリー映像 (CM) の先頭から 30 フレーム目のものとしている。これは、クエリー映像作成は本システム外で作成するものであり、人間の手操作による映像の切り出しによることを想定し、スタート位置の 1 秒以内の誤差を考慮したものである。もし先頭フレームを代表画像とすると、一つ前の CM の最後の部分である可能性が高くなるからである。また、代表画像の作成には、JMF (Java Media Framework) を使用している関係 (現時点で、JMF が MPEG2 へ未対応) で、本システムで扱う MPEG ストリームは、全て MPEG1 としている。ただし、代表画像の表示をしないのであれば、画紋情報抽出モジュールと検索モジュールは、MPEG2 でも動作する。(但し、4:2:0 のみ)

3.5 MPEG7 インスタンスについて

本システムにおける検索クエリー作成では、抽出した画紋情報を MPEG7 の仕様に従った

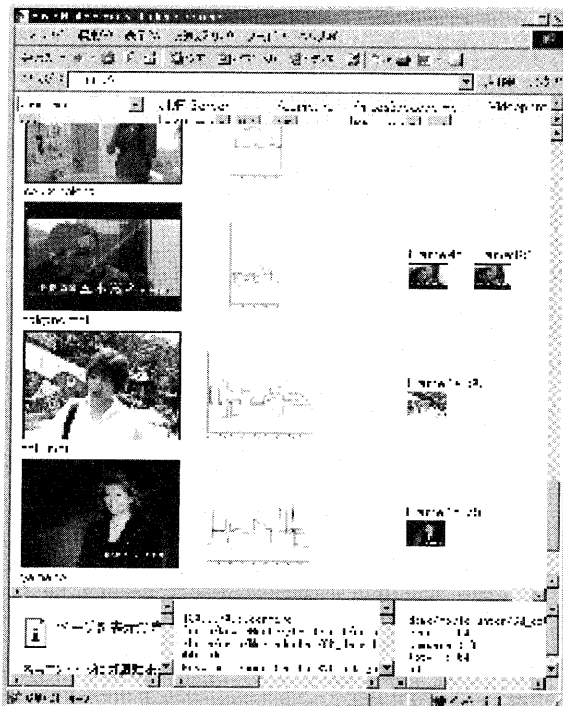


図 3 動画検索結果表示の例

形で記述するインスタンスを生成することができる。生成される MPEG7 インスタンスは、MPEG7 の完全記述版 (Complete Description) で記述することにし、画紋情報は MPEG7 Part3 Visual Descriptor で定義されている ColorLayout Descriptor (各成分 DCCoeff のみ) を使用して記述する。(厳密には、ColorLayout Descriptor は、ACCCoeff まで必須であることに注意)

本システムによって生成させた MPEG7 インスタンス例を以下に示す。..... は、紙面の都合上記載を省略した部分である。

```
<?xml version="1.0" encoding="UTF-8"
standalone="yes"?>
<Mpeg7 type="complete"
xmlns="http://www.mpeg7.org/...."
xmlns:xsi="http://www.w3.org/....">
<ContentDescription
xsi:type="ContentEntityType">
<MultimediaContent
xsi:type="VideoType">
<Video id="tom">
<MediaRelTimePoint
timeBase="/Mpeg7/Content....">
0
</MediaRelTimePoint>
<MediaLocator
xsi:type=
"TemporalSegmentLocatorType">
```

```
<MediaUri>....</MediaUri>
</MediaLocator>
<VideoSegment>
<TemporalDecomposition
criteria="spatialKey"
gap="true" overlap="true">
<StillRegion id="frame0">
<MediaRelTimePoint
timeBase="/Mpeg7/....">
1
</MediaRelTimePoint>
<VisualDescriptor>
<ColorLayout numOfCCoeff="1"
numOfYCcoeff="1">
<YCcoeff>
<YDCCoeff>156</YDCCoeff>
<YACCCoeff/>
</YCcoeff>
<CbCoeff>
<CbDCCoeff>281</CbDCCoeff>
<CbACCCoeff/>
</CbCoeff>
<CrCoeff>
<CrDCCoeff>246</CrDCCoeff>
<CrACCCoeff/>
</CrCoeff>
</ColorLayout>
</VisualDescriptor>
</StillRegion>
....
</TemporalDecomposition>
</VideoSegment>
<MediaIncrDuration
timeUnit="PT1N30F">
459
</MediaIncrDuration>
</Video>
</MultimediaContent>
</ContentDescription>
</Mpeg7>
```

3.6 画紋情報の表示について

本システムでは、操作性の向上と共に、システムユーザに対して、MPEG ストリームから抽出した画紋情報を視覚的にわかりやすい形態で表示させるために、表示サーバにおいて 2 種類のグラフの生成表示機能を提供している。一つは、横軸に時間軸をとり、縦軸に画紋値 (Y 値, Cb 値, Cr 値) をとり、画紋情報の時間変化の様子を表示させている。二つ目は、横軸が時間軸のみの 1 次元のグラフであり、各時刻におけるフレームの平均色で示した棒グラフを軸方向に並べて表示させたもので、見た目にはカラーバーのようにみることが出来る。また、時間軸のサイズを、動画再生プレイヤー (WindowsMediaPlayer6.4) の再生位置を示すインディケータの領域のサイズに合わせ、上下に並べて表示させることにより、動画像の再生位置と画紋値のカラーバー表示の時間軸上の位置を揃えて表示させるようにした。これにより、動画像の各フレームにおける代表色の時間

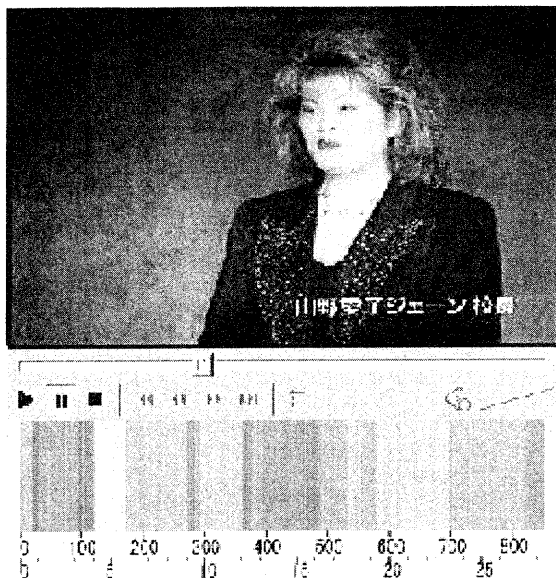


図4 画紋情報のカラーバーによる表示

的变化が、大変わかりやすく表示させることができた。カラーバーによる表示例を、図4に示す。

このグラフの生成・表示機能は、XML の一種である SVG (Scalable Vector Graphics)[4] を用いて表示させている。

4. 考察

4.1 システムの処理速度について

本システムにおける動画像 (CM) 検索に要する時間は、入力する約 650 秒のテレビ放送の MPEG ストリーム (MPEG1, 352 × 240, 1.5Mbps) に対して、画紋情報抽出から CM 検索 (検索クエリーは 21 個の CM) 等の処理を経て表示クライアントでの検索結果表示完了までに、約 60 秒かかる。本システムでは検索部分を、我々の従来システムにおける画紋情報によるアルゴリズムを基本的にそのまま流用していることから、文献 [2] での検討結果より処理時間のうち検索に要している時間は僅かであり、その処理の大半を Software MPEG Decoder における Decode 処理 (画紋情報抽出) に費やしており、全体の処理時間は、検索サーバの CPU 速度に大きく依存している。このことから検索サーバ内における検索モジュールは、システム全体の全処理時間のうち、画紋抽出モジュールからの画紋情報抽出結果の待ち時間が大半を占めていることになる。つまり、本システムにおける検索クエリー CM 数は、暫定的に 21 個としているが、その数を大幅に増加さ

せても、本システムにおいては、CPU の速度とは関係なく対応できることになる。

また、システム全体の処理時間は、検索サーバの CPU 速度に依存しているが、本システムの現状 (PentiumIII 1GHz) においても、入力ストリームのリアルタイム再生の 10 分の 1 以下の時間で処理していることから、単純計算ではあるが、理論上同時入力 MPEG ストリームを 10 本、つまり 10 チャンネル分までは、リアルタイムに CM 検索を行うことが可能である。

4.2 MPEG7 インスタンスについて

画紋情報の数値データのみでは、15 秒間の映像の場合、そのファイルサイズが約 7k バイト程度であるが、画紋情報を MPEG7 に従って記述 (完全記述版: Complete Description) すると、約 200k バイト程度と約 30 倍程度となった。本システムでは、メタ情報として画紋情報 (ColorLayout Descriptor) しか記述していないので、インスタンス中で、オーバーヘッドとしての記述部分の割合が高かったことが、その差を大きくしたものと思われる。但し、絶対的に MPEG7 インスタンスのファイルサイズが大きいわけではないので、本システムにおける処理時間には、影響していないと思われる。

5. むすび

本稿では、我々の研究グループが従来から提唱している「画紋情報」を利用した、動画像検索システムの実装について報告した。本システムでは、MPEG7 に準拠したインスタンス文書を利用したことで、MPEG7 を使った応用システムの実現例を示すことができた。今後の課題としては、検索対象となるテレビ映像の入力方法の改良や、検索クエリーの映像種類の拡大等が挙げられる。

謝辞

本研究は、通信・放送機構早稲田リサーチセンター 高橋克直氏 (現松下技研) らの研究成果を引き継いで実施したものである。ここに謝意を表する。

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像検索方式に関する検討," 信学技報, IE98-83 (1998-11).
- [2] 高橋克直, 富永英義, 杉浦麻貴, 横井摩優, 寺島信義, "特徴的な動画像の画紋情報を用いた高能率動画像検索法," 画像電子学会誌, 第 29 巻, 第 6 号 (2000)
- [3] ISO/IEC JTC1 SC29/WG11, "Multimedia Content Description Interface (ISO/IEC 15938:MPEG7) Final Committee Draft," Mar., 2001
- [4] <http://www.w3.org/TR/2000/CR-SVG-20001102/>