

映像コンテンツ同定のための特徴量抽出手法

横山 和正[†] 高木 真一^{††} 小舘 亮之^{†††} 富永 英義[†]

[†] 早稲田大学 工学部 電子情報通信学科

〒1169-8555 東京都新宿区大久保 3-4-1

^{††} 早稲田大学 大学院 国際情報通信研究科

〒169-0051 東京都新宿区西早稲田 1-3-10

^{†††} 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

E-mail: †yokoyama@tom.comm.waseda.ac.jp

あらまし 近年のブロードバンドネットワークの普及により、映像コンテンツのネットワーク上での流通基盤が整いつつあり、映像コンテンツの検索、同定技術の必要性が高まりつつある。本稿では、映像コンテンツ中のカットの色情報から抽出し、カットの比較を実現する特徴量 MAV (Macroblock Address Value) を定義し、その抽出手法を提案する。そして MAV を用いたカットの比較結果を利用することで、時間編集された映像コンテンツ同士の比較が実現できることを示す。

キーワード 映像コンテンツ, 同定, 動画像検索, MPEG

A Significant Information Extracting Method for Video Content Identification

Kazumasa YOKOYAMA[†], Shin'ichi TAKAGI^{††}, Akihisa KODATE^{†††}, and Hideyoshi TOMINAGA[†]

[†] Department of Electronics, Information and Communication Engineering, Waseda University
Okubo 3-4-1, Shinjuku-ku, Tokyo, 169-8555 Japan

^{††} Graduate School of Global Information and Telecommunication Studies, Waseda University
Nishiwaseda 1-3-10, Shinjuku-ku, Tokyo, 169-0051 Japan

^{†††} Graduate School of Global Information and Telecommunication Studies, Waseda University
Nishiwaseda 1-3-10, Shinjuku-ku, Tokyo, 169-0051 Japan

E-mail: †yokoyama@tom.comm.waseda.ac.jp

Abstract Today, the necessity to retrieve and identify the video content is being put into place by complete equipment of video content circulating bases on network. In this paper, we define the significant information MAV (Macroblock Address Value) and propose an extracting method of this. MAV is extracted from color information of the video content's cut, and the use of MAV make the comparison of cut possible. Then we show a possibility of trimming video contents comparison using a comparative result of cut.

Key words Video content, Identification, Video retrieval, MPEG

1. 研究の背景と目的

MPEG に代表される画像符号化技術の向上により、我々の身の回りに符号化された画像データが広まりつつある。また、PC の性能の向上に伴い、PC 上での映像の編集、取り込みも比較的容易に行え、インターネットの通信速度の向上により、画像ファイルのネットワーク上での流通も可能になりつつある。

この様に、一般ユーザレベルでも容易に画像の取得、編集、配信ができるようになると、ネットワーク上やデータベース上に、様々な種類の映像コンテンツが混在する状態になることが予想される。種々の映像コンテンツの中には、映像コンテンツを1つの作品としてみると、人が見れば同じ作品と見させるような映像コンテンツ同士が複数存在する可能性があり、これは、著作権の管理や、データベース管理を行う上では、望ましくない。その為、映像コンテンツを検索、識別する技術が今後ますます重要になってくる。

映像コンテンツから特徴量を抽出し、検索する研究は、様々な手法が提案され、研究されてきた。[1]~[6] しかし、それらの多くは、膨大な量の映像コンテンツの中から、ユーザが求めるある特定の1つのシーンを検索するものであり、シーンが集まった映像コンテンツ全体の検索を行うものではない。

そこで、本稿では、映像コンテンツ全体の同定手法について検討する。

本稿では、以下、上記の機能を実現する為に有意な情報の抽出手法について論ずる。2. においては対象とする映像コンテンツについて、3. では、映像コンテンツの同定方法、4. ではカットの比較方法についての検討を記し、5. で提案手法の説明と、実験による評価、考察を、6. でまとめを行う。

2. 対象とする映像コンテンツについて

本稿では、映像コンテンツに対する編集や、状態の変化には、以下のものがあると考えられる。

- 時間的な編集を受けている
(最初、最後が切れている、間が抜けている)
- フレーム中の色情報が一部変化している
(字幕付き、伝送中に色ずれが起きている)
- フレームサイズが異なる
(拡大、縮小されている)

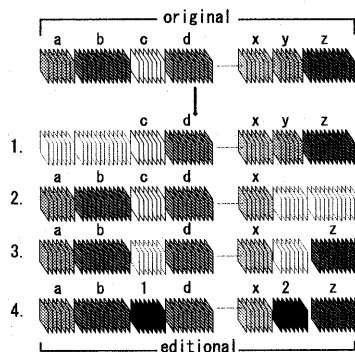


図1 オリジナルと時間的に変化した映像コンテンツの関係

色ずれや字幕、フレームサイズの違い等には既存のシーンの検索方法を映像コンテンツ全体に適用することで対応することができる[1],[3]。時間編集を受けたものは、シーンの検索方法を映像コンテンツ全体に適用しても対応することはできない。

そこで、今回は対象とする映像コンテンツに、特に図1のような、元は同じ映像コンテンツだが、

- (1) 最初が切れている
- (2) 最後が切れている
- (3) 間の映像が一部抜けている
- (4) 間に他の映像が混じっている。

等、時間的な編集を受けた映像コンテンツの同定手法を提案し、評価する。

また、映像コンテンツの種類は、時間編集をよく受けるものとして、映画やドラマ等を想定する。また、色ずれなどの色の変化は想定外とする。

3. 映像コンテンツの同定方法

映像コンテンツは、一般的にカット点と呼ばれるカメラの on, off で区切ることができる。このカット点の間の映像をカットと定義し、映像コンテンツをカットの集合体と見なす。

この様に、映像コンテンツをカットの集合体と考えた場合、映像コンテンツに時間的な編集が加わっても、共通のカットが残ることが考えられる。例えば、図2の様に、ある映像コンテンツ A が a,b,c,...,x,y,z というカットで構成されているとする。この時、映像コンテンツ A に時間的な編集が加わることで、映像コンテンツ A 中の一部のカットが無くなり、新たな映像コンテンツ A' が発生したとする。この場合、映像コンテンツ A' は映像コンテンツ A 中のカット

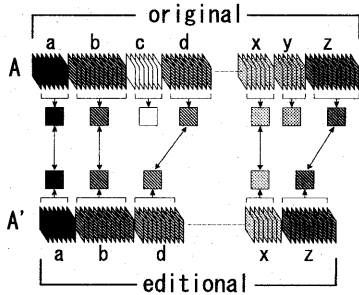


図2 オリジナルと時間的に変化した映像コンテンツの同定

の一部が無くなった映像コンテンツであるので、図2の様に、カットcやカットyがなくなったとしても、他のカットは映像コンテンツAと同じ状態であり、映像コンテンツAと映像コンテンツA'の間にいくつかの共通のカットが存在することになる。

そこで、映像コンテンツをカット点で分割し、2つの映像コンテンツのカットから特徴量を抽出して比較することで、共通のカットの割合を調べ、時間的な編集の影響を少なくして、映像コンテンツを同定することができる。

また、カット点の分割には、1ピクチャ単位でカットを分割する χ^2 検定法を使用する[7]。

4. カットの比較方法の検討

4.1 画紋を用いた比較方法

カットを比較する手法の一つとして動画の特徴情報を用いた動画検索を応用する手法について検討する。

文献[1]では、動画から抽出する特徴情報を画紋と定義している。その画紋を用いた動画の比較の手順を以下に示す。

- (1) 比較したい動画の双方から1フレーム毎に輝度、色差の平均値を算出する
- (2) 双方から取り出した各1フレームの輝度、色差の平均値を時系列に並べて特徴量とする
- (3) 双方の特徴量のマッチングを行う

ここで、輝度、色差の平均値を1フレーム毎に取り出すのは、圧縮動画を完全に複合することなく取り出せるマクロブロックのDC成分から輝度、色差の平均値を算出し、処理の高速化を図っているためである。

4.1.1 画紋を用いた手法による実験

映像コンテンツ中のカットの特徴量として、画紋

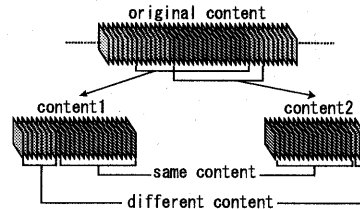


図3 実験用の動画の生成

表1 特徴量の比較結果

	共通カット	正解数	誤検出数	未検出数
洋画1	35	16	65	14
アニメ1	13	11	21	2

を用いる手法について実験した。

実験には、映像コンテンツ2種類を用意し、各作品について、図3のように元となる映像コンテンツから2つ、互いに2、3秒取り出し位置のずれた時間長約30分の動画を取出す。そして取り出した各動画をカット点分割、15秒以上のカットから画紋を抽出し、画紋の比較のアルゴリズムで比較を行った。

ここで比較に用いるカットの長さを15秒以上としたのは、文献[1]において、画紋を用いた動画の検索が15秒以上の動画において最も検索精度がよいことが示されていたためである。比較実験の結果を表1に示す。

4.1.2 実験結果に対する考察

画紋を用いた動画の検索手法は、文献[1]では18~58個の正解の検出に対して0~3個の誤検出、0~1個の未検出と、高い精度の検索を実現していたが、表1の様に誤検出が多い結果となった。

ここで、CMと映画の1カットの画紋の時間軸波形は、図4、5のようになる。CMは短時間でも色の変化が激しいのに対して、映画のカットは色の変化が少ないことが確認できる。

元々画紋は、CMなどの短時間だが色の変化の大きな映像を対象としていた。しかし、図5のように本稿で対象とする、映画などのカットは、色の変化が少ない。画紋の比較方法では、色の時系列の変化を特徴量としてマッチングを取っているため、映画のような色の変化が少ない動画では、特徴量の変化が少なくなってしまい誤検出が増えることになる。

4.2 カットの特徴に対する考察

文献[1]での実験の結果を踏まえ、映画などのカットの特徴について考察する。

一般的に、映画やドラマ等の映像コンテンツ中のカットには、次のような特徴がある。

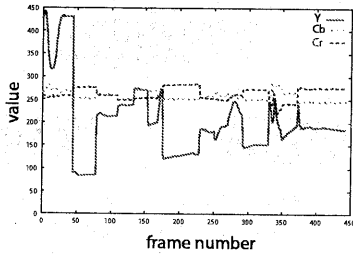


図4 CMの画紋の時間軸波形

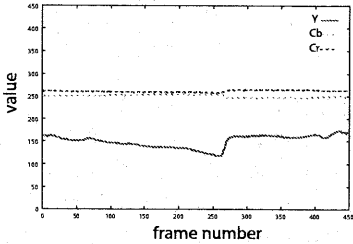


図5 映画のカットの画紋の時間軸波形

表2 映画のカットの長さの分布 [%]

コンテンツ名	総カット数	カットの長さ [sec]				
		5~10	10~15	15~20	20~30	30~
アニメ1	173	72.8	16.2	4.6	5.2	1.2
アニメ2	138	76.8	10.9	9.4	1.4	1.4
洋画1	74	59.5	14.9	6.8	5.4	13.5
洋画2	59	59.3	8.5	6.8	6.8	18.6
洋画3	97	46.4	11.3	9.3	23.7	9.3

カット中ではフレーム間の色の変化量は少ない
 カットはカメラの on, off に相当するため、カメラの高速な移動やカメラの前を物体が高速に通り過ぎたりしない限り、カット点間でのフレーム間の色の変化量は少ない。

カットの長さは比較的短い

映画やドラマなどは、多数のカメラで撮影を行った後、編集をして映像コンテンツを作成するため、カメラの切り換わりが多くなり、検出されるカット1つ1つは短くなる。表2に映画の χ^2 検定法を用いたカット点分割の結果を示す。

4.3 特徴量抽出時の問題点

画紋を用いた手法の結果及び前節のカットの持つ特徴を考慮すると、特徴量を取り出す際の問題点として、以下のものが挙げられる。

- フレーム間の色の変化は少ない
- カットが短いので特徴量が少ない
- Iピクチャの位置がずれるとカット内のフレームのデータが変わる

- カットの分割をIピクチャ単位で行う為、同じカットでもIピクチャの位置がずれているとカット内のフレームのデータが変わる

以上の問題点を考慮すると、カットから取り出す特徴量に要求される条件は、

- フレーム間の色の変化量に関係ない
- カットの長さに関係ない
- Iピクチャのずれの影響が少ない

となる。

5. 提案手法

前節の議論を踏まえた上で、映画のカットのような色の変化の少ない動画像同士の比較に適する特徴量を提案する。

5.1 カットの特徴量

カット中の、各マクロブロックのカット内での平均値の集合を特徴量とする方法を提案する。特徴量抽出の手順を以下に示す。(図6)

- (1) カットの中からIフレームを取り出す
- (2) カットの中のIフレームのマクロブロック毎の輝度、色差の平均値を求める
- (3) 各マクロブロック毎の輝度、色差の平均値の集まりを特徴量 MAV(Macroblock Average Value) とする

5.2 特徴量の比較方法

特徴量同士の比較方法に平均誤差を用いる。平均誤差を用いた比較方法は次のようになる。

まず、前節の図6の方法を用い、比較したいカット双方から図7のような2つの MAV, A と B を取り出す。ここで特徴量の各マクロブロック毎のカット内の輝度の平均値の値を各々 $a_1, a_2, a_3, \dots, a_{MBN}, b_1, b_2, b_3, \dots, b_{MBN}$ とする。但し、ここで MBN はマクロブロックの数 (Macro Block Number) である。すると、この二つの MAV の平均誤差 ME_Y は、次式のようにになる。

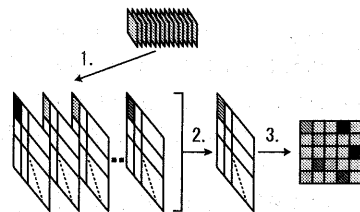


図6 特徴量の取り出し方

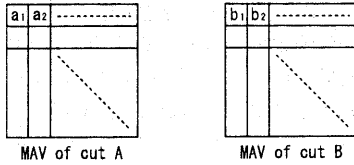


図7 カットの特徴量

表3 特徴量の比較結果

	共通カット	正解数	誤検出数	未検出数
洋画1	35	35	4	0
アニメ1	13	13	2	0

$$MEY = \sum_{i=0}^{MBN} \frac{((b_i - a_i) - ave)^2}{MBN} \quad (1)$$

ave は MAV 同士のマクロブロック毎の誤差の平均値であり、次式で示される。

$$ave = \sum_{i=0}^{MBN} \frac{(b_i - a_i)^2}{MBN} \quad (2)$$

同様の方法で、色差についても ME_{Cb} , ME_{Cr} を求める。そして、2つの MAV を比較して平均誤差 MEY , ME_{Cb} , ME_{Cr} の値が全て閾値以下となったときに、2つのカットが一致すると判定する。今回は閾値を輝度を 50, 色差を 5 とした。

5.3 特徴量の評価

5.3.1 画紋を用いた手法との比較

画紋を用いた手法と精度を比べる為に、画紋を用いた手法の時と同条件下で実験を行った。

その実験結果を表3に示す。

5.3.2 実験結果に対する考察

表1と表3より、画紋を用いた手法に比べて正解数が上がり、また、誤検出は従来手法の 6.15% に減り、未検出は 0 個になっている事が確認できる。

洋画1, アニメ1ともに誤検出があるが、これは図8の様にカット点分割の際に片方の動画像では1つのカットとして判定されたがもう片方では複数のカットとして判定され、この複数のカットが他方の1つのカットと一致したと判定してしまったことが原因である。

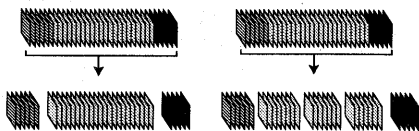


図8 カット点分割の失敗例

表4 編集を加えた映像コンテンツの詳細

	MAV の数	時間長 [sec]
洋画 1CM	71	1046
洋画 2CM	113	1367
洋画 3CM	131	1240
洋画 4CM	169	1353
アニメ 1CM	137	1236
アニメ 2CM	146	1194

表5 比較結果 (MAV の数の割合 その1) [%]

	洋画 1CM	洋画 2CM	洋画 3CM
洋画 1(126)	42.9	1.79	0.00
洋画 2(116)	0.00	50.0	0.00
洋画 3(185)	0.00	0.00	61.4
洋画 4(169)	0.00	0.00	0.00
アニメ 1(224)	0.00	0.00	0.00
アニメ 2(225)	0.00	0.00	0.00

5.3.3 提案手法の評価

前節で、提案手法が画紋を用いた手法よりもカットの比較の精度がよいことを示したが、15秒以上のカットだけでは映像コンテンツ全体を比較するにはカットの数が少ない。

そこで、3秒以上のカットを対象として検証実験を行った。

本稿では、6種類の映像コンテンツを使用した。その全てについて、オリジナルの映像コンテンツの一部を編集し、テレビ放送で流されている CM 合計 300 秒を挿入した映像コンテンツを用意した。各映像コンテンツ 3 秒以上の MAV の数と映像コンテンツの時間長を、表4に示す。但し、使用した映像コンテンツは、MPEG1 で符号化された、画像サイズ 352[pixel] × 240[line], ビットレート 1800[kbps] のものを使用した。

また、各映像コンテンツの 3 秒以上のカットから MAV を取り出し、映像コンテンツの同定実験を行った。結果を表5, 6, 7, 8に示す。

ここで表5, 6は、各映像コンテンツの 3 秒以上の全カット数に対する一致したカットの数の割合である。各映像コンテンツ名の横に 3 秒以上のカットの総数を示した。

また、表7, 8は各 CM 入りの映像コンテンツの時間長の和に対する、比較において一致したカットの時間長の和の割合である。各映像コンテンツ名の横にオリジナルから切り出した映像の割合 [%] を示す。

5.4 結果に対する考察

表5, 6, 7, 8, いずれにおいても、異なる映像コンテンツ間の誤検出は、洋画 1, 洋画 2CM 間での誤

表6 比較結果 (MAVの数の割合 その2) [%]

	洋画 4CM	アニメ 1CM	アニメ 2CM
洋画 1(126)	0.00	0.00	0.00
洋画 2(116)	0.00	0.00	0.00
洋画 3(185)	0.00	0.00	0.00
洋画 4(169)	60.6	0.00	0.00
アニメ 1(224)	0.00	65.4	0.00
アニメ 2(225)	0.00	0.00	72.4

表7 比較結果 (時間長の割合 その1) [%]

	洋画 1CM	洋画 2CM	洋画 3CM
洋画 1(71.3)	57.7	10.6	0.00
洋画 2(78.1)	0.00	54.0	0.00
洋画 3(75.8)	0.00	0.00	47.4
洋画 4(77.8)	0.00	0.00	0.00
アニメ 1(75.7)	0.00	0.00	0.00
アニメ 2(74.9)	0.00	0.00	0.00

表8 比較結果 (時間長の割合 その2) [%]

	洋画 4CM	アニメ 1CM	アニメ 2CM
洋画 1(71.3)	0.00	0.00	0.00
洋画 2(78.1)	0.00	0.00	0.00
洋画 3(75.8)	0.00	0.00	0.00
洋画 4(77.8)	55.2	0.00	0.00
アニメ 1(75.7)	0.00	56.5	0.00
アニメ 2(74.9)	0.00	0.00	62.3

検出以外存在しない。それに対し、同種類の映像コンテンツ間の同定においては、表5, 6, 7, 8ともに際立って高い値を示しており、洋画1, 洋画1CM間の値も、洋画1, 洋画2CM間での誤検出の値よりも高い値を示している。

このことより、提案手法が、映像コンテンツの同定を確認するには、十分な精度があることが確認できる。

表5, 7において、洋画1と洋画2CM間で誤検出が発生している。この誤検出を起こす原因となったカットは、いずれも長さが比較的に長く、暗い場面であった。色が全体的に暗く、長い場面である場合、MAVの輝度、色差ともに値が平均化されてしまい、特徴点が失われ、他の色が暗く、長い場面と誤検出を起こしたと考えられる。

6. まとめ

本稿では、著作権管理やデータベース管理の為に映像コンテンツ同定手法として、オリジナルの映像コンテンツと、それを時間編集した映像コンテンツ同士を比較し、一致していると判断する同定手法について検討した。

その方法として、コンテンツをカットに分割し、

カット同士の一致した割合を求めることでコンテンツの同定を判断する手法を提案した。その際に、カットの特徴量としてカットのマクロブロックごとの平均値の集合をMAVと定義し、MAVのカットの比較に対する有用性を示した。また、MAVを用いることで、映像コンテンツ同士の同定が行えることも示した。

7. 今後の課題

今後の課題として以下のことが挙げられる。

色ずれなどの色の変化への対応

実際に編集された映像コンテンツが流通する際は、伝送や編集機器の影響で色ずれが発生することは十分考えられるので、色ずれにも対応させる必要がある

暗い場面での誤検出の解消

暗くて時間長の長いカットが誤検出を引き起こす問題について対処する必要がある

閾値の理論的裏付けに基づいた設定

一致と判定する閾値を実験的に輝度を50、色差を5としたが、理論的な裏付けに基づいた閾値の設定が必要である

カット比較時の誤検出を考慮した評価

提案手法の評価において、カットの比較時の誤検出を考慮していない。誤検出の割合と誤検出を考慮した比較結果の検証が必要である

文 献

- [1] 高橋克直, "有意情報抽出による画像情報へのアクセス手法の研究," 平成12年度早稲田大学博士論文, (2001)
- [2] 粕谷英司, 前田鏡二, 宮森恒, 富永英義, "圧縮映像中のパラメータを利用した高速照合とその検索手法の提案," 情報研報, (1997)
- [3] 関本信博, "逐次フレーム入力されるクエリーによる時系列データの類似区間スボッティング検索," 電子情報通信学会論文誌, pp.747-757(2001)
- [4] 村野井亮治, ZHAO J, 早坂里奈, 伊藤雅仁, 松下温, "ショット識別子による検索を利用したデジタル動画像における著作権保護システムの構築," 情報処理学会シンポジウム論文集, (1998)
- [5] 長坂晃明, 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, pp.1831-1837(1998)
- [6] 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," 電子情報通信学会論文誌, pp.2735-2744(2000)
- [7] 長坂晃明, 田中譲, "カラービデオ映像における自動索引付け法と物体探索法," 情報処理学会論文誌, vol.33, pp.543-550(1992)