

MPEG 映像からのテロップ検出方法の検討

A study on detecting telop from MPEG video

倉橋 誠

Kurahashi Makoto

要 旨 映像コンテンツの効率的な視聴を可能とするために映像構造化を行う手段として、映像中のテロップを検出し、その映像の特徴として利用することが有効である。一方、近年デジタル放送の普及などにより、MPEG などに代表されるデジタル映像が普及している。そこで、MPEG 形式で圧縮符号化された映像からテロップ検出を行う方法を開発した。これは、MPEG パラメータである DCT 係数に注目し、その高周波成分の値から、元の画像の強いエッジの有無を判定することでテロップ検出を行う方法である。

この方法により、従来のマクロブロック符号化方式に着目したテロップ検出方法に比べて、テロップを表示している領域を高精度で検出することができた。

Summary For viewing a lot of video content efficiently, it is useful to segment each video program into several scenes automatically. An effective approach to achieve this requirement is using captions superimposed on video images as keys.

In this paper, we propose a method for detecting captions from MPEG videos by checking high frequency elements of DCT coefficients encoded in MPEG streams.

キーワード : テロップ検出, MPEG, H.264, 映像構造化, デジタル放送, DCT

1. まえがき

近年のストレージの大容量化により、ユーザが録画し、保存できる映像コンテンツの量が增大している。限られた時間で大量のコンテンツを効率よく視聴するために、これらのコンテンツを構造化し、整理や検索を効率的に行うことが課題となっている。

コンピュータ処理により自動的にコンテンツの内容を認識し、映像の構造化を行うための有用な手掛かりとして、映像中に含まれているテロップを利用することが考えられている。これ

は、テロップがコンテンツの意味を端的に表している場合が多く、テロップ以外の特徴量を用いるよりも的確に映像の意味を抽出できるためである。さらに、テロップの文字内容を認識しなくとも、表示されているテロップの位置、大きさといった特徴や存在自体を、映像コンテンツの意味を捉える手掛かりとして利用することも考えられている。

一方、放送のデジタル化が進み、送信される映像信号がベースバンドの信号から、MPEG-2 や H.264 に変わりつつある。筆者らは、このよ

うな状況をふまえ、代表的なデジタル圧縮符号化方式であるMPEGを対象とした、デジタル映像からのテロップ検出方法の検討を行った。

2. MPEGからのテロップ検出

2.1 概要

圧縮符号化されたデジタル映像コンテンツからテロップを検出する方法は、一旦元映像にデコードしてエッジなどの特徴抽出を行う方法と、符号化されているパラメータをそのまま利用する方法に大きく分けられる。前者は、デコードさえ行えば、従来のテロップ検出方法を活用でき、どのような圧縮符号化形式の映像にも対応できる利点があるが、映像のデコードや画像処理に比較的多くの計算を必要とする。一方、後者は、完全なデコードは必要なく、さらに圧縮符号化時に算出した映像の特徴パラメータを利用できるため、計算量を抑えた効率的な検出処理が可能である。

デジタル映像からのテロップ検出には、実用性の観点から計算量の削減が求められている。このため、MPEGなどで圧縮符号化されたデータから直接検出処理を行うことが望ましいと考えられる。

2.2 従来研究

MPEG形式の映像からテロップを検出する技術は、既にいくつかが発表されている。

一つは、動き補償なしでフレーム間相関を利用して符号化されている領域をテロップ候補とし、そのような領域の形状判定によってテロップ領域を検出する方法⁽¹⁾である。他に、DCT係数の時間的な変化からテロップの出現を検出し、その後の出現位置のマクロブロックの符号化方式やDCT係数から表示位置の特定やテロップらしさの判定を行う方法⁽²⁾がある。これらの方法により、テロップ検出をMPEG映像から直接行うことが可能となっている。

しかし、既存の方法には問題点もある。文献(1)では、テロップ以外の動かない被写体をテロップとして誤検出するなどの問題があり、ま

た文献(2)では、ブロックの行単位での時間的な変化が大きい場合をテロップの出現と判定するため、幅が短いテロップは検出できない。また、巨大なテロップの出現とカット点の区別を付けにくく、特殊効果を伴うテロップに対応することも難しい。

このように従来方法では検出精度に問題があり、出現の仕方などにさまざまな特徴のあるテロップへの対応が難しいという課題があった。

一方、テロップ検出を利用したアプリケーション⁽³⁾も発表されている。文献(3)では、ニュース番組で表示されるテロップを検出し、位置や大きさといった特徴からニュース項目の冒頭部分で表示されているテロップであるかを判定する。これを利用して、各ニュース項目の頭出しや、ニュース項目を選択して再生するインタラクティブ再生を行うことができる。

2.3 本研究での検出方法

空間領域でのテロップの縁、すなわちテロップと背景との境界部分には、エッジ(急激な輝度や色差値の変化)が存在する。テロップは視覚を利用した情報伝達手段であり、見やすさを確保する必要があることから、これはほとんどのテロップに共通した特徴である。従来は、非圧縮映像からのテロップ検出方法として、このようなエッジを利用したものが多く使われてきた。

本方法でも同様に、テロップのエッジに注目する。MPEGでは、画像は8×8ピクセルのブロックと呼ばれる単位で、その周波数成分がDCT係数により表現されている。テロップのような急峻なエッジのある画像を周波数空間に変換すると、高周波成分の絶対値が高くなることが知られている。図1は、境界付近で輝度が急峻に変化している画像と緩やかに変化している画像のDCT係数を比較したものである。DCT係数は、右、下ほど高周波の成分であり、図1では表記が明るいほど絶対値が大きな係数であることを表している。図1より、輝度の変化が急な画像の方がDCT係数の高周波の成分にまで、大きな値が出現していることがわかる。この性質を利

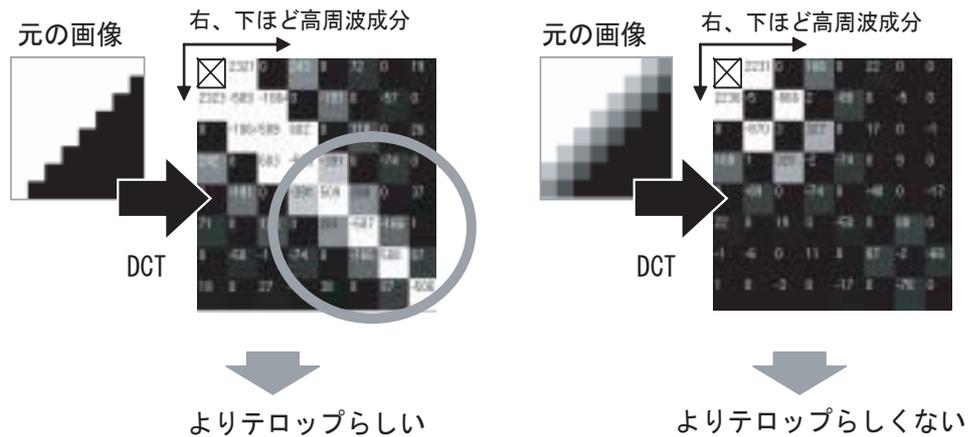


図 1 エッジと DCT 係数

用して、高周波成分に重みをおいて周波数成分値を評価することで、ブロックに対してエッジによるテロップらしさの判定を行う。

このように本方法では、DCT 係数により急峻なエッジの有無を推定することにより、テロップが表示されている領域の検出精度向上を図った。また、この判定を複数のフレームに渡って行い、出現時の特殊効果や表示中の背景などの変動による影響を受けにくい、強固なテロップ検出を目指した。

3. 方法・実装

本章では、MPEG からのテロップ検出の提案方法について述べる。全体としては、テロップの候補領域判定、テロップ識別という 2 段階の処理を行う。

3.1 テロップ候補領域判定

テロップ候補領域判定では、1 フレームを構成するブロックごとにテロップが表示されているかどうかを判定し、フレーム内においてテロップが表示されている領域を判定する。これは、ブロック内にどの程度急峻なエッジが存在するかで判定する。

フレーム I_n を構成するブロック b_{xy} ($0 < x < \text{Width}/8, 0 < y < \text{Height}/8$) が、テロップの表示されているテロップブロックであるかを判定するには、まず b_{xy} の 8×8 の 2 次元 DCT 係数を、

周波数帯別に n 個の領域 a_1, a_2, \dots, a_n に分割する。たとえば、DC 成分である $(0, 0)$ の側から同心円状に、10 の領域に分割する。図 2 にこの関係を示す。

⊗	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	2	3	4	4	5	6	7
3	3	4	4	5	6	7	8
4	4	4	5	6	6	7	8
5	5	5	6	6	7	8	9
6	6	6	7	7	8	8	9
7	7	7	8	8	9	9	10

図 2 DCT 係数周波数帯分割

領域 a_f ($0 < f < n$) のそれぞれの中で絶対値が最大の $a_{f_{\max}}$ を領域の代表値とし、それに対して重み付けをする。重み付けには、あらかじめ周波数帯ごとの重み付けの加減を定めたベクトル w を用意する。これは、係数の高周波成分を重視するため、高周波の領域ほど高い値とする。

次に、式(1)により、ブロックのエッジによるテロップらしさの評価値 v を決定する。

$$V = \sum_{f=0}^n |a_{f_{\max}}| \cdot w_f \quad \dots (1)$$

なお、フレーム内の同位置のブロックが時間的に連続してエッジブロックと判定された場合、DCT係数を前後のフレームで比較し、変化が大きい場合は文字が切り替わったと判断する。この場合も、テロップと関わりの深い高周波成分に重点を置いて評価する。

DCT係数の高周波成分値以外の特徴を併用することもできる。たとえば動きベクトルを特徴として用いてもよい。

テロップが画面上に静止している場合、そのテロップを含むマクロブロック(16 × 16 ピクセルの、予測の最小単位)は、テロップという特徴的な映像を効率よく符号化するため、P/Bフレームにおいては動き補償なしでフレーム間予測を利用するマクロブロックタイプとなっている可能性が高い。また、一定方向に一定速度で流れているテロップを考えると、そのテロップを含むマクロブロックは、図3に示すように、同様な方向と長さの動きベクトルによる動き補償を行なっている可能性が高い。動きベクトルがこのような条件に適合している場合に、そのマクロブロックがテロップらしいと判断することができる。

これらより、エッジと動きベクトルの両方がテロップとしてふさわしい領域を、テロップ候

補領域と判定する。その結果、該当ブロック(マクロブロック)が候補領域に含まれているか否かを表す二値の行列を出力する。なお、二つの条件の合算方法によって、結果はブロック単位またはマクロブロック単位となるが、以降は両者とも広い意味で捉えてブロックと呼ぶ。

3.2 テロップ識別

テロップ候補領域を検出した後に、各テロップを識別する。つまり、テロップ候補領域の空間的・時間的分布から、どの部分が一本のテロップであり、それがいつからいつまで表示されていたかを特定する。これは、大きく分けて次の2段階の処理で行う。

1段階目は、表示中のテロップを特定する形状判定の段階である。ここでは、ある時点でのテロップ候補領域の分布から、矩形状のテロップらしい固まりを検出する。そしてその領域を、1本の表示中のテロップであると仮に判定する(図4)。このとき、テロップ候補領域が矩形に近い形状をなしていない部分は、形状がテロップとしてふさわしくないと判断し、以降の判定対象から除外する。

2段階目は、テロップの消失を確認することにより、一本のテロップの存在を確定する段階である。すなわち、 I_{n-1} フレームで判定した矩

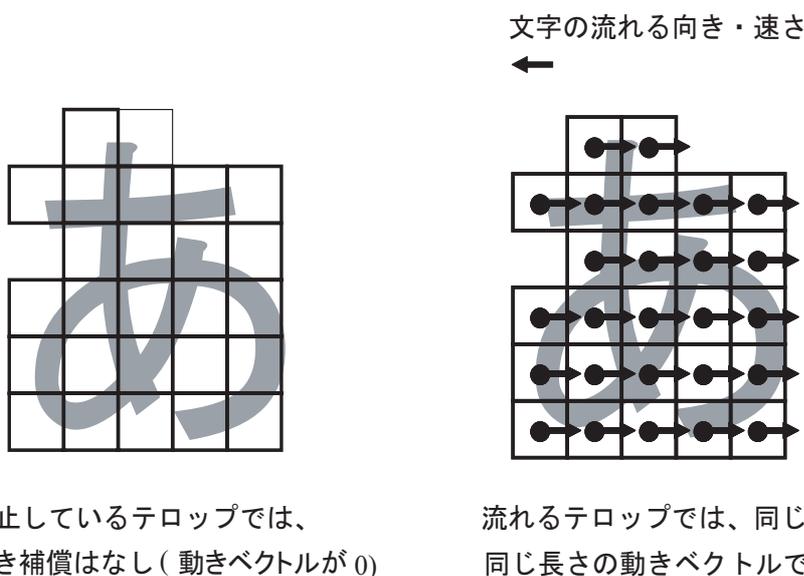


図3 動きベクトル

形領域を構成するテロップブロックで、 I_n フレームでテロップブロックではなくなり、なおかつそのブロックが過去に一定以上連続してテロップを表示している矩形領域であると判定されていた場合に、 I_{n-1} フレームまで表示されていた一本のテロップが消失したと判定する。なお、今回テロップ候補領域と判定されたブロックでも、連続表示中に文字の変化があったと判定している場合は、消失した場合と同様の処理を行う。

3.3 実装

実装はPC上で行った。

実験に用いたMPEG映像は等間隔で1フレームが挿入されているものと仮定し、テロップ表示領域判定とそれに次ぐテロップ識別は、1フレームが出現するタイミングで行っている。P/Bフレームでは動きベクトルがテロップらしいかを観測し、1フレームの間で集計する。その結果を、1フレームが出現したときの表示領域判定で利用する。これにより、DCT係数をデ

コードするのは1フレームのみでよく、負荷を低く抑えられる。全体の流れは図5に示すようになる。

なお、想定している典型的なイントラフレームの挿入間隔は、30fpsの映像において15フレーム毎、すなわち0.5秒に1枚である。

4. 評価

検出結果の評価は、次の二段階で行った。

第一の評価は、テロップ候補領域検出結果の評価である。実際にテロップが表示されているマクロブロックの集合 T_a と、検出したテロップ表示ブロックの集合 X を比較し、テロップ表示ブロックの検出精度を測定するものである。評価は再現率 $|T_a \cap X|/|X|$ と適合率 $|T_a \cap X|/|T_a|$ で表す。ここで、 $|X|$ は集合 X の要素数を意味する。再現率は、実際にテロップが表示されている領域に対して検出結果が占める割合を表し、値が大きいほどもれなくテロップをカバーできていることを意味する。適合率は、テロッ

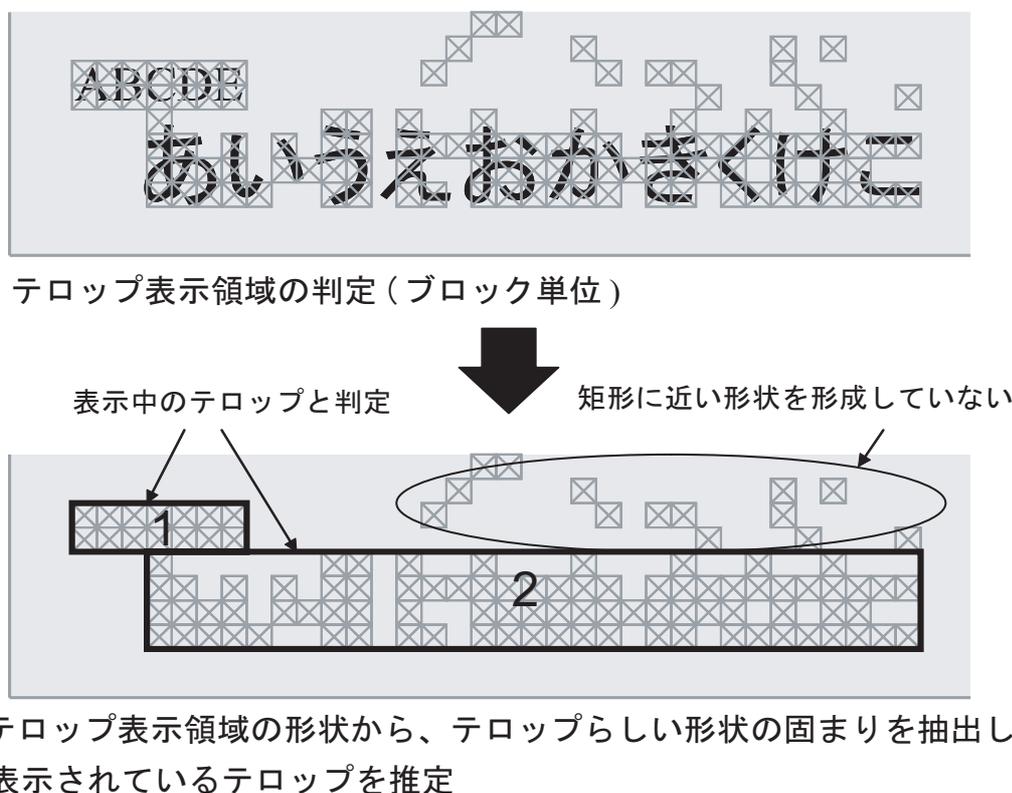


図4 形状判定

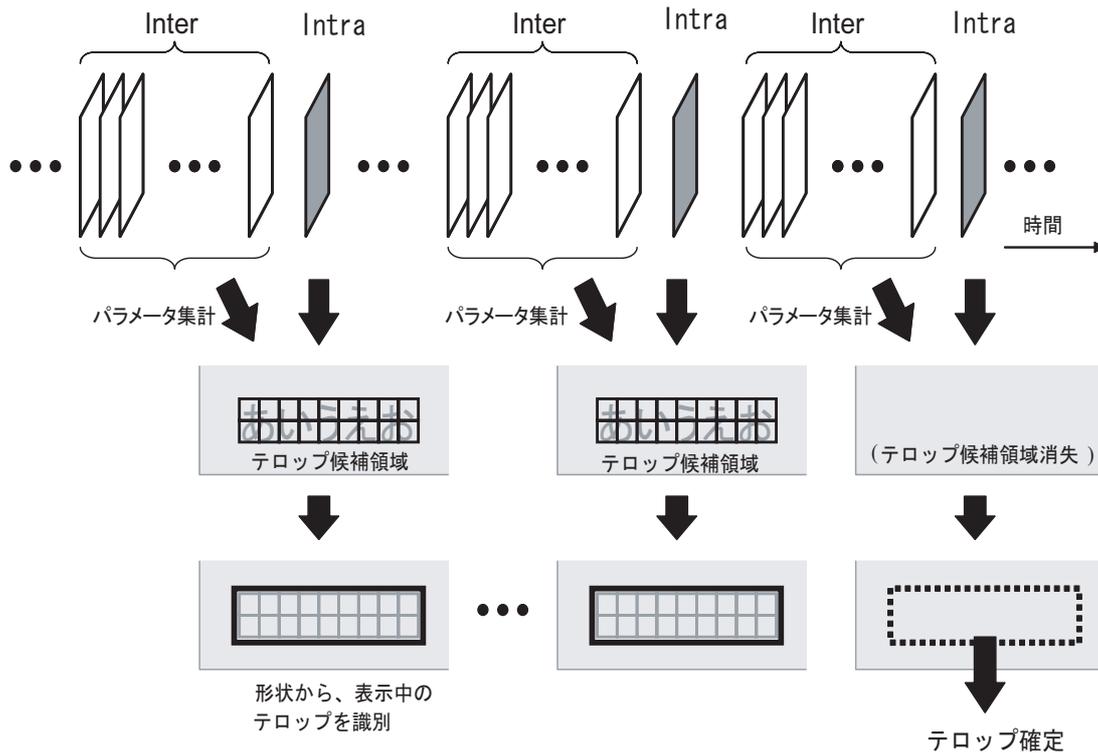


図5 全体の流れ

ブ領域として検出した領域に占める正解ブロックの割合を表し、値が大きいほど誤検出が少ないことを意味する。

この評価は、次の各方法で比較した。

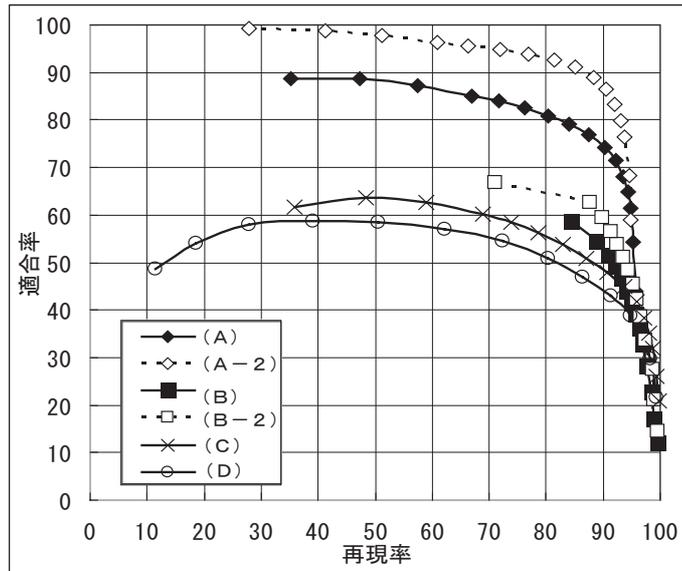
- (A) DCT 係数重み付け + 動き評価による判定
- (B) 動き補償なし + フレーム間相関による判定 (従来方法)
- (C) 「DCT 係数重み付け」のみによる判定
- (D) 「DCT 係数重み付けなし」のみによる判定

なお、(A) (B) の結果に孤立点の除去など簡単な処理を加えたものをそれぞれ (A-2) (B-2) とした。それぞれの方法で、判定基準 (閾値) を変化させ、その都度再現率および適合率をプロットしたグラフを図6に示す。このグラフは横軸が再現率、縦軸が適合率であり、両者を両立するほど (グラフ上で右上端に近づくほど) 精度が高い方法であるといえることができる。

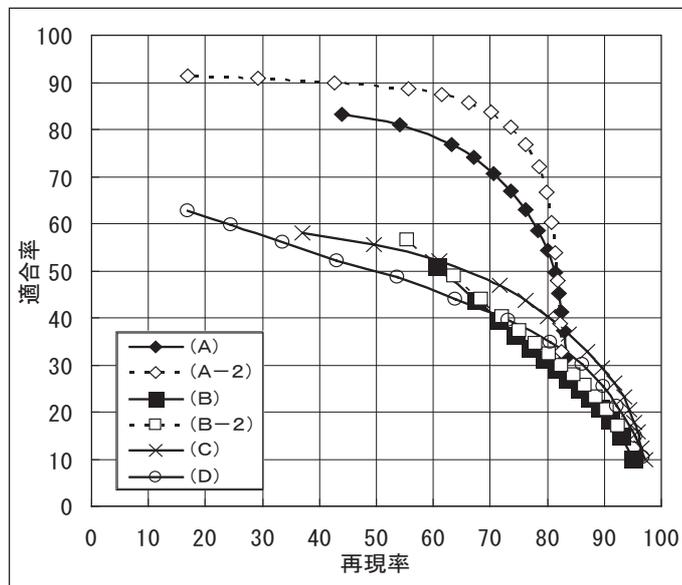
第二の評価は、テロップ識別で得られる結果の評価である。第一の評価の (A) と (B) による判定結果を利用し、形状判定と消失判定を行って

テロップの識別を行った。実際に表示されたテロップの集合 T_1 と、1本のテロップであると識別した検出結果の集合 Y を比較し、テロップの検出精度を測定した。評価は再現率 $|T_1 \cap Y| / |Y|$ と適合率 $|T_1 \cap Y| / |T_1|$ で表す。再現率は、実際に表示されたテロップに対する検出できたテロップの割合であり、値が大きいほど表示されたテロップをもれなく検出できたことを意味する。適合率は、テロップとして検出した結果に対する正解の割合を表し、値が大きいほどテロップの誤検出が少ないことを意味する。この結果を表1に示す。

評価対象としたコンテンツは、解像度が、720 × 480、5Mbps の MPEG-2 方式で、アナログ地上波放送を録画したものである。第一の評価ではニュース番組と音楽番組の合計2時間分の動画を使用し、第二の評価では、それにニュース番組と情報バラエティ番組をさらに2時間分加えたものとした。いずれの評価も正解データは、人間が目視検査を行い、判断した数値を用いている。



(a) ニュース番組



(b) 音楽番組

図6 テロップ表示領域判定結果

表1 テロップ検出評価結果

番組	テロップ数	テロップ領域判定方法	正検出数	全検出数	再現率	適合率
ニュース番組1	164本	本方法	139本	183本	85.8%	76.0%
		従来方法	115本	188本	71.0%	61.2%
ニュース番組2	337本	本方法	289本	362本	85.8%	79.8%
		従来方法	206本	450本	61.1%	45.8%
音楽番組	248本	本方法	146本	173本	58.9%	84.4%
		従来方法	68本	266本	27.4%	25.6%
情報 / パラ エティ番組	431本	本方法	359本	666本	86.9%	58.6%
		従来方法	166本	459本	40.2%	36.2%

5. 考察

第一の評価では、どちらの番組でも(A)で最も高い適合率が得られた。特にニュース番組では、再現率80%台を維持しながら適合率も80%台が得られた。このニュース番組では、エッジが急なテロップを多用しており、この検出方法が期待しているテロップの特徴に合致しているためである。また、(C)と(D)との比較から、どちらの番組でも高周波成分への重み付けにより精度が向上していることが確認できた。なお、いずれの番組においても(A)の再現率の最高値が(B)~(D)と比べて低くなっている。これは、適合率を重視した固定パラメータを採用しているためである。テロップ候補領域の誤検出を少なくすることで、形状判定の精度向上を図っている。

テロップ候補領域判定では、急峻なエッジがあつて動かない被写体を誤検出することが多かった。たとえばスタジオや建物などの背景がこれに当たる。しかしこのような領域は、文字部分に比べてエッジの集中が少なかったり、形状が矩形らしくなかったりすれば、形状判定で除去することができる。そのため、これは第二段階のテロップ識別に必ずしも誤識別をもたらす問題点ではない。なお、形状判定方法の改良は今後の課題である。逆に、第二段階のテロップ識別に特に大きな影響を与えるテロップ候補領域判定段階での問題点は、連続して表示されるテロップの切り替わりが検出できない場合があることである。

次に、テロップ識別の段階で検出漏れや誤検出を起こす原因の大部分は、以下の2種類であった。一つの原因は、複数のテロップが近接して表示されている場合、近接しているテロップをまとめて一つのテロップである判定してしまうことである。どちらか一方のみが消失した場合、そのテロップは検出漏れとなってしまう。他の原因は、消失時にフェードアウトのような効果のかかるテロップに対して、消失判定が対応していないために検出漏れが生じるとい

うものである。

非圧縮映像からのテロップ検出でピクセル単位でエッジの判定を行うのに比べると、MPEGからの検出では8×8ピクセルのブロック単位での判定となるため、どうしても粗い精度でのテロップ検出になってしまう。ただし、応用によって、テロップ検出にどれほどの精度が必要であるかは異なる。たとえば音楽番組では、歌詞の切り替わりを検出することができなくても、テロップが表示されていることや大体の大きささえわかれば、歌唱シーンの検出は可能である。一方文献(3)の応用では、テロップをピクセル単位に近い精度で位置・大きさによって分類した上で、ニュースの冒頭テロップを選別する。そのような空間的な精度をMPEGからの検出のみで確保するのは、現状では困難であるといわざるを得ない。この問題の対策としては、たとえば、MPEGからの検出で範囲を限定した上で、画像の一部分のみをデコードし、エッジ判定などにより詳細なデータを抽出するといった方法でMPEGからの検出を活用することも考えられる。

6. まとめ

映像構造化で利用する映像の特徴として有用なテロップの検出を、標準的な圧縮符号化方式であるMPEG映像から行う方法を提案した。画像の高周波成分に注目することにより、従来の方法よりも高精度でテロップ検出を行うことができた。

デジタル映像の圧縮符号化形式には種類がいくつかあるが、時間的相関(フレーム間予測)や空間的相関(周波数変換)といったデジタル映像圧縮符号化の基礎的な考え方を利用した検出方法であれば、さまざまな形式の映像に対応することも可能である。

今後は、テロップ表示領域の形状判定の方法をさらに検討を進めるほか、同一位置での文字の切り替わり検出や、テロップ消失時の特殊効果への対応を行い、検出精度をより高めたい。

参 考 文 献

- (1) 佐藤, 他: “ MPEG 符号化映像からの高速テロップ領域検出法 ”, 電子情報通信学会, Vol. J81-D-II, pp.1847-1855, 1998
- (2) 加藤, 他: “ MPEG ビデオからのテロップ検出に関する一検討 ”, 情報処理学会, 研究報告・オーディオビジュアル複合情報処理, 32-2 pp7.-12, 2001
- (3) 宮里, 他: “ テロップを用いたニュース番組の自動ハイライト作成 ”, FIT2003 (第2回情報科学技術フォーラム), 一般講演論文集・第三分冊, pp.75-76, 2003
- (4) 倉橋: “ MPEG 符号化映像からのテロップ検出方法に関する一検討 ”, FIT2004 (第3回情報科学技術フォーラム), 一般講演論文集・第三分冊, pp.63-64, 2004

筆 者

倉 橋 誠 (くらはし まこと)

所属: 研究開発本部 総合研究所 情報メディア技術研究部

入社年月: 2003年4月

主な経歴: 総合研究所にてマルチメディア情報技術の研究業務に従事