I-029

MPEG 符号化映像からのテロップ検出方法に関する一検討
A study on detecting captions from MPEG videos

倉橋　誠†
Makoto KURAHASHI

１．はじめに

近年のネットワークを通じた動画配信やデジタル放送の本格化により、デジタル映像がより身近となった。また、ストレージの大容量化により、記録保存されるコンテンツ量も増大する。その視聴を支援するため、コンテンツを構造化し、整理や検索を効率化することが課題となっている。

映像コンテンツを自動的に構造化するためには、コンテンツの内容を認識することが必要となるが、その一手法として、映像コンテンツの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いことや、映像処理によりコンテンツを理解するよりも遙かに容易な処理で、テロップから映像の意味を抽出できるためである。

本稿では、このような状況で重要性が増大しているデジタル映像からのテロップ検出を、代表的な圧縮符号化方式である MPEG を対象として検討した。

２．関連研究

MPEG 形式の映像データ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。たとえば動き補償なしでフレーム間相関を利用して符号化された領域の集中領域をカウントし、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がある[1]。また、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評価を行う方法もある[2]。これらの方法により、従来は MPEG 映像をデコードした後にベースバンドより行っていたテロップ検出を、MPEG 映像から直接行うことが可能となっている。しかし、従来手法では、高い精度は期待できず、出現の仕方など様々な特徴をもつテロップへの対応が難しいといった問題点があった。

一方ベースバンド処理ではあるが、テロップ検出を利用したアプリケーションも発表されている。たとえば[3]は、ニュース番組中で表示されるテロップを検出し、その位置や大きさ等の特徴を利用して、それが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。

３．方法と実装

3.1 検出方法

空間領域におけるテロップの境界部分には、エッジ(急激な輝度や色差値の変化)が存在する。これは、視覚を利用した情報伝達手段であるテロップの見やすさに関わるため、多くのテロップに共通する性質である。特に、自然画像における典型的なエッジ部分に比べて、人工的に挿入したテロップの境界エッジは、より急峻なものである傾向が強いと考えられる。このような急なエッジを含む画像は、周波数領域では高周波成分に絶対値の大きな値が現れることが知られている。ここで提案するテロップ検出方法は、この性質を利用し、MPEG 映像のイントラ符号化フレームで符号化されている DCT 係数の高周波成分に、絶対値の大きな値が現れているかを評価することで、そのブロックがテロップ領域であるかの判定を行うものである。これに加え、インター符号化フレームでは、動き予測の状態から、そのマクロブロックがテロップらしいかの評価を行う。

本手法では、DCT 係数より急峻なエッジの有無を推定することにより、既存手法に対してテロップ領域検出の精度向上を図る。また、この判定を連続的に行うことで、出現の仕方や表示中の変動に対して強固なテロップ検出を目指す。

3.2 実装

評価対象に、30fps の映像を、1GOP 当たり 15 フレーム(うちイントラ符号化フレームが 1 フレーム)でエンコードした映像を想定し、DCT 係数の評価を、イントラ符号化フレームの周期で行うものとする。

イントラ符号化フレーム I_n を構成するブロック $b_{xy}(0 \leqq x \leqq Width/8、0 \leqq y \leqq Height/8)$ が、テロップの表示されているテロップ領域であるかを判定する。b_{xy} の 2 次元 DCT 係数 $\mathbf{c}$ を、周波数帯別に n 個の領域$\{\mathbf{a}_1, \mathbf{a}_2,\ldots,\mathbf{a}_n\}$に分割する。たとえば、次の行列 $\mathbf{A}$ を使用し、同じ数値が振られた位置をそれぞれ一つの領域とし、$\mathbf{a}_1,\ldots,\mathbf{a}_{11}$ の 11 の領域に分ける。

$$\mathbf{A}=\begin{pmatrix} 0, 1, 2, 3, 4, 5, 6, 7 \\ 1, 1, 2, 3, 4, 5, 6, 7 \\ 2, 2, 3, 4, 4, 5, 6, 7 \\ 3, 3, 4, 4, 5, 6, 7, 8 \\ 4, 4, 4, 5, 6, 6, 7, 8 \\ 5, 5, 5, 6, 6, 7, 8, 9 \\ 6, 6, 6, 7, 7, 8, 9, 9 \\ 7, 7, 7, 8, 8, 9, 9, 10 \end{pmatrix}$$

次に、領域毎の重み付け値のベクトル $\mathbf{w}$ を用意する。これは、係数の高周波成分への値の出現を評価するため、高周波の領域ほど高い重み付け値とする。

領域 $\mathbf{a}_f$ の中で絶対値が最大である要素を a_{fmax} とし、ブロック b_{xy} のテロップらしさ評価値 v を、次のように定義する。

$$v = \sum_{f=0}^{n} w_f \cdot \left| a_{fmax} \right|$$

そして、テロップを決定する閾値 Thr に対して v>Thr の時、b_{xy} がテロップ領域であると判定する。この結果、イントラ符号化フレーム I_n に対して該当領域がテロップであるか否かを表す二値の行列を出力する。

最終的に、インター符号化フレームで判定したテロップらしさの評価をカウントしたものと、イントラ符号化フレームの DCT 係数による判定結果を総合し、テロップ領域を判定する。

† パイオニア(株) 研究開発本部 総合研究所

４．評価

評価は、次の 2 通りの視点で行った。

第 1 は、テロップ領域の検出に成功したかを評価するものである。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** と、テロップ領域として検出したマクロブロックの集合 **X** を比較し、各方法のテロップ表示領域の検出精度を測定するものである。評価は再現率$|\mathbf{T}\cap\mathbf{X}|/|\mathbf{T}|$と適合率$|\mathbf{T}\cap\mathbf{X}|/|\mathbf{X}|$で表す。

第 2 は、ある時点で表示されているテロップの検出に成功したかを評価するものである。テロップが存在するフレームを **F**、そのうちテロップが表示されているマクロブロックのうち一定以上をテロップ領域として検出できたフレームを $\mathbf{F}_s$、テロップを含まないマクロブロックのうち一定以上をテロップ領域として誤検出したフレームを $\mathbf{F}_w$ とする。また、テロップの存在しないフレームで、テロップとして検出したマクロブロック数が一定以上であるものを $\mathbf{N}_w$ とする。評価は再現率$|\mathbf{F}_s|/|\mathbf{F}|$と適合率$|\mathbf{F}_s|/(|\mathbf{F}_s|+|\mathbf{F}_w|+|\mathbf{N}_w|)$で表す。

これらの評価を、以下の 4 つの方法で、閾値の条件を変えながら比較した。

(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
(C) DCT 係数重み付けによる判定
(D) DCT 係数重み付けなしでの判定

(A-2)と(B-2)は、(A)(B)に孤立点の除去など簡単な処理を加えたものである。第 1 及び第 2 の評価の結果をそれぞれ図 1 と図 2 に示す。

評価対象としたのは、ニュース番組と音楽番組の合計 2 時間分の動画データである。解像度 720×480、5Mbps で、MPEG2 方式でエンコードしたものである。

５．考察

第 1 の評価では、いずれの番組でも(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%台を得られた。音楽番組では、ニュース番組ほどの再現率と適合率の両立ができなかった。本手法はテロップに急峻なエッジが発生することを期待しているため、その条件に合致する見やすさ重視のテロップが比較的多いニュース番組と、そうでない音楽番組の差が結果に表れている。また、どちらの番組でも、(C)は(D)と比べて高い精度が得られたことから、高周波成分に注目することで精度向上が達成できたと言える。

第 2 の評価では、(A)～(D)いずれの方法でも、第 1 の評価結果を反映した結果となっている。

評価結果全般において(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パラメータ設定としたためである。目的に応じて再現率を重視する方向へのパラメータ調整も可能である。

検出結果を見ると、スタジオ背景等急なエッジのある領域を誤検出することがあった。このようなテロップ以外のエッジ領域は、標準的なテロップと比べてエッジの密度が低い場合が多いことから、検出密度の低い領域を取り除くことで、ある程度除外できると考えられる。実際、結果の(A-2)、(B-2)では、孤立点の除去により精度向上が実現している。しかし、過度に行うと大きなテロップの検出漏れが増加してしまう。文字の大きなテロップは、テロップ内にエッジの密度が低いブロックが存在するためである。今後は、このような問題点に対応するとともに、テロップを区別して検出する方法を検討したい。

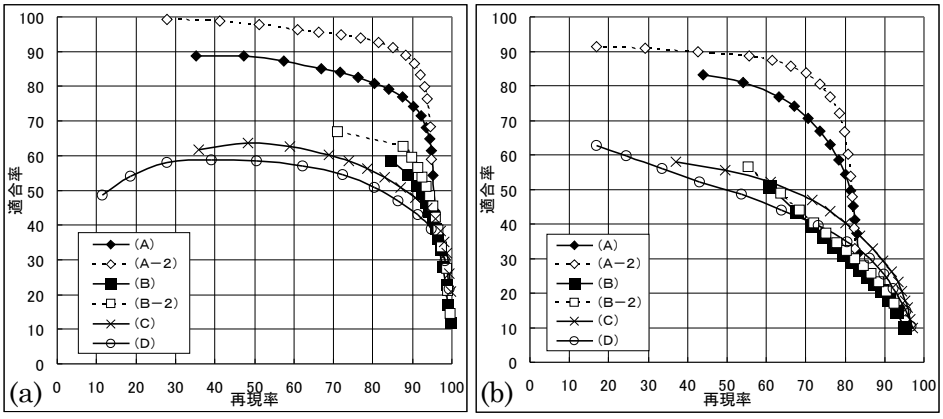

図 1 テロップ表示領域の検出精度の評価　(a) ニュース番組　(b) 音楽番組

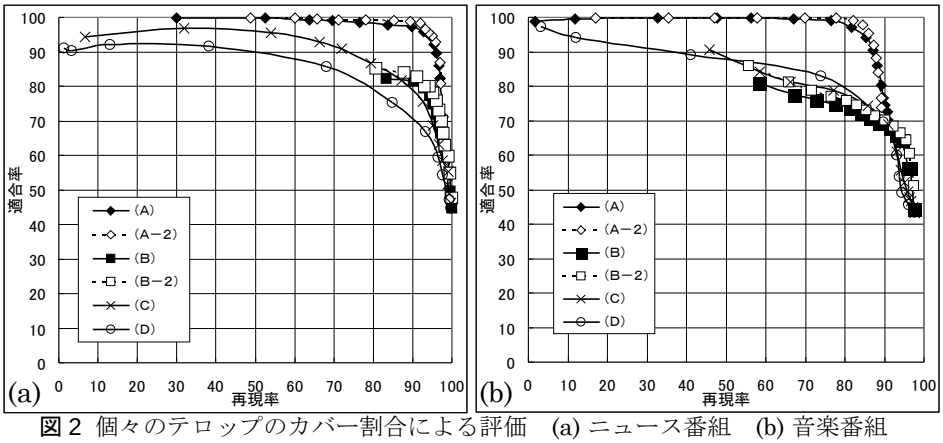

図 2 個々のテロップのカバー割合による評価　(a) ニュース番組　(b) 音楽番組

参考文献

[1] 佐藤ほか ”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] 加藤ほか “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] 宮里ほか “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, pp.75-76, 2003