オーディオビジュアル複合情報処理　35－6
(2001. 12. 13)

ニュース索引のための MPEG からのテロップ検出に関する検討

加藤 晴久　　柳原 広昌　　中島 康之

KDDI 研究所
〒356-8502 埼玉県上福岡市大原 2-1-15

E-mail: {hkato, yanap, nakajima}@kddlabs.jp

あらまし 映像検索の一助として，MPEG 圧縮コンテンツから高速にテロップを検出する方式について検討および提案を行ってきた．本研究では，より検索性に優れた索引を生成するため，テロップ検出の対象をニュース映像におけるニュース項目の文字列に限定し，インデックス生成のためにテロップの分類を行うことで効果的なテロップ検出する方式を検討した．

キーワード MPEG，テロップ検出，インデキシング

A Study on caption detection from MPEG coded data for news indexing

Haruhisa KATO, Hiromasa YANAGIHARA, and Yasuhiro NAKAJIMA

KDDI R&D Laboratories Inc.
2-1-15 Ohara Kamihukuoka-shi Saitama 356-8502, Japan

E-mail: {hkato, yanap, nakajima}@kddlabs.jp

Abstract We proposed the fast caption detect algorithm from MPEG coded data. This paper proposes the caption detection for news indexing. The captions are distinguished in order to create the effective news index. The caption geometric information and caption appearance time are used to determine how the caption is effective for index.

Key words MPEG, caption detection, indexing

1. はじめに

複数のコンテンツを含む映像データベースから特定のフレームを検索する場合には，膨大な手間と時間がかかる．このため，映像データベースに対しては効率的なインデックス生成が求められる．適切なインデックスはコンテンツの概略を把握したい場合にきわめて有効である[1]．

映像コンテンツのインデックス生成は従来から人の手作業によって行われてきた．しかし，この作業は非効率であることからインデックス生成の自動化が求められている．インデックス生成において最も重要なことは，抽出されたフレームだけでコンテンツ全体を把握できるように構築することである．そのため，映像データベースの検索キーとして，映像コンテンツを過不足なく表現できる特徴的なフレームを選択する必要がある．

一方で，多数の映像からなるデータベースを構築する際に，情報量の多い映像は圧縮して保存されている．計算機の高性能化や映像の入出力環境が整いつつあるいま，すでに大量の映像が蓄積されていると考えられる．これらの資産を有効に活用するためには，圧縮されているデータを取り扱う必要がある．

しかし，従来の特徴抽出アルゴリズムを自動インデックス生成の方式として利用する場合には，圧縮されているデータを復元し，もとの輝度レベルに戻さなければならない．これは復元した映像を展開するスペースが必要となる上，復元処理にかかるコストも無視できない．

これらの問題を解決する方法として，圧縮コンテンツから符号化されたデータを直接操作する方式を提案ならびに検討されている[2-5]．この方式は映像の特徴としてフレームにオーバーレイされる文字列（テロップ）を利用する．ただし，符号化データをもとにテロップを検出することを第一の目標に掲げているため，テロップを含むフレームは抽出されても，抽出されたフレームの集合はインデックスとしてコンテンツの概要を把握するのに有効か否かの判断までは行っていなかった．

本稿では，より検索性の優れたインデキシングを前提にして，テロップ検出の対象をニュース映像におけるニュース項目の文字列に限定し，さらに精度良く検出する方式を検討する．

2. 従来のテロップ検出方式と問題点

MPEG 符号化データを完全に復号せずに，部分的な復号または符号化データそのものを直接操作することでテロップを検出する方式が文献[2,3]で提案されている．これらの方式ではマクロブロックの符号化モードがテロップによって特定のモードになりやすいことを利用している．符号化モードの種類によってカウンタを増減させ，閾値を越えたマクロブロックをテロップ領域の候補と判断する．

テロップによって選択されやすい符号化モードはDCT 係数が符号化され，長さ 0 以上の動きベクトルを持たない(no MC coded)．しかし，この符号化モードはテロップ以外の静止した領域によっても選択されやすいため，誤検出の主な要因となっている．

抽出されたテロップ領域からはテロップの大きさ，テロップの重心位置，テロップ形状などの幾何的な情報を取得できる．しかし，出現しているテロップの変化については従来の検出方式では正確に検知できない．例えば，図 1 に示すように台詞やコメントなどの複数のテロップが時間的に連続して表示される場合では，マクロブロックタイプのカウントでテロップの存在は検出できてもテロップ文字列の変化を検知できない．テロップ文字列の長さが変化すれば検出も可能であるが，同程度の文字列長の場合は切り分けが難しい．

テロップが出現するフレームをインデックスとして利用する際には異なるテロップは全て抽出することが望ましい．ランダムアクセスポイントとしては連続して出現するテロップの最初に出現したフレームだけをインデックスに利用すればよいが，テロップが変化する全てのフレームからインデックスを構築することでコンテンツの全容を把握しやすくなる．また，切り分けが可能になれば個々のテロップが出現している時間を同定することができ，インデックスとしての有効性判定にも応用できる．

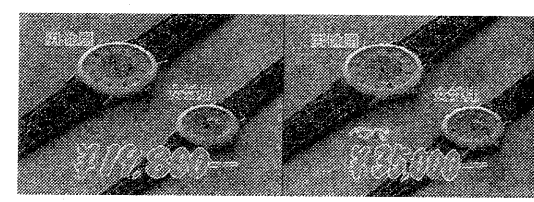

図 1 従来法で切り分けが困難な一例

3. インデックスとして効果的なテロップ

3.1 ニュース映像に表れるテロップの特徴

ニュース映像中に現れる文字列は，編集によってインポーズされた文字列のほかに，説明を記したクリップや撮影背景にある看板や標識など多岐にわたる．本稿では背景に存在する文字列は映像の主体的な内容とリンクしない場合が多いため検出の対象から外す．

テロップ内容の意味的な分類としてニュース項目，台詞や説明，人名や地名の 3 種類に分ける．ニュース映像の内容を把握するという前提でテロップを検出するとき，ニュース項目を示す文字列はニュース内容を簡潔に表し，このフレームを見るだけでニュースの内容を把握できる．さらにニュース項目は各ニュースの冒頭に現れ，このフレームにランダムアクセスすればそのままニュースを閲覧できるため，ニュース項目を最も重要な検出対象にする．

ニュース項目の形状として以下のような特徴が挙げられる．

- フレームに対してセンタリング
- テロップ幅が比較的長い
- 上下左右対称矩形
- 表示時間が比較的長い

ニュース項目の文字列が出現する仕方は一瞬で現れることは稀であり，ワイプ，ディゾルブ，フェードなど他の映像効果を伴っているものが多い．文字列がワイプしている途中のフレームを抽出してもインデックスとしては不完全であるので，文字列に施された映像効果が完了し，定常状態に至ったフレームを抽出する必要がある．

また，台詞や説明を表すテロップは文字量自体が最も多く，フリップによる解説やインタビューの台詞や外国語の日本語訳などはニュース内容を詳細に伝えるので，インデックスとしてふさわしい．

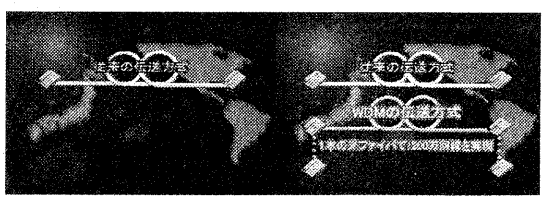

図 2 複数テロップの出現過程

台詞や説明の形状として以下のような特徴が挙げられる．

- フレームに対して左揃え
- テロップ幅が比較的長い
- 非対称形状が多い
- 表示時間が比較的短い

ただし，図 2 のように複数のテロップが一度に現れず少し間を置いてから全体が表示される場合は，一部のテロップだけが表示されているフレームをインデックスから省く必要がある．

上記特徴から，ニュース項目と台詞の相違は文字配置の揃え方と表示時間にあると考えられる．

最後に，撮影の対象となる人名や地名はシーンが切り替わった直後に現れることが多い．台詞や説明と比較するとニュース内容を示す情報量は多いとはいえない．人名や地名のテロップには以下のような特徴が挙げられる．

- フレームに対する配置は任意
- テロップ幅が比較的短い
- 非対称形状が多い
- 表示時間が比較的短い

さらに縦書きで現れることも多く，中段左右に出現するテロップのほとんどが人名や地名を示すものである．

3.2 ニュース映像に表れるテロップの配置

ニュース映像に表れるテロップについて出現位置と内容を目測で計測した．6 局の映像から 6 時間分のニュース映像を使用した結果，対象となるテロップ総数は 1636 個存在した．

ニュース映像で用いられるテロップの出現位置を調査した．表 1 にフレームを 9 分割した領域ごとのテロップ出現確率を求めた結果を示す．テロップの重心位置は下段中央が約半数を占め，最も多い．次に下段右が 10%以上の確率を占め，下段全体でテロップ出現確率の 7 割近くを占める．特にニュース項目に関しては，すべて下段に位置していた．

中段の左右の領域に出現するテロップに関しては縦書きの文字列で地名，人名や役職名を示すことが多い．中段中央に位置するテロップは中央だけで生じた例は少なく，他の領域にテロップを伴う場合が多い．

表 1 ニュース映像の位置毎のテロップ出現確率

	右	中央	左	合計
上段	6.23%	0.98%	9.54%	16.75%
中段	3.91%	5.62%	5.66%	15.19%
下段	7.21%	49.0%	11.85%	68.06%
合計	17.35%	55.60%	27.05%	100%

上段中央に表示されるテロップは出現回数が最も少ない．上段左右では時刻やニュース項目が小さく表示されていることが多く，出現している時間が比較的長い．

よって，本稿ではテロップの重心が下段中央に位置するニュース項目が最もインデックスとして相応しいと仮定し，台詞等のテロップと切り分けを実現することがインデックス生成に重要であると考える．

4. 提案するテロップ検出方式

4.1 テロップ候補の選定

テロップ検出方式については,従来方式と同様にP, B ピクチャのマクロブロックの符号化モードと参照フレーム間距離に応じて計数する．ただし，テロップの連続性を考慮して GOP 毎の符号化モードカウンタのリセットは行わない．代わりにテロップが消失したときでも即応性を持たせるため，GOP 毎に係数全てをある値で割る．I ピクチャまでに累積された計数からテロップ候補を判定する．計数のとりうる最大値の半分を閾値とし，カウンタが閾値を超えるマクロブロックをテロップ候補とする．

テロップ候補に対しては文字特性として AC 成分の密集度を判定する．文字による密集したエッジはブロックの全ての AC 成分に大きく影響する一方，マクロブロックタイプによる計数で抽出される領域には文字領域と静止領域が存在する．文字領域には複雑なエッジが多数存在することから，AC 成分の水平方向と垂直方向それぞれの絶対値部分和を求め，2 つの部分和がともに閾値を越えることをテロップ候補の条件とする．文字が一部でもかかっているマクロブロックはテロップ候補にするために，4 組の DCT 係数のうち一つでも条件を満たせばマクロブロック全体をテロップ候補とする．

抽出されたテロップ候補は孤立点を排除した後に同一領域を形成し，領域の幅と高さが閾値より短い領域を排除する．また，テロップ領域が領域内外で輝度に大きな違いがあることを利用して，テロップ候補領域の境界ラインにおける DC 差分の分散が閾値よりも小さいときはテロップ候補から外す．

4.2 テロップ領域の出現消失判定

過去の I ピクチャで検出された領域と現在の I ピクチャで検出された領域が同位置に存在する場合を考慮し，2 つのテロップが同一の文字列か判断を下す．連続表示されるテロップでは文字の色や大きさ，フォントは統一されており，領域全体では文字列の変化が捉えられない．たとえば，文字列を並び替えた場合を考えると，平均や分散などの基本統計量に変化は現れない．よって，フレーム間のテロップの変動を判定するためにはテロップ領域の個々のブロックについて変動を捉える必要がある．

しかし，ブロック単位では文字の背景の変動に敏感になる．領域全体から見ると小さな変動でも，ブロック単位の判定では局地的な誤差に大きく左右される．このため背景の変動が大きいと，同一のテロップを異なるテロップとして誤認識する可能性がある．

領域全体またはブロック単位のいずれの判定でもテロップの出現による変動やテロップ出現後の定常性を誤認識する恐れがある．そこで，個々のブロック変動とテロップ領域全体の変動を同時に判定するため，差分 2 乗和を利用する方法が考えられる．しかし大きな差分値が一つでも存在すると，この特徴量も大きく影響を受ける．

よって，代表値に対する検定のひとつである Wilcoxon の符号順位和検定を利用する．このノンパラメトリック検定はテロップ候補領域に生じる誤差の分布を仮定することなく推定できる．また，標本数が小さくて分布に正規性が保証されない場合でも利用できるので，小さなテロップ領域にも適用が可能である．

過去に検出されたテロップ領域に対して，連続する I ピクチャ間の AC 絶対値部分和が大きいブロックのフレーム間 DC 差分をもとめる．差分値が 0 でない個数を Nとし，差分の絶対値が小さいほうから順位を付け，差分値の正負ごとに順位の和を求める．和の小さなほうを検定統計量 Tとする．検定統計量 Z_0は式 1 で与えられ，Nが十分大きいときは正規分布に従う．

$$Z_0 = \frac{|T - N(N+1)/4|}{\sqrt{N(N+1)(2N+1)/24}} \qquad \text{(式 1)}$$

Nが小さいときは Wilcoxon の符号順位検定統計量分布を参照する．

有意水準αが有意確率 $P=\Pr\{|Z|>Z_0\}$ より大きいとき変化が生じたと判断する．2 群の代表値に差はないとした帰無仮説が棄却された場合は，過去のテロップ領域は消失し新たなテロップ領域が出現したと判断する．

帰無仮説が採択された場合，過去の I ピクチャに存在したテロップ文字列に変化は起きておらず，定常性が保たれていると判断する．このとき，現在の I ピクチャで検出された領域で過去の領域と重ならない部分は現在の I ピクチャで検出もれと判断し，現在のテロップ領域に加える．この結果，現在の I ピクチャのテロップ領域が過去のテロップ領域を包含する場合は，単一のテロップがまだ出現過程にあると判断し，過去のテロップ領域は削除し現在の領域を抽出する．

また現在の I ピクチャで，はじめて形成された領域については現在の領域に対して同様の検定を行い，テロップ出現に伴う変化が見られない場合は領域を削除する．何らかの変化が認められると判定された場合はテロップ領域を確定する．このとき，他に重なり合わないテロップ領域が存在する場合は，個々のテロップは出現が完了していてもフレーム全体としてはまだ出現過程にあると判断する．一連のテロップ出現フレームについて最後のテロップ出現フレームがインデックスとして有効であるので，出現過程のフレームは削除し最後の出現フレームだけをインデックスとして選択する．

4.3 テロップのインデックス有効性判定

検出されたテロップはインデックスとしての有効性を判定する．ニュースコンテンツのインデックスとして最適なフレームはニュース項目を示す文字列を含むフレームとしたため，その特徴を備えたテロップを優先的に選択する．

テロップが人名や地名を表すものであるか否かについては領域の大きさで判定が可能である．領域の大きさと矩形度を調べるため，ラインに対する占有率 r_s を式 2 のように定義する．

$$r_s = \prod_{i=1}^{t_h} \frac{c_i}{m} \qquad \text{(式 2)}$$

ここで，t_h はテロップ領域の短辺方向の長さを表す．領域を構成するマクロブロック数をラインごとに求め，c_i $(1 \leqq i \leqq t_h)$とする．1 ラインのマクロブロック数 m に占めるテロップ領域の割合の総積を求める．領域が大きいもの，矩形に近いものほど比率が大きくなる．

次にテロップの位置に対する評価として，テロップ形状に対して外接長方形の長辺方向の重心位置を求め，フレーム幅の中心線までの距離の比を利用する．比率 r_c を式 3 から求め，1 に近いほど有効であると判断する．

$$r_c = 1 - \left| \frac{b_x}{w/2} - 1 \right| \qquad \text{(式 3)}$$

ここで，w と b_x はそれぞれ画素単位でのフレーム幅と外接長方形の長辺方向の重心座標を示す．

テロップ形状の対称性はテロップ領域の重心位置と外接長方形の中心との距離の比を利用する．外接長方形の中心 b_x と重心位置 t_x の比率 r_b を式 4 から水平垂直それぞれに求め，1 に近いほど有効であると判断する．

$$r_b = 1 - \frac{|t_x - b_x|}{b_w/4} \qquad \text{(式 4)}$$

ここで，b_w は外接長方形の幅または高さを示す．

テロップの出現時間については，一定時間までは有効性への寄与率が高いが出現時間が長すぎるテロップは情報量が乏しい．しかし，長時間出現するテロップはフレームの 4 隅に小さく現れることが多い．よって，上記の 3 判定で順位付けが困難な領域についてのみ出現時間による判定を行い，出現時間が長いほど有効であると判断する．

5. 実験結果

評価実験として，1.2Mbps の MPEG-1 ビデオ形式で圧縮された 30 分のニュース映像に対して検出を行った．効率的なインデックス生成のため，画面の下 3 分の 1 のみを検出範囲に指定した．テロップ総数 120 個のうち検出範囲に現れたテロップ数は 73 個で，内訳はニュース項目が 5 個，台詞および説明が 43 個である．

テロップ検出によって抽出された 75 個に対して，過剰検出は 14 個，検出もれは 12 個であった．テロップの正検出率は 83.6%になる．このときニュース項目は全て検出できており，台詞および説明のテロップのうち 36 個が検出できた．

次に，検出された 75 個のテロップについてインデ

ックスとしての有効性を判定し，ニュース項目のテロップを抽出した．

図3はテロップ領域のラインに占める占有率 r_s を示す．ニュース項目が出現したシーンは矢印で表している．上位6フレームには5つのニュース項目すべてと台詞のフレームが1フレーム含まれていた．

テロップの分類においてテロップ領域の大きさだけで判断した場合は，ニュース項目よりも複数行になった台詞のテロップほうが大きい場合がある．実際に上位6フレームの中でテロップ領域が最も大きいのは台詞のテロップであった．さらに台詞のテロップ領域が対称形をしていたため，形状情報のみでは図4の台詞のテロップが最もニュース項目の特徴を満たしていると誤認した．しかし，ニュース項目と台詞の最も異なる点は表示時間であるため，正確な表示時間で順位付けを行うことで，ニュース項目と台詞を識別することが可能であった．特に連続表示される台詞は一つ一つの表示時間がニュース項目に比べて短いため，テロップの表示されている正確な時間が有効性判断に効果的である．

占有率 r_s が小さなものに対しては台詞のテロップがフレームに対して左揃えであることが多く，重心位置と形状の対称性で，台詞と名称のテロップを識別することができる．形状の対称性比率 r_b が小さいものほど台詞のテロップとして抽出する．ただし，テロップ領域が1行からなる領域に対しては対称性比率が意味をなさないことから，対称性より先にフレームに対する位置判定を優先する必要がある．

また，インデックス有効性判断により過剰検出した場合でもニュース項目の特徴に沿わなければ除外できるので，検出もれを起こさないように閾値を設定すると全体のテロップ正検出率は向上すると考えられる．

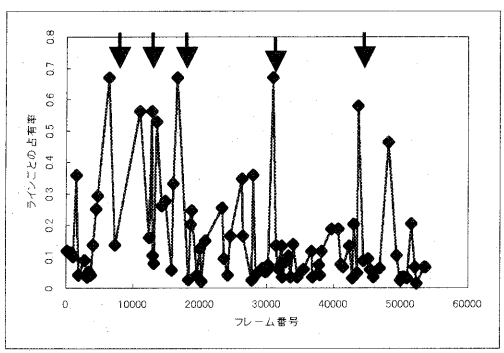

図 3 フレーム幅に対するテロップ領域の占有率

6. まとめ

MPEG符号化データからのテロップ検出において抽出されたテロップのインデックスとして有効性を判定する方式を検討した．テロップのインデックスとしての有効性をテロップ領域の形状情報と表示時間に求め，ニュース項目を示すテロップが持つ特徴との類似性を有効性判定に利用した．インデックスの構成に際しテロップ出現だけでなくテロップの消失を検出し，正確なテロップの表示時間を得る方式を提案した．インデックスの有効性を判定し上位から提示することで，利用者の要求に応じた精度のインデックス生成が可能となる．

文　献

[1] 佐藤隆，新倉康臣，谷口行信，阿久津明人，外村佳信，浜田洋，“MPEG 符号化映像からの高速テロップ領域検出法，” 信学論(D-II), vol. J81-D-II, no.8, pp.1847-1855, 1998.

[2] Zhong, Y, Zhang, H, Jain, A.K., “Automatic Caption Localization in Compressed Video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no.4, pp.385-392, April 2000.

[3] Kim, K.I., Jung, K., Park, S.H., Kim, H.J., “Support vector machine-based text detection in digital video,” *Pattern Recognition*, vol.34, No.2, pp.727-529, Feb. 2001.

[4] 加藤晴久，中島康之，柳原広昌，” MPEG ビデオからのテロップ検出に関する一検討，” 情処研報，vol.2001, No.19, pp.7-12, 2001.

[5] Shigenobu Aoki, “ウィルコクソンの符号順位検定，” *http://aoki2.si.gunma-u.ac.jp/lecture/Average/mpsr-test.html*