オーディオビジュアル複合情報処理 32－2
（2001．3．2）

MPEGビデオからのテロップ検出に関する一検討

加藤晴久　　中島康之　　柳原広昌

KDD 研究所
〒356-8502 埼玉県上福岡市大原 2-1-15
{hkato, nakajima, yanap}@spg.kddlabs.co.jp

本稿では圧縮された映像データに対して符号化パラメータを直接操作することにより，テロップ出現フレームと位置情報を抽出する方式を提案，検討した．フレーム間差分による検出処理とマクロブロック符号化タイプによる検出処理はともに計算処理が軽く，且つ，互いに補い合うことができる．2 つの異なる処理の組み合わせにより，従来の符号化タイプの検出だけでは困難だった静止映像からのテロップ検出精度が向上することを確認した．実験では MPEG-1 で符号化された映像を用い，85%のテロップを正しく検出できた．また、検出処理時間は復号処理時間の約 5 倍の速さで完了した．

A Fast Caption Detection from MPEG Coded Data

Haruhisa Kato, Yasuyuki Nakajima, Hiromasa Yanagihara

KDD R&D Laboratories Inc.
2-1-15 Ohara Kamifukuoka Saitama 356-8502, JAPAN
{hkato, nakajima, yanap}@spg.kddlabs.co.jp

This paper proposes a fast caption detection algorithm from MPEG coded data. Hierarchical caption detection is performed in order to achieve a fast and accurate detection. In the proposed algorithm, caption candidates are firstly selected using inter I-picture difference. Then coding modes and motion vector information of P- and B-pictures are used to determine spatial and temporal positions of caption appearance. In the experiment using MPEG-1 sequences, it has been shown that 85% of caption can be correctly detected. Furthermore, the detection speed is less than a quarter of decoding process.

1. はじめに

複数の映像コンテンツから目的の映像を参照するには適切な索引が必要である．索引はコンテンツの概要を簡単に掴めるように，特徴的な静止画フレームから構成される．従来の索引作りは人手を介して行われているため，巨大な映像データベースの構築が進むにつれて索引作りが困難となってきた．この問題を解決する方法として，キーフレームの抽出を自動化する技術への要望が高まっている．

また検索に限らず，映像データベースの 2 次利用を目的とした知的構造化の研究が始まっている．コンテンツの構造化とはコンテンツの要約や構成を把握することであり，映像データベースの階層的な構築，分類の一助となる．その中で，コンテンツ内容理解として映像中に現れる文字情報の解析技術が注目されている．映像にオーバーレイされる文字情報は内容の注釈，強調，補足などの付加情報を表す．この映像内容とリンクした文字情報によって，映像コンテンツの概略を的確かつ簡潔に理解することができる[1]-[5]．

一方で，情報量の膨大な映像の蓄積には圧縮処理が不可欠である．そのため国際標準の映像圧縮技術が広く普及し，今日の膨大な映像コンテンツの利用を促進させている．しかし，圧縮された映像データは伝送，蓄積用途には優れている一方，先に述べた映像の検出および要約理解などの 2 次利用には取扱いが煩雑になる．つまり，圧縮された映像の解析には復号過程が必要であり，検出や解析の処理以外にも時間的，計算量的に大きなコストがかかる．これを解決する方法として，圧縮符号化されたデータを直接操作することで復号過程を経ずに映像を解析する方式が提案されている．

本稿では前記の要求を満たす一方式として，圧縮された映像データの符号化パラメータを直接操作することで，キーフレームとしてのテロップ出現フレームを高速に検出し，そのフレーム内における位置と形状情報を抽出する方式を提案する．

2. 従来のテロップ検出方式

従来の圧縮符号化情報を利用したテロップ検出法として，文字列と背景との明確なエッジ部を構成する領域をテロップと仮定した方式が報告されている[6]．この方式は MPEG-1 を対象とし，I ピクチャの DCT 係数情報のみを利用している．P ピクチャや B ピクチャの情報を一切利用していないことから，I ピクチャの間隔が長くなるにつれて時間的な検出解像度が低くなる．また，一般に I ピクチャよりもフレーム枚数が多い P，B ピクチャの情報を利用することで，検出率向上に改善の余地が残っている．

P，B ピクチャの符号化情報を利用した方式としては，MPEG-2 を対象としたテロップ検出法が報告されている[7]．この方式はマクロブロックの符号化モードに応じて計数を行う．計数カウンタが閾値をこえる領域を形状判断することで，テロップ領域を抽出する．この方式はテロップ領域外にランダムな動きベクトルが多数存在することを仮定している．低解像度の MPEG-1 映像などでは動きベクトルが相対的に小さくなるため，MPEG-2 を対象としたときほど高い検出率は望めない．特に，カメラが固定されているニュース映像は背景に動きベクトルが存在しないため，テロップ以外の領域を過剰検出することが多くなる．

3. 検出対象とするテロップ

テロップとは映像中に上書き挿入される映像編集の一方式である．本来は映像に挿入されるものすべてをテロップと呼ぶが，本稿では映像に挿入された静止文字列をテロップとして検出する．スクロールし続ける字幕等は今後の検出対象とした．また，テロップ文字情報を検索キーとして利用することを想定して，時刻表示等の小さなテロップは検出対象外とした．

4. 2 段階テロップ検出方式

検出処理速度の向上のために，検出処理を段階的に適用する．初めは検出処理の時間間隔を粗く設定し検出判定を行う．この判定でテロップの候補となった領域についてのみ，時間間隔をフレーム単位まで細かく設定し詳細な検出処理を行う．

本稿では映像の圧縮符号化方式として MPEG を利用し，時間解像度の粗い判定は複数の I ピクチャの情報をもとにテロップ検出を行う．I ピクチャによる検出でテロップの出現が確認された場合，フレーム単位での判定を P および B ピクチャの情報を用いてテロップ出現を確定しテロップの位置を抽出する．以上の 2 段階の検出方式からフレーム内でのテロップ位置を抽出するとともに，1 フレーム精度でテロップ出現を検出する．

4.1. Iピクチャによる検出判定

テロップは比較的短時間で出現し，その後数秒間持続して表示される．しかし，出現開始から定常状態に入るまでに数フレームを要するテロップの出現を検出するためには，離散的にフレームを比較し変動を捕らえる必要がある．離散的に検出判定を下すため，フレーム内符号化されているIピクチャを利用する．

Iピクチャ単位での検出はテロップ出現に伴う変化を調べ，その後の位置に対する定常性を判定する．図1のようにIピクチャ I_n にテロップが出現するとき，I_n に対して過去のIピクチャ I_{n-1} とフレーム全体の比較によって変化が認められた領域を注目し，その領域についてのみ複数の未来のIピクチャ I_{n+k} において変化が収まるか否かを検討する．

テロップ文字列は一般に高輝度の色，特に白色がよく用いられており，背景画像と見分けがつきやすいよう配慮されている．この特徴はDCT係数DC成分が高い値を持つことに対応するので，テロップ候補を選定するのに利用できる．

一方DCT係数AC成分については，ACの多寡は文字と背景が織り成すエッジ領域の有無に対応するが，符号化単位が8x8画素のブロックに分割されているため，テロップ文字がブロックの境界に位置する場合はAC成分にエッジの存在を示す値は現れない．また，文字とブロックの相対的な大きさにも依存し，文字サイズが大きくなるにつれてAC成分の1ブロック当たりに占める割合は減少する．文字サイズに依存することは同時に空間解像度に依存し，空間解像度が高いほどブロックに対して文字が大きくなるためAC成分による判定は困難となる．一方で計算量的については，全体的に計算負荷が小さい直接符号化操作の中で，AC絶対値部分和計算が占める割合は比較的大きい．よって，本方式はAC成分を利用せず，計算負荷の削減を図った．

テロップ出現に伴う変化の発生判定には，DCT係数のDC成分判定基準として用いた．大まかな変化を捉えるため，粗い階級値の輝度ヒストグラムから判断する．テロップは画面に対して水平または垂直に現れると仮定して，ヒストグラムを縦横のマクロブロック1ライン毎に求める．次に I_n と I_{n-1} のヒストグラムの差分絶対和が閾値を超えるラインをテロップ領域候補とする．但し，テロップ候補がフレームの大部分を占める場合はシーンチェンジやディゾルブなど，テロップ以外の要因による輝度変化として対象フレーム I_n での検出処理を終了する．さらに候補となったラインに対して，ブロックごとにDCのフレーム間差分を求め，閾値を超えるブロックをテロップ候補とする．

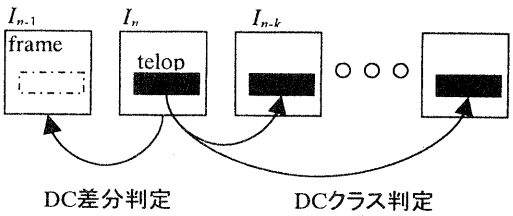


図1 Iピクチャ判定

次にフラッシュなどによる一瞬の変動をテロップと誤検出しないため，I_{n+k} 以降のIピクチャに対し，テロップ候補の定常状態を判定する．定常状態は最低でも数秒間にわたると仮定する．変化の収束判定は文字領域がマクロブロックの一部にのみ存在するときでも収束を見極める必要がある．特にテロップ自体は静止していても，背後で物体の動きやカメラワークによる変動が生じる可能性があり，単純な過去との差分では判断を誤りやすい．また，テロップは撮影後の映像にオーバーレイされるので光の散乱などの外乱に影響を受けず，テロップの輝度はまったく変化しない．よって，テロップ領域同士の時空間的な近傍での相関が変化しないことを利用する．

時空間的相関の同一性にはDC成分の輝度を利用したクラス分類を用いる．クラス形成にはマクロブロックの4つのDC成分に対して，それぞれが閾値を超えるか否か判定する．4つのDCに対する判定結果からマクロブロックを16種類のクラスに分類する．ただし，全てのDC成分が閾値を下回る場合は低輝度領域と判断し，テロップ候補から除外する．残る15種類のクラスを使って，複数の未来のIピクチャについてクラスが全て一致するとき，テロップ出現に伴う変化が収束したと判断する．すなわち，I_n のクラスが I_{n+k} のクラスと一致するマクロブロックをテロップ候補とし，一致しないマクロブロックはテロップ候補から除外する．このテロップ候補はDC差分判定でマイナス方向変化が生じテロップが消失したと判

断されるか，もしくはシーンチェンジなどにより DC 差分判定で全画面に変化が生じたと判断されるときまで保持する．

4.2. P, B ピクチャによる検出判定

I ピクチャの検出で抽出されたテロップ領域候補に限定して，P，B ピクチャの情報を利用したフレーム単位の検出を行う．各フレームに対してマクロブロックの符号化モードの分布からテロップの位置を抽出し，次に動き予測情報の時間的参照方向から出現したフレームを特定する．

P，B ピクチャ特有の情報としてマクロブロックの符号化モード情報と動き予測情報がある．一方で，定常状態に入ったテロップは完全に静止すると仮定しているため，静止テロップを含むブロックは動きベクトルを持つことはない．さらに文字のテクスチャは誤った動き予測を生じさせにくく，符号化モードでは動きベクトル情報を持たない no MC coded，Skip，Intra のいずれかに対応すると考えられる．よって，I ピクチャ判定で収束が始まる区間に対して，符号化モードの分布をテロップ判定に利用する．しかし，動きベクトルは時間的参照距離が短いと見かけ上の動きが小さく，動領域にも動きベクトルが与えられないことがある．逆に時間的参照距離が遠くなるにつれて静止領域にもランダムな動きベクトルが割り振られやすくなる．エンコーダによっても符号化モードの選択結果が異なるので，動き予測情報の時間的参照距離を符号化モードに対する信頼性情報として利用する．no MC coded, Skip, Intra であれば参照フレームまでの時間的距離に比例した数を係数カウンタに加算し，そうでなければ時間的距離に反比例した数をカウンタから減算する．I ピクチャごとに累計を取りまとめ，閾値以上の値を持つ領域をテロップ候補とする．同時にカウンタをリセットする．

次に，テロップ出現フレームの検出は動き予測情報の時間的参照方向からフレーム単位で判定する．まず I ピクチャ判定で変化が起ったと判断された GOP に対して，連続する B ピクチャ（以下，B ピクチャ群という）のテロップ領域とテロップの出現フレームには図 2 に示すような関係が成り立つと仮定する．

- 出現フレームが B ピクチャのときは，B ピクチャ群に順方向だけでなく逆方向動きベクトルも存在する(図 2 上段)

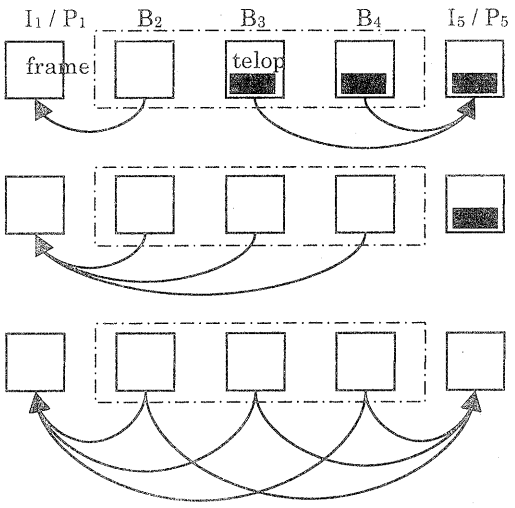


図 2　テロップ出現フレームと動き予測の関係

- 出現フレームが B ピクチャのときは，B ピクチャ群に順逆方向の切り替わりが一度だけ存在する(図 2 上段)
- 出現フレームが I，または P ピクチャのときは，B ピクチャ群に順方向動きベクトルのみ存在する(図 2 中段)
- B ピクチャ群に両方向予測が存在する場合はテロップが出現していない(図 2 下段)

I ピクチャ判定でテロップの出現が検知された場合，該当する GOP 内部の P，B ピクチャについて動きベクトルの時間的参照方向を調べる．テロップの出現フレームは，複数の連続した B ピクチャの中で上記の仮定を満たすフレームとする．

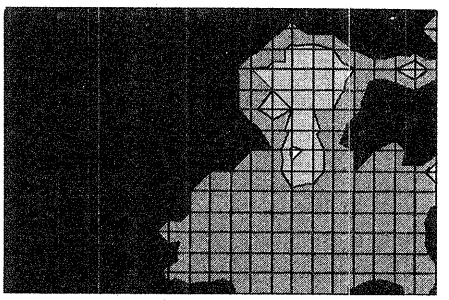
但し静止したテロップを仮定しているので，動きベクトルの長さはほぼ 0 となる．更に，テロップ出現後の GOP に対しても，同位置のマクロブロック毎に，動きベクトルの有無を確認する．明確な長さを持つ動きベクトルを持つマクロブロックは，テロップ候補から除外する．

最後にテロップを検出する動機を考慮し，天気予報の気温情報や時刻表示などのテロップを取り除くため，領域が小さい候補を排除し，テロップ検出を終了する．

5. シミュレーション結果

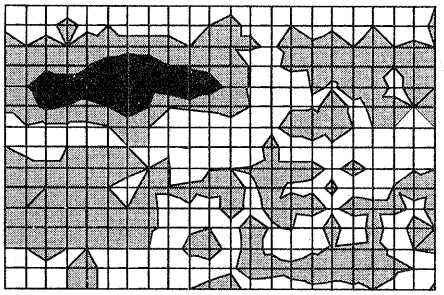
映像の圧縮方式は MPEG-1 Video を用い，合計 1 時間のニュース映像を対象とした．実験に使用した映像（図 3）は Canopus 社製 MPEG-1 リアルタイムエンコーダで変換され

図 3 原画像



□ −20−−10 ▒ −10−0 ▒ 0−10 ■ 10−20

図 4 符号化モードカウンタ



■ −1000−−500 ▒ −500−0 □ 0−500

図 5 DC 差分結果

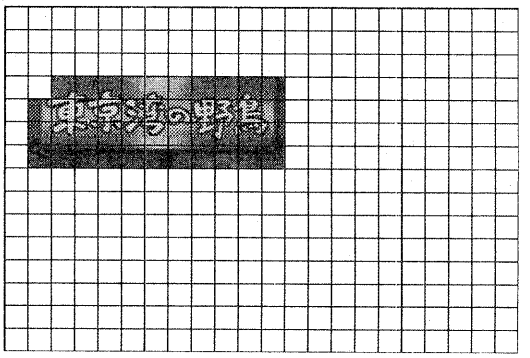


図 6 テロップ位置検出結果

た 2.0Mbps の MPEG-1Video ストリームである(SIF 形式 352x240,M=3,N=15). 空間的な検出解像度は動きベクトル情報を持つマクロブロック単位である.

図 3 はフレームの左上部にテロップが現れた原画像である. 図 4 はテロップが出現したのちの 1GOP の符号化モードを集計した結果を示している. 重み付けは参照フレームまでのフレーム枚数そのものを用いた. 色の濃い領域ほど動きベクトルが存在しないことを表し, 静止したテロップである可能性が高い. しかし, テロップの位置に相当する領域を含みながらテロップ以外の領域も同時に存在し, この情報のみではテロップの位置判定は困難である. 図 5 は 2 つの I ピクチャによる DC 差分結果を示している. I ピクチャ判定はある I ピクチャに対して, 過去 1 枚, 未来 2 枚の I ピクチャを利用した. 変動があった領域のうち定常状態に収まった領域において出現したテロップを捕らえることができた. 図 4 と図 5 の結果を重ねあわせ, 整形した最終結果が図 6 である.

表 1 テロップ検出結果

正検出率	未検出率	誤検出率	検出速度
75(84.3%)	14(15.7%)	8(9.64%)	797.7fps

一方, カメラワークが存在する他のシーンではまったく逆の結果が得られた. すなわち I ピクチャ判定では DC 差分判定がフレーム全体を捕らえてしまうが, DC クラス判定と P, B ピクチャ判定でテロップを正しく捕らえていた. よって, 動きのあるシーンでも静止したシーンにおいても 2 つの検出方式のいずれかがうまく機能し, テロップを正確に抽出することができた.

テロップ検出結果を表 1 に示す. 映像に出現したテロップは計 89 個である. 検出したテロップ数 83 個のうち、正しいものは 75 個であった. フレーム内のテロップの位置と形状に関しては、図 6 のようにほぼ正確に抽出することができた. 一方, 検出にかかる計算時間は, MPEG-1 の復号処理が平均 162.4fps であるのに対し, 検出処理は平均 797.7fps で, およそ

1/5 の時間で検出できた．テロップ出現フレームについて検出率の評価には次式を用いて算出した．

正検出率＝(正検出数)/(実テロップ数)x100

未検出率＝(未検出数)/(実テロップ数)x100

誤検出率＝(誤検出数)/(正検出数+誤検出数)x100

6. 考察

一度出現したテロップはカメラワークや映像効果の影響を受けず，実時間にして数秒間は完全に静止している．I ピクチャは一般に 12～15 フレーム間隔で配置されているため，テロップ領域以外では I ピクチャ間の相違が如実に現れた結果となった．また DC クラス判定は緩やかな判定ではあるが，テロップの定常性を捉えながら背景の変動に対してロバスト性を持つ．さらに DC クラス判定を複数枚の未来フレームに適用することで検出精度を高めることが可能である．

また，参照フレームとの時間的距離が近いほど物体の動きは小さくなるため，no MC coded の出現確率が高いことを符号化タイプの分布から確認した．時間的距離による信頼性情報を加味した符号化タイプのカウント方式は，no MC coded の不要なカウントアップを抑え，テロップを内包するブロックを的確に選択する．一方で他のエンコーダではフレームごとの符号化モード分布に特徴的な差は現れず，重みを均一にした場合と同じ結果が得られた．よって，重み付けカウント方式はエンコーダが正確な動き予測を行わないほど効果を発揮すると考えられる．

I ピクチャ間のフレーム間差分による検出処理とマクロブロック符号化タイプによる検出処理はともに計算処理負荷が軽く，且つ，互いに補い合うことができる．フレーム間差分処理の組み合わせにより，符号化モードの検出だけでは困難だった静止映像からのテロップ検出精度が大幅に向上した．

一方で，過剰検出したフレームの特徴として挙げられるのは，カメラワークや映像効果であった．特にシーンチェンジとワイプに反応した過剰検出が大半を占めた．ワイプの後のシーンが静止している場合は I ピクチャ判定の条件を満たすことが要因と考えられる．未検出に関しては，I ピクチャに出現途中のテロップが現れるとき，DC 差分判定で補足できないことがあった．また、シーンチェンジとともにテロップが出現している場面で未検出が多く見受けられた．

検出に失敗した要因の多くはシーンチェンジに関わる場面であることから，今後は他の映像効果とテロップ出現の差別化を検出する方法を検討することが求められる．

7. おわりに

圧縮符号化データ形式の映像コンテンツに対し，符号化データ上での直接処理操作により，高速にテロップを検出する方式を提案した．本方式は符号化データ上での時間的推移を把握し，テロップの出現を検出するとともに符号化方式の分布からテロップの位置及び形状情報を抽出した．また，本方式の検出精度について検討を行った．

参考文献

[1] 中島康之, 堀裕修, 塩原敏充, "キーワード画像抽出による動画像サマリの作成," 情処全大, F-10-2, pp.91-92, 1994.

[2] Anil Jain and Kalle Karu, "Learning Texture Discrimination Masks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.18, No.2, 1996.

[3] S.W.Lee, D.J.Lee, H.S.Park, "A New Methodology for Grayscale Character Segmentation and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.18, No.10, pp.1045 - 1050, 1996.

[4] Yu Zhong, Kalle Karu, Anil K.Jain, "Locating Text in Complex Color Images," *Pattern Recognition*, Vol.28, No.10, pp.1523 - 1535, 1995.

[5] K.Y.Jeong, K. Jung, E.Y.Kim, H.J.Kim, "Neural Network-based Text Location for News Video Indexing," *IEEE International Conference on Image Processing*, Vol.3, pp.319 - 323, 1999.

[6] Yu Zhong, Hongjiang Zhang, and Anil Jain, "Automatic Caption Localization in Compressed Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22, No.4, pp.385 - 392. 2000.

[7] 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.