

MPEG-2映像からのニュース番組高速構造化

青木 恒[†] 児玉 知也[†] 岩田 達明[†] 山口 昇[†]

[†]株式会社 東芝 研究開発センター マルチメディアラボラトリー

〒212-8582 神奈川県川崎市幸区小向東芝町1

E-mail: †{hisashi.aoki,tomoya.kodama,tatsuaki.iwata,noboru1.yamaguchi}@toshiba.co.jp

あらまし 本技術報告では、類似ショットの繰り返し登場を利用してニュース番組を話題ごとに整理する手法について提案し議論する。著者らはデジタル放送、デジタルビデオなどで用いられている MPEG-2 Transport Stream 形式の映像音声のうち、Video Elementary Stream の I-picture 中に含まれるブロックの直流成分のみを利用することにより、フルデコードよりも5倍程度高速に処理が完了することを示す。また本手法では話題分割に際しては類似ショットの登場傾向より対話部を推定することにより、本来、話題の先頭でないショットの除去が可能であることを実験を通して示す。

キーワード ニュース、類似ショット、MPEG-2

High-Speed Topic Organizing of News Programs in MPEG-2 Format

Hisashi AOKI[†], Tomoya KODAMA[†], Tatsuaki IWATA[†], and Noboru YAMAGUCHI[†]

[†] Multimedia Laboratory, Corporate R&D Center, Toshiba Corporation

1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki-shi, Kanagawa 212-8582 Japan

E-mail: †{hisashi.aoki,tomoya.kodama,tatsuaki.iwata,noboru1.yamaguchi}@toshiba.co.jp

Abstract The authors propose and discuss a topic organizing system that structurizes news shows on TV. The system handles audio-visual data in MPEG-2 Transport Stream format. In this system, DC components of blocks in I-pictures are used to analyze repetitious appearance of similar shots. The analysis is done approx. 5 times faster than decoding every pixel in every frame. Also, the authors introduce an algorithm to detect dialog scenes. The experiments show that the algorithm effectively eliminates wrong top shots for news topics.

Key words news, similar shot, MPEG-2

1. はじめに

近年、家庭においてもテレビ放送番組をデジタルデータとして録画できる装置が急速に普及している。ランダムアクセス可能な映像データは、ビデオテープの時代には実現が困難だった様々な新機能の提供をもたらしている。

その一方で、コンテンツ(映像内容)に踏み込んで提供される機能という視点で見ると、EPG(電子番組表)を用いた録画予約や録画済番組のラベリング、CM検出といったものに限られている。「観たい番組の観たい場面を探し出す」ために、テープ時代と同様に早送り再生を行わなくてはならない中において、映像解析応用への期待は低くない。

映像認識、音声認識を利用したテレビ番組データの構造化は、これまでも様々な手法が研究、提案されてきたことは言うまでもない。そうした高度なメディア処理を伴う方法の多くは、コンテンツ中で表現されているものが「何であるか」をテキス

トでラベリングする取り組みであり、人間が自然と体得している認識アルゴリズムへの挑戦ともいえる。それだけに、より人間の理解に近いラベリングが期待できる一方で、複雑で計算量の多い処理を必要とする傾向にある。

著者らは、より低計算量で高速に動作する映像構造化システムを提供することによって、装置の処理能力の高低に関わらず映像へのアクセシビリティを一層向上させることが可能であると考えている。本報告では、上記のようなデジタル録画機器の映像形式として主流である MPEG-2 映像を対象とし、ニュース番組を一例とした高速構造化手法の提案と議論を行う。

2. 背景

番組コンテンツに対して認識処理を施すことによって内容理解を行い、構造化に供する取り組みとしては、音声認識を用いるもの [1] や、映像字幕認識を用いるもの [2]、映像からの人物認識を行うもの [3] や、これらの複合利用 [4] などがある。

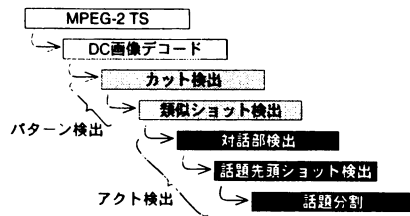


図1 処理の流れ
Fig. 1 Process flow

しかしながら前章で議論したように、これらは知的なラベリングや構造化が期待できる反面、計算量が多くなる傾向にあるため、処理装置に望まれる仕様が高くなったり、低性能の装置においては長い処理時間を要したりする恐れがあるなど、家庭などで普通に使用されるまでにはなお解決すべき問題が多い。

一方、現在のデジタル・ビデオ・レコーダにも EPG から情報を得て、録画済みの番組にタイトルや出演者などの情報を関連させて記録するものがある。また、クローズドキャプションなどテキスト情報として放送と同時に供給されるデータを用いて構造化を試みた例もある [5]。

しかし現在提供されている EPG は番組単位であり、コンテンツの構成に即した情報を得ることはできない。また、クローズドキャプション放送や字幕放送、データ放送による発話テキストの提供については、すべての番組に付与されているわけではない。生放送では発話と表示のタイミングのずれがあるため同期が必要、話者識別が保証されない、テキストベースの知識処理が必要、など、簡便かつ効果的に利用できる環境は限定されると言わざるを得ない。

著者らは、MPEG-2 で圧縮された映像の動き予測ベクトルがカット点において特徴的な挙動となることを利用して高速にカット検出を行い、一方で MPEG-2 映像から I-picture の各ブロックにおける直流成分だけを取り出した小画像 (以下、簡単に「DC 画像」と称する) を用いて類似ショットを検出し、ドラマや映画の構造化を行った [6]~[8]。

この手法は、組織化された対話シーンが存在するドラマや映画の場合には構造化に一定の効果があることが上記論文で示されているが、ニュースのように対話が中心でない番組に対してはそのまま応用することはできない。

類似ショットの繰り返し登場に着目し、ニュース番組におけるアンカーパーソンのショット検出や野球中継でのイベント検出を試みた例もある [9]。ニュース番組に関しては、上記論文では最も頻繁に登場する類似ショットがアンカーパーソンのショットであるという仮定に基づいている。しかしながら 2 人以上のアンカーパーソンが出演する番組では検出されないアンカーパーソンが生じる恐れがある。また、スタジオに出演するゲストや記者とアンカーパーソンの対話が存在した場合、単純に最も頻繁に登場したショットをアンカーパーソンとしてしまうと、対話中のアンカーパーソンをも検出されるため、ブラウジングの上では対話の途中で細切れにインデックスが付与される恐れもあるなど、改善の余地がある。

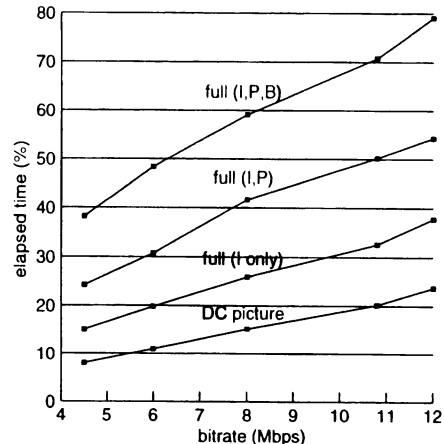


図2 デコード方法、ビットレートと処理時間 (映像時間に対する比)
Fig. 2 Elapsed time v. bitrate and decoding method

3. 処理手順と負荷

提案手法では、図1のような手順で処理を行う。処理対象データとしてはデジタル放送や D-VHS デッキなどで用いられている MPEG-2 System Transport Stream (以下 MPEG-2 TS) 形式のファイルを用いた^(注1)。手順は大きく 3 段階であり、第 2 段階、第 3 段階については既述の著者らの文献 [6] に倣って「パターン検出」「アクト検出」と称する。

3.1 DC 画像のデコード

まず MPEG-2 TS を Demultiplex し、Video Elementary Stream (ES) から DC 画像を取り出す。原画像が 720×480 画素 (SDTV) であれば、各ブロックの直流成分から生成される DC 画像は 90×60 画素となる。この段階でデコードが Video ES の一部に限定されるため、処理が大幅に高速化される [8]。

図2には、様々なビットレートの MPEG-2 TS からデコードを行った際の処理時間を示した。グラフは上から、すべてのフレームの全画素のデコードを行って後段の処理に必要なように 90×60 画素にサブサンプリングした場合 (full (I,P,B))、I ピクチャ・P ピクチャのみ全画素デコードとサブサンプリングした場合 (full (I,P))、I ピクチャの全画素デコードとサブサンプリング (full (I only))、そして DC 画像を直接デコードした (DC picture) 場合である。

図2の実験に用いた映像は、原則的に $M = 3, N = 15$ の GOP 構成でエンコードされたものであり [10]、すべて同じ番組から生成されたものである。ビットレートは実測値で、ファイルサイズを映像時間で割ったものである。縦軸は映像時間に対する処理時間の比率で、著者らが開発した MPEG-2 Demultiplexer および MPEG-2 Video ES Decoder の処理に要した時間を示している。なお、実験は Intel Celeron 733MHz / RedHat Linux 8.0 上で行ったものである。

DC 画像を用いることにより、交流成分の復号および逆

(注1): ただし暗号化されていないもの

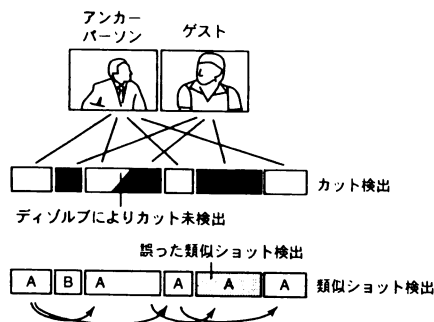


図3 カット未検出が引き起こす問題 (イメージ)
Fig.3 Problem caused by misdetection of cut

DCT(離散コサイン変換)の処理を省略できるため高速化が実現する。また、DC画像は 8×8 画素のブロックの直流成分であり、単純に縦横8画素ごとにサブサンプリングした画像と違って 8×8 画素値の平均がほぼ反映されているというメリットもある。

この結果から分かるように、DC画像を用いることによりフルデコード時よりも2.5~4.8倍の速度で処理が完了することがわかる。最も高速な4.5Mbpsの場合では、20分の番組に対するDC画像デコードは映像時間の12.6倍速の96秒で終了する。最も高いビットレートの映像の場合でも映像時間の4.0倍速のデコードが可能であるなど、低性能の処理装置でもリアルタイム処理が十分可能であることを示している。

3.2 パターン検出

色相ヒストグラムと輝度成分のモザイク画像を用いてDC画像間の類似検定[8]を行い、これに基づいてカット検出と類似ショット検出を行う。

すなわち、色相ヒストグラムについては $C_b - C_r$ 空間にプロットした座標 (C_b, C_r) が反時計回りに C_b 軸となす角 θ を求め、 $[\theta_i, \theta_{i+1})$ の範囲にある画素の数をカウントしたヒストグラムを作成し、 $[\theta_i, \theta_{i+1})$ の範囲の画素数を h_i 成分とする色相特徴ベクトル \vec{h} を作成する。DC画像 m と n について類似検定を行う際には、 h_m と h_n のユークリッド距離が既定値を超えるかどうかを根拠にする。

一方、輝度成分はDC画像のY成分からモザイク化した画像を作成し、各モザイクの輝度値を b_i 成分とする輝度特徴ベクトル \vec{b} を作成する。DC画像 m と n について類似検定を行う際には各成分の差分を検定する d を下記のように定義し：

$$d_i = \begin{cases} 1 & |b_{m,i} - b_{n,i}| \geq b_{thres} \\ 0 & |b_{m,i} - b_{n,i}| < b_{thres} \end{cases} \quad (1)$$

$\sum_i d_i$ が規定値を超えるかどうかを類似検定の根拠にする。

カット検出は着目点を挟んで数フレームずつを用いた類似検定を行い、非類似となる組み合わせが一定数以上の場合に着目点がカットであると判定する。これにより、簡易的な方法ながらフラッシュや大きなオブジェクトの横切りなど、突発的で瞬間的なカットの誤検出を抑制できるほか、トランジションによる時間幅のある場面切り替えを検出することができる。

表1 パターン検出(カット検出+類似ショット検出)の処理時間

Table 1 Elapsed time for pattern determination

	番組数 (放送回)	平均ショ ット長 [秒]	平均処理時間 [映像時間比%]	ショットあたり 処理時間 [秒]
ニュース	7(9)	3.21	0.38	0.012
全番組	23(30)	2.84	0.35	0.010

2つのショットの類似検定を行う際には、番組先頭に近い側のショットから終端の数枚と、番組終端に近い側のショットから先頭の数枚のDC画像を取り出し、上記のカット検出と同様の方法を用いて、類似となる組み合わせが一定数以上の時に類似ショットとして同一のラベルを付与する(類似ショットを同一ラベルでグループ化する)。

ただし類似ショット検出、およびアクト検出においては、過分割のカットよりも未分割のカットの方が処理精度に大きな影響を与える。これは、非類似ショットと類似ショットが誤って結合ショットとなった場合、そこが接合点となって2つの類似ショットグループがマージされる恐れがあるためである。例えばアンカーパーソンとゲストの対話において、2人間のカメラ切り替えが検出されなかったとすると、アンカーパーソンとゲストが同じ類似ショットのラベルを付与されることになり、アクト検出において両方もがアンカーパーソンショットに採用されたり、両方もが却下されたりすることが生じうる(図3)。

このため、カット検出のパラメータは敏感な側に設定している。なお、上記のような問題を防ぐための検証法も著者らによって提案されている[8]が今回は類似ショットの検証は行っていない。

表1には23番組(同番組を複数本処理したものもあるのでこれを加えると30本、合計21時間31分)に関してパターン検出に要した処理時間を示す。ニュース番組であることによって全番組に比べてごくわずかな処理時間の変動はあるものの、平均すると30分のニュース番組を映像時間の263倍速の6.9秒で処理することを示しており、DC画像のデコードまでに要する時間に比べるときわめて高速に処理が完了する。また、前節で示したようにDC画像のデコードまでの処理過程は原映像のビットレートに依存するが、パターン検出以降は依存しない。なお、処理に用いた装置の性能は前節と同じである。

3.3 アクト検出

「理想的な」ニュース番組ではアンカーパーソンのショットのみが類似ショットとして繰り返し出現するので、出現回数が最も多い類似ショットを先頭にショットを時間順に列挙すればニュースの構造化は可能である[9]。より実際のニュース番組を類似ショットの繰り返し登場を利用して構造化する場合、第2章で議論したように、複数のアンカーパーソンや、同じアンカーパーソンでもカメラアングルが異なるために2つ以上の類似ショットグループに分かれて属してしまっただけのショットをもニュース項目先頭として検出しなければならない。このためには出現回数が「最も多い」ではなく、「一定数以上」という条件に代えて、より多くの類似ショットグループをアンカーパー

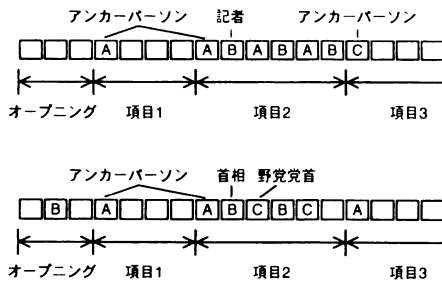


図4 対話部が引き起こす問題 (イメージ)
Fig. 4 Problem caused by dialog scene

ン・ショットの候補として採用する一方、登場の間隔や、ショットの長さなどに基づく条件で絞り込まなければならない。

著者らの実験では、出現回数やショットの長さを条件としたとき、アンカーパーソンのショットから最も弁別が困難な類似ショットはニュース番組中で記者やゲストと交わされる対話の部分であった。対話部分ではアンカーパーソンの、対話相手(記者やゲストなど)も頻りに登場するほか、どちらのショットも発言量によって長くも短くもなる(図4上)。対話相手のみであれば、その類似ショットグループが登場する時間分布の小ささで却下することも可能かもしれないが、この対話が数分に及ぶ場合、もっぱらニュース項目を短く伝える役割のサブアンカーパーソンのとの区別がつかなくなる。また、対話中のアンカーパーソンのが残ってしまうのでは、対話が細かいニュース項目に分割されてしまうという問題がある。

それ以外にも国会討論などがニュース項目の場合、質問者と回答者が対話状に登場する上、その日のニュースのメインピックの一つとしてオープニング中で紹介されることもある(図4下)。このような場合、登場回数、ショットの長さ、登場時間分布などの指標を見る限りではアンカーパーソンのショットと同じ振る舞いをするようになる。

そこで著者らは図1に示したように、類似ショット繰り返し登場の傾向からまず対話部を検出し、アンカーパーソンのショットと判定されたショットグループであっても、対話部に属するものを除外することで誤ったアンカーパーソンのショット検出を低減した。

対話部は以下のように検出する。

- 1) いずれかの類似ショットグループに属するショット2つを選択する
- 2) 上記の区間の「対話度数」が最大になるように、終端側のショットを探索する
- 3) 対話度数の最大値が規定値を超える場合には上記の区間を対話部と定義する

「対話度数」とは、あるショットからあるショットまでの時間区間が対話部であるかどうかを検定するために著者らが定義した指標である。ショットaからbまでの対話度数は:

$$I_{ab} = \left(\sum_{i=a}^b \rho_{ab,i} \lambda_i \right) \left(\sum_{i=a}^b \rho_{ab,i} \right) / \left(\sum_{i=a}^b \lambda_i \right)^2 \quad (2)$$

で定義される。ここで λ_i はショットiの時間長、 $\rho_{ab,i}$ は、ショット区間[a,b]においてショットiの類似ショットが存在する場合には1、それ以外の場合には0となる。対象区間内で対話に貢献しているショットが長い時間を占め、かつ、ショットが速いテンポで切り替わっている時に対話度は大きな値となる。

対話部を定義した後、対話部を除く類似ショットの登場回数(アンカーパーソンのならば頻りに登場する)、ショットの長さ(長いものが存在する)、登場時間分布(長い時間範囲に登場する)といった条件から、アンカーパーソンのショットと考えられる類似ショットグループを複数選択し、その中でも対話部に属さないものをアンカーパーソンのショットとして検出する。こうして検出されたアンカーパーソンのショットを各ニュース項目の先頭として話題分割を行う。最初に登場するアンカーパーソンのショット群はオープニングとみなす。なお、本手法による検出結果は次章で説明する。

前節までと同じ装置でニュース番組のアクト検出を実行したところ、20分~76分間の7番組(9放送回)に対して処理時間は0.01~0.02秒と、測定のはば下限値であった。

以上の結果から、MPEG-2 TSからDC画像を用いて仮に30分のニュース番組を構造化する場合、DC画像の作成までに143秒、パターン検出までに6.9秒、アクト検出はほぼ一瞬で終了することがわかる(原映像4.5Mbpsのときの理論値)。著者らの実験システムではDC画像の作成とパターン検出を同時に行い、この両処理をリアルタイムで入力されるMPEG-2 TSに対してCPU使用率7%で動作している。例えばパーソナルコンピュータに映像を入力しながらリアルタイムでパターン検出までを行い、終了後にアクト検出を一瞬で実行することにより、見かけ上は映像完了とほぼ同時に構造化が完了することになる。

4. 実験

これまで述べてきた手法によりニュース番組の話題分割を行った結果を図5~10に示す。各ショットアイコンの上に表示するのはショット番号である。

図5は、ニュース番組(1)の処理結果で、上述の対話部検出を行う前の状態である。アンカーパーソンの中央に映っているショット19から、やや左に寄って映っているショット22までは実際には同じショットであるが、アンカーパーソンの動きが激しいため、簡易カット検出で4つのショットに分割され、それぞれが別々の類似ショットグループに属している。このため、両者がニュース項目の先頭とされてしまっている。197と199も同様である。また、335から347までは前のニュース項目についてのアンカーパーソンのスタジオ解説者の対話であるが、これも別々のニュース項目とされている。

ショット739から857までは、スタジオにゲストを呼び、アンカーパーソンを交えてトークを繰り返しているが、この部分は細分化され、この対話部分だけで26のニュース項目になってしまっている。ショット942はコマーシャル中のショットであるが、誤った類似ショット検出によりアンカーパーソンの(ショッ

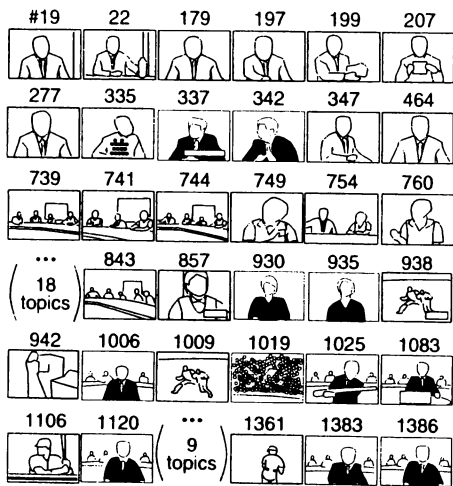


図5 対話部検出前の実験結果 (1)
Fig. 5 Topic-leading shots (before dialog determination)

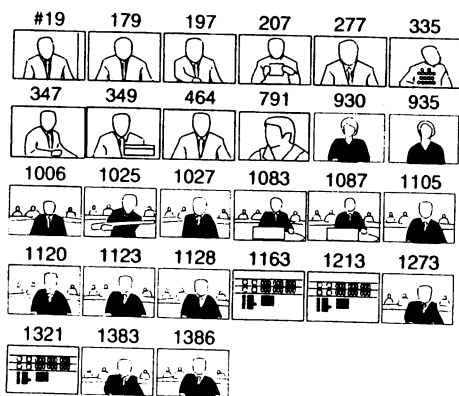


図6 対話部検出後の実験結果 (1)
Fig. 6 Topic-leading shots (after dialog determination)

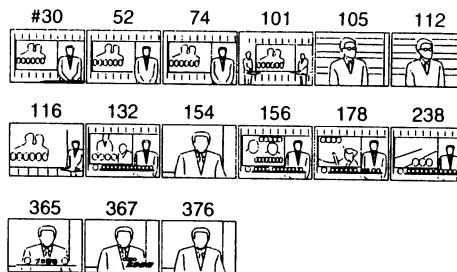


図7 対話部検出前の実験結果 (2)
Fig. 7 Topic-leading shots (before dialog determination)

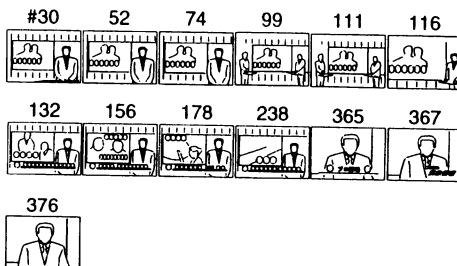


図8 対話部検出後の実験結果 (2)
Fig. 8 Topic-leading shots (after dialog determination)

ト 935 など) と類似とされてしまい、誤分割に寄与している。1006以降のスポーツコーナーでは、打席の異なる野球のフィールドが一つのカメラアングルから撮影されているためにこれらが類似ショットとなり、登場頻度と時間的登場範囲の条件を満たしてしまい、アンカーパーソンと誤認識された。

これに対して、図6は対話検出により上記の問題が解決していることがわかる。図6において、実際にはアンカーパーソンではないにも関わらず、ニュース項目先頭と誤認識されているのはショット791, 1163, 1213である。このうち野球のスコア表である1163と1213は、下記のような理由で必ずしもエラーとはいえない。本手法では、アンカーパーソンショットが連続して登場する場合、最後尾のアンカーパーソンショットを次の話題の先頭とみなし、それ以外は前項目のラップアップ(まとめ)と解釈している。この番組ではプロ野球の結果が5試合報じられており、スコア表は登場パターンから5つともがアンカーパーソン(=話題先頭)ショットの候補となっている。しかし、3試合についてはスコア表に引き続いてスポーツキャスターが顔を出しているため、スコア表はラップアップの属性が付けられた。ショット1163と1213については、スポーツキャスターの登場なく別の試合に移行するケースであり、ここに話題区切りが挿入されることはユーザの利便性を考えると(インタフェース上は)誤りとは断じられない。なお、対話中のショット791が登場したのは誤りである。

同様に、図7と図8にはニュース番組(2)を処理した結果を、それぞれ対話部検出前、対話部検出後の2通りで示した。図7では、スタジオでの記者とのトークが話題区切りに誤って採用

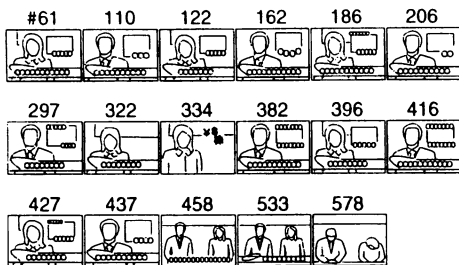


図9 実験結果 (3)

Fig. 9 Detected topic-leading shots

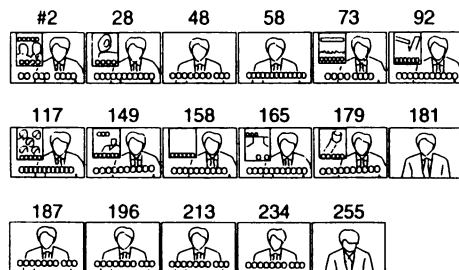


図10 実験結果 (4)

Fig. 10 Detected topic-leading shots

され、ショット 105,112 が誤検出されている。図8では、これらの問題は解消している。

図9にはニュース番組(3)の処理結果を、図10にはニュース番組(4)の処理結果を示す(対話部検出後の最終結果のみ)。ニュース番組(4)ではすべての話題区切りが過不足なく検出されている。なお、図10のショット2,28,73,92,117,149,158,165,179は一つの類似ショットグループ、それ以外は別の類似ショットグループであり、2つの類似ショットグループがアンカーパーソン(=話題先頭)ショットとして検出されている。

5. おわりに

以上、述べてきたように、著者らは MPEG-2 Transport Stream 形式の映像を対象に、Video ES の I-picture/ブロック直流成分のみから生成される DC 画像を用いることにより、Intel Celeron 733MHz プロセッサを用いた実験のベストケースで映像時間の 8%程度、すなわち 12 倍速 (30 分番組の場合、2 分 30 秒) という高速な処理でニュース番組を話題ごとに分割する手法を開発した。これは MPEG-2 TS の全フレームを全画素についてデコードして行った場合に比べて 4.6 倍の高速さである。

本手法においては、ニュース番組中に存在する対話部分に対し、「対話回数」という指標を用いることにより効率的に除去し、対話部から話題先頭となるショットが大量に生成されるのを抑制した。対話部の除去により、スポーツコーナーにおける野球フィールドの繰り返し登場など、実際には「対話」ではない非話題先頭ショットをも除外することが可能になった。

本手法により、ニュース番組から観たい話題を探し出してラ

ンダムアクセスするための表示インターフェースの提供などが可能になり、今後ますます増えると考えられるデジタル映像の家庭などにおける利用の利便性を向上させる一助になることが期待できる。

今後は、今回のように比較的低性能の処理装置においても高速に構造化処理が完了できるという条件を保ちつつ性能を上げるために、別のフレーム類似判定手法の導入を検討するほか、著者らが提案した類似ショットの「確からしさ」を示す指標の導入 [6] など、まず類似ショット検出の面での性能向上をはかる。

一方で、ニュース番組に限定せず、多岐に渡る放送番組において構造化が可能なものの識別と、他ジャンルの構造化手法についても取り組んでいく予定である。

文 献

- [1] 金出 武雄, 佐藤 真一, "Informedia: CMU デジタルビデオライブラリプロジェクト," 情処学会誌, Vol.37, No. 9, pp.841-847, 1996.
- [2] 堀 修, 三田 雄志, "テロップ認識のための映像からのロバストな文字部抽出法," 信学論 D-II, Vol.J84-D-II, No.8, pp.1800-1808, 2001.
- [3] S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming and Detecting Faces in News Videos," IEEE Multimedia, Vol.6, No. 1, pp.22-35, 1999.
- [4] 井手 一郎, 山本 晃司, 浜田 玲子, 田中 英彦, "ショット分類に基づく映像への自動索引付け手法," 信学論 D-II, Vol.J82-D-II, No.10, pp.1543-1551, 1999.
- [5] 小野 晋太郎, 井手 一郎, 坂井 修一, 田中 英彦, "キャプションを複合的に利用したニュース文からの重要文抽出," 第 62 回情処全国大会 No.8M-3, Vol.2, pp.329-330, 2001.
- [6] H. Aoki, O. Hori, and S. Shimotsuji, "A Shot Classification Method of Selecting Effective Key-frames for Video Browsing," Proc. ACM Multimedia 96, pp.1-10, 1996.
- [7] 金子 敏充, 堀 修, "動きベクトル符号量を用いた MPEG 映像からの高速カット検出," 信学会研究報告, PRMU86-100, 1996.
- [8] 青木 恒, 堀 修, "繰り返しショットの統合による階層化アイコンを用いたビデオ・インターフェース," 情処論文誌, Vol.39, No.5, pp.1317-1324, 1998.
- [9] 菅野 勝, 植村 幸市, 柳原 広昌, 中島 康之, "ショット遷移に基づく映像の意味的インデクシングに関する検討," 情報科学技術フォーラム (FIT2003) 一般講演論文集第 3 分冊 J-088, pp.377-378, 2003.
- [10] ISO/IEC 13818-2 "Information Technology - Generic Coding of Moving Picture and Associated Audio: Video," International Standard, 1995.