

InverseDirector: 対話検出によるニュースとバラエティ番組の自動構造化

青木 恒

東芝 研究開発センター マルチメディアラボラトリー

ビデオの内容を一覧で把握し、観たい場面に素早くアクセスするためには内容構造に即した表示インターフェースが必要である。筆者はテレビ番組で発生する類似ショットの繰り返し登場から場面の対話活性を測定する指標を導入し、映像中の対話区間の自動検出を行った。これを用いて、ニュースやバラエティ番組を話題ごとにまとめあげることができる。本論文では、素材の配置に基づき話題区切りを推定する、いわばディレクターの逆作業を行う InverseDirector システムについて議論する。

InverseDirector: Topic Organizer for News and Variety Shows

Hisashi Aoki

Multimedia Laboratory, Corporate Research and Development Center, Toshiba Corporation

This paper describes the “InverseDirector” system, which analyzes the structure of TV program from the repetitious appearance of similar shots. The author introduces an interactivity measurement of video segment, that denotes the activity of conversation by calculating frequency and occupancy of similar shots in the segment. With this measurement, the dialog scenes are found and the topic groups are effectively defined. Experiments show the system provides a structured viewing interface for news and variety shows.

1 はじめに

近年、ハードディスク・ビデオ・レコーダやテレビ録画できるパーソナル・コンピュータの急速な普及により、家庭においても放送番組のデジタル録画が一般的になりつつある。また、こうした機器の普及は、録画済み番組タイトルの一覧やタイムシフト再生などデジタル録画特有の機能提供も相俟って、従来のビデオ録画・視聴スタイルそのものをも刷新しつつあることは言うを待たない。



図 1: 本システムの画面生成結果

一方、デジタル放送やインターネットを通じて提供されている電子番組表 (EPG) を利用して予約録画す

ることにより、番組タイトルやジャンル情報、出演者などが同時に記録でき、視聴時にこれらの情報を提示したり、ユーザの録画履歴から嗜好を推定して自動録画したりできるなど、ビデオテープ録画と比較して番組検索のインターフェースも飛躍的に改善されている。

しかしながら普及機器で利用できる番組記述情報 (メタデータ) は現在のところ番組単位で付与されている。したがって「観たい番組」を探し出すためのインターフェースは実現したが、「観たい場面」を探し出すインターフェースの自動提供には至っていない。

目的の場面を探しやすくするための構造化には、1) シーン記述データを利用する、2) 映像、音声データを解析する、という2つのアプローチが考えられる。

1) は、MPEG-7, XML などでのシーン記述が提案されており [1]、テキストでの効率的な構造化や要約表示が実現でき [2]、将来の放送での実現可能性がある。現行放送でもクローズドキャプションなどで字幕テキストが提供されており、それを用いて構造化に供することもできる [3]。しかし現状では構造化を想定したシーン記述つき放送は一般視聴者が利用できるに至っていない。字幕放送についても本質的に発話をテキスト化したものであり、画面から理解できる情報 (話者の交代など) は省略されているなど、番組構

成の自動理解に利用するためには2)と同様の高度な知識処理を必要とする。

2)は、過去に録画したビデオなど、シーン記述の存在が期待できないコンテンツにも適用できるメリットがある。音声認識により発話をテキスト化し、検索キーワードとして提供したり[4]、画像処理によって映像中の人物を認識して登場人物のラベリングを行ったり[5]、字幕部分を認識してテキストラベルに利用したり[6]、あるいはそれらを複合的に利用したり[7]といった方法がある。

これらの方法はメディア理解を伴うため、人間の認知に近い高度なラベリング、インデクシングが期待できる反面、計算量が多くなる傾向にある。このため、高い処理能力を持った機器上での動作が前提となるなど、家電機器のように処理性能が限定された装置への搭載までには解決すべき問題が多い。

青木らはMPEG-2映像から軽微な計算量で抽出可能な縮小画像を解析し、類似ショットの繰り返し登場を検出することで映画やドラマなどの構造化と表示効率向上を試みた[8]。テレビ番組の映像解析例ではないが、後藤は音楽の「サビ」を自動検出することに成功し、楽曲の特徴的な部分に直接アクセスできるようなインタフェースの提供を可能にした[9]。映像や音楽など、時系列データを鳥瞰するインタフェース提供のためにはメディアデータの構成推定が有効であるが、青木らや後藤のように映像、音声そのものの認識をすることなく、類似区間の繰り返しパターンを利用することで、**ディレクターが行う演出作業の逆**(すなわち、シナリオに即して素材を構成配置するのではなく、素材の構成配置からシナリオを推定すること)ができる。

本論文で筆者は類似ショットの繰り返しパターンを利用して番組全体や部分区間の対話の活発さ(Interactivity)を示す指標を導入し、これを用いてテレビ番組中の対話区間を推定する方法を提案する。これによって、従来手法で取り組まれたよりも多様な構成を持ったニュース番組の話題構造を推定するとともに、本手法を応用してバラエティ番組を話題構造に即して構造化することが可能であることを示す。

2 対話検出の必要性と応用可能性

2.1 ニュース番組構造化の課題と対策

既出の青木らの方法では、映画やドラマの構造化を主眼にし、類似ショットの登場範囲を推定場面とした。つまり、まず類似ショット同士をグループ化して

それぞれにグループA,B,C…というラベルを付与する。グループA,Bの時間的登場範囲にオーバーラップがあるとき、それらを足し合わせた時間範囲を推定場面とするものである。

しかし、この方法をニュース番組に適用すると、アンカーパーソン(メイン司会者)のショットが番組の先頭近くから終端近くになで分布し、番組のほとんど全体が一場面となってしまふ。

菅野らは類似ショットのうち最も高い頻度で登場するものをアンカーパーソンショットと仮定する方法での構造化を検討した[10]。しかし、筆者の実験によると、すべてのアンカーパーソンショットが一つの類似ショットグループに属するケースは一部に限定される。2人のアンカーパーソンが出演するニュース番組では、実際のアンカーパーソンショットは、それぞれのワンショット、ツーショットと最低でも3パターンある。メイン司会者が一人のニュース番組であっても何通りかのカメラアングルがあるなど、より多くの構成に対応するためには発展が必要である。

さらに、スタジオの記者やゲスト、現場の記者などと対話したりすることもある。単にアンカーパーソンのショットを話題の区切りとするだけでは、このような対話部分で聞き手として登場するアンカーパーソンをもすべて話題の先頭となり、ニュース項目がコマ切れになる。

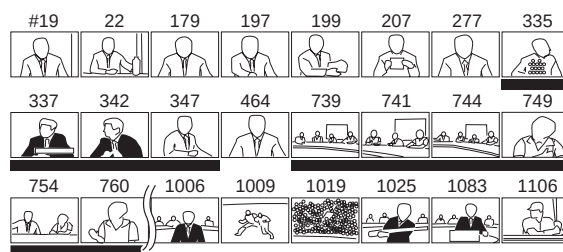


図 2: アンカーパーソン検出の失敗結果†

図2はニュース番組に対し、登場頻度が上位の類似ショットがアンカーパーソンショットであるという条件で話題先頭を抽出した例である†。最上段は正しくメイン司会者が抽出されている。しかしショット335から347まではスタジオ解説者とメイン司会者のトーク、ショット739以降、省略記号まではゲストと司会者のトークであるが、これら部分的な対話シーンでも高頻度で類似ショットが登場するため、話題の先頭と無関係なショットがアンカーパーソンショットと誤判定されている。

このような問題を解決するためには、ニュース番組

†図ではイラストとしたが、実験結果は実際の処理画面を再構成したものである。以下†を付した図はすべて同じ

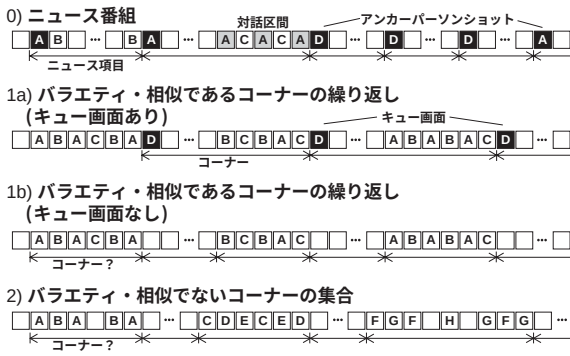


図 3: 番組種ごとの類似ショット登場モデル

中の対話部分を検出し、対話部分は除外した上で、類似ショットの繰り返し登場パターンに基づくアンカーパーソンショット検出を行い、話題の区切りを決定する必要がある(図3のケース0参照)。

2.2 バラエティ番組構造化への応用

従来、映像解析による番組構造化は映画、ドラマ、ニュース、あるいはスポーツ番組などに関して取り組まれてきたが、バラエティ番組の中にも明確な構造をもって構成されているものがあり、図3の1a),1b),2)の3つに大別できる。

1a),1b)は、ある話題に関するスタジオ出演者と司会者によるトーク場面と、スタジオ外で録画、編集された情報ビデオとの2パートの交互登場によって構成されるもので、クイズ番組やトーク番組、情報番組などで顕著である。歌番組についても、ゲストと司会者のトーク場面と、そのゲストの歌唱場面との2パートで構成されるものが多い。

「類似ショット登場の観点で相似」であるコーナーが繰り返されるバラエティ番組の中でも、コーナータイトル画面(以下、「キュー画面」と称する)が類似ショットとして繰り返し登場するもの(1a)と、そうでないもの(1b)がある。キュー画面は一般的にその他の類似ショットとは異なる類似ショットグループに属し、対話部分から独立している。したがって、高い頻度で登場する類似ショットのうち、対話部分への関与が低いものを検出することで、キュー画面、ひいてはコーナーの検出を行うことができる。

キュー画面の存在しない番組(1b)については、一つ一つの対話部分そのものが話題単位のまとまりであると仮定することで、ある程度の構造化を行うことができる。2)についても同様で、番組が「トーク場面(対話部分)」と「情報ビデオまたは歌唱場面(非対話部分)」の交互登場で構成されている場合には対話部分や非対話部分への頭出しが可能になるだけでも、

話題に即したアクセスが容易になるものと思われる(4.4節参照)。

3 対話度数と対話区間の定義

3.1 対話度数

上述のように、類似ショットの繰り返し登場パターンから対話部分の検出を行うことは番組の構造化に関して有意義である。そこで筆者は、動画像中の時間区間に対する対話の活発さを量的に評価するための「対話度数(Shot interactivity)」という指標を導入し、これが一定値以上になるような区間を順次定義していくことにより、対話部分の検出を試みた。

対話度数に求められる要件は以下のようなものと仮定した：

- 1) 対象となる時間区間中で、対話に費やされている時間が長いほど、高い対話度数になる
- 2) 対象となる時間区間中で、話者の推移が活発であるほど、高い対話度数になる

1)は、対象となる時間区間が実際の対話部分にどれだけかかっているかが度数の高低に反映されるべき、という条件であり、この高低を示す**対話密度** d をまず定義する。2)は、キャッチボールのようにテンポよく話者が切り替わっているかが度数の高低に反映されるべき、という条件であり、この高低を示す**対話速度** v を次に定義する。ショット a から b までの対話度数 I_{ab} は $d_{ab} \times v_{ab}$ によって表されるものと定義する。

対話密度 d_{ab} および対話速度 v_{ab} は実際には以下のように計算する：

$$d_{ab} = \left(\sum_{i=a}^b \rho_{ab,i} \lambda_i \right) / \left(\sum_{i=a}^b \lambda_i \right) \quad (1)$$

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

ここで λ_i はショット i の時間長、 $\rho_{ab,i}$ は、ショット i がショット区間 $[a, b]$ において対話に貢献しているか否かに応じて1または0の値をとる係数である。

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 ρ_i は1、それ以外の場合には0となる(図4)。したがって、ショット区間 $[a, b]$ の外にのみショット i の類似ショットが存在する場合は、 ρ_i は0となり、対話密度におけるショット時間、および対話速度におけるショット個数の計数に加味されない。

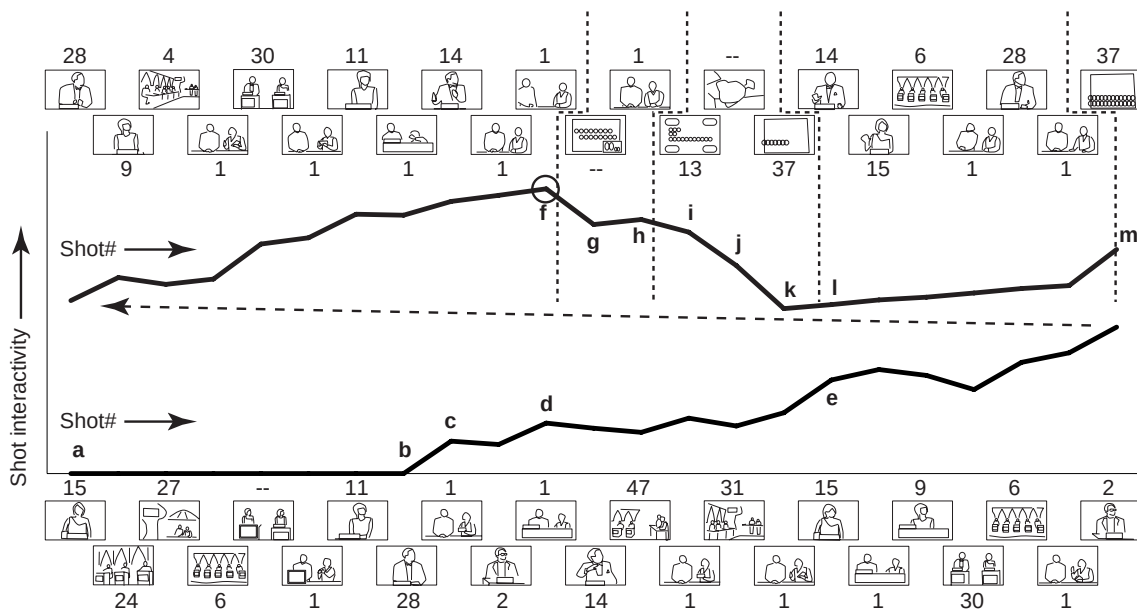


図 5: バラエティ番組に適用した対話度数の推移†

	Shot p	q	r	s	t	u	v
	A	C	A	B	B		
$\rho_{pr,i}$	0	0	0				
$\rho_{ps,i}$	1	0	0	1			
$\rho_{pt,i}$	1	0	0	1	0		
$\rho_{pu,i}$	1	0	0	1	0	0	
$\rho_{pv,i}$	1	0	0	1	1	0	1

図 4: ρ の定義

対象区間に類似ショットが存在しないものを計数から除外する理由は、(後述のように) この対話度数を番組全体にも適用できるようにするためである。いわば、区間 $[a, b]$ だけをひとつの「ミニ番組」とみなし、その番組中の対話の活発さを評価している。区間 $[a, b]$ の外にのみ類似ショットが存在するようものは、もし $[a, b]$ 間のみを切り出した映像ソースから処理したとすれば本来類似ショットが発見されなかったものである。このため、このようなショットは活発さへの寄与から除外している。

このようにして対話密度は対象区間内で対話に貢献しているショットが占める時間的割合を示す。対話速度は概ね単位時間当たりのショット切り替わりの頻度を示す。

図 5 には、バラエティ番組中の一部について上述の対話度数 I を計算した例を示す。横軸はショットの登場順であるが、スペースの都合上、前半部と後半部に分けて折り返し、重ねあわせ表示している。縦軸はショット a から横軸上の各ショットまでの対話度数で、前半部と後半部は同スケールである。各アイコンの上下に示してある数字は類似ショットグループの ID 番号であり、「--」と記述されているものは番組内に

類似ショットが発見されなかったことを示す。

区間 $[a, b]$ までは区間内に繰り返し登場する類似ショットがないため、対話度は 0 となっている。区間 $[a, c]$ で初めてグループ ID 「1」のショットが 2 回登場するため、対話度が正値を持ち始める。以降 d, e など、複数登場するショットが増えるたび対話度数が増加する。

図 5 の区間では、ショット f までで一つの知識情報に関するスタジオトークが終了する。ショット g で要点と出演者の評価点数をまとめるテキスト画面が表示され、 h で司会者が「それでは次の…」とつなぐ。ショット i は、次の知識情報の投稿者の名前を表示するテキスト画面で、 j, k でこの項目に関するビデオが流される。 l からは再びスタジオに戻り、いま提示されたビデオに関連するスタジオトークとなる。トークは m の前まで継続し、再び本項目の詳細情報のビデオが流される。ショット m の後のショットは詳細情報ビデオであり、類似ショットが登場しないために対話度は減衰してゆく。

このように一つの話題に関する対話区間 $[a, f]$ の終端において対話度も最大になる。これを利用して対話区間を定義することが可能である。

3.2 番組の対話性測定

対話度は、対象区間を「ミニ番組」と見なした際の対話の活発さを示す指標であるが、番組全体に対して対話度を計算すれば番組そのものの「対話性」を算出することができる。

図 6 はニュース、バラエティ、およびドラマ(アニ

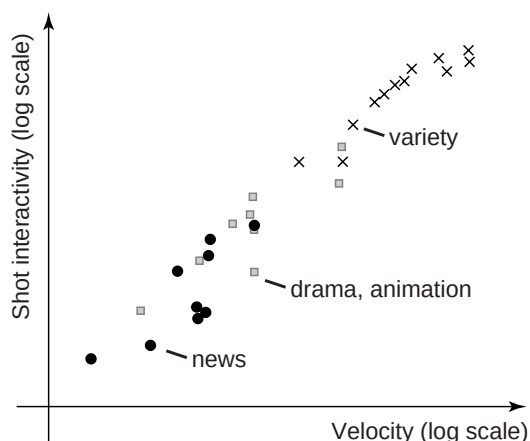


図 6: 対話度数の番組種との相関

メーション、時代劇を含む)23 番組、のべ 30 本分、合計 21.5 時間について対話度数と対話速度をプロットしたものである。

ドラマに関しては番組依存性が高いものの、バラエティ番組は突出して対話性が高く、ニュース番組は低い、という仮定に沿った傾向が観測される。ここに挙げた番組種だけを入力する場合に限っては、対話度数が一定値以上であるといった条件を用いることにより、バラエティ番組を他から自動弁別する可能性を示すものである。

3.3 対話区間の決定

3.1 節のように、あるショットを起点とし、対話度数が最大となるような終端ショットを探索することで対話区間を推定できるが、最大値が小さい場合には真の対話区間ではない場合がある。例えば対象区間中できわめて離れた間隔で類似ショットが 1 組だけ登場するような場合 (「A ————— A」など) である。そこで対話度数に閾値 I_{thres} を決定し、最大値が I_{thres} を超えるか否かによって対話区間であるかどうかを判定する必要がある。

前節に示したように番組全体の対話度数を計算した際、バラエティ番組は顕著に高い数値を取ることがわかった。そこで、図 6 のバラエティ番組のみの対話度数に関して、平均から標準偏差を減算したものを閾値 I_{thres} と定める。

なお、対話度数の計算対象区間の先頭、末尾がいずれかの類似ショットグループに属する場合、その外側にいずれの類似ショットグループにも属さないショットを付け加えても対話度数は常に下がる。したがって対話度数の最大値の計算は範囲の両端がいずれかの類似ショットグループに属する場合に限って行えばよい。対話区間の定義は実際には以下ようになる：

- 1) 番組先頭から、いずれかの類似ショットグループに属するショットを探索する
- 2) 類似ショットグループに属するショットを開始点として、番組終端に向かって対話度数を計算する
- 3) 対話度数の最大値と I_{thres} とを比較する
- 4) 3) の結果、対話区間であるとされる場合には、2) で最大値を生成したショットまでを対話区間とする
- 5) 4) の結果が対話区間であれば対話区間の終端の次のショットから 1) を行う
対話区間でなければ 2) の開始点の次のショットから 1) を行う

4 構造化実験

4.1 解析対象データ

処理対象となる映像は Aoki らの方法 [11] と同様、MPEG-2 映像データから I-picture を取り出し、各ブロックの直流成分のみから構成される 90×60 ピクセルの画像を入力データとする。各画像からは色相ヒストグラム各ビン*の値をもつ特徴量ベクトルと、輝度画像をモザイク化した画素値からなる特徴量ベクトルを作成する。色相ヒストグラムはフレーム内のオブジェクトの微動や輝度の変化にロバストであり、輝度モザイクは画面の大まかなレイアウト変化を検出可能にする。また、直流成分を用いた画像は元の圧縮映像に比べて少ないデータ量でありながら端的に内容を示しており、デコード処理も軽量である [12]。

この 2 つの特徴ベクトルを用いて簡易的なカット検出と類似ショット検出処理を行う。カット検出は注目点の前後から数フレームずつを取り出し、前側と後側のフレームの類似度を特徴ベクトルを用いて判定し、類似である組み合わせが一定数未満である場合に注目点をカット (場面の切り替わり) と定義する。カットによって区切られた連続するフレームの集合をショットと称する。

次にショット同士の類似性を判定する。2 ショットが類似であるかどうか判定するためには、先頭に近い側のショットの終端から数フレームと、終端に近い側のショットの先頭から数フレームずつを取り出し、

*ヒストグラム計算では色相値全体を小範囲に分割し、各範囲に何個の画素がプロットされたかを計数するが、ビンとはその小範囲のこと

簡易カット検出と同様の類似判定を行う。2ショットが類似である場合には、同じラベルを付与してゆく。

こうして番組全体に対してカット検出によるショット定義、および類似ショットグループ定義(ラベリング)が終了した段階のデータを、次節以下のように構造化する。なお、図3のケース0)は4.2節の、ケース1a)は4.3節の、ケース1b)および2)は4.4節の方法で処理する。適切な処理方法の自動選択については4.5節で議論する。

4.2 ニュース番組

3.3節のように対話区間を決定し、対話区間以外について各類似ショットグループごとに所属ショット数や分布時間、所属ショットの最長時間などの諸量を計算する。それぞれの諸量に対し、規定の条件を満たすものをアンカーパーソンショットとし、それを開始点として分割することによって話題ごとのまとめ上げをおこなう。番組冒頭から最初にアンカーパーソンショットが登場するまでの区間はオープニングとする。

アンカーパーソンショットが連続する場合には、連続の最後のショットを新たな話題の開始と見なし、それ以外の連続アンカーパーソンショットは前項目に属すラップアップ(まとめ)と判定する。対話区間にアンカーパーソンショットの類似ショットが含まれる場合、最初に登場するアンカーパーソンショットを便宜的に新たな話題の先頭とする。

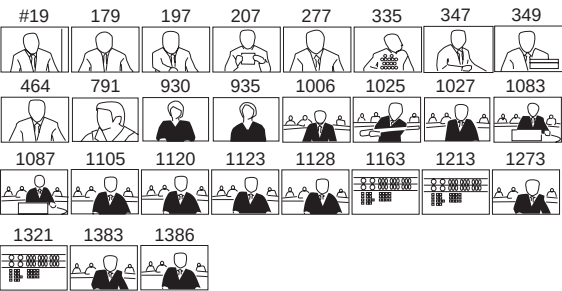


図 7: ニュース番組の項目先頭ショット抽出結果†

図7は対話区間検出を行ったうえで図2と同じニュース番組を構造化した結果で、各ニュース項目の先頭一覧である。ゲストとの対話シーンからは話題先頭として誤検出されたショット791のみが表示され、それ以外はショット464までのメイン司会者、930,935の女性アナウンサー、1006以降ではスポーツキャスターが先頭となり、図2のように野球ニュースの途中が先頭とみなされている部分はない。1163,1213,1321ではスコア表が先頭となっているが、この部分ではスポーツキャスターを挟まずにプロ野球の試合を順

次報じている。アンカーパーソンのショットではないが、このショットの次から別の試合の報道が始まるので、試合別に区切る上ではアンカーパーソン同様、依然として重要な手がかりの一つといえる。

4.3 コーナー検出可能なバラエティ番組

3.3節のように対話区間を決定し、各類似ショットグループごとに、その所属ショットの全体数に対して、いずれかの対話区間に登場しているものの割合を対話寄与数として計数する。対話寄与数が一定値以下である類似ショットグループのうち、所属ショット数、分布時間などの諸量が規定の条件を満たすものをコーナータイトルとし、それを開始点として分割することによって話題ごとのまとめ上げを行う。

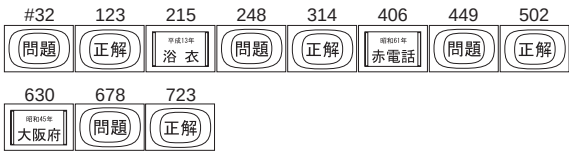


図 8: バラエティ番組の項目先頭ショット抽出結果†

図8はクイズ番組を本手法で構造化したときの各コーナーの先頭一覧ショットである。問題、正解、次の項目テーマ画面のすべてが正しく検出されている。

4.4 コーナー検出不可のバラエティ番組

3.3節のように対話区間を決定し、対話区間同士で時間的に近接しているものや、同じ類似ショットグループに属すショットを含む割合から対話区間を接続し、大まかなコーナー分けを行う。ただし、これは類似ショットを用いたコーナー分割が困難なものを補助的にまとめあげるものであるため、分割構造は番組種によって大きく依存する。

図9には、歌番組を本手法を用いてコーナー推定した結果を示す。軸上の刻みがシステムが推定したコーナー境界であり、長方形は目視による番組の意味的境界である。先頭を除く24個の意味的境界に対し、時間ずれのない検出が9箇所(○)、誤差10秒以内の検出が7箇所(△)で、これらを合計すると67%の正解率であった。

4.5 コーナー検出成否の自己判定

EPG等を用いて番組がバラエティであることが既知であったとしても、4.3/4.4節で述べたようにコー

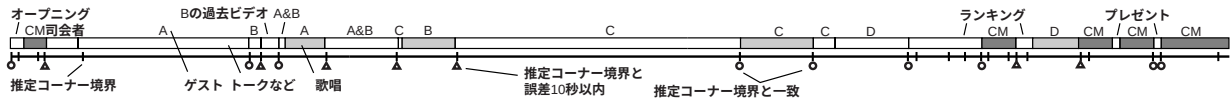


図 9: コーナー概要推定結果†

ナー検出が可能であるか否かによって異なる処理手順を用いる必要がある。これは以下のような方法によって簡易的に判定することが可能である。

まず、3.2 節に示したような方法でバラエティ番組であるということがわかったら、コーナー検出が可能であるか否かに関わらず、4.3 節の方法でコーナー分割を試みる。その結果、コーナータイトルとされたショットの時間的登場分布が正当なものであるかどうかを見る。具体的には、最も簡単な指標として最初のコーナータイトルが登場した時刻が、番組全体の長さに対してどの程度の位置にあるかを観測する。

一般的な構造から考えて、番組全体時間の 4 分の 1 以上経過後にコーナータイトルがあることは少ないと推測される。

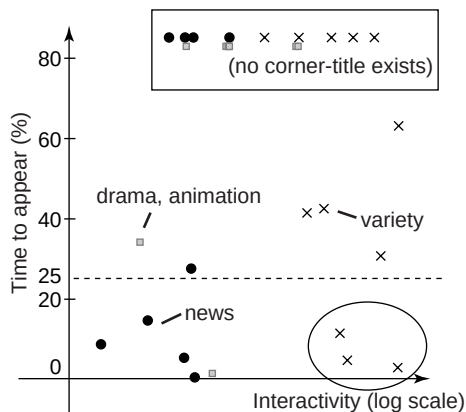


図 10: 最初のコーナーが登場する時間

図 10 には、横軸に既出の対話度数、縦軸には番組全体の長さに対して最初のコーナータイトルが登場する時刻の割合がプロットしてある。本指標は本来、バラエティであるということが既知である番組に対して用いるものであるが、参考のために図 6 と同じように他のジャンルを含む 23 番組をプロットした。図 10 の上方にあるプロット点は、コーナータイトルが検出できなかった番組である。バラエティ番組のうち、下方の楕円に囲んだ部分にある 3 番組はそれぞれ、「問題」「正解」の CG、「はいテーマです」とアナウンサーが紹介するショット、「〇〇県・×山△彦さんからの知識情報」という画面がコーナータイトルになりうるような番組である。楕円の外にプロットされた番組はこうしたコーナータイトルを持たない種類の番組である。

図でわかるように、コーナータイトルが検出できる番組は番組全体時間の 4 分の 1 程度以下に最初のコーナータイトルがあることがわかる。

5 構造化表示インタフェース

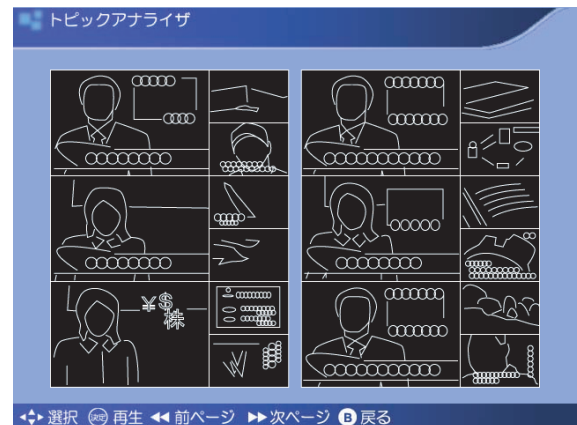


図 11: ニュース番組の表示結果†

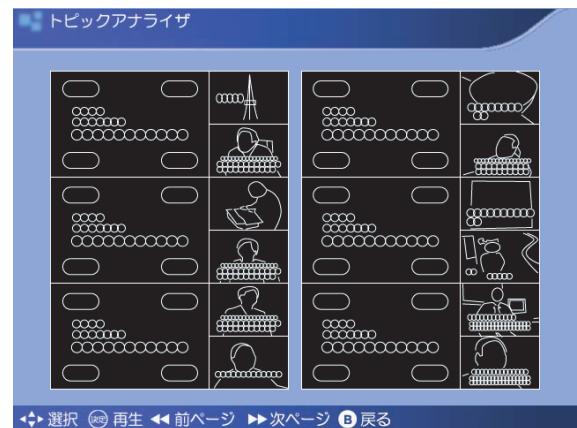


図 12: バラエティ番組の表示結果†

本手法によるニュース/バラエティ番組の構造化結果を表示するのには、単純には話題先頭のショットを代表フレームとして一覧すればよい。しかしながら、より話題内容の一覧性を高めるため、4.1 節のデータのみから得られるショットの諸量を用いて筆者は以下のようなインタフェースを作成した。

話題項目ごとに時間長の長いショット上位 2 つを選択し、サブアイコンとして先頭ショットの代表フレーム (メインアイコン) に併記する。ただし、サブアイ

コンの表示順は出現時間順とする。また、同じ類似ショットグループに属すショットがメインアイコン、サブアイコンに重複して登場しないようにする。

このようにした表示結果が既出の図1に示してある。また、このほかのニュース番組(図11)、図5のバラエティ番組を構造化表示した例(図12)も示す。ニュースにおいては、ニュース項目の先頭である男女のアンカーパーソンショットがメインアイコンとして表示されており、アイコン中の文字からもニュースの概要を知ることができる。ニュース項目の字幕表示がないものや読解が困難なものは、サブアイコンが概要を知る手がかりとなる。バラエティ番組の場合は、すべてのメインアイコンはコーナー先頭である情報投稿者画面になっている。

こうした構造化および表示方法により番組内容を話題別に一覧できるほか、興味ある話題項目に直接アクセスしたり、ある項目の視聴途中でも別の項目先頭にジャンプしたりするインタフェースの提供が可能になる。

6 おわりに

以上述べたように、筆者はテレビ番組の全体および一部について対話の活発さを評価する指標「対話度数」を導入した。この指標はニュース、バラエティ番組の対話区間検出に用いることができるほか、対象番組のジャンルが不明な場合でも、トークを中心に構成された番組であるかどうかを数値として測定できることを示した。また、対話区間の検出により、ニュース/バラエティ番組それぞれをより効果的に話題ごとに分割する手法についても述べ、分割された映像の一覧表示方法の一例を提案した。

これにより、ユーザが番組全体の内容を一瞥で視認できる映像ブラウジング・インタフェースが提供できるほか、各話題の概要も容易に知ることができるようになる。今後は、より多くのジャンルに対する自動構造化方法を検討するほか、さらに多くの番組データを用いることでニュース/バラエティ番組に関しても、より精度の高いコーナー分割手法に取り組む。

また、代表フレームの選択方法についても課題が残されている。番組固有の構造知識を用いることによって、例えばクイズ番組の正解部分を一覧表示に盛り込むことも可能にはなるだろうが、そのように「考える楽しみ」を奪うような提示方法には疑問が残る。番組の要点を一覧表示でどこまで開示するのが視聴者にとって望ましいか、あるいは表示方法をどのように選択させるインタフェースが必要かという議論も

進めていきたい。

参考文献

- [1] ISO/IEC 15938: “Information Technology – Multimedia Content Description Interface” (2001).
- [2] 山本 他:XMLを用いた後整理型マルチメディア・アルバム・システム, インタラクション 2001 論文集 pp.143-144 (2001).
- [3] 小野 他: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).
- [4] 金出 他: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).
- [5] Satoh, S., et al.:Name-It: Naming and Detecting Faces in News Videos, *IEEE Multimedia*, Vol.6, No. 1, pp.22-35 (1999).
- [6] 堀 他: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).
- [7] 井手 他: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).
- [8] 青木 他: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・インタフェース, 情処論文誌, Vol.39, No.5, pp.1317-1324 (1998).
- [9] 後藤:SmartMusicKIOSK: サビ出し機能付き音楽試聴機, インタラクション 2003 論文集 pp.9-16 (2003).
- [10] 菅野 他: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).
- [11] Aoki, H., et al.: A shot classification method of selecting effective key-frames for video browsing, *Proc. ACM Multimedia 96*, pp.1-10 (1996).
- [12] 青木 他:MPEG-2 映像からのニュース番組高速構造化, 情処研究報告 2003-AVM-43, pp.151-156 (2003).