

テロップを用いたニュース映像のトピック分割

関口明紀[†], 森田啓義[†], I Gusti Bagus Baskara Nugraha[†][†] 電気通信大学 情報システム学研究所 ネットワークシステム学専攻

1. はじめに

近年, 予め好みのニュース番組を指定しておくで毎回自動録画し, リモコンの専用ボタンを押すことでいつでも当該ニュース番組の再生をするテレビが市販されている。

また, 携帯機器での映像の録画再生が容易になってきていることから, 保存された最新ニュース映像を移動中にみることといったニュース映像の視聴形態がより身近なものとなってきている。

このように録画されたニュース映像をユーザーの都合の良い時に視聴する形態が浸透しつつあるが, ニュース映像を先頭から視聴するだけでは閲覧性が悪く, ユーザーが興味をもったニューストピックだけを選択・視聴可能とすることが出来れば利便性が高まる。

こうした課題を解決すべく, ニュース番組を自動的に構造化し, 好みのニューストピックへ簡単にアクセスする手法が提案されている[1]。

ニュース映像をトピック毎に分割する研究は従来より行われており, 用いる情報の種類は概ね, ①アナウンスショット, ②テロップ, ③字幕放送テキスト, ④音声データ, ⑤トピックの変化点検出用情報, の5種類に分けられる。

アナウンスショットを検出することでトピック分割をする手法としては, トピックがアナウンスショットから始まるという前提のもと, ニュース映像における各ショットから類似ショットを検出し, 一定数以上の出現数を有する類似ショット群をアナウンスショット候補とする方法がある。特にコメンテーターとアナウンスとの対話部をアナウンスショット候補から除く手法[2], ショット群のクラスタリングに加え, 顔検出結果を用いる手法[3]などがあるが, 実際のニュース映像では, トピックの先頭がアナウンスショットから始まらないことも多々ある。

テロップを用いる手法としては, ニュース映像から検出したテロップの中から, テロップサイズや位置, 表示時間に基づいてニュース冒頭に出現する「冒頭テロップ」を検出することで, トピックの先頭を見つける手法[1], [4]がある。しかし, 冒頭テロップを見つけることが出来たとしても, 全てのトピック先頭に冒頭テロップが出現するとは限らない。

字幕放送テキストを用いる手法としては, 形態素解析によって得た名詞列の語義属性(一般, 人物, 場所・組織, 時相)を解析し, 文毎に作成した属性別の出現頻度付きキーワードベクトルの類似度からトピック分割をする手法[5]があるが, 主要局により放送されるニュース映像のうち字幕放送テキストが付加されているのは40%~50%にすぎない。また, デジタル放送では字幕テキストとアナウンスの発話のずれが大きくなっており, 正確なトピックの境界を決めることは困難である。音声データを用いる手法もあるが, その成功の可否には音声認識精度の影響が大きいといえる。

他にも, トピックの変化点検出用の情報を予め付加して

おく方法[6]が提案されている。しかし, こうしたトピックの境界に関する情報を実際のニュース映像に付加して放送することは未だ実現していない。

本論文では, 各ニューストピック内で表示される, アナウンスショット, トピック対応テロップ, 冒頭テロップについて, その出現にパターンが存在することを利用してトピック境界検出を行う方法を提案する。3つの情報の出現パターンを利用することで, アナウンスショットや冒頭トピックの出現から始まるトピックだけでなく, トピック対応テロップの出現から始まるトピックについてもトピック境界を検出可能となる。さらに, トピック境界の検出に3情報の出現パターンとの一致を利用することで, トピック境界の誤検出を防ぐことが出来る。

ここで, トピック対応テロップとは, トピックの放送中にほぼ継続して表示されるテロップであり(図1), 人物や場所の名称, 時刻, 人物の発言内容とは表示時間やテロップサイズ等によって区別されるものである。



図1 トピック対応テロップ

また, 冒頭テロップとは, トピックの冒頭に出現するテロップのことであり, 例えば図2のようなテロップを指す。

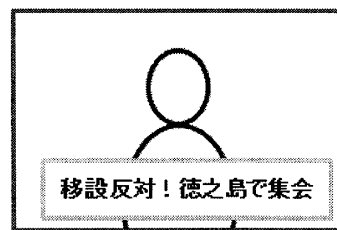


図2 冒頭テロップ

提案手法では, ニュース映像から, ①テロップを検出する処理と, ②アナウンスショットを検出する処理を行い, ③それらの処理結果から, 出現パターンを利用してトピック境界検出を行い, ④検出結果を用いてトピックの先頭にアクセスするファイルを生成する, という流れでトピック分割処理を行う(図3)。

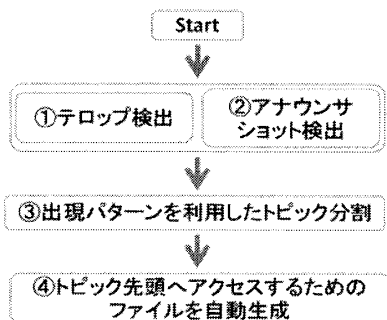


図3 提案するトピック分割手法の全体フロー

本論文では、第2節にテロップ検出等各ステップの詳細を、第3節では、実験結果を、第4節で考察、第5節にまとめを記載する。

2. 提案手法詳細

2.1 テロップ検出

テロップ検出を行う手法は、画像を復号後にエッジを検出することで行う手法と、MPEG データに含まれる種々のパラメータを利用する方法がよく用いられている。本研究では、対象とする映像が MPEG 符号化されていることを前提としているため、MPEG 符号化パラメータを利用する手法を提案する。

文献[7]では、符号化パラメータの中でも、DCT 係数と、動きベクトルを利用したテロップ検出方法が提案されているが、特定の時点のテロップ候補領域分布のみから、テロップ領域の判断を行うという問題があった。提案手法では、この問題を、テロップ検出を、テロップ領域の座標や表示開始フレームナンバーを同じテロップの重複がないように「テロップ候補リスト」へ登録する処理(登録処理)と、「テロップ候補リスト」に登録されているテロップ領域について表示が継続しているか判定する処理(表示継続判定処理)と、「テロップ候補リスト」に登録されているテロップ領域について、領域座標を修正する必要があるかどうかを判定する処理(領域修正処理)とを並行的に行うことで解決する(図4参照)。

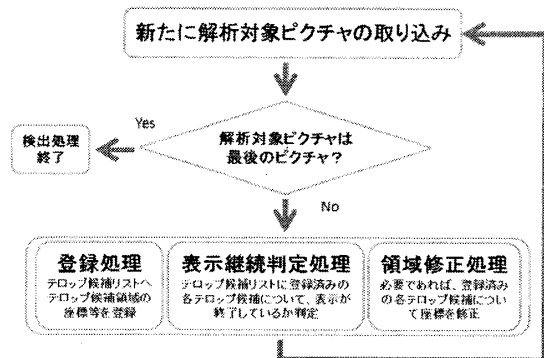


図4 テロップ検出フロー

登録処理では、動きベクトルの大きさが0、且つ DCT 係数の絶対値を閾値処理し、エッジが多く存在すると判別された MB の集合を「テロップ候補領域」として検出し、

時間的に連続するフレームにおいて、およそ同一の領域が複数回連続して「テロップ候補領域」として検出された場合に、「テロップ候補リスト」に登録する処理を行う。

表示継続判定処理では、「テロップ候補リスト」に登録済みの各テロップ候補領域について、その領域内における動きベクトルの大きさが0、且つ DCT 係数からエッジが多く存在すると判別された MB 数の割合が閾値以内であるかどうか毎フレーム毎に判定していく。そして、閾値以内であれば表示が継続していると、閾値未満であれば当該テロップ候補が消失したと判定する。消失時に、「テロップ候補」の表示継続時間が閾値以上であれば当該「テロップ候補」を、「テロップ」であると判別し、テロップ領域座標、表示開始・終了フレーム No 等の情報を記録する。

領域修正処理では、テロップ候補リストに登録済みのテロップ候補領域について、各候補領域毎に、領域座標を修正すべきかどうかの判定を行う。判定は、現処理中のフレームにおけるラベリング結果から分かる新たなテロップ候補領域と、テロップ候補リストに登録済みの候補領域とが重複するかどうか、また、重複するのであれば両候補領域の大きさの変化具合、両候補領域中の動きベクトルの大きさが0、且つ DCT 係数の絶対値からエッジが多く存在すると判別された MB 数の割合はどちらが大きいかといった点を基に行う。

なお、修正処理によって、テロップ出現時に偶然テロップ領域の一部のみを検出してしまった場合でも、正しい領域を検出出来る可能性が高まり、表示継続判定処理の精度が上昇すると共に、正しいテロップ画像が取得しやすくなる。

図5はテロップ領域修正処理の例である。図中青色の領域はテロップが存在する可能性が高いと判定された MB である。図の左側の画像は時刻 T におけるテロップ候補領域、右側の画像は時刻 $T + \Delta t$ におけるテロップ候補領域であり、テロップ領域の一部のみをテロップ候補領域として検出した場合でも、正しい領域に修正されていることが分かる。



図5 テロップ候補領域修正処理

2.2 アナウンサショット検出

アナウンサショットの検出には、前述したように顔検出結果の利用が近年主流となってきているが、提案手法でも顔検出の結果からアナウンサショットを検出する。

処理の流れは、解析対象の動画像に対し、最初にステップ①から④までの処理を行う。すなわち、①カット点検出を行い、ショットの先頭フレームにおいて顔検出を行う、②続いて、画面中にアスペクト比調整用の帯が存在するかどうかの判定を行う、③2.3で記載した手法によりテロップ検出を行い、それぞれのショットについて、テロップの有無や存在する場合はおおまかな位置を記録する、④カット点~続くカット点間で1フレームあたりの平均動き量を算出する(図6)。

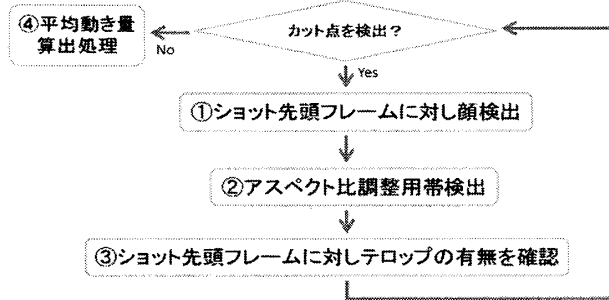


図6 アナウンサショット検出パラメータ取得フロー

動画像の終わりまで、当該処理を行ったら、上述した①～④の結果から、ショットの先頭フレーム画像内で顔検出がされているか、検出されている場合はその顔の座標と大きさは閾値以内か、画面中に帯は存在しないか、平均動き量は閾値以内か、テロップは存在しているか、存在している場合は画面中のどのあたりか、といった事項を条件として、アナウンサショットを見つける。

①の顔検出は、OpenCVのライブラリを用いて行った。OpenCVでは、人間の正面顔画像を学習させた識別器と、そのパラメータをXMLファイルにより提供しており、当該XMLファイルを読み込み、オブジェクト検出用の関数を用いることで画像から正面顔画像を抽出し、その座標を取得することができる。アナウンサショット判別のパラメータとして、顔領域の左上座標と顔領域面積を用いている(図7)。

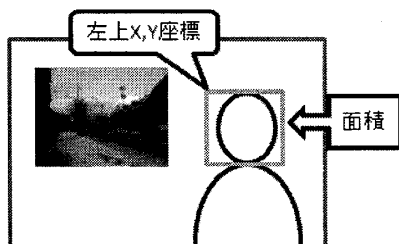


図7 アナウンサショット検出結果例

ところで、ニュース映像では、画面中に、アスペクト比を調節するための帯領域が表示されることが多い(図8)。一方、トピックの先頭となるアナウンサショットでは、アスペクト比調整用の帯が入ることはなく、顔検出結果から人物が写っているショットが検出された場合に、この帯の有無を判断することで、当該ショットがトピックの先頭となるアナウンサショットであるかどうかの判別をすることができる。アスペクト比調整用帯の有無の判定は、MBタイプを画面の左端及び右端からそれぞれ所定MB分カウントしていき、帯領域の場合に現れるSkippedタイプのMB数が閾値以上かどうか判定することで行った。

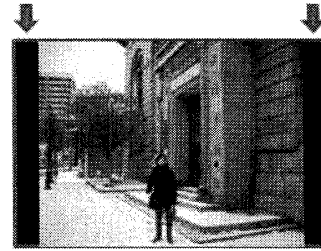


図8 アスペクト比調整用の帯領域

さらに、アナウンサショットでは比較的画像全体に変化がないため、ショット内に含まれる各フレーム毎の動きベクトルの値が小さいものとなる傾向がある。このため、ショット毎にフレームの平均動き量を算出していくことで、そのショットがアナウンサショットであるか判別する際の一助とすることが出来る。ただし、アナウンサショットでも動画像を含むものも存在するので、画面全体の動き量を考慮するのではなく、画面の所定ラインより下方部分での動き量を考慮するようにした。

また、トピックの先頭となるアナウンサショットでは、番組ロゴや時刻表示を除いてテロップが全く表示されていない状態から始まる。その後冒頭テロップが表示されることが多いが、先頭フレームではテロップは表示されていない。そこで、ショット先頭中にテロップが出現しているかどうかを判定することでトピック先頭に関するアナウンサショットであるかどうかを判定している。

2. 3 各種情報の調査結果について

ニュース番組には、5分～10分程度のものをいれると、NHK、NNN(日本テレビ)、JNN(TBS)、FNN(フジテレビ)、ANN(テレビ朝日)、TXN(テレビ東京)系列といった主要放送局で、1日あたり、合計50～60程度の番組がある。

今回は、そのうち、朝の時間帯と夜の時間帯、及び日中頻繁にある短いニュース、計17番組について調査をした。

(1) 冒頭テロップの有無について

16番組で冒頭テロップの存在が確認された。ただし、全トピックで冒頭テロップが入るのは短いニュース番組だけであり、30分以上の番組ではサブトピックを含めると、トピック先頭に冒頭テロップが入るのは半分に満たない。

(2) アナウンサショットの有無について

調査対象の全17番組でトピック冒頭がアナウンサショットである場合があることを確認した。ただし、アナウンサショットを介さずにトピックが変化する事例も多々ある。また、アナウンサショットでの背景やアナウンサ人数や、トピックの紹介をするアナウンサが1つのニュース中に変化することが多々あることがわかった。

アナウンサショットの背景は、大きくわけて、3パターン存在し、①ほぼ静止(スタジオ)、②ほぼ静止+ニュース動画、③報道部、であった。

また、アナウンサショットへ移行する際にみられる効果としては、瞬時に切り替わるもの、アングルの変化を伴うものが存在する。

背景については、番組時間長に関係なく①～③のパターンが存在したが、時間の短い番組では、アナウンサショットへの移行は全てのケースで瞬時に切り替わるものであ

った。一方、時間が30分以上の番組では、ときどきアングルの変化を伴いながらアナウンサショットへ移行する番組、毎回アングル切り替えを伴った移行する番組があった。

(3) トピック対応テロップの有無

17番組中、15番組でトピック対応テロップの存在を確認した。トピック対応テロップの位置は画面の左上、右上、左下、右下の4パターンが存在していた。なお、トピック対応テロップの位置は、1番組中に変化することもあった。

(4) トピック中における出現パターン

詳細にはもっと細かく分けることが出来るが、概ね以下の4パターンに分けられる。

・パターン1

時間の短いニュース番組に多くみられるパターンで、トピックが冒頭テロップの存在するアナウンサショットに始まり、続くショットでは、トピック対応テロップが出現しつつ、トピック内容を説明するショットが続いていく。

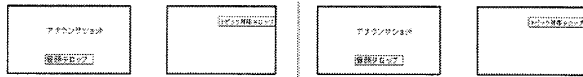


図9 パターン1

・パターン2

単独トピック連続紹介時・階層構造を有するトピックに多いパターンである。トピック連続紹介とは、アナウンサが様々なニュース項目を紹介する報道形態である。

こういった場合には、一番最初にアナウンサショットが入り、連続して各トピック（もしくはサブトピック）が紹介されていくため、図10のようなパターンになる。

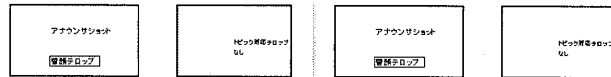


図10 パターン2

・パターン3

冒頭テロップから始まるものの、トピック対応テロップの存在しないトピック構造である(図11)。長時間の番組に入ることがある。

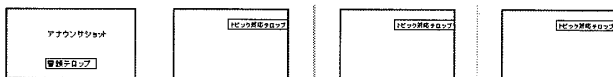


図11 パターン3

・パターン4

アナウンサショットから始まるものの、冒頭テロップが存在しない場合である(図12)。



図12 パターン4

3. 実験結果

3.1 実験環境

ニュース映像の解析には、OSがMacOSX10.4のPCを用いた。そして、上記したテロップ検出、アスペクト比調整用帯検出等の各種処理はMplayerのffmpegライブラリを用いて行った。具体的には、livavcodecのmpegvideo.cに各種検出用のプログラムを追加することで、debugモード時に行うようにしている。

Mplayerを用いている理由としては、debugモード時に特定の引数を指定することによって、DCT係数の抽出が比較的簡単にでき、また、動きベクトルを再生画像に重畳して表示する再生モードが用意されているといったように、各種のパラメータの取得が容易であることから用いた。

また、アナウンサショット検出のための顔検出は、OpenCV2.0のhigui等のライブラリを用いて行った。当該処理は、テロップ検出等と同じくdebugモード時に行った。

3.2 テロップ検出結果

NHK, 日本テレビ, TBS, フジテレビ, テレビ朝日, テレビ東京の5分~15分程度の短いニュースで、いずれもCMなしのものを用いた。また、天気予報コーナーについては検出対象外とした。

・正検出とする条件

テロップ領域の6割以上の領域をテロップ領域として検出し、テロップ表示中にロックを外さないこと。また、テロップイン点、アウト点の誤差が1秒未満である場合に正検出とした。

・未検出とする条件

テロップ領域の6割未満の領域をテロップ領域として検出や、テロップ表示中に誤ったテロップの消失判定をした場合、テロップイン点、アウト点の誤差が1秒以上である場合は未検出とした。

・誤検出とする条件

時刻、ロゴ等、背景の一部等、テロップ以外のものを検出した場合に誤検出とした。

・評価項目

Precision (適合率) = 正検出数 / (誤検出数 + 正検出数), Recall (再現率) = 正検出数 / 正解数, の値をもって評価を行った。当該数値はテロップ検出を含み検出結果の評価手法として一般的に用いられているものである。表1は各映像に対してトピック検出を行った結果である。

表1 トピック検出結果

VideoNo	正解数	正検出数	誤検出数	未検出数	recall	precision
1	26	23	3	3	0.88	0.88
2	9	6	0	3	0.67	1.00
3	42	38	2	4	0.90	0.95
4	24	21	0	3	0.88	1.00
5	11	7	0	4	0.64	1.00
合計	112	95	5	17	0.85	0.95

未検出事例としては、1フレーム中に存在する2つのテロップを1つと判断してしまったもの、時間的に連続してテロップが瞬時に変化する場合、1つのテロップとして判定してしまったもの、テロップ領域の一部しか検出できなかったものがあつた。誤検出事例としては、背景の一部や、時刻表示をテロップとして認識した場合がある。

3.3 アナウンサショット検出結果

・評価対象映像

NHK, 日本テレビ, TBS, フジテレビ, テレビ朝日, テレビ東京の5分~15分程度の短いニュース。基本的にはテロップ検出対象と同じ映像を用いたが、No17の映像がアナウンサショットが1つしかないといったような間

題があったため、アナウンサショットを多く含む No3 (30 分), No15 (20 分) に対してもアナウンサショット判定を行った。

・正検出とする条件

トピックの開始に関するアナウンサショットを検出した場合を正検出とした。

・未検出とする条件

トピックの開始に関するアナウンサショットを検出できなかった場合を正検出とした。

・誤検出とする条件

トピックに関係のないアナウンサショットを検出した場合や、同じアナウンサショットを複数検出した場合を誤検出とした。

・結果

テロップ評価と同じく、Precision (適合率) = 正検出数 / (誤検出数 + 正検出数), Recall (再現率) = 正検出数 / 正解数, の値をもって評価を行った。

表2は各映像に対してアナウンサショット検出を行った結果である。

表2 アナウンサショット検出結果

VideoNo	正解数	正検出数	誤検出数	未検出数	recall	precision
1	4	4	1	0	1.00	0.80
2	2	2	0	0	1.00	1.00
3	6	6	6	0	1.00	0.50
4	3	3	1	0	1.00	0.75
5	1	1	0	0	1.00	1.00
6	10	5	4	5	0.50	0.56
7	12	12	0	0	1.00	1.00
合計	38	33	12	5	0.87	0.73

誤検出の原因となったものの多くは、アナウンサショット以外で顔検出がされているもの、アナウンサショットではあるものの、トピックに対するコメントを述べているといったように、トピックの先頭に関係がないものであった。

また、未検出であった映像は、顔検出をしていないもの、番組ロゴをテロップとして検出していたために、「テロップの有無」条件で未検出となったものがあつた。

3.4 トピック分割結果の表示

現在のプログラムでは、トピックパターンを用いるところまでは実装出来ておらず、アナウンサショット検出結果と、トピック対応テロップを見つけた結果から、HTMLファイルを自動生成するようにしている。

(1) トピック分割方法

アナウンサショット検出結果と、トピック対応テロップを見つけた結果からトピックの境界を検出する。

①アナウンサショット検出結果の利用

2.4に記述した手法により、人物が検出され、更に所定の顔領域条件を満たし、また、帯領域やショットの先頭フレームにテロップの存在しないショットを、トピックに関するアナウンサショットとして検出する。本来であれば検出結果に対し、トピックパターンを考慮した上で、トピックの先頭に関するアナウンサショットであるか判断するのであるが、現在は、検出結果をそのまま利用している。

②トピック対応テロップの利用

2.3に記載した手法で見つけたトピックのうち、所定時間以上表示されているトピックについて、トピック対応テロップであると判断する。トピック対応テロップは、表示対象に関する情報を説明するテロップ (例えば、「今朝

9時官邸前」や、人物の名前) や、台詞テロップに比べて圧倒的に長時間表示されるため、単に時間的に閾値を設けるだけで簡単に見つけることができる。

(2) 分割結果の表示

分割結果を、テレビや携帯機器といった多くの機器で表示可能な形式であるhtml形式で出力する。

HTML5では、ブラウザの種類に応じて再生可能なファイルは異なるものの、予め動画の指定個所の再生を指示する関数が用意されているため、検出したトピック先頭のサムネイルへのローカルアドレスと、サムネイルが選択されたときに動画像の所定秒数箇所を指定する関数などを自動生成することが比較的簡単に出来る。図13は話題数の最も多かったVideoNo7について、自動生成したhtmlファイルの表示例である。また、図14は別の表示例であり、映像全体の長さに対応させたスクロールバーの下方にトピック境界に関するサムネイル画像の有無に応じた球状の画像を表示させている。図14の表示形態では、スクロールバー上のスライダー図形をマウスで移動させることに応じて、表示するサムネイルを変更することができる。

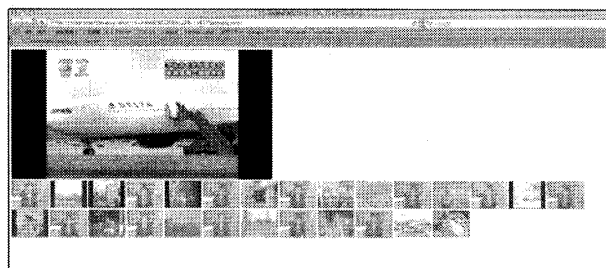


図13 htmlファイルの表示例



図14 htmlファイルの表示例2

4 考察

冒頭テロップ検出結果だけを利用するのではなく、トピック対応テロップの検出結果を利用することによって、冒頭テロップの検出結果のみを利用しただけではわからない、トピックについても検出された。例えば、図13のサムネイル一覧の2段目右端から3枚が1つのトピック(各地の年末行事)に関するサムネイルであるが、その中に、「文化遺産の大掃除」のサブトピックと、「正月用達磨の作製」のサブトピックがあり、どちらのサブトピックへもアクセス可能となっている。

現時点では、トピックの境界に関するアナウンサショットとして検出されたアナウンサショットや、検出されたトピック対応テロップを、単純にトピックの境界と判断しているが、トピックには階層構造を有するものが存在するので、それだけでは、どの階層におけるトピック境界かわか

らないといった問題が生じる。階層構造を有するトピックとしては、例えば、M 党の支持率低下という大きなニューストピックの下に、首相の贈与問題や、党内の意見不一致といった小さなトピック（以下、「サブトピック」とする）について連続して報道される場合が挙げられる。

また、トピックの境界に関するアナウンサショットとして検出されたアナウンサショットや、検出されたトピック対応テロップを、単純にトピックの境界と判断するだけでは、時間的に連続して検出されたトピックの境界に関するアナウンサショットについて、本当にトピックの境界に関するアナウンサショットであるのか判定出来ないという問題がある。

今後は、2.5で述べた、トピック対応テロップ、冒頭テロップ、アナウンサショットの出現パターンを利用することで、こうした問題を解消していきたい。

また、テロップ検出、アナウンサショット検出の誤検出、未検出のうち、特にトピックの境界判定に関するものについては改善する必要がある。

トピック対応テロップの誤検出事例として、時間的に連続したテロップを1つのテロップと判断したもの、アナウンサショットの未検出・誤検出事例として、アナウンサショットがアングルの変化を伴っている場合の未検出があった。今後は、[8]に記載された手法の適用等を通して改善を図ることも検討する。

5. まとめ

本論文では、ニュース映像に必ず存在している、アナウンサショットとテロップを用い、トピックの先頭に冒頭テロップが出現しない場合についてもトピック分割可能な手法を提案した。

従来手法で提案されている冒頭テロップ検出だけでなく、トピック中、長時間表示されるトピック対応テロップ、アナウンサショットを検出することによって、単に冒頭テロップを検出するだけではわからない、サブトピックについても検出することが可能となった。

今後は、冒頭テロップ、トピック対応テロップ、アナウンサショットの時間的な出現パターンを用いたトピック変化点検出を実装することによって、トピック変化点の検出精度を向上していきたい。

参考文献

- [1] 宮里肇：“冒頭テロップ検出によるニュース番組の自動構造化” PIONEER R&D Vol14.No1 pp72-78
- [2] 青木恒，児玉知也，岩田達明，山口昇：“MPEG-2 映像からのニュース番組高速構造化” 情報処理学会研究報告 2003-AVM-43 pp151-156,20031212
- [3] Akira Yanagawa, Winston Hsu, Shih-Fu Chang：“Anchor Shot Detection in TRECVID-2005 Broadcast News Videos,” Columbia ADVENT Technical Report, 2005
- [4] 加藤晴久，柳原広昌，中島康之：“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” 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] 2001(121), pp31-36, 20011213
- [5] 井出一郎，孟洋，片山紀生，佐藤真一：“大規模ニュース映像コーパスの意味構造解析” 電子情報通信学会技術研究報告．PRMU，パターン認識・メディア理解 103(296)，13-18, 20030902
- [6] 上原龍也，堀川将幸，住田一男：“見たいシーンにすぐアクセスできる番組ナビゲーション技術” 東芝レビュー 55 巻 10 月 pp17-19,2000.10
- [7] 倉橋誠：“MPEG 映像からのテロップ検出方法の検討” PIONEER R&D Vol15.No1,pp1-9
- [8] 樫尾隆亮，仲野豊，吉田俊之：“マルコフモデルを用いたテロップ候補領域に対する軟判定”，電子情報通信学会技術研究報告，SIS，スマートインフォメディアシステム 107(237)，pp45-48, 20070921