

冒頭テロップ検出によるニュース番組の自動構造化

Automatic segmentation of News programs by detecting the title open caption

宮里 肇

Hajime Miyasato

要旨近年TV放送の多チャンネル化や、HDD、DVD-RWといった蓄積メディアの低廉化、大容量化に伴い、ユーザが多くのコンテンツを手軽に蓄積できるようになった。しかしユーザがコンテンツ視聴に使える時間は限られている。そのため、見たいコンテンツだけを素早く見ることのできるシステムの実現が期待されている。

筆者は、冒頭テロップを検出することで、ニュース番組をトピックごとに自動構造化する手法を考案した。構造化されたニュース番組は、トピックごとのメニュー表示が可能となるので、ユーザは興味を持ったトピックのみを短時間で視聴できる。本稿では、このニュース構造化手順を説明し、実験結果を示す。

Summary Users have access to many TV programs easily in recent years with the benefit of large and cheap storage media, such as HDD, DVD-RW. However users are too busy to watch TV. Therefore, a system to play users' favorite content quickly would be useful.

The author proposes a method for segmenting News programs into every topic by detecting the title open caption. Segmented news program can generate a topic menu, then the user can see their favorite topic quickly by using the topic menu. In this paper, the author explains how to segment News programs, and shows the experimental results.

キーワード：自動構造化，ニュース番組，冒頭シーン，要約再生，テロップ

1. まえがき

近年TV放送のデジタル化や多チャンネル化に伴い、ユーザが入手できるコンテンツの数は増加の一途をたどっている。またハードディスクやDVD-RWといった蓄積メディアの低廉化、大容量化と相まって、ユーザが記録保存できるコンテンツの量は膨大なものになりつつある。しかしユーザがコンテンツ視聴に使える時間は限られている。そのため、見たいコンテンツだけを効率良く探すことができるシステムの実現

が期待されている。

このようなシステムを実現するには、まずコンテンツの構造化が必要である。例えばコンテンツのシーンごとにタイトルやコメントを付け、それをメニュー表示することで、ユーザはコンテンツ全てを視聴しなくても全体を把握できる。さらにメニュー項目から興味を持ったシーンを選択し、そのシーンからコンテンツを再生するアプリケーションを開発することも可能である。しかしコンテンツの構造化を人手で

行なった場合，非常に手間がかかるため，処理の自動化が強く望まれていた。そこで筆者は，ニュース番組を自動構造化する手法について研究を行なった。

2. ニュース番組の自動構造化

2.1 ニュース番組の一般的構造

図1に一般的なニュース番組の構成を示す。図示したように，一つのニュース番組は通常複数のニュース項目(以降トピック)から成る。さらに各トピックは大きく分けて「冒頭シーン」と「詳細シーン」の2つから構成される。冒頭シーンとは，アナウンサーがトピックの概要について説明するシーンである。一方詳細シーンとは，トピックについて細かく説明するシーンで，取材映像をバックにニュース内容を解説するシーンや，ニュース現場からのレポートシーンなど，多くの場合，複数のシーンから成っている。この2つのシーンは「冒頭シーン」「詳細シーン」の順に出現するので，各トピックの冒頭シーンを検出すれば，ニュース番組をトピックごとに構造化することが可能となる。

2.2 従来のニュース自動構造化手法

ニュース冒頭シーンを自動検出するための研究は，従来から広く行なわれている⁽¹⁾⁽²⁾。これらの研究では，冒頭シーン検出を実現するのに，以下のような特徴を抽出している。

- ・ 画面中央にアナウンサーの顔がある。
- ・ アナウンサーがニュース原稿を読み上げる。
- ・ 冒頭シーンの前後には，無音部分がある。
- ・ 冒頭シーン同士は，画面の類似度が高い。

しかし，これらの特徴を抽出するには，顔画像

認識や音声認識，話者認証といった複雑な処理を必要とする。さらにこれらの特徴は，ニュース番組によっては当てはまらない場合も多い。例えばアナウンサーの顔は画面中央にあるとは限らず，画面隅に存在することもある。冒頭シーン同士は画面の類似度が高い，という点も同様である。

2.3 本研究でのニュース自動構造化手法

ニュース番組を自動構造化するにあたり，筆者はニュース番組内の「冒頭テロップ」に注目した。冒頭テロップとは図2のように，ニュースの冒頭シーンに出現するタイトルテロップ(この例では「日経平均反発 1万5,000円台回復」のテロップ)のことである。このような冒頭テロップは，ニュース番組ごとにフォントや出現位置などの違いはあるものの，ほぼ全てのニュース冒頭シーンに見られるものである。そこで筆者はニュースに出現するさまざまなテロップから，このような冒頭テロップが自動検出できれば，ニュースの冒頭シーンを検出することが可能と考えた。



図2 冒頭テロップ

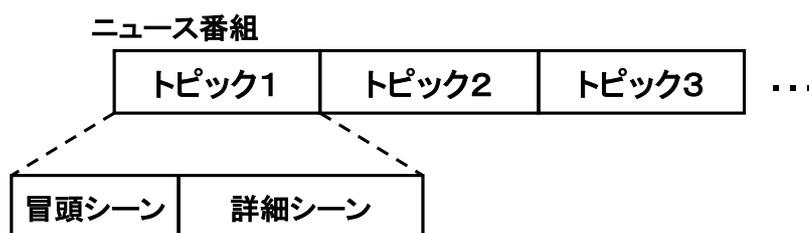


図1 ニュース番組の一般的な構造

3. 冒頭テロップ検出手法

冒頭テロップ検出手法を述べる。冒頭テロップ検出は、「テロップ検出」「テロップ分類」「冒頭テロップ判定」の3ステップで行なう。

3.1 テロップ検出

最初に、ニュース映像からエッジ(隣り合う画素の輝度差が急峻な部分)を検出する。画面内エッジ検出はラプラシアン法を用いる。つまり座標 (i, j) の画素の輝度を $f(i, j)$ とし、ラプラシアンオペレータ行列を A (式1)としたときの $\left| \sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} a(x, y) f(x, y) \right|$ の大きさにより、座標 (i, j) のエッジ判定を行なう。さらに得られたエッジを時間軸方向に観測し、位置が変わらないエッジ(静止エッジ)を検出する。ニュースのテロップは基本的に静止エッジで構成されるため、静止エッジの検出によりテロップ検出が可能となる。なお、画面内をスクロールするテロップは静止エッジとならないが、このようなテロップは冒頭テロップではないため無視する。

$$A = \begin{pmatrix} a(i-1, j-1) & a(i, j-1) & a(i+1, j-1) \\ a(i-1, j) & a(i, j) & a(i+1, j) \\ a(i-1, j+1) & a(i, j+1) & a(i+1, j+1) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (\text{式1})$$

次にエッジブロックの検出を行なう。エッジブロックとは、画面をブロック単位(8x8画素)

で分割したとき、内部の静止エッジが一定数以上のブロックを指す。テロップが消失するとき(つまり静止エッジが消失するとき)、エッジブロックもまた消失する。よって消失のタイミングが同じエッジブロックを、近距離のもの同士でまとめて整形(矩形化)することで、テロップ領域を特定できる。元映像からテロップ領域を特定するまでの流れを図3に示す。

テロップエッジ(テロップと背景画像の境界)は輝度の明暗がはっきりしているため、得られたテロップ領域の中から中間輝度エッジの除去を行い、テロップエッジのみを高精度に抽出する。具体的には、まず特定されたテロップ領域内で再度エッジ検出を行い、次にエッジ画素の輝度ヒストグラムを作成し、それを高輝度/中間輝度/低輝度に分類する。

分類には、大津の自動閾値選定法⁽³⁾を用いた。これは画像の輝度ヒストグラムを適応的に高/低輝度領域に二値化するものである。今回はこの手法を二段階適用し、図4のようにエッジ画素を三つの領域に分類した。テロップ領域のエッジのうち、この中間輝度からなるものを除去することにより、テロップエッジのみを高精度に抽出できる。

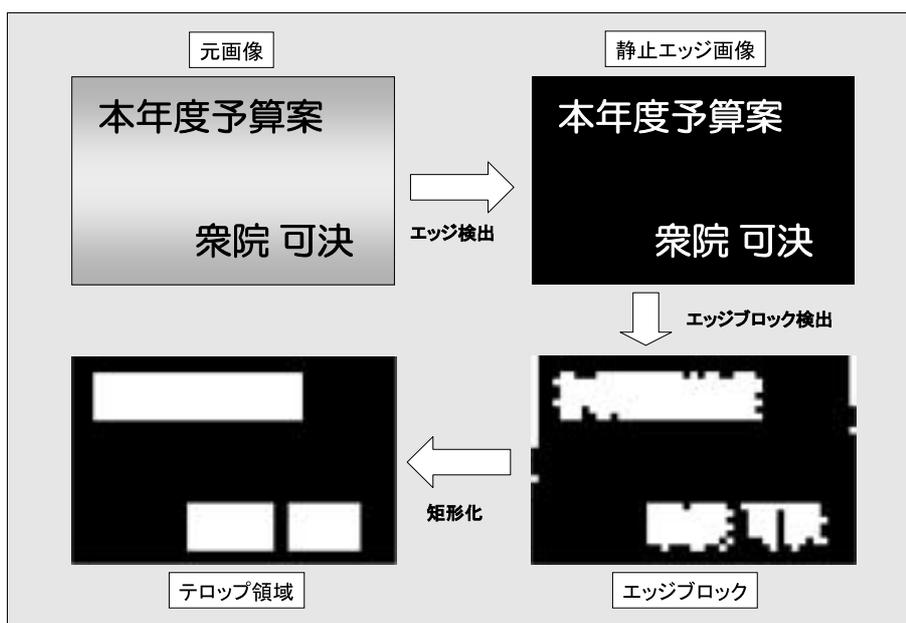


図3 テロップ領域の検出手法

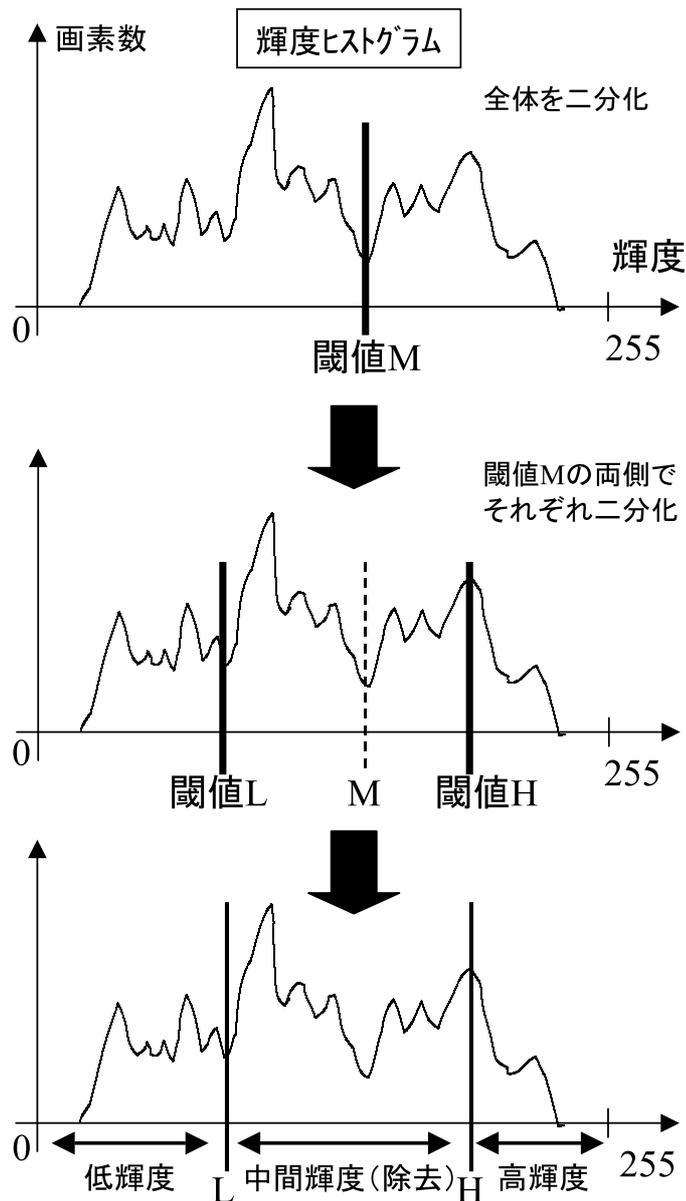


図4 輝度ヒストグラムの分類

3.2 テロップ分類

得られたテロップエッジを元に文字部分の抽出を行い、テロップ情報を取得する。ここでテロップ情報は、次の三要素からなる。

- ・ テロップ位置
- ・ 文字サイズ
- ・ 出現フレーム

次に検出された全てのテロップについて、テロップ位置と文字サイズのペアが等しいもの同士でテロップを分類する。

3.3 冒頭テロップ判定

検出された全てのテロップに対し、それが冒頭テロップであるかどうかの判定を行なう。冒頭テロップの判定にあたり、まず冒頭テロップの性質を明確にする必要がある。本研究では冒頭テロップの性質を以下のように定義した。

性質1：文字サイズが比較的大きい。

性質2：毎回ほぼ定位置に表示される。

性質3：複数回表示される。

性質4：画面下部に表示される。

性質 5：各冒頭テロップ間は，一定以上の時間間隔がある。

性質 6：ニュースの前半に出現しやすい。

これらの性質について補足する。

性質 1：

・ニュース番組には冒頭テロップの他に，アナウンサなどの人名テロップ，会話テロップ，CM テロップなどさまざまなテロップが存在する。冒頭テロップはトピックの内容を端的に分かりやすく視聴者に示すことを目的としているので，これら他のテロップと比べて目立たせる傾向がある。

性質 2：

・トピックごとに冒頭テロップの位置を変更することは殆ど無い。

性質 3：

・少なくとも 10 分以上のニュース番組であれば，トピックは複数あるため，冒頭テロップもまた複数存在する。

性質 4：

・冒頭シーンは，アナウンサのバーストショットとなる場合がほとんどである。そのため，アナウンサの顔を隠さないよう，冒頭テロップは通常画面下部に表示される。

性質 5：

・冒頭シーンの後には必ず詳細シーンが続くため，各冒頭テロップの間には一定以上の時間間隔がある。

性質 6：

・ニュース番組の後半には，天気予報など，ピック以外のコーナーが多く設けられる傾向がある。よって，冒頭テロップは，ニュースの前半に出現する可能性が高い。

先ほど分類したテロップ集合を，冒頭テロップの持つこれらの性質と照合することで冒頭テロップ判定を行なう。この結果，最も照合率の高いものを冒頭テロップとして決定する。

4. 実装

前述した冒頭テロップ検出手法をアルゴリズム化し，以下の PC 環境にて実装した。

- ・ OS : Microsoft Windows 2000
- ・ CPU : Intel Pentium4 2.0GHz
- ・ メモリ : 512MByte

また製品への機能搭載も考慮し，組込ベース (DSP) の実装も行なった。

- ・ DSP : TI 社製 TMS320C6711 (動作クロック 150MHz, 900MFLOPS)
- ・ メモリ : 外部 RAM 16MB
- ・ コードサイズ : 約 200KB

DSP 実装では，PC とほぼ同等のアルゴリズムを用いてリアルタイムでのテロップ検出および分類を実現した。DSP の負荷は 50% 程度であり，必要なメモリサイズ (実行コードを除く) はニュース 1 時間につき 2KB 程度であった。

なお処理の最適化を行なうことにより，処理負荷やコードサイズ等はさらに軽減できると考えられる。

5. 実験

実際のニュース番組から冒頭テロップを検出する実験を行なった。まず代表的なニュース番組を元に，表示位置や文字サイズの一致条件など各種パラメータの最適化を行った。そこで良好な結果が得られたことを確認したのち，この方式やパラメータの妥当性を見るため，在京 6 局で現在放送されている番組時間が 30 分以上のニュースを対象とし，冒頭テロップ検出実験を行った。結果を表 1 に示す。なお表中で

再現率 = 良検出数 / 冒頭テロップ数

適合率 = 良検出数 / (良検出数 + 誤検出数)

である。再現率は最小 72.7%，平均は 92.3% であった。

6. テロップを用いたメニュー表示

検出した冒頭テロップ情報を用いることで，ユーザが興味を持ったトピックのみを素早く視聴できるアプリケーションが実現できた。実現

例を図5に示す。

まず元となるニュース映像から、冒頭テロップ情報に基づいて冒頭テロップ画像を切り出しておき、それをボタンとした番組メニューを作成する(図の左側)。ここで各ボタンには、対応する冒頭シーンの再生開始時間を登録しておく。このようなアプリケーションを利用することで、ユーザはメニューから興味のあるトピックを選択し、そこからトピックを視聴できる。

番組メニューのボタンとして冒頭テロップの画像自身を用いるため、文字認識などの処理が

不要となる。

7. まとめ

冒頭テロップを検出することにより、ニュース番組の自動構造化を行なった。

処理内容はテロップ検出がメインであるため、従来のニュース自動構造化に関する手法と比べて処理負荷の軽いシステムとすることが出来た。今後は、精度改善など手法の改良と、新たな自動構造化手法について研究を進め、有効な手法を探っていきたいと思う。

表1 冒頭テロップ検出結果

放送局	ニュース	冒頭テロップ数	良検出	誤検出	再現率	適合率
A局	A-1	45	37	8	82.2%	82.2%
	A-2	7	6	1	85.7%	85.7%
	A-3	7	7	5	100.0%	58.3%
	A-4	36	34	5	94.4%	87.2%
	A-5	15	14	2	93.3%	87.5%
B局	B-1	8	8	3	100.0%	72.7%
	B-2	12	12	31	100.0%	27.9%
	B-3	11	8	1	72.7%	88.9%
	B-4	11	10	8	90.9%	55.6%
C局	C-1	5	5	37	100.0%	11.9%
	C-2	16	14	9	87.5%	60.9%
D局	D-1	15	15	0	100.0%	100.0%
	D-2	37	36	3	97.3%	92.3%
	D-3	7	7	1	100.0%	87.5%
E局	E-1	4	4	11	100.0%	26.7%
F局	F-1	12	12	9	100.0%	57.1%



図5 冒頭テロップを用いたメニュー表示

参 考 文 献

- (1) 有木: " DCT 特徴のクラスタリングに基づくニュース映像のカット検出と記事切り出し", 電子情報通信学会論文誌, Vol. J80-D- , No. 9, pp. 2421-2427, 1997.
- (2) 斎藤, 有木: " ニュース画像のデータベース化に向けて - ニューススタジオの映像検出と記事切りだし", 画像電子学会, 第149回研究会, 95-04-04, pp. 13-16, 1995.
- (3) 大津: " 判別および最小2乗基準に基づく自動しきい値選定法", 電子情報通信学会論文誌 Vol. J63-D, No. 4, pp. 349-356, 1980.
- (4) 宮里, 田畑: " テロップを用いたニュース番組の自動ハイライト作成", FIT (情報科学技術フォーラム) 一般講演論文集, 第3分冊, pp. 75-76, 2003.

筆 者

宮 里 肇 (みやさと はじめ)

所属: 研究開発本部 総合研究所 ストレージシステム研究部

入社年月: 1999年4月

主な経歴: MPEG-7 応用の研究開発を経て, 現在総合研究所にて, マルチメディアコンテンツ応用開発に従事。