

映像中の情報を用いたニュース映像要約手法の研究

永橋功丞[†] 宮岡伸一郎[†]東京工科大学 メディア学部メディア学科[†]

1. はじめに

近年、テレビの多チャンネル化や、インターネット動画の登場により、映像コンテンツの爆発的な増加がみられるようになってきた。しかしながら、それらの映像をすべて視聴する時間を確保することは現実的でないことも事実である。

膨大な映像コンテンツを効率よく視聴するためには、映像コンテンツ中の必要部分のみ抜き出し視聴できることが望ましい。

本研究では映像コンテンツ、特にニュース映像の自動要約を実現するシステムを構築する。

ニュース映像では、すべての映像を必要とされる場面は非常に少ない。ニュース映像の視聴者にとっては、自分に必要な情報においてのみ詳細が必要であり、それ以外の情報については概要のみ把握できれば良いと考えられる。そのために、ニュース映像中の各トピックの概要説明部分を抽出することで、ニュース映像の要約を実現することを目標とする。

2. 重要部分定義

宮里による研究^[1]を参考に、ニュース番組の重要部分の定義を行った。ニュース映像の構成は一般的には複数のトピックによって構成され、ニュースのトピックごとに、トピックの概略を説明する「冒頭部」と、そのトピックを詳細に説明する「詳細部」に分けられる。

「冒頭部」では、あるトピックに対して、おおよその説明がアナウンサーによってなされる。対して、「詳細部」は「冒頭部」で示されたトピックに対して、深い説明がなされる。

本研究ではニュース映像の全体を把握することのできる映像要約を目的としているため、「冒頭部」を重要部分と定義し、抽出することで映像要約を実現する。

「冒頭部」には共通して、「詳細部」と異なる以下のような特徴が存在する(図1)。



図1 冒頭部の例

(1)被写体となる人物は1人、ないしは2人。

“Summarizing Method for News Programs based on Video Information”

[†]Kousuke NAGAHASHI, Shinichiro MIYAOKA
School of Media Science, Tokyo University of Technology,
1404-1 Katakura-machi, Hachioji-shi, Tokyo 192-0982 Japan

- (2)被写体となる人物は、画面上でほぼ固定。
- (3)カメラはほぼ固定。
- (4)画面下部にテロップが挿入される
- (5)カットは一定以上の時間をもつ

これらの特徴を検出することで、重要部分の判定を行う。

3. 重要部分抽出手法

3.1 カット点検出手法

本研究では重要部分抽出の処理単位として、映像のカットを用いる。カット点の検出には、グレースケール化したフレームの2次の差分値を用いる。1次差分値を用いる場合、パン、およびチルト動作による誤検出の可能性があるが、パンおよびチルト動作時は一定速度で動くことが多いため、2次の差分値を利用することでこの影響を除外できると考えられる。

差分値は式(1)、式(2)によって求められる。

$$S'_t = S_t - S_{t-1} \quad (1)$$

$$S_t = \sum_x \sum_y |I(x, y, t) - I(x, y, t-1)| \quad (2)$$

ここで

t :再生フレームの時刻

$I(x, y, t)$: t フレームにおける画素 (x, y) の濃淡値である。
カット点における差分値の変化を次の図に示す(図2)。

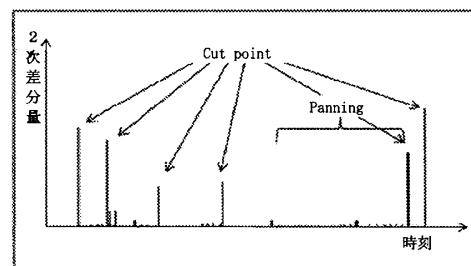


図2 カット点における差分変化グラフ

求められた差分値に対し、 $S'_t > 30$ の閾値を設定し、閾値を満たすものをカット点として抽出した。

重要部分は一定以上の長さをもったカットであるため、これにより検出されたカットのうち、8秒以上の長さをもつカットを、重要部分候補のカットとして検出をする。

3.2 人物領域認識手法

画面上の人物領域を認識する手法として、肌色領域抽出による手法を用いる。この手法には武岡らによる研究^[2]を参考にした。

ニュース番組においては人物が登場する際、顔部分がフレーム中に存在する。画面上の肌色領域を抽出し顔の存在を推定することで、人物の存在数、人物の動きの有無を推定する。

本研究では動画から取得したフレームをYCbCr表色系に変換し、肌色領域抽出を行った。

肌色画素の閾値を式(4)のように設定し、抽出を行った。結果を図3に示す。

$$\begin{aligned} 80 < Cb < 120 \\ 140 < Cr < 160 \\ 100 < Y \end{aligned} \quad (4)$$

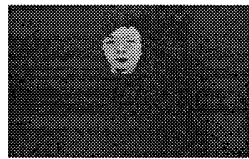


図3 肌色領域の抽出

得られた肌色領域のカット内での移動量を求め、移動量が一定以下となるカットを重要部分候補として検出した。

3.3 カメラワーク・テロップ検出手法

映像中のカメラワークとテロップの認識の手法について述べる。本研究ではオプティカルフローを用いて、この2つの処理を同時に行う手法を考案した。

重要部分ではカメラワークが存在せず、画面下部にテロップが挿入されるという特徴があるため、画面上部にはオプティカルフローが存在せず、画面下部約1/4の領域にのみ強いオプティカルフローが出現する。この特徴を検出することで、重要部分の判定を行う。

ただし、画面上部のフローを検出する際、出演者の顔の動きなどにより適切な結果を得ることができない。この問題を解決するために、画面中央80%の領域内のフローを除外した(図4)。

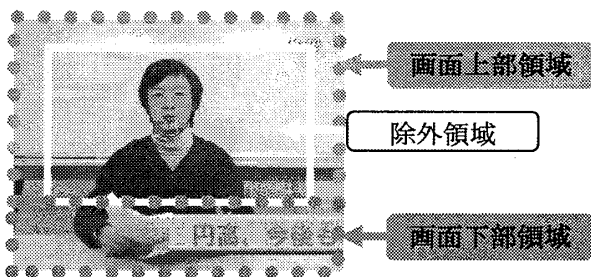


図4 オプティカルフローの領域設定

画面上部、下部それぞれのフローの平均を計算し、下部にのみ強いフローが現れた場合、そのカットを重要部分候補として検出する。

4. 実験・評価

4.1 カット点検出実験

実際のニュース映像を用いて、カット点検出を行った実験結果を次に示す(表1)。

表1 カット検出数

	実際の カット数	再現率 (良検出数)	適合率 (総検出数)
番組A	78	100%(78)	92.8%(84)
番組B	122	97.6%(119)	88%(127)

カット点検出実験の結果、再現率は比較的良好な結果が得られたが、適合率はそれほど高くはない。これは番組中のVTRなどでフラッシュが焚かれた際に、カット点と誤認し、誤検出となってしまう場合が多いためである。

4.2 映像要約実験

人物認識、カメラワーク・テロップ検出を利用した映像要約の実験結果を以下に示す(表2)。

表2 映像要約実験結果

	実際の 重要部分数	再現率 (良検出数)	適合率 (総検出数)
番組C	9	88.9%(8)	88.9%(9)
番組D	17	76.5%(13)	72.2%(15)

映像要約実験では、番組によっては良好な結果を得ることができた。

番組Dでは番組の後半中の重要部分に、人物以外の肌色領域が画面中に存在してしまい検出漏れの要因となった。また、番組C、Dともにオプティカルフローの誤検出が多く、重要部分検出にあまり寄与していない結果となった。

5. おわりに

本研究では、映像中から得られる情報のみを利用した、ニュース番組の自動要約を行った。

実験の結果、カット点の分割は良好な結果が得られたものの、生成した要約映像については検出漏れが多く、良い結果を得られていない。

今後は顔検出やテロップ検出の精度向上を図り、要約映像生成の精度向上の手法を探っていきたいと思う。

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, No.1, pp.72-78, 2004
- [2] 武岡 さおり, 尾崎 正弘, 足立 義則: 個人認証のための顔画像抽出と顔方向の自動認識. 名古屋女子大学紀要, 50, (人・社), pp145-151, 2004