

カット・字幕・カメラワークに基づく映像カテゴリ分類と映像要約への応用に関する研究

白石 大輝

指導教員：村上 和人

1. はじめに

近年、自動映像要約に関する研究が進められている。Ronfard らの先行研究[1]では、映像をショットと呼ばれる単位に分割し、ショットごとに代表フレームを決定するという方法を提案している。また、田中らは[2]で、野球映像に特化した方法を提案している。しかしこれらの先行研究には、①編集者の意図が反映されにくい、②限定的な映像にしか適用できない、という二つの問題点がある。

そこで本研究では、映像のカテゴリ分類を行い、それぞれ異なる要約方法を用いるというアイデアのもとに、3つの映像特徴を用い、5つのカテゴリに対して映像要約を行った。そして、30人の被験者に手動映像要約を行ってもらった結果を真値とし、自動映像要約の結果と比較すると、約6割類似しているという成果を得た。

2. 基本アイデア

本研究では、映像特徴の中から「カット」、「字幕」、「カメラワーク」に着目した。それに基づき、カテゴリ分類を行い、自動映像要約に利用する。その概要を図1に示す。

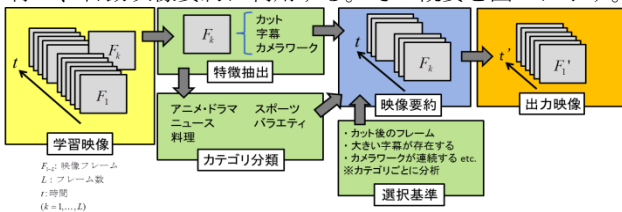


図 1. 研究概要

使用する3つの特徴は、先行研究[3]～[5]を参考に取得する。まずカットは、カットが存在するフレーム間は、その差異が大きいことを利用する。フレーム間差分が閾値以上で、そのフレーム間をカットであると判断する。

字幕は、映像のフレームを \$A \times B\$ のブロックに分割し、ブロック内のエッジ密集度が閾値以上であり、それを満たすブロックが一定時間以上継続する場合存在するとする。

カメラワークは、カメラワークが存在するフレームはある方向へのフローが大きいことを利用する。オブティカルフローを求めて4方向へ分類し、それぞれフローの総和を求める。そして最大の物とその他の物の差分を取り、その値が閾値以上でカメラワークがあると判断する。

3. 研究手法

3.1 カテゴリ分類の手順

本研究では、多数存在する映像カテゴリの中から、代表的と思われる5つのカテゴリ「アニメ・ドラマ」、「スポーツ」、「ニュース」、「バラエティ」、「料理」に注目し、これらの分類を目指す。まず、カテゴリ分類におけるパラメータ決めのための学習を行う。

＜学習手順＞

Step. 1 学習データの特徴抽出

カット、字幕、カメラワークを検出、それらが映像内に含まれる割合を計算する。

Step. 2 学習データから領域分割を行う

取得した特徴を正規化し、ボロノイ分割を行う。

次に、学習した領域分割を利用し、カテゴリ分類を行うための手順を示す。

＜映像特徴に基づくカテゴリ分類の手順＞

Step1. 入力データの特徴抽出

カテゴリ分類の対象となる入力データの特徴を抽出、映像中に含まれる割合を算出する。

Step2. カテゴリの決定

入力データがボロノイ分割されたどの領域に属するか調べることで、カテゴリを決定する。

3.2 映像要約の手順

30人の被験者に手動による映像要約を行ってもらい、その際どの映像特徴に注目が集まったのか、カテゴリごとに分析を行った。その結果、表1のようになった。

表 1. 被験者が注目する映像特徴

アニメ・ドラマ	カット、カメラワーク：注目度高
スポーツ	カメラワークが連続する部分：注目度高 字幕：注目度低
ニュース	カメラワーク：注目度低
バラエティ	字幕：注目度高 カメラワーク：注目度低
料理	小さい字幕：注目度高

また、以下に自動映像要約の手順を示す。

＜映像特徴に基づく自動映像要約の手順＞

Step1. カテゴリ分類を行う

入力データを5つのカテゴリに分類する。

Step2. 入力データの特徴抽出

3つの特徴を抽出、それらが映像内に含まれる割合を算出する。

Step3. 各フレームの重要性を定める

各フレームの重要性を重要度と名付け、以下の計算で求める。ここで、重要度を \$I\$、各特徴による重要度を \$I_{1\sim3}\$、それを正規化したものを \$I'_{1\sim3}\$、各特徴の重要度における重みを \$w_{1\sim3}\$ とする。

$$I = (w_1 I'_1 + w_2 I'_2 + w_3 I'_3) / (w_1 + w_2 + w_3) \quad (1)$$

式(1)における重みを、表1を基に変化させ、カテゴリごとに異なる要約方法を適用する。

Step4. 重要性が高いフレームを選択する

式(1)の結果から、先行研究[6]におけるフレーム選択方法を用い、重要フレームを抜き出す。

4. 実験

4.1 カテゴリ分類の評価実験

学習データを各カテゴリ10本の計50本、実験データを各カテゴリ10本の計50本とし、カテゴリ分類を行った結果を表2に示す。それぞれ、何本の映像のうち何本カテゴリ分類を成功したかを表している。

表 2. カテゴリ分類の結果

全体	50本中38本	アニメ・ドラマ	10本中6本
スポーツ	10本中7本	ニュース	10本中7本
バラエティ	10本中10本	料理	10本中8本

4.2 映像要約の評価実験

手動による要約映像と自動による要約映像がどれだけ類似しているか比較することで評価を行う。より多くの被験者から選ばれたフレームを真値とし、自動要約映像がどれだけ類似しているか示す割合を類似率と表現する。

実験を行うにあたり、式(1)における重要度算出式の重みを、表1を参考に以下のように定める。

アニメ・ドラマ	:	$w_1 = 1.5$	$w_2 = 1.0$	$w_3 = 1.5$
スポーツ	:	$w_1 = 1.0$	$w_2 = 0.5$	$w_3 = 2.0$
ニュース	:	$w_1 = 1.0$	$w_2 = 1.0$	$w_3 = 0.5$
バラエティ	:	$w_1 = 1.0$	$w_2 = 1.5$	$w_3 = 0.5$
料理	:	$w_1 = 1.0$	$w_2 = 2.0$	$w_3 = 1.0$

これらより、各カテゴリ10本、計50本の映像に対し自動映像要約を行い、類似率を算出した結果が表3のようになる。

表3. 自動映像要約の結果

カテゴリ分類が成功した映像	58%
カテゴリ分類が失敗した映像	40%

カテゴリ分類が失敗した映像より成功した映像の方が高い値を示しており、カテゴリ分類を行った方がより良い結果を得ることができると分かる。

5. 考察

5.1 カテゴリ分類の学習機能

4節におけるカテゴリ分類では、学習回数は各カテゴリ10回であった。学習回数の増加による精度向上が見られれば、学習機能を備えたカテゴリ分類方法だといえる。そこで、各カテゴリ学習回数を1回～40回で変化させ、カテゴリ分類の正答率を確認する。図3に結果を示す。

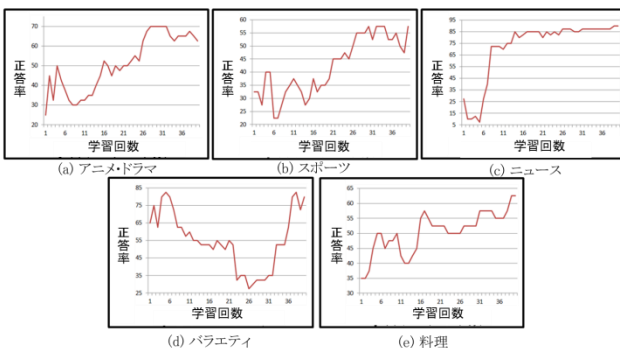


図3. カテゴリ分類の正答率

どのカテゴリにおいても、学習回数の増加に伴い、カテゴリ分類の正答率が向上していることが分かる。バラエティカテゴリのみ一度低下した後向上しているが、これは、学習データにおける特徴を含む割合にバラつきがあるためである。しかし、最終的に向上しているということは、やはり学習回数が増えれば精度が向上するといえる。ゆえに、本研究におけるカテゴリ分類は、学習機能を備え、使用すればするほど精度が向上する手法だと言える。

5.2 使用特徴の有効性

本研究は、「カット」、「字幕」、「カメラワーク」に基づき行ってきたが、それら全てに有効性があるか確認する必要がある。そこで、使用特徴の違いによる、カテゴリ分類の正答率の変化を確認した。図4～6に、学習回数を変化させた際の、入力データのカテゴリ別におけるカテゴリ分類の正答率を示す。

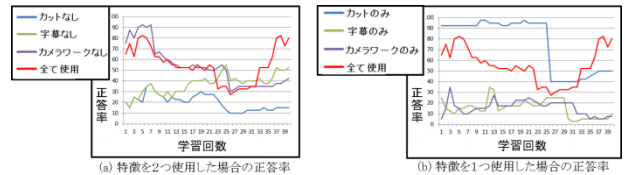


図4. バラエティの使用特徴の違いによるカテゴリ分類正答率

図4(a)を見ると、学習回数が増えた後では、カットを使用しない場合の正答率が、3つ全てを使用する場合と比較して大きく低下していることが分かる。また、図4(b)を見ると、カットのみを使用した場合は、他の特徴のみを使用した場合よりも正答率が高くなっていることが分かる。これらより、カットの有無によって正答率に影響を及ぼすので、カットの有効性は認められたといえる。

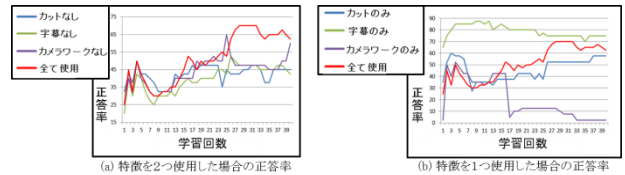


図5. アニメ・ドラマの使用特徴の違いによるカテゴリ分類正答率

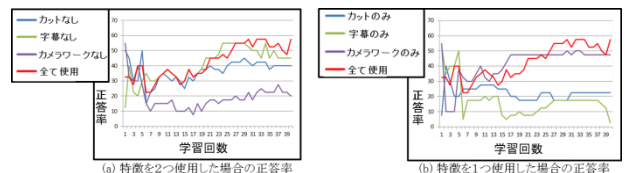


図6. スポーツの使用特徴の違いによるカテゴリ分類正答率

同様に字幕に関しては図5より、カメラワークに関しては図6より有効性が認められる。これらより、本研究で使用した3つの特徴「カット」、「字幕」、「カメラワーク」は、カテゴリ分類に有効な特徴といえる。

6. おわりに

本研究では、映像特徴「カット」、「字幕」、「カメラワーク」を用い、それらがカテゴリ分類、自動映像要約に有効であることを明らかにした。また、学習回数を増やすことでカテゴリ分類の精度が向上したことから、学習機能を備えたカテゴリ分類手法を実現したといえる。

本研究で利用した特徴は3つ、対象としたカテゴリは5つであるが、今後はより多くのカテゴリに対し、より高い精度で分類、映像要約を行える手法を目指す必要がある。今後の課題として、①重要度付加における各特徴の重みの決定方法に、学習機能を備え精度向上を図ること、②カテゴリの追加と細分化、③映像特徴の取得における精度向上、ということが必要である。

参考文献

- [1] R. Ronfard, "Shot-level description and matching of video content", Proc. SPIE, vol. 3229, pp. 70-78, 1997
- [2] 田中 孝志, 小方 善貴, 片岡 充照, 黒木 修隆, 沼 昌宏 "野球中継番組におけるテロップ情報を用いたダイジェスト映像自動生成" 映像情報メディア学会技術報告 33(54), 25-28, 2009-12-03
- [3] 長坂 晃朗, 田中 謙 "カラービデオ映像における自動索引付け法と物体探索法" 情報処理学会論文誌 Vol.33 No.4, 1992
- [4] 新井 啓之, 桑野 秀豪, 倉掛 正治, 杉村 利明 "映像中のテロップ表示フレーム検出方法" 電子情報通信学会論文誌 2006/6 Vol.J83-D-II No.6
- [5] 三浦 宏一, 浜田 玲子, 井出 一郎, 坂井 修一, 田中 英彦 "動きに基づく料理映像の自動要約" 情報処理学会論文誌 Vol.44 No.SIG 9(CVIM 7), 2003
- [6] 白石 大輝, 村上 和人 "映像要約のための画像特徴抽出に関する一考察", ITE Technical Report vol.35, No.20, ME2011-58