

画像処理を用いた映像の時間的構造化

6D-8

村田 充弘†
筑波大学

中村 裕一†
理工学研究科†

大田 友一†
電子・情報工学系†

1 はじめに

現在、マルチメディア情報処理の分野においても画像情報を扱う必要性が増し、静止画像だけではなく動画像（映像）を扱うことが一般的になりつつある。映像では、静止画とは異なり、時間的な変化—例えば動きや話し

の展開など—を表現することができる。また、映像に記録されているカメラ操作やカットのつなぎ方、テロップの表示方法などに撮影者の意図が込められている。本研究では、画像処理によって、これらの手がかりを抽出してカット間の関係や意味を推定し、それを基に映像を構造化することを試みた。映像の素材としては、撮影者の意図の明確なCM映像を用いた。

2 映像の時間的構造とそのモデル化

映像は、「カット」の組み合わせで構成されており（詳しくは3.1を参照）、映像の時間的構造はカットの時系列を基本として考えることができる。さらに、カットはそれぞれ意味—例えばCM映像における商品の機能説明—を持っており、その組合せによって映像全体の意味的構造が形成されている。

そこで本研究では、各カットが映像で果たす役割を分類し、これらのカットの種類や順序などの組合せパターンを解析することにより、映像を構造化する。そのために、まず、約150本のCMを手動で解析し、典型的なパターンをモデル化した。

2.1 CM映像のカット間の関係モデル

まず、商品を最もアピールする場面の前後でCMを区切り、表1に挙げるような特徴および出現位置に着目し、カットの分類を行った。その結果、カットの状態を表2のように定義すると、これらの特徴の有無との間に、表3に挙げるような対応関係があることが分かった。ただし、表中の「○」は画像中に存在する、「×」は存在しない、「*」はdon't careを表す。

表 1: CM映像中の特徴

特徴	意味
人物	映像中に映る登場人物のこと。
テロップ	画面中に出る文字のこと。

次に、表3から、図1に示すようなカット間の関係モデルを作成した。図中の「テロップ」は、次のカットにおいてテロップがある場合にはその矢線の先の状態が接続することを示し、「~テロップ」はテロップがない場合に接続する状態を示す。なお、これらの状態の変化は「背景の変化」があることを前提条件としている。

表 2: CM映像のカットの状態

状態	意味
タイトル・商品名	シリーズ化しているCMにおけるタイトル、商品名を提示。
エピソード	商品とは直接には関係ない映像。商品との印象を重ね合わせるための映像などもここに含まれる。
商品外観	商品の外観を提示。
機能説明	商品の機能や、使用方法、特徴などを提示。
決め	人物が商品の使用後の感想やキャッチコピーなどを言うなどして、商品の印象をアピールする。
商品説明	商品を最もアピールする場面。テロップ、音声を伴う。
念押し	商品説明があった後に出てくる「決め」。商品の印象をさらに強調。
企業説明	商品を提供する企業の紹介。

2.2 CM映像の主張とモデルの関係

商品のアピールを行っているCM映像について、その目的、主張、媒体の点に関して分類を行うと、図2のようになる。

例えば、商品の「存在」を主張として「商品のアピール」をするCMの場合、「外観」を媒体にする場合と「名前」を媒体にする場合があり、「名前」としては、商品の「名称」自体と商品の「案内」によってアピールする

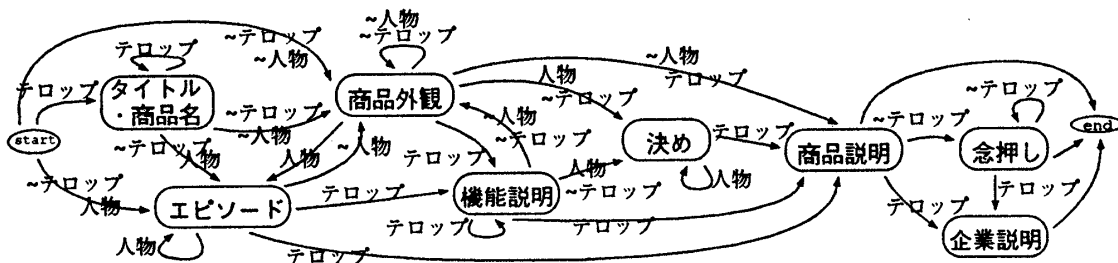


図 1: CM映像のカット間の関係モデル

Temporal structure analysis of movie using image processing techniques

Mitsuhiro MURATA†, Yuichi NAKAMURA†, Yuichi OHTA†

Master's program of Sciences and Engineering†, Institute of Information Sciences and Electronics†, University of Tsukuba

表 3: CM 映像中の特徴とカットの状態の対応関係

a) 商品説明前			
人物	テロップ	出現位置	対応する状態
*	○	先頭のみ	タイトル・商品名
○	×		エピソード
×	×		商品外観
*	○	先頭には出ない	機能説明
○	×	末尾のみ	決め

b) 商品説明後			
人物	テロップ	出現位置	対応する状態
○	*		念押し
×	○	末尾のみ	企業説明
×	×		(このようなカットはない)

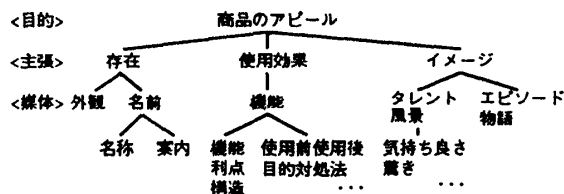


図 2: CM 映像の意味の分類結果

場合の2種類があることを示す。主張毎にCMを分類すると、表4に示すような共通点があることが分かった。

表 4: CM 映像の主張毎の共通点

主張	共通点
存在	必ず商品外観を提示するカットが存在する。
使用効果	必ず商品の機能説明を行うカットが存在する。
イメージ	エピソードカットのみで、商品外観や機能説明のカットが一つもない。

これより、映像中のカット間の関係から、映像の持つ主張が推定できると考えられる。

3 映像の解析のための画像処理

前章で述べたような映像の構成を計算機で自動的に解析するために、まず、カットの境界であるカット変わりの検出を行い、その次に、関係モデルの入力となっている人物の検出、背景変化の検出を行う。

3.1 カット変わりの検出

ここでは、我々が開発したカット変わり検出法 [1] によって表5に示すようなカット変わりの検出を行う。

表 5: カット変わりの手法とその特徴 [2]

手法	特徴
瞬間的切り替え	1フレームで画像全体を入れ替える。
フェードイン	暗い画面から画像を徐々に表す。
フェードアウト	画像を徐々に消して暗い画面にする。
ディゾルブ	画像をフェードアウトすると同時に次の画像をフェードインする。

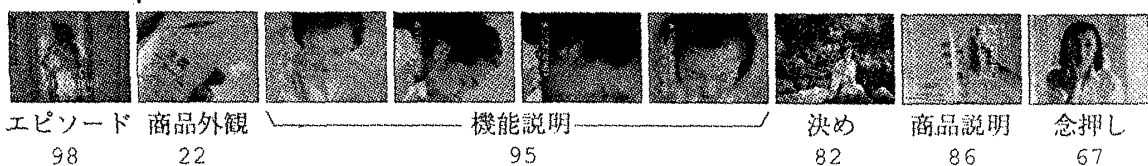


図 3: CM 映像のカットの状態の推定結果

3.2 人物の検出

CM 映像において、人物は商品の説明や商品イメージのアピールなどの役割を果たしている。その結果、2.1で述べたような構造化を行うための重要な手がかりとなる。人物の検出は顔領域を検出することで行う。まず、(R-B)色の画像から適当な肌色の存在範囲でしきい値処理して二値化し、ラベル付けを行って画像を領域分割する。そのうち、画面の幅を1:1に内分し、高さを1:1.6に内分する点に近い領域を顔領域とする。

3.3 背景変化の検出

CM 映像の場合、使われている場面の数はそれほど多くなく、場面が変化したかどうかは、背景が変わったかどうかで判定できる。ここでは、画面の端から50ピクセルの幅の領域を背景領域とし、背景領域で使われている色間の距離がしきい値を越えたとき、背景変化が起こったと判定する。

4 映像の時間的構造の推定

入力は、15秒(450フレーム)のCM映像とした。3で述べたように、まず、カット変わり検出を行い、カットに分割する。次に、各カット毎に人物の検出を行い、人物の有無を判定する。文字の有無に関しては手入力したデータを用いた。これを基に、図1に示す関係モデルを用いて、各カットの状態を求める。また、表4の関係をj用いて、映像の主張を推定する。ここでは、「商品外観」「機能説明」「エピソード」に割り当てられたカット長の割合から、各主張の強度を推定する。

実験結果を図3に示す。各カットの下に推定された状態およびカット数を示す。この結果からこのCMの主張は、「イメージ」0.46(98/215)、「存在」0.10(22/215)、「使用効果」0.44(95/215)となった。

5 まとめ

映像の時間的構造を自動的に解析する方法としてカット間の関係モデルを提案した。CM映像の入力に対し、自動的にカット変わりを検出してカットに分割し、人物を自動的に検出して、手入力によるデータを付加し、時間的構造を推定した。今後の課題としては、顔領域の色の範囲の自動設定、文字の有無の自動判定、音声情報の付加などが挙げられる。

参考文献

- [1] 村田, 中村, 大田: “映像におけるカット変わりの自動検出 — フェードイン、フェードアウト、ディゾルブの場合 —”, 情報大全 50(前), 6D-7, 1995
- [2] 栗田, 高橋, 竹下: “画像のソフトウェア”, テレビジョン学会編, 画像エレクトロニクス講座 8, コロナ社, 1980