

# 字幕の語義に基づくニュース映像の 分類・索引付け手法

## Classification and indexing method for news video based on semantic attributes of captions

井手 一郎  
Ichiro IDE

山本 晃司  
Koji YAMAMOTO

田中 英彦  
Hidehiko TANAKA

東京大学大学院 工学系研究科  
Graduate School of Engineering, The University of Tokyo

### 1 はじめに

様々な媒体を通じて提供されるマルチメディア、特に映像情報の量は膨大であり、将来の再利用や検索を考慮すると、内容を反映した索引付けを行なって蓄積しておく必要がある。しかし、膨大な量の映像に対して、内容を反映した索引付けを行なうためには相応の労力を要し、従来のように人手で行い続けるのは現実的ではない。本研究ではこのような需要を背景として、映像への自動的な索引付け手法を提案する。また、再利用や検索の需要の大きさという点から、ニュース映像を対象とする。

映像への自動的な索引付けに関する研究は、旧来の画像に関する詳細な背景知識に頼る手法のほか、近年、映像に様々な形で付随する自然言語情報を利用する試みが行なわれている。その中でも CMU における Informedia プロジェクト [4] は既に商用化水準に達しつつあるが、これを始めとする多くの類似研究において、索引語は話題単位、あるいは出現時点(発声・表示された時点)に、出現頻度や珍しさなどの基準に基づいて付与されることが多く、映像内容を忠実に反映して、かつ時間的粒度の細かい索引付けはあまり行なわれていないのが現状である。

筆者らは、ショットをいくつかの典型(会見・集合など)に分類することにより、映像内容をおおまかに推定した後、各典型毎に必要な索引語を字幕などから抽出する手法を提案し、簡単な実験において評価した [2]。これにより、ショットという時間的粒度の細かい区間への、映像内容に忠実な索引付けが可能となる。同様の手法は中村ら [3] も提案しているが、これらのように典型的ショットへの分類では、分類基準を天下一的に与えるため、限られた分類にとどまらざるを得ない。

そこで本研究では、字幕の語義の自然言語的分類に基づいてショットを分類し、各分類に属す映像から、自動的に分類基準(分類を代表する画像的特徴量)を統計的に収集する手法を提案する。また、ひとたびこのような分類基準が得られれば、字幕が存在しないショットに関しても、特徴量ベクトルの類似度から、どの分類にどの程度の確からしきで含まれるかを求めることにより、より汎用的な索引付け手法となり得る。

#### 1.1 用語の定義

一般に映像は、図 1 に示すような階層構造として定義される。本研究で主に扱うショットとは、画像的に切れ目のない連続した静止画像(フレーム)の集合のことであり、ほぼ

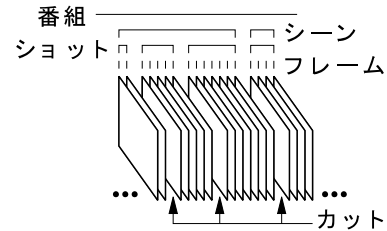


図 1: 映像の構成と用語の定義

同一映像内容が継続する最小素片とみなせる。また、ショット同士の境界をカットと呼ぶ。

### 2 字幕の語義に基づく索引付け

以下に示すような 2 段階の手順により、索引付けの自動化を行う。

#### 1. 各語義分類を代表する画像的特徴量の収集

(1) 字幕の語義分類を解析するとともに、(2) ショットの特徴量ベクトルを計算し、(1) で得られた語義分類を代表する特徴量へ統計的に反映させる。

#### 2. 各ショットの分類

字幕そのものの他に、ショットの特徴量ベクトルと各語義分類を代表する特徴量との類似度を求め、類似度の高い語義分類を索引として付与する。

以下の 2 節で各々について詳しく記す。

#### 2.1 語義分類毎の画像的特徴量ベクトルの収集

以前から言語学の分野において、語を概念(意味)により分類する試みが行われてきた。本研究では、その中でも定評のある分類語彙表 [5] を用いて、字幕の語義を分類する。分類語彙表は、36,780 語が 798 項目に分類されている辞書であり、カタカナ語などの収録数は少ないものの、ニュース字幕に含まれる語を処理するには概ね十分な語彙をもつ。

図 2 に、分類語彙表の各分類項目を代表する画像的特徴量を統計的に収集する様子を示す。

この例が示すように、以下の手順に従い、十分な量の字幕付きショットを解析することにより、字幕の語義分類と画像の特徴量ベクトルとの組を求め、各分類項目を代表する特徴量を統計的に収集する。

#### 1. 字幕を解析

日本語の字幕の場合、末尾に存在する名詞に注目することにより、字幕全体の属性が解析できる [2]。そこで、字幕を形態素解析し、末尾が名詞である場合、それが分類語彙表中のどの分類項目に含まれるかを調べる。

## 映像データ (ショット単位)

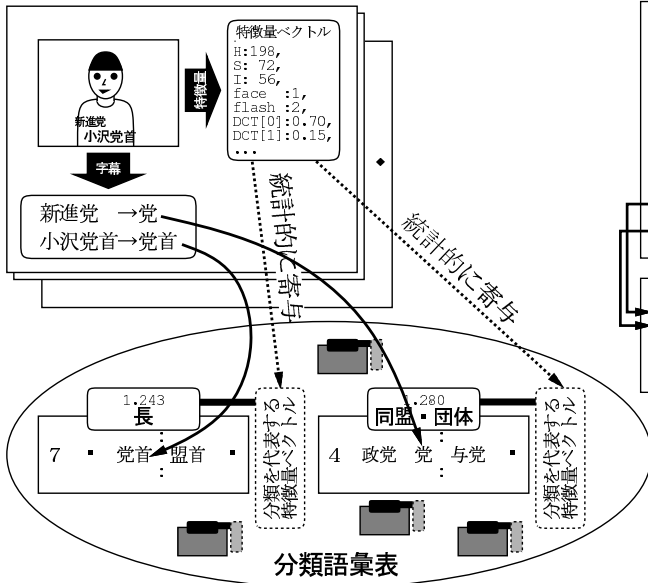
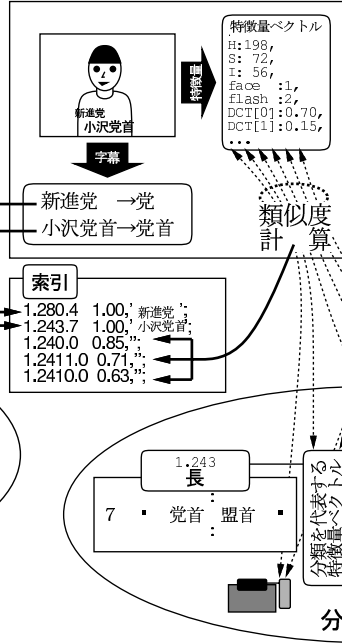


図 2: 各分類項目を代表する特徴量ベクトルの収集

## 字幕がある場合



## 字幕がない場合

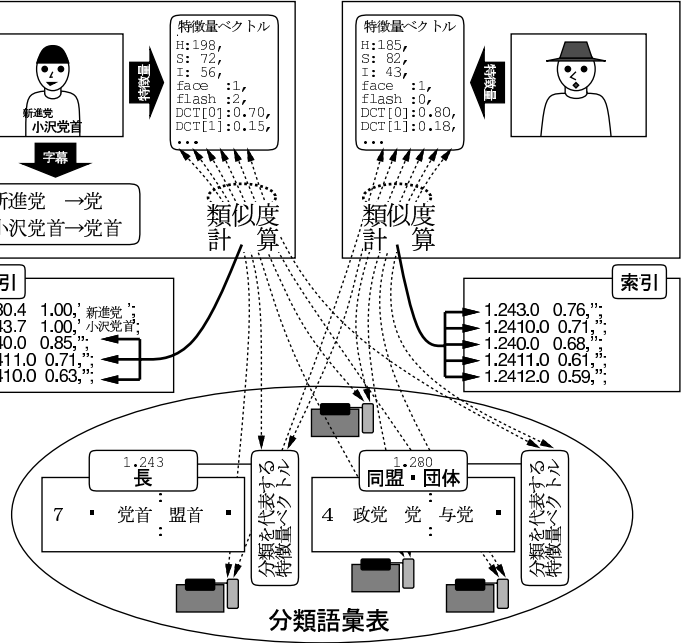


図 3: ショットへの索引付け

### 2. 特徴量ベクトルを抽出

分類を行うのに十分な次元数をもつ特徴量ベクトルの各変数の値を求める。詳細は [1] を参照。

### 3. ショットが属する分類項目を代表する特徴量へ反映

各分類項目を代表する特徴量へ、新たに 2. で求めた特徴量ベクトルを統計的に反映させる。

### 2.2 ショットへの索引付け

図 3に、ショットへの索引付け法を字幕の有無に場合分けして示し、以下に各々について記す。なお、実際の日本語ニュース番組を解析した結果、全体の 6 割程度のショットに字幕が存在した。

#### 2.2.1 字幕が存在しない場合

ショットを代表する特徴量ベクトルと、分類語彙表中の各分類項目を代表する特徴量との類似度を計算し、一定の閾値以上の分類項目番号と類似度を列挙し、索引とする。

このようにすることにより、字幕が存在しないショットについても、内容を推測し、便宜的な索引付けが可能となる。さらに類似度を用いることにより、検索時の問い合わせに対して、提示結果の確信度を定量的に示すこともできる。

#### 2.2.2 字幕が存在する場合

基本的には字幕が存在しない場合と同じだが、字幕そのものも分類番号とともに併記しておく。また、分類語彙表中に存在しない字幕の場合も、そのまま記述しておく。

このようにすることにより、固有名詞などの分類語彙表からは得られない重要な情報も漏らさずに、索引付けすることが可能となる。

### 3 おわりに

本手法は依然実験段階にあり、適切な分類を行うために必要な特徴量の種類および、語義分類を代表する特徴量の表現法の検討を経て、本格的な有効性の評価は今後の課題である。

しかし、各ショット分類基準の抽出 (語義分類毎の特徴量の抽出) から、索引付けにいたるまで、完全に自動的な処理が可能である点および、字幕が存在しないショットについても内容を推測した分類が可能である点は非常に有効であると考えられる。

また、現時点では日本語ニュース番組を対象にしているが、欧米の番組の場合には字幕の使用が内容的にも量的にも著しく限定されているため、今後は音声 (を書き下したものの i.e. クローズドキャプションなど) を始めとする他の自然言語情報の、本枠組における有効な利用法も検討したい。

### 参考文献

- [1] 山本 晃司, 井手 一郎, 田中 英彦: 「言語情報を基準にした学習による映像分類手法」; 1998 年電子情報通信学会総合大会 (Mar. 1997 本大会一般講演にて発表予定).
- [2] 井手 一郎, 山本 晃司, 田中 英彦: 「ショットの分類に基づく映像データへの自動的索引付け」; 情報処理学会第 56 回全国大会 (Mar. 1997).
- [3] Nakamura, Y., Kanade, T.: "Semantic Analysis for Video Contents Extraction -Spotting by Association in News Video-"; Proc. The Fifth ACM International Multimedia Conf., pp.393-402 (Nov. 1997).
- [4] 金出 武雄, 佐藤 真一: 「Informedia: CMU デジタルビデオライブラリプロジェクト」; 情報処理, Vol.37, No.9, pp.841-847 (Sep. 1996).
- [5] 国立国語研究所: 「国立国語研究所言語処理データ集 5 分類語彙表 [フロッピーディスク版]」; 秀英出版 (Dec. 1993).