

言語情報を基準にした学習による映像分類手法

Image classification method based on linguistic information referred learning

山本 晃司
Koji YAMAMOTO

井手 一郎
Ichiro IDE

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

蓄積された映像情報を効率的に再利用するには、映像内容の類似したものを分類し、整理されていることが望まれる。このような映像の分類手法として、既知のいくつかの典型的な映像の集合を用いて画像的な特徴量空間を構築[1]したり、特徴をルールで記述[2, 3]し、未知の映像の特徴量をこれらの特徴量空間やルールと比較することによって、どの典型的な映像の集合に属するかを判別する手法が提案されている。しかし、これらの手法は典型的な映像を人手によって決めているため、分類できる映像の種類が限られてしまう。

そこで、本稿では言語情報が付随した映像を対象として、言語情報を基準に自動的に学習用の分類を作成し、画像的特徴量空間を構築する手法を提案する。

2 言語情報を基準にした学習

映像にはテレビ番組における字幕や音声、画像の説明文など、言語情報が付随するものが多く見られる。これらの言語情報から、画像の内容を示す単語を得ることができれば、分類語彙表をもとに類似した単語をもつ映像の集合を自動的に分類することができる。分類語彙表とは、類似した意味を持つ単語が系統的な項目に分類された辞書である。分類された映像の集合は内容的に類似しているため、各集合がどのような画像的特徴量を持つかを学習し、知識として持つことにより、未知画像の特徴量と比較し、内容的に類似した集合に分類することができると考えられる。

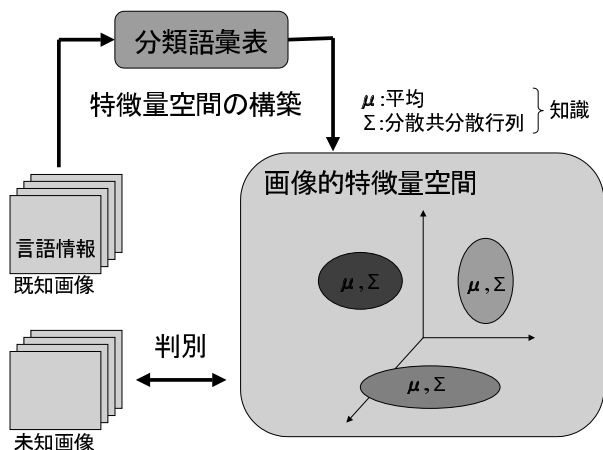


図 1: 言語情報を基準にした学習

3 ニュース映像のショット分類への応用

数ある映像情報の中でも、ニュース映像は資料的価値が高く、データベース化が望まれる分野である。また、字幕や音声などの言語情報も付随する。そこで、ニュース映像を対象として、本分類手法を実装、評価する(システムの全体像は文献[4]に詳しい)。

3.1 特徴量空間の構築

特徴量空間構築の際に用いる言語情報としては字幕を用いる。ただし、技術的情報(「資料」、「報告」など)や発言を書き下した字幕は除外する。また、スタジオのキャスターのショットやCGのショットなど、字幕が明らかに画像内容を示していないショットも処理対象に含めない。画像的な特徴量としてはオブジェクトの中でも比較的抽出が容易かつ重要な意味を持つ顔の情報[2]や映像の概要を表現する低周波成分のDCT係数などの構造情報、色ヒストグラム分布などの色情報を用いて、特徴量空間を構築する。

3.2 画像の類似度の比較

画像の類似度の判別は特徴量空間に分布する各分類番号のクラスと未知画像の特徴量ベクトルのマハラノビス距離 D を用いる。

$$D^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

ただし、クラスの特徴量の平均を $\boldsymbol{\mu}$ 、分散共分散行列を $\boldsymbol{\Sigma}$ 、未知画像の特徴量を \mathbf{x} とする。

4 まとめ

映像に付随する言語情報を基準に自動的に学習用の分類を作成し、画像的特徴量空間を構築する手法を提案した。本手法を用いることにより、映像を人の感覚に近い基準で、分類できるようになると考えられる。今後はニュース映像を対象として、本手法を評価し、有効性を確認したい。

参考文献

- [1] 孟, 佐藤, 坂内: “事例を利用した映像のシーン分類手法の一提案”, 第55回情処全大, 4K-8, 2, pp.162 (1997-09)
- [2] 山本, 井手, 田中: “ニュース映像中のショット分類に関する研究”, 第55回情処全大, 4K-7, 2, pp.160-161 (1997-09)
- [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 (1997-11)
- [4] 井手, 山本, 田中: “字幕の語義に基づくニュース映像の分類・索引付け手法”, 1998 信学総合大会 シンポジウム講演 (1998-03)