

字幕の語義に基づくニュース映像分類

井手 一郎, 浜田 玲子, 田中 英彦, 坂井 修一

{ide,reiko,tanaka,sakai}@mtl.t.u-tokyo.ac.jp

東京大学大学院 工学系研究科*

1 はじめに

放送される映像量の増加につれ、再利用や検索を考慮して、それらを整理して蓄積する必要性が高まっている。とりわけ、重要性和利用価値の点から、ニュース映像へ索引付けを行う価値は高い。現在、この作業は主に人手に頼っているが、増加するデータ量に対応するためにも、きめ細かな検索に要求される精度を満たすためにも、自動的な索引付けが望まれる。

筆者らはニュース映像に対して、映像とそれに附随する言語情報の双方を利用した自動的な索引付けの実現を目指している。Informedia プロジェクト [1] の News-on-Demand システム [2] に代表されるように、同様の手法によるニュース映像に対する自動的な索引付けの注目すべき試みがいくつか行われている。しかし、それらの索引付け戦略は、主に語の出現頻度や単なる出現に基づいており、キーワードと映像内容との対応が必ずしも保証されず、映像データへの索引付け手法としては不十分である。この問題を解決するために、筆者らはショットの分類に基づく索引付け手法 [3] を提案してきた。この手法は、ショットを数通りの画像的に典型的な分類に分類し、典型的な内容に応じた意味属性をもつキーワードを附与するものであり、Nakamura ら [4] によっても同様の試みが行われている。しかし、この手法は分類規則を天下一的に与えるために、典型的な分類の数が少なくならざるを得ない。

そこで現在、まず教師映像データから映像分類規則を学習し、その後一般の入力映像を分類し、その分類に応じた索引付けを行う手法を開発中である [5]。学習過程において、分類基準として映像に附随する言語情報の概念分類を用いることにより、典型的分類の数が増えるうえに、言語情報が附随しない映像に対しても、画像的特徴量のみから大雑把な索引付けが可能になる。本発表では、このような索引付けを実現するための第一段階である映像分類手法を紹介し、概念分類と画像的特徴量の対応を調べる簡単な予備実験を行った結果を示す。

* "News Video Classification Referring to Semantics of Captions"
Ichiro IDE, Reiko HAMADA, Hidehiko TANAKA and Shuichi SAKAI
Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

2 字幕の語義に基づく映像分類

2.1 教師映像からの分類規則の学習

図 1 に分類規則の学習過程を示す。まず、映像分類規則を学習するために、言語情報（ここでは字幕を利用）を含むショットを解析する。各ショットからは画像的特徴量ベクトルが抽出され、附随する字幕の語義 — 具体的には、分類語彙表 [6] の分類項目番号 — に基づいて分類される。「分類規則」と呼ぶものの、実際には各概念分類項目に対応付けされたショットの特徴量ベクトル群から得られる統計的な情報である。教師映像データの学習を経た後には、各概念分類項目への分類規則が学習される。学習に用いる手法としては、主成分分析や記憶に基づく推論 (MBR) などを検討中である。

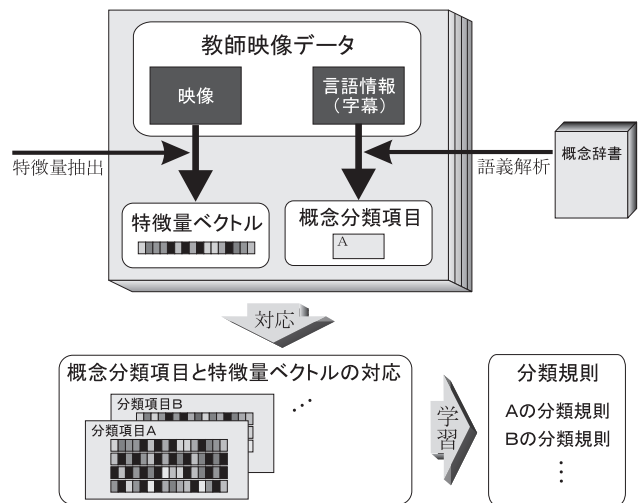


図 1: 教師映像からの分類規則の学習

2.2 予備実験: 概念分類と画像的特徴量の対応

分類規則の学習の予備実験として、75 分間のニュース映像について、字幕の概念分類と画像的特徴量の対応を調べた。画像的特徴量ベクトルの要素としては、1 つの変数 — ショットの先頭フレーム中に存在する比較的大きい顔領域の個数 — のみを設定した。表 1 に上記の条件に従って得られた、顔の個数と、対応する概念分類項目のうち累積度数で上位 30% に入るものの対応を示す。

表 1: 予備実験の結果：概念分類項目と画像的特徴量の対応（累積度数上位 30%）

顔の 個数	概念分類項目の見出し			
	人間に関するもの	集合に関するもの	場所に関するもの	その他のもの
なし	—	「意見, 決定, 調査, 承認」	「地名」	「単位」
1 個	「長」 「人名」	—	「事務所, 市場, 駅」 「地名」	—
2 個	「身分」 「人間」	「宣言, 報告, 噂」	「社寺, 学校」 「家, 宿, 教室」 「地名」	「単位」 「数」 「原理, 規則, 方法, 制度, 習慣, 計画」 「金銭」
3 個 以上	—	「話, 議論, 批評, 説明」 「約束, 交渉, 賛成」 「意見, 決定, 調査, 承認」 「議会」 「集会, 出欠」	「地名」	—

なお、一部の概念分類項目の見出しは、分類語彙表のものを短縮して示した。

この対応表を見ると、

- 人間に関する概念分類項目は 1, 2 個の顔と対応
- 集合に関する概念分類項目は 2 個以上の顔と対応
- 場所に関する概念分類項目は顔の数と無関係に分布

し、顔の個数と概念分類項目との間に良い対応関係が見られる。場所に関する概念分類項目が顔の数と無関係に分布したのは、この実験ではそれらを分類するような変数を特徴量ベクトルの要素として設定しなかったためであり、適当な変数の設定により、その他のものとともに分類可能になると思われる。また、顔がない場合に、集合に関する「意見, 決定, 調査, 承認」という概念分類項目が含まれたのは、人々が集合している映像であっても、正面からの映像でないために顔領域が抽出できなかったのが原因である。

3 おわりに

本発表では、附随する言語情報に基づくニュース映像の分類手法を提案し、簡単な予備実験を通じてその有効性を検討した。この時点では本手法は一見、顔画像と人物名を対応付ける Name-It システム [7] に類似しているようにも見えるが、用いる画像的特徴量とデータ量の拡大により、分類できる対象が拡がり、その可能性と汎用性を高めることができると思われる。

今後は、提案手法により獲得した映像分類規則を用いた自動的索引付けについての検討も進め、最終的には高度なキーワード抽出と索引付けを目指す。

参考文献

- [1] The Informedia Project, <http://www.informedia.cs.cmu.edu/>.
- [2] Hauptmann, A. G. and Witbrock, M. J.: Informedia News-on-Demand: Using Speech Recognition to Create a Digital Video Library, *Proc. AAAI'97 Spring Symposium on Intelligent Integration and Use of Text, Image, Video and Audio Corpora*, pp.120-126, May 1997.
- [3] 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, 第 56 回情報処理学会全国大会論文集 (2), pp.263-264, Mar 1998.
- [4] Nakamura, Y. and Kanade, T.: Semantic Analysis for Video Contents Extraction —Spotting by Association in News Video—, *Proc. ACM Multimedia'97*, pp.393-402, Nov 1997.
- [5] 井手 一郎, 山本 晃司, 田中 英彦: 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 年電子情報通信学会総合大会論文集 (情報・システム 1), pp.403-404, Mar 1998.
- [6] 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, 1993.
- [7] Satoh, S., Nakamura, Y. and Kanade, T.: Name-It: Naming and Detecting Faces in Video by the Integration of Image and Natural Language Processing, *Proc. IJCAI-97*, pp.1488-1493, Aug 1997.