

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

5Q-6

井手 一郎, 山本 晃司, 田中 英彦

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

東京大学大学院 工学系研究科*

1 はじめに

1.1 概要

様々な媒体を通じて提供されるマルチメディア、特に映像情報の増加に伴い、再利用や検索を容易に実現するための自動的な索引付けの需要が高まっている。

本研究では、ショット（連続した映像の最小素片…1.2で定義）単位の分類に基づき、分類毎にキーワードを選択的に付与することにより、ショットという時間的に短い区間に、映像内容に則した索引付けを試みた。その結果、索引付けの需要が高いニュース番組に対し、会見などについて人物を6割、集会などについて場所を9割以上の精度で索引付けすることができた。

1.2 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できる。

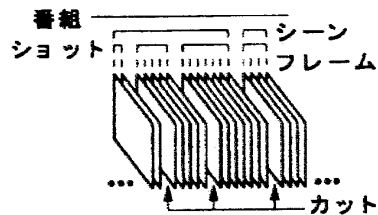


図1: 映像の構成と用語の定義

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 ショット分類に基づく索引付け

2.1 全体構成

一般にニュース番組は、図2に示すような典型的なショットが度々出現する。また、話題の先頭ではキャスターがタイトルと共に映ることが多い。そこで本研究では、ニュース映像から図2に示す3つの典型的ショットを簡単な画像処理により分類し、各々に適した索引付けを行う。図3にその手法の全体構成を示す。

索引付けには、映像内容を端的に表す字幕の語義を解析してキーワード候補としたものから、各ショット分類

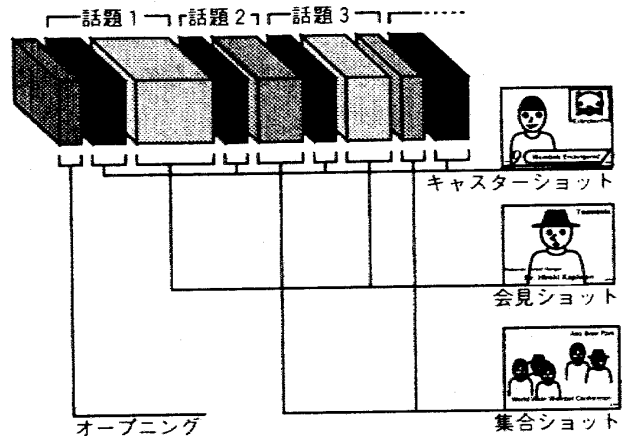


図2: ニュース番組の構成

の映像内容に適した語義をもつものを選択的に付与する。同様の試みは中村ら[1]も行っているが、本研究では日本語ニュース映像を対象にしているため、自然言語の扱いが異なる。

なお、画像処理と自然言語処理の詳細については[4]と[3]を参照されたい。

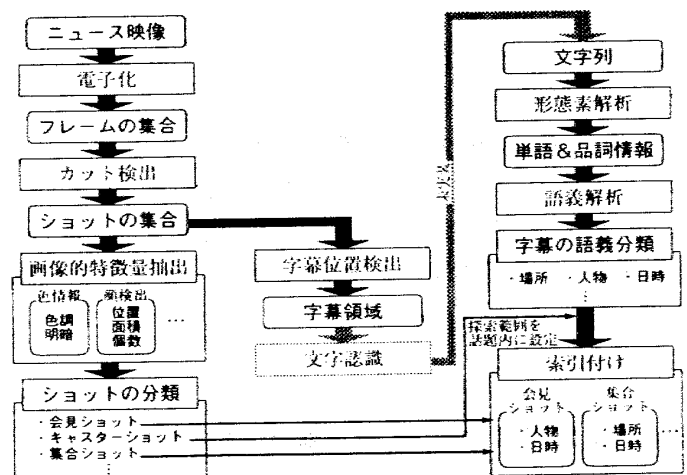


図3: ショット分類に基づく索引付け手法の全体構成

2.2 ショットの画像的分類

図3に示すように、ニュース映像を画像的特徴量に基づき以下の3つの典型的なショットとその他に分類する。

・会見ショット

顔領域を抽出し、一定の領域に顕著なものが1, 2個存在するものを分類する。最終的には次の「キャ

*"Automatic Indexing to Video based on Shot Classification"
Ichiro Ide, Koji Yamamoto, Hidehiko Tanaka
University of Tokyo, Graduate School of Engineering,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

スターショット」に該当するものを省く。

- キャスターショット

「会見ショット」間の色情報に基づくクラスタリングの結果、最大のクラスタをなすものを分類する。多く話題の境界に出現し、索引付けの際に必要なニュースの構造を把握するのに必要となる。

- 集合ショット

顔領域を抽出し、全画面に顕著なものが3個以上存在するものを分類する。

この基準による各ショットの検出精度は、「会見ショット」が7割程度、「キャスターショット」が100%、「集合ショット」が3割程度であった。

2.3 字幕の語義分類

字幕や音声をそのまま話題全体や出現したショットのキーワードとして索引付けする手法も行われているが、前者は索引付け単位の時間的粒度が荒く、後者はキーワードと映像内容の対応が保証されないため、ショット単位に映像内容に則した索引付けを行う立場からは不十分である。

そのような索引付けを行うためには、字幕の語義をある程度理解する必要がある。字幕の出現位置を中心としたニュース字幕の語義解析手法[2]もあるが、本研究では番組に依存しない解析手法として、末尾の名詞に着目した字幕の語義解析手法[3]を採用し、字幕の語義を人物、場所、日時、その他のいずれかに分類する。この手法はおよそ8~9割程度の精度で解析が可能であり、実用上は問題のないことが分かっている。

2.4 ショット分類別の索引付け

以上のようにして、ショットを画像的特徴量に基づいて分類し、一方で字幕を末尾の名詞の種類に基づいて語義解析した結果から、各ショット分類に適した索引付けを以下のように行う。

- 会見ショット

「会見ショット」では、一般に大寫しになっている人物が話題になっている可能性が高い。そこで、キーワードとしては、「人物」と「日時」を索引付ける。

- 集合ショット

「集合ショット」では、特定の人物というよりは、どのような集合であるか(会合名や開催場所)が話題になっている可能性が高い。そこで、キーワードとしては、「場所」と「日時」を索引付ける。

索引付けの際は、索引付けするショットから話題の冒頭にある「キャスターショット」まで遡及的に辿り、最も近くにある該当する語義のキーワード候補を選択する。

3 実験

2.4に示した方針により各ショット分類に索引付けを行った結果の正答率を表1に示す。なお、ここでは索引付け手法を単独で評価するために、ショット分類と語義解析は完全になされているものと仮定した。

ショット分類	人 物	場 所	日 時
会見ショット	63%	—	12%
集合ショット	—	97%	32%

表 1: ショット分類別索引付け成功率

場所に関する索引付けはほぼ完全に行えたが、人物と日時に関しては予想以上に悪い結果となった。

しかし、人物と日時とでは精度が悪い原因が異なる。日時の精度が悪いのは、そもそも日時に関する字幕そのものが少ないためである。一方、人物に関しては、例えばインタビューを行っている「会見ショット」では、2人の人物が交互に登場しても、字幕は最初の1回にしか表示されないことが多く、その後の人物索引付けに支障を来している。このような誤りは、話題よりもさらに細かい番組構造を解析して、不連続な同一映像内容のショットに対して最初の字幕から得られるキーワード候補を適用することにより解消される。

4 おわりに

本発表では、画像的特徴量による映像の分類と語義解析による字幕の分類を統合的に利用した、ニュース映像への自動的索引付け手法を紹介した。依然各々の分類精度の向上が要求されるが、映像内容に則して時間的粒度の細かい索引付け手法として、有望である感触が得られた。

今後は、現在は恣意的かつ固定的な映像の分類基準を統計的に収集する手法の検討及び、より細かい番組構成の解析を通じた索引付け精度の向上を行う。

参考文献

- [1] Nakamura, Y., Kanade, T.: "Semantic Analysis for Video Contents Extraction -Spotting by Association in News Video-"; Proc. The Fifth ACM Intl. Multimedia Conf., pp.393-402 (Nov, 1997).
- [2] 渡辺 靖彦, 岡田 至弘, 長尾 眞: 「TVニュースで用いられるテロップの解析」; 情報処理学会研究会報告 96-NL-116, Vol.96, No.89, pp.107-114 (Nov, 1996).
- [3] 井手 一郎, 田中 英彦: 「末尾の名詞に着目したTV放送中の字幕の意味属性解析」; 情報処理学会第55回全国大会予稿集, Vol.2, pp.91-92 (Sep, 1997).
- [4] 山本 晃司, 井手 一郎, 田中 英彦: 「ニュース映像中のショット分類に関する研究」; 情報処理学会第55回全国大会予稿集, Vol.2, pp.160-161 (Sep, 1997).