

山本 晃司 井手 一郎 田中 英彦

{kyama, ide, 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

放送の多チャンネル化などに伴い、増加する一方の映像情報から、自分に必要な情報を取得したり、蓄積された情報を有効に再利用するには、映像に索引をつけて、データベース化されていることが望まれる。特にニュース映像は資料的価値が高く、日々最新の情報が提供されるため、自動的な索引付けによるデータベース化が望まれる分野である。このようなデータベースは、Informedia プロジェクトの News-on-Demand [1] のように、実験的にサービスの提供が行われているものもあるが、必ずしも映像内容に応じたきめ細かい索引付けが行われているわけではない。

本稿では映像内容を反映させた索引付け機構として、ニュース映像を構成するショットをいくつかの典型的なショットに分類し、その内容に応じた索引付けを行う方法を提案する。また、典型的ショットの分類に関する予備的な実験及び評価を行う。

2 ニュース映像の構成

一般的なニュース番組の構成は図1のように1つの話題はスタジオのキャスターのショットから始まり、いくつかの取材映像を経て、次の話題の先頭であるキャスターのショットに戻るといった流れの繰り返しにより構成される。図中、各ブロックはショットを表しており、同じ色のブロックはシーンを表す。ただし、ショットとは連続したフレームの集合を示し、シーンとは画像的あるいは内容的に連続したショットの集合を示すものとする。

ニュース映像に特化した映像データベースを構築する場合はこの番組の構成を考慮することによって、スタジオのキャスターのショットを認識し、番組を話題単位に分割して索引付けを行うことができる [2]。しかし、話題単位での索引付けは必要とするショットのみを検索することが困難であるため、映像情報に注目したデータベースにおいてはショットあるいはシーン単

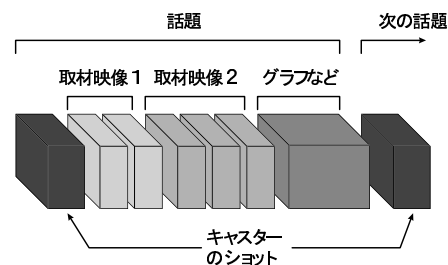


図1: ニュース映像の構成

位で索引付けされていることが望まれる。

3 ニュース映像中のショット分類

3.1 典型的ショット

ニュース映像を構成するショットはキャスターショット以外にもインタビューのショットなど、同様な状況下で撮影されたものがいくつか存在する。このようなショットを典型的ショットと呼ぶこととする。これらの典型的ショットは色情報など、なんらかの画像的な特徴量が類似しているため、比較的簡単な画像処理を用いて、分類が可能であると思われる。分類に用いる画像的特徴量は色情報などのほか、空間的な構造を用いることができる。特にニュース映像では人物を撮影した映像が多いため、画面内での顔の位置や個数、大きさなど、抽象度の高いモデルを用いることが有効であると考えられる。

表1にニュース番組中に見られたキャスターショットを含めた典型的ショットの例及び番組中に登場した割合を示す。

表1: 典型的ショットの例

典型的ショット	割合
会見、インタビュー	24%
集会、会議	16%
キャスター(スタジオ)	9%
CG(グラフなど)	4%

3.2 ショット分類に応じた索引付け

ニュース映像を典型的ショットに分類することによって、そのショットの内容に応じた索引項目を用いることが可能となる。例えば、インタビューや会見のショット

*Shot Classification of News Images,
Koji Yamamoto, Ichiro Ide, Hidehiko Tanaka,
University of Tokyo, Graduate School of Engineering,
7-3-1 Hongou, Bunkyo-ku, Tokyo 113, Japan

トであれば、話者の名前は重要な索引項目となるが、集会や会議など大勢の人々が映っているようなショットであれば、話者の名前より、むしろ、場所や時間が重要な索引項目となる。各ショットにおいて、有効であると考えられる索引項目例を表 2 に示す。これらの索引となる情報は字幕など、言語情報として与えられるので、自然言語処理を用いて処理し、索引付けする [3, 4]。

表 2: 典型的ショットとその索引項目例

典型的ショット	索引の項目
キャスター	話題のタイトル
会見、インタビュー	話者、場所、時間
会議	議題、場所、時間
CG(グラフなど)	表題

4 ショット分類の実験

ショット分類の予備的な実験として、会見ショットの分類を行った。また、会見ショットはキャスターショットと画像的特徴量が類似しているため、両者を区別するために会見ショットとして得られたショットからのキャスターショットの分類を行った。

画像はファイルとして蓄積されたものを用いる。画像サイズは 320×240 、色数は 24bit である。

4.1 会見ショット

会見のショットやキャスターのショットは画面内に話者の顔が比較的大きく映っているという特徴を持つ。そこで、各ショットの先頭フレームにおいて、顔領域となる肌色領域を抽出し、その重心の位置が図 2 の斜線部分で示される範囲に収まるものが 1~2 個のものを検出する。肌色領域の抽出は画像の色情報を HSV 空間に変換し、 H (色相) 及び S (彩度) が一定範囲内にあるものを抽出することにより行う。

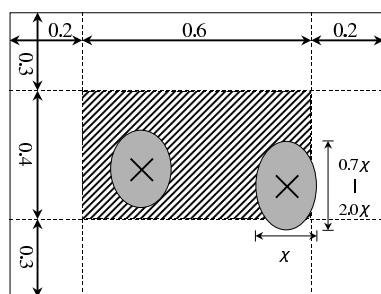


図 2: 検出条件

4.2 キャスターショット

キャスターショットは会見ショットに含まれて検出される。そこで、上記の条件のもと会見ショットとして検出されたショット間の相互の類似度を調べ、類似度が高いものの組合せを求める。このような組合せのう

ち、番組全体に渡って一様に登場するものをキャスターショットとして検出する。類似度は画像を R, G, B 各 4 階調で表現される 64 色に減色し、その色ヒストグラムを χ^2 検定を用いて比較することにより求める。

4.3 結果

結果を表 3 に示す。尚、ニュース番組中の全ショット数は 211 ショットである。

表 3: 実験の結果

会見 ショット	該当するショット	50
	正検出	34
	誤検出	13
キャスター ショット	検出洩れ	16
	該当するショット	19
	正検出	19
	誤検出	0
	検出洩れ	0

単純な画像的特徴量を用いているにも関わらず、比較的良好な結果を得られた。特に、キャスターショットの検出は再現率、適合率共に 100% を得られ、本稿で用いた手法が有効であることが確認できた。会見ショットの検出においても再現率、適合率共に 7 割程度を得られている。会見ショットの検出洩れの原因は検出条件の設定が適切でなかったことによるものが多い。

5 おわりに

本稿ではニュース映像中のショット分類とそれに基づいた索引付け手法を提案し、実験によりその有効性を確認した。現在、会見ショットの検出条件は手動で設定しているが、学習を伴った識別アルゴリズムを用いることにより、より適切な条件を求められると思われる。今後はこれらの精度の向上や他のショットの検出への対応を行う。また、本稿で用いた映像は特定のニュース番組のものであったため、より広範囲なニュース映像を対象に調査し、多くのニュース番組に対して有効な手法となるようにする必要がある。

参考文献

- [1] A. G. Hauptmann, M. J. Witbrock, M. G. Christel: "News-on-Demand: An Application of Informedia Technology", D-lib magazine (1995-09), <http://www.cnri.reston.va.us/home/dlib/september95/nod/09hauptmann1.html>
- [2] 山本, 浅野: "ニュース映像データベースの索引付け手法の一提案", 第 51 回情処全大, 6S-10, 2, pp.251-252(1995-09)
- [3] 井手, 田中: "画像・言語情報の統合的利用による映像データの自動的インデクシングの試み", 言語処理学会第 3 回年次大会予稿集, pp.485-488 (1997-03)
- [4] 井手, 田中: "末尾の名詞に着目した TV 放送中の字幕の意味属性解析手法", 第 55 回情処全大, 6J-04 (1997-09)