5M-08

ニュース映像における人物領域と背景領域の個別特徴量解析による人物と場面の追跡・索引付け手法

井手一郎　浜田玲子　坂井修一　田中英彦

{ide,reiko,sakai,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

放送される映像量が増えるにつれ，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が高まりつつある．なかでも，重要性と利用価値の点からニュース番組への索引付けの意義は大きい．現在この作業は主に人手に頼っており，増大する量への対処と，きめ細かな検索要求のための質を考えると，自動化の需要が大きい．

筆者らは画像とそれに附随するテキストの双方を参照することによりこの問題に取り組んでいる．Informedia プロジェクトの News-on-Demand システム [3] をはじめ，同様の手法によるニュース映像への自動索引付けの試みは数多くある．しかし，それらの多くは自動索引付けの需要をある程度満たしはするが，索引付け手法は，単純な語の出現や統計量に基づくもので，索引と画像内容の対応を必ずしも保証していないという問題がある．

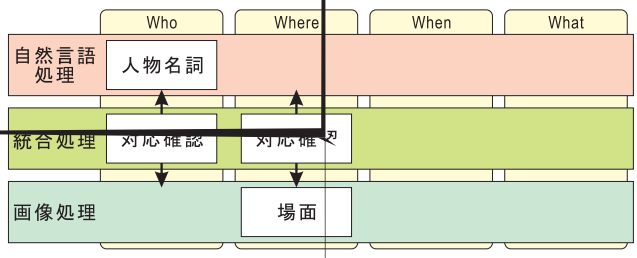
本稿ではこのような問題意識のもとに，索引と画像内容の対応を保証する索引付け手法の画像処理部分として必要になる，同一人物・場面の追跡及び，画像特徴量に基づく場面推定について述べるとともに，背景領域の画像特徴量の類似度に基づく簡単な場面推定実験の結果について報告する．

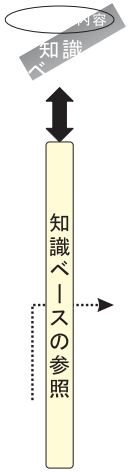
2　索引付け機構の全体像

図 1に提案機構の概念図を示す．本機構では，4 つの異なる意味属性；いつ（when），どこで（where），誰が（who），何を（what），に分けて索引付けを行う．実際の検索要求が，「**ある時**に**誰か**が**どこか**で**何かを**している映像が見たい」という形式のものでることは容易に想定されるので，このような属性に分かれた索引は妥当であると考えられる．

このような機構をもつシステムとして，Name-It システム [2] があるが，対応関係を考慮してはいるものの，人物名と顔画像との対応に限定されている．

また本機構では，画像特徴量と画像内容との対応関係を記述した知識ベースを導入する．ここで知識とは，明





2: 領域分割による人物と場面の追跡と推定.

画像中の背景領域の[redacted]の追跡と場面推定を行い[redacted]き，同一人物の[redacted]を行う．また，行為[redacted]定は，各領域[redacted]合して，いくつかの典型的ショット分[redacted]行うことで実現する．

---

* “Tracking and indexing people and scenes by regional image feature analysis on news video”
Ichiro Ide, Reiko Hamada, Shuichi Sakai, Hidehiko Tanaka
Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

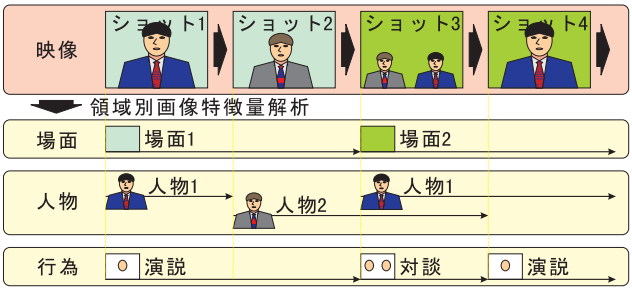


図 3: 人物と場面の追跡.

3.1 同一人物と同一場面の追跡

まず，人物と場面の追跡に関しては，図 3に示すように，領域毎の画像特徴量に基づき，人物（前景，who）と場面（背景，where）をショットの境界を越えて追跡する．ここでは画像を（1）頭部,（2）胴体,（3）背景の 3 つの領域に分割する．そして，頭部領域の類似度，つまり顔の特徴と，胴体領域の類似度，つまり着ている服の特徴に基づき同一人物を追跡し，背景領域の類似度で同一場面を追跡する．他の目的のための人物追跡は既存の研究でも行われているが，ここでは良好な照明条件で正面顔を撮影することが多いというニュース映像の特徴を利用して，動物体追跡などの複雑な処理は行わない．

3.2 知識ベースに基づく画像内容の推定

人物と人物索引，場面と場所・組織索引との対応には，画像特徴量と画像内容との関係に関する知識が必要である．このために，画像特徴量と画像内容に関する知識ベースが必要となり，多量の訓練画像から統計的に知識を獲得する．このような関係を獲得するのは一般には非常に困難と思われるが，類似した状況で撮影された類似した画像内容を多く含むニュース映像においては，ある程度特定のジャンルに限定すれば可能と思われる．現在，主に閣議や国会の画像からなる国内政治にジャンルを絞ることを考えている．

図 3の下部には，行為（what）に関する索引付けに利用する典型的ショット分類も示されている．ニュース映像は類似した状況で撮影された類似した画像内容（典型的ショット）を多く含むため，ショット分類により映像中の出来事がおおよそ推定できる．ショットは画像特徴量に基づき，[1] で採用されいてる,（1）演説,（2）集合,（3）その他の人物といったような，いくつかの典型的な分類に分類される．このような明示的条件に基づくショット分類も，知識ベースの利用による行為内容推定の一部と考えることができる．

3.3 場面の推定実験

実際のニュース映像から目視で選択した 60 の国内政治関連画像（設定した場面の種類を表 1に示す）と同様に選択された実験用画像（各 1 枚ずつ）の背景領域との比較の類似度による画像内容推定実験の結果を示す．特徴量としては RGB 色空間において 16 階調，共起距離 1〜4 の色共起分布（Color Correlogram）[4] を用い，特徴量の類似度は特徴量ベクトルのなす角の余弦を用いた．なお，領域分割も目視により行い，分割を行わない遠景の画像については，背景領域の代わりに全領域での比較を行った．

表 1: 場面推定実験のために設定した場面.

（a）	国会議場	遠景（分割なし）
（b-1）	閣議室前室	遠景（分割なし）
（b-2）	〃	人物アップ（分割あり）
（c-1）	報道官会見	遠景（分割なし）
（c-2）	〃	人物アップ（分割あり）

この結果,（c-1）を除いたものについて，類似度第 1 位に正しい場面の画像が出現した．また，類似度上位 3 件中では,（c-1）を除いたもの全てにおいて，上位 5 件中では,（a）,（b-2）,（c-2）において，正しい場面の画像が過半数を占めた．

4 おわりに

本稿では，人物に注目して，画像を人物（前景）領域と背景領域に分離して画像特徴量の解析を行うことにより，人物，場面，行為の追跡と推定を行う手法を提案した．また，背景領域の画像特徴量の類似度に基づく場面推定実験において，良好な結果が得られた．今後は，実際に国内政治にジャンルをしぼった自動的な処理と，自然言語処理の結果を利用した索引付けを行っていく．

参考文献

[1] 井手一郎，山本晃司，浜田玲子，田中英彦：ショット分類に基づく映像への自動的索引付け手法，信学論（D-II）Vol.J-82-D-II, No.10, Oct 1999 掲載予定.

[2] Satoh, S., Nakamura, Y, Kanade, T.: Name-It: Naming and detecting faces in news video, *IEEE Multimedia*, Vol.6, No.1, pp.22-35, Mar 1999.

[3] Wactler, H. D., Hauptmann, A. G., Witbrock, M. J.: Informedia News-on-Demand: Using speech recognition to create a digital video library, *CMU Tech. Rep.*, CMU-CS-98-109, Mar 1998.

[4] Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., Zabih, R.: Image indexing using color correlograms, *Proc. IEEE Computer Vision and Pattern Recognition Conf. '97*, pp.762-768, 1997.