

シーンデータベースを用いた画像認識の検討

5Q-1

孟 洋†, 佐藤 真一‡, 坂内 正夫†

†東京大学生産技術研究所, ‡学術情報センター

1 はじめに

画像を認識する場合、シーンが特定できれば、認識対象の認識が容易になると考えられる。これは、シーンが限定されれば、映っているオブジェクトの種類や認識対象の映り方などが限定できるため、シーンに応じた認識モデルや認識手法を用いることが可能となるためである。

一方、人間のシーンの認識能力について考えてみると、シーンを少し見ただけで、全体的な雰囲気からそのシーンを判別できる場合が少なくない。つまり、シーンはオブジェクトを明確に認識することなしに同定できる可能性があるといえる。

そこで、現在、我々は、より柔軟で効果的な画像認識の実現を目指し、一部のテレビ画像のように作為的に作られ画面構成がほぼ定まっているような画像を対象として、シーンの同定を利用した画像認識手法について検討を行なっている。

本稿では、シーンの同定のためにシーンデータベースを用いた画像認識手法の枠組について報告する。

2 シーン DB を用いた画像認識

テレビ画像などの作為的に撮影された画像では、カメラ位置の制約や演出意図などの点から、画面内でのオブジェクトの映る大きさや位置がほぼ定まっていると考えられる。このため、予め代表的なシーンのカット画像を用意し、データベース化しておけば、対象画像とそのカット画像の特徴量を比較することにより、対象画像の

シーンやカットを同定できる可能性がある。シーンやカットが同定できることでの利点としては、以下の点があげられる。

- シーンやカットの情報によりトップダウン解析が可能
- シーンやカットに応じた認識モデル、及び、認識手法が利用可能

つまり、データベース内の代表的な各カットに対し、そのカットに依存したオブジェクトの認識モデルを持たせることで、認識モデルの単純化がはかれ、これにより、認識モデルの設定が容易、かつ、認識処理時間の短縮、認識率の向上が可能となる。

これは、例えば、スポーツ画像での人間の映り方を考えてみると、大きく顔だけが映っている場合や小さく全身が映っている場合、また、ユニフォームを着ている場合や着ていない場合など様々な状態がありえるため（図1）、これらを単純に人間という枠でモデル化するのは大変困難なものとなるが、サッカーなどとシーンが設定できれば人間の映り方も限定され、モデル化が行ないやすくなるということを意味している。

シーンデータベースの構成を図2、また、画像認識の過程を図3に示す。画像認識は、はじめに、シーンやカットの同定を行ない、その後、その結果に対応した認識モデルによりオブジェクト認識を実行するという形で行うことになる。



図1: 人間が映っている画像の例

A Study of Image Recognition using Scene Database

Hiroshi MO†, Shin'ichi SATOH‡ and Masao SAKAUCHI†

†Institute of Industrial Science, University of Tokyo

‡National Center for Science Information Systems

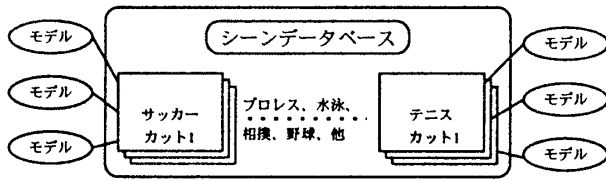


図 2: シーンデータベースの構成

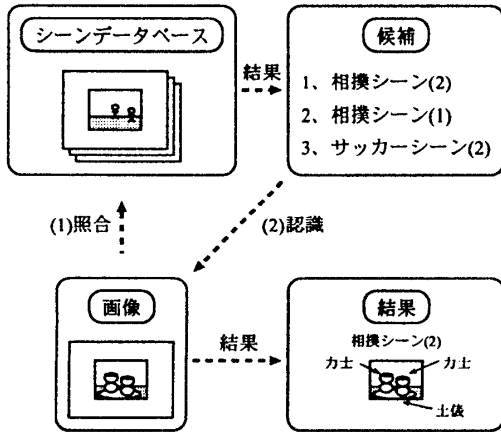


図 3: 画像認識の過程

3 類似画像検索手法

シーンデータベースを用いたシーンの同定は、対象画像に対して類似画像検索を行なうことで実現される。通常、類似画像検索は、色情報や形状情報、テクスチャ情報、位置情報などの特徴量を用いて行なわれる。シーン同定のための類似画像検索も基本的には同様の方法で行えば良いと思われるが、特に以下の点に注意し検討する必要がある。

- オブジェクト特徴の多様性の影響
 例えば、同じ人間であっても服装により色情報などが変化する
- 動物体の位置や姿勢の違いの影響
 例えば、同じシーンであっても人間の位置などが変化する
- 撮影条件やカメラパラメータの違いの影響
 例えば、同じシーンであっても映る向きや大きさなどが変化する

4 簡単な実験例

現在、スポーツシーンを対象とした画像認識についての検討を行なっている。ここでは、スポーツシーンにおける人間の認識の簡単な実験例を示す。実験に用いたシーンデータベースは、水泳、ゴルフ、サッカーなど7種類のスポーツ約80枚のカット画像からなるもので、類似画像検索手法としては、画像全体の色相ヒストグラムをベクトル化し、画像間の距離をそのベクトルのなす角の大きさで評価する方法 [1] を用いた。相撲画像において力士を認識した場合の結果を図4に示す。



(a) 相撲画像 (b) 力士

図 4: 相撲画像における力士の認識

5 おわりに

本稿では、シーンデータベースを用いた画像認識手法について概説し、簡単な実験例により、その可能性を示した。

この手法は、千差万別の画像を扱うにはむいていないが、テレビ画像などを扱う場合には有効な手段となるのではないかと考えている。

今後は、シーンを同定するために要求される類似画像検索手法や、この認識手法の効果などについて検討を行なっていく予定である。

参考文献

[1] Y.Yaginuma and M.Sakauchi: "Multi-Purpose Interface for Still/Moving Image Retrieval", SPIE Applications of Digital Image Processing XVII, Vol.2298, pp.260-267(1994)