

## シーンの状況モデルを用いた画像認識の検討

5F-10

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

一般に、大きさや向きなどの認識対象の画像への映り方は、シーンの状況やカメラ位置などの違いにより、様々に変化する。このため、この映り方の違いを考慮した認識が行える必要がある。以前、認識対象の映る大きさの違いに対応するためのモデルの一構成方法として、距離モデルの考え方を示した [1]。しかし、この方法では、近景、中景、遠景と映る大きさ別に特徴をモデル化するため、情報の少ない遠景側の認識知識が設定しにくいなどの問題があった。

人間が画像を認識する場合を考えてみると、認識対象が明確に判別できないような場合においても、周囲の状況などからある程度の認識が可能である。これは、人間が、認識対象に関する知識とともに、シーンの状況に関する知識も利用して、総合的な判断を行なっているためである。

そこで、サッカーシーンにおける人間の認識を例に、認識対象に関する知識とともに、シーンの状況に関する知識も利用した画像認識の方法について検討を行なった。

## 2 距離モデル

一般の画像では、その撮られ方により、認識対象の映る大きさは変化する。このため、認識対象が大きく映っている場合には認識対象の局所の特徴を、また、小さく映っている場合には全体的特徴や周囲環境を認識に用いるなど、認識対象の映っている大きさにより認識に用いる情報を変える必要が生じる。距離モデルとは、こ

のような認識に用いる情報の違いを映る大きさ別にモデル化し、同一領域に対する各モデルの満足度を数値的に Dempster の結合則 [2] により統合することで、大きさに依存しない認識を実現しようとするものである。人間の場合、次に示すような近景、中景、遠景の三モデルでモデル化する。

- (1) 近景モデル → 人間頭部のモデル
- (2) 中景モデル → 人間全身のモデル
- (3) 遠景モデル → 場における人間のモデル

## 3 シーンの状況モデル

テレビ画像などの作為的な画像では、演出意図やカメラ位置の制約などから、認識対象の映り方や位置関係などカットの構図がほぼ決まっている場合が多い。このため、カットの構図から、場や認識対象の認識、評価を行なうことが可能となる。そこで、シーンの状況モデルとして、場の知識などを含めたカットの構図に関する知識を用いた認識について考えてみる。

図 1 に、サッカーシーンの代表的なカットの例を示す。このようなカットに対し、次に示すような内容を記述することで、図 2 に示すようなカットモデルを定義する。

- (1) カットの同定するための特徴量  
「まとまったグリーン領域が存在」
- (2) 場のモデル  
「まとまったグリーン領域はフィールド」
- (3) 認識対象の映る大きさ、位置、数の関係  
「遠景の人間が複数存在」
- (4) 場と認識対象の関係  
「フィールド上に人間は存在」

A Study of Image Recognition Using Model of Scene Situation

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

†Institute of Industrial Science, University of Tokyo

†National Center for Science Information Systems

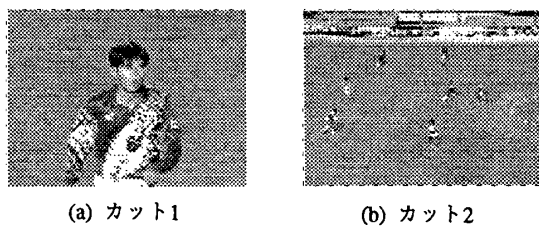


図 1: サッカーシーンでのカットの例

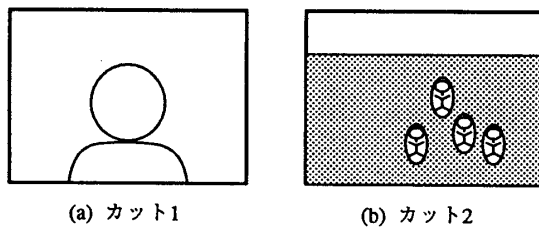


図 2: カットモデルの例

#### 4 認識

認識は、図 3 に示すプロセスにしたがって行なう。はじめに、画像を直接評価し、カットの同定を試みる。カットの同定ができる場合には、そのカットモデルにより場などのラベル付けを行なう。次に、ラベル付けの結果を用いて、距離モデルによる認識を実行する。これにより、認識対象の候補領域を求める。最後に、求めた候補領域とカットモデルとの整合性をチェックし、必要ならば候補領域の再探索を行ない、最終的な認識結果を求める。

具体的には、カット 2 のような画像の場合、最初に、まとまったグリーン領域を評価することで、モデルのカット 2 を同定し、フィールドのラベル付けを行なう。次に、距離モデルにより人間の認識を試みる。ここで、フィールド領域の情報から遠景側の人間の認識が容易になる。そして、最後に、候補領域とモデルとの整合性をチェックし、候補領域の評価値を調整する。基本的に、抽出された候補領域の評価は、すべて各領域に対する評価値の調整により行なう。

カットモデルを用いることで、次に示すような効果が期待できる。

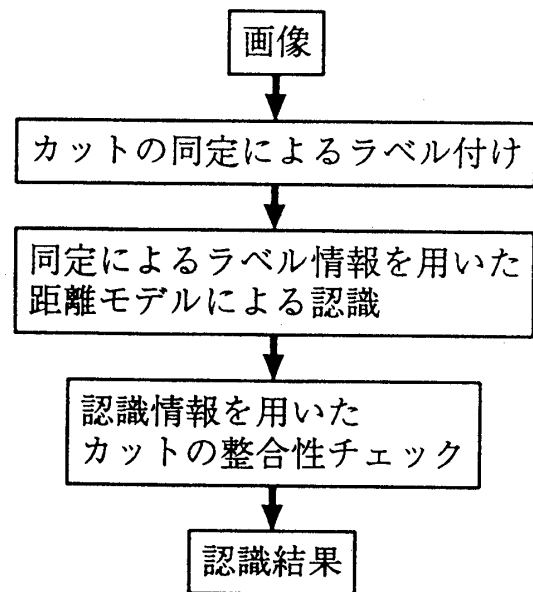


図 3: 認識プロセス

- (1) より確実な認識対象の認識が可能
- (2) 場の認識が可能
- (3) カットモデルに基づく候補領域の探索が可能

#### 5 おわりに

本稿では、サッカーシーンにおける人間の認識を例に、シーンの状況に関する知識を利用した認識について概説した。この方法は、より複雑で種類の多い一般的なシーンに対して、すぐに適用することは難しいかもしれないが、より柔軟な認識を実現する上での一つの考え方になれば幸いである。今後は、一般的なシーンの状況のモデル化の可能性、及び、その効果についての検討を行なっていく予定である。

#### 参考文献

- [1] 孟洋, 佐藤真一, 坂内正夫, "多層のモデルを用いたスポーツシーンからの人間の抽出," 情報処理学会第 47 回全国大会, 7L-1, 1993.
- [2] Shafer G., "A Mathematical Theory of Evidence," Princeton Univ. Press, 1976.