

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

7L-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

一般に、計算機による物体の認識は、あらかじめ設定したその物体に関する知識に基づいて行われる。このため、物体に関する知識をいかに記述するかが重要な問題となっている。

画像上での物体の見え方を考えてみると、同じ物体でも、カメラとの距離や場の違いなどで、その物体の画像への写り方は異なる。したがって、物体が大きく写っている場合には、その物体の局所的特徴が、また、小さく写っている場合には、その物体の全体的特徴や周囲環境が、その物体を表す特徴となるため、「物体の見え方」の記述は画像上での物体の写る大きさにより複数必要となると考えられる。

一方、シーンを理解する上で重要な要素の一つに人間がある。これは、多くの場合、人間の存在、及び、行動がシーンの内容を決定するためである。人間領域の抽出に関する研究としては、顔領域の抽出を目的としたもの [1] などが行なわれているが、画像上での色々な人間の見え方に対応できるような方法については検討されていない。

そこで、人間の見え方を画面上での大きさ別に記述する「距離モデル」、及び、周囲環境として画面構成を記述する「カットモデル」を用いたサッカーシーンからの人間の抽出について検討を行った。

2 距離モデルとカットモデル

2.1 距離モデル

画像上に写る人間の大きさは、カメラの焦点距離を固定とした場合、距離の二乗に反比例する。このため、カメラの近くにいる人間は大きく細部まで写るため、顔を構成する目や口の検出により、人間であることを認識できる。一方、遠くにいる人間は小さく細部まで写らないため、人間の全体像や動きの検出により、人間であることを認識する必要がある。距離モデルとは、前述の「人間の見え方」を画像上での大きさ別に記述するもので、近景、中景、遠景

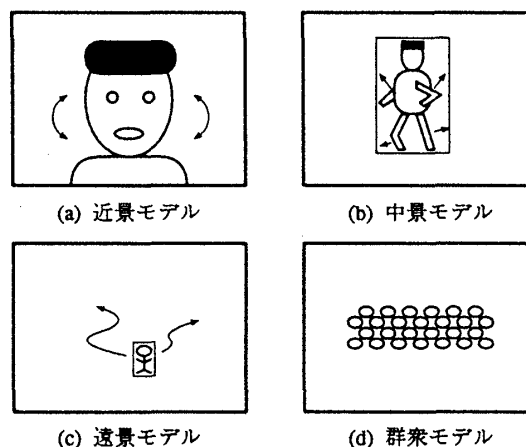


図 1: 各距離モデルの対象

の三モデル、及び、群衆モデルを考えることができる。各距離モデルの対象を図 1 に示す。

- 近景モデル → 人間頭部の見え方モデル、及び、その遷移モデル。
各顔の向きに対して構成部品の形状、色、位置関係、また、各顔の向きの遷移関係を記述する。選手のアップに対するモデルである。
- 中景モデル → 人間全身の見え方モデル、及び、動作のモデル。
人間全身の概略的形狀や構成部品の形状、色、位置関係、また、人間の動作によるそれらの変化を記述する。これは、人間領域を囲む外接長方形の形状変化や内部の色分布により記述する。選手の動作が確認できる場合に対するモデルである。
- 遠景モデル → 場における人間の動きモデル。
人間の場における動きを記述する。これは、場の移動速度やその変化により記述する。選手がフィールド上で動いている場合に対するモデルである。
- 群衆モデル → 群衆の見え方モデル。
群衆の見え方をテキスト特徴で記述する。これは、群衆領域の色分布やパワースペクトル分布などにより記述する。観客が写っている場合に対するモデルである。

Extraction of Human from Sports Scene using Multilayer Model

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

†Institute of Industrial Science, University of Tokyo

†National Center for Science Information Systems

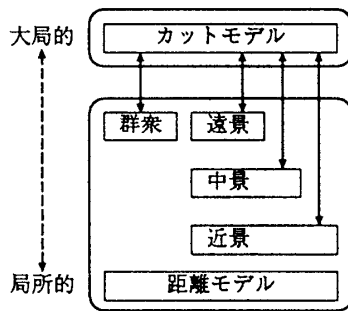


図 2: 各モデルの関係

2.2 カットモデル

カットモデルとは、画面上での各距離モデルにおける人間の配置関係、及び、背景などの周囲環境を記述するモデルである。このモデルは、距離モデルの上位に位置し、各距離モデルの処理結果の統合管理を行なう。サッカーシーンの場合、人間の配置関係とは「フィールド上に選手は存在する」、「画面の前面下ほど近景の人物がいる可能性が高い」「フィールドの周囲の観客席に群衆は存在する」などの知識で記述される。

3 認識過程

各モデルは、以下に示す過程にしたがって並列に認識を実行する。各モデルにおける認識では、Dempster&Shaferの確率理論 [2] を利用し、確信度を算出する。そして、その結果を統合することで最終結果を求める。各モデルの関係を図 2 に示す。

3.1 各モデルによる認識

○ 近景モデル

人間の特徴的な肌色を利用して、人間の頭部候補領域を抽出する。エッジ分布から、目、口領域の抽出を行ない、各候補領域の確信度を求める。

○ 中景モデル

カットモデルにより定義されるフィールド領域を分離し、動領域を追跡することで、人間候補の領域を抽出する。人間領域はフィールド領域にかかり、外接長方形が縦長であるなどの知識から、各候補領域の確信度を求める。

○ 遠景モデル

中景モデルと同様に、人間候補領域を抽出する。速度、動きから、各候補領域の確信度を求める。

○ 群衆モデル

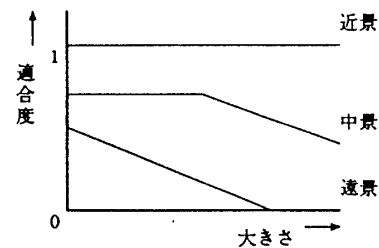


図 3: 人間の写る大きさとモデル適合度

画像をブロック分割し、色ヒストグラム分布、パワースペクトル分布から、群衆人間領域を求める。

3.2 認識結果の統合

同一領域に対して近景、中景、遠景モデルによる認識結果が求まる。これらの統合を行なうことで、人間らしさそれなりの認識が可能となる。

はじめに、それぞれのモデルでの認識において、各知識の満足度を証拠として考え、Dempster&Shaferの確率理論を利用し、{人間である、人間ではない}の各確率を求める。次に、この確率を図 3 に示すようなモデル適合度を考え、各モデルの認識能力を考慮した確率へ変換する。ここで、モデル適合度とは、各モデルの認識の有効度を表すもので、人間が大きく写っている場合には近景モデルが適合しなければならない、人間が小さく写っている場合でも遠景モデルのみで確実な人間抽出はできないということを示すものである。モデル適合度を考慮した新しい確率は、{人間である、人間ではない}の各確率に、適合度を乗じることで求める。そして、変換後の確率に対し Dempster の結合法則を適用することで、最終結果を求める。

4 おわりに

今回、画面上での大きさ別に人間の見え方を記述する「距離モデル」、及び、周囲環境としての画面構成を記述する「カットモデル」の考え方について概説した。今後は、人間候補領域の抽出方法、動きのモデル化や評価方法などについて、更に検討を行なっていく予定である。

参考文献

- [1] 安居院猛, 長尾智晴, 中嶋正之, "静止濃淡情景画像からの顔領域の抽出," 電子情報通信学会論文誌, Vol. J74-D-II, No. 11, pp. 1625-1627, 1991.
- [2] Shafer G., "A Mathematical Theory of Evidence," Princeton Univ. Press, 1976.