色頻度・局所エッジ成分を総合的に用いた物体検出手法とサッカー映像における選手検出への応用

内海 慎久 † 浜田 玲子 † 井手 一郎 ‡ 坂井 修一 †† 田中 英彦 ††

† 東京大学大学院 工学系研究科, ‡ 国立情報学研究所

†† 東京大学大学院 情報理工学系研究科

〒 113-8656 東京都文京区本郷 7-3-1

Tel: 03-5841-7413

{utsumi,reiko,sakai,tanaka}@mtl.t.u-tokyo.ac.jp, ide@nii.ac.jp

一般に，領域分割や動的輪郭抽出手法など物体検出手法の多くは輝度情報を基にすることが多い．従って，物体によっては，輝度に落したときに背景色に吸収されてしまうため，手法の適用以前に検出精度が低下する問題がある．これを解決するため，本稿では，色頻度と局所領域におけるエッジ情報を総合的に取り入れたファジィ評価関数を定義し，それに基づき画素単位で領域判別を行なう手法を提案する．これにより，輝度情報では得られなかった物体領域を補完することが可能になる．また，サッカー映像における選手領域の検出に本手法を応用し，評価実験を通して有効性を確認した．

色頻度, 局所エッジ成分, ファジィ演算, サッカー映像, 選手検出

Object Detection by Associating Color Frequency with

This paper presents a new technique for detecting and tracking players in soccer image sequences. Generally, most object detection techniques, such as boundary-based approaches and region-based approaches, refer to intensity. Since when intensity is referred, foreground pixels tend to be absorbed by background pixels, object segmentation accuracy declines before further processes are performed. To overcome this problem, fuzzy evaluation function that associates color frequency with local edge information and performs region distinction per pixel is proposed. Using this function, object regions ignored by the intensity information is complemented. This technique is applied to the detection of the player region in soccer videos. We confirmed the validity of our method through an evaluation experiment.

Color, Local Edge, Fuzzy Evaluation Function, Soccer Video, Player Detection

1　はじめに

近年の放送媒体の多様化に伴い，大量の映像データが家庭に配信され，またデータベースとして蓄積されつつある．これらのマルチメディアデータを再利用し有効に活用するため，映像の索引付けや検索に関する研究が昨今盛んに進められている．ここで我々はサッカー中継映像を対象とし，映像からの選手検出手法について述べる．

サッカー映像の検索においては，各イベントに含まれる特徴的な画像を検出するなどの比較的簡単な処理で実現することができる．また，イベント発生時のカメラワーク [7] や歓声などの音声情報 [8] を用いることで，さらに精度良くイベントを検出する試みもなされている．

一方，「パス」「シュート」などの特定動作や特定選手の検索においては，映像から選手やライン・ボールなどのオブジェクト（物体）をそれぞれ検出し，動作や位置関係の変化に基づいて選手の動きを分類するなどの処理が必要になる．

しかしながら，サッカー映像における物体検出は困難な問題である．映像からの物体検出の研究は，コンピュータビジョンの分野では古くから行なわれてきた．しかしそれらの研究は，固定カメラを用いる等の限定された環境下で実験評価したものが多い．従って，テレビサッカー映像のように，移動するカメラにおいて複数物体がランダムに移動する映像に対して適用することは困難である．さらに，上記の手法の多くは，計算量や情報量の削減のために輝度情報を基にすることが多い．そのため一部の物体は，輝度に落としたときに背景色に吸収とである．これは，従来の物体検出手法で用いられてきた輝度濃淡画像に比べ，適切に物体／背景領域を判別できるように画素を配分した画像である．従って，本手法で得られた濃淡画像を従来の検出手法に適用すれば，検出精度の向上が期待できる．本稿では，さらに提案手法をサッカー映像における選手領域の検出に応用し，評価実験を通して有効性を確認した．

以下，2 章では従来の物体検出手法の問題点及び提案手法について述べる．3 章ではサッカー映像における選手検出手法について，同様に従来手法と提案手法について詳述する．4 章では評価実験について述べ，提案手法の有効性を明らかにする．

2　物体検出手法

2.1　従来手法の問題点

物体検出手法では，入力画像として輝度濃淡画像が用いられることが多い．しかし，輝度情報に変換する際に色情報が失われるため，一部の物体領域が背景領域に吸収されてしまう問題がある．これは，検出手法を適用する以前に検出精度を下げる要因になっている．

さらに，画像内の強い輝度差（エッジ）の影響で，比較的輝度差の小さい物体領域の検出に失敗することがある．例えばサッカー画像では，選手領域は輝度差が大きいために検出しやすいが，ライン領域の場合，輝度差は選手領域のそれに比べて小さいため，しばしば検出漏れの対象になる．

2.2　提案手法の概要

これらの輝度濃淡画像の問題点に対し，本手法では，色頻度及び局所エッジ成分に着目した濃淡画像を生成することで解決する．狙いとしては，色頻度を用いることで輝度情報だけでは得られなかった領域を取得し，局所領域でのエッジ成分に着目することで，ライン領域など，画像内で比較的エッジの小さい物体領域を補完することにある．

具体的には，3 次元 RGB ヒストグラムから得られる

本手法では，ファジィ演算を基にした評価関数を定義する．通常の演算では絶対値が演算対象になるが，ファジィ演算では各属性の「度合い（評価）」を演算対象とする．従って，色頻度やエッジ成分といった異なる属性同士の演算に適用することができる．

以下ではファジィ演算について述べ，次に評価関数の詳細について述べる．

2.3.1 ファジィ演算

曖昧さを取り扱う理論をファジィ理論という．ファジィ理論は，例えば制御を行なう場合など，曖昧さを無視できないような状況下で用いられる．

ファジィ理論でもっとも中心的役割を示すものは，ファジィ集合である．ファジィ集合は，通常の集合を拡張したものである．ある事実にどの程度あてはまるかという度合いを [0,1] の定義域で表すことで，曖昧さを

集合の度合いは [0,1] で正規化したものなので，ファジィ演算の結果も必ず [0,1] に収束する．

から，ある事項を評価するために必要な属性を見つけ出すという問題，あるいは，いくつかの代替案を評価するという意思決定の問題に応用できる．従って，本研究の

提案手法では，画素 $I(i,j)$ が物体領域に属する度合

$$S(i,j) = \max(F(i,j), \min(E_{\{R,G,B\}}(i,j))) \quad (1)$$

F，E はそれぞれ I における色頻度，局所エッジ成分の評価である．F，E，S は [0,1] の値をとるが，1 に近いほど物体領域，0 に近いほど背景領域に着目する画素が属することを意味する．F，E の max 値をとることで，ラインの切目などの，背景色と類似していてもエッジの存在する画素の評価が上がる．

また，局所エッジ成分 E は，各 RGB 成分での局所エッジの度合いを min 演算している．これは，エッジには全 RGB 成分に画素差がみられる，という仮定に基づいている．すなわち，全 RGB 成分に画素差がみられないものは，雑音と仮定している．従って着目画素がエッジならば，min をとってもエッジ成分は残る．一方，雑音の場合，min をとることで画素差を吸収でき，評価 E を抑えることができる．

2.3.3 色頻度の正規化

サッカー映像において，頻度の少ない色ほど物体領域であり，頻度の多い色ほど背景領域であると考えられる．よって，色頻度が少ないほど F を上げる．

まず各 RGB 値を N 段階に粗量子化した後，RGB3 次元ヒストグラム H_N を計算する．これより，$I(i,j)$ における $F(i,j)$ を式 (2) で定義する．

$$F(i,j) = \begin{cases} 0 & H_N(i,j) > 2\overline{H_N} \\ 1 - \frac{H_N(i,j)}{2\overline{H_N}} & otherwise \end{cases} \quad (2)$$

ここで，$\overline{H_N}$ は H_N の平均値である．$F(i,j)$ は [0,1]

る．本手法では，3×3 画素の局所領域内におけるエッジ成分の度合いを求めることで，ラインの切目など，小さなエッジ成分をもつ物体領域を補完する

着目画素を $I(i,j)$，周辺画素を $C_n(n = 0,1,\ldots,7)$ として，エッジ成分の度合い d を式 (3) で正規化する．

これより，$I(i,j)$ の局所領域におけるエッジ $E(i,j)$ を式 (5) で定義する．

$$E_K(i,j) = \frac{\sum_{n=1}^{8} d_{Kn}(i,j)}{8} \quad (5)$$

すなわち，周辺画素 C へのエッジ成分を平均することで，物体領域周辺や雑音の領域での評価を抑えている．E は [0,1] で正規化された値であり，F 同様，値が大きいほど物体領域，小さいほど背景領域であるといえる．

2.4 領域判別

S は [0,1] の値なので，255 倍することで 8bit の濃淡画像を生成することができる．さらにその濃淡画像に対し，判別閾値法による 2 値化を行うことで，物体領域を検出する．

本手法をサッカー映像に適用した例を図 1，2に示す．各図の（左）は，提案手法と比較するため，輝度濃淡画像に Sobel フィルタをかけた後に，判別閾値法で 2

ある．また，図2（左）ではラインの切目が目立つが，フィルタをかけた際に，ラインのエッジ成分が背景領域に吸収されてしまったためである．

は，色情報を特徴量に用いることによって適切に物体領域を検出できたこと，局所領域のエッジに着目することで，画像全体的には小さなエッジ成分をもつ領域を補完することができたこと，などが理由にあげられる．

3　選手領域検出への応用

3.1　提案手法の概要

選手などの移動物体検出（追跡）への主なアプローチとして，境界ベース (Boundary-based) と領域ベース（Region-based）が挙げられる．境界ベースの代表例としては，SNAKES を時空間に適用した Active contour モデル [5] などがあり，これらは多くの追跡問題に用いられている．ただし，正確な初期状態を設定する必要があり，スポーツ映像においてその都度設定することは現実的ではない．

に与えるクラスタの形状が結果に影響して本来の領域境界以外の部分で分割されることなどがあげられる．

本手法では，大量の映像を処理するという意味合いから，なるべく計算量を抑えるため，ヒューリスティックな情報を用いて選手領域を検出する．さらに，複数フレームにおける処理結果を用いて精度の向上を目指す．

などの不要領域を検出した後，フィールド領域を取得しておく．次に，フィールド領域と 2 値化された画像を比べることでフィールド内の物体領域を検出する．ここで，物体領域と選手・ライン・ボールを指す．

さらに，領域の大きさ等から隠れ（選手またはラインとの重なり）のない選手領域を取得し，追跡を順次行なって残りの選手領域を取得する．

3.2　不要領域検出

3.2.1　画像周辺領域

一般に，サッカー映像はカメラ動きが激しい．そのため，キャプチャリングの際に画像の周りにブレが残り，選手領域の誤検出につながることがある．よって本手法では縦 5 画素，横 6 画素の周辺領域を予め不要領域とした．

3.2.2　フィールド外領域

ここでは，観客席などの，選手が含まれない領域を入力画像から除去する．スポーツ映像では，フィールドは一定の色範囲をもつことから，色を使っておおよそのフィールド領域を先に求める．

まず，Smith の 6 角錐モデルに基づき，色相 H を取得する．次に H 値を 2 つの閾値 θ_{min}，θ_{max} を用いて大まかに緑成分を含む領域を取り出す．

$$\theta_{min} \leq H(i,j) \leq \theta_{max} \tag{6}$$

さらに，観客席にも緑成分が含まれることがあるので，面積の一番大きな領域を取り出す．具体的には緑領域に対しラベリングを行い，最大領域を取得する．次にその領域の輪郭を取り出し，輪郭内の領域をフィールド領域とする．これと入力画像との差分によりフィールド外領域を除去する [4]．

3.2.3　得点欄・時刻欄・その他の字幕

3.2.2項で得られたフィールド領域には，得点欄・時刻欄などの字幕が含まれることがあり，これらは誤検出の原因になるので取り除く必要がある．

サッカー映像における得点欄・時刻欄の特徴としては

- 試合によって出現位置が変わるが，同試合においては同じ位置に出現する
- 試合中に不定期に出現する
- 目立たせるために，比較的大きな輝度値を持ち，出現中の輝度値の変化はほとんどない
- 基本的に横方向に出現する

に変換した後，Sobel フィルタをかけて 2 値化

[redacted]

を除去

3. フレームを 10pixel×10pixel の小領域に分割し，エッジの密集度が閾値以上である領域を字幕領域候補とする
4. 字幕領域候補が横につながる部分を字幕領域として検出

以上の処理により，フィールド領域から得点欄・時刻欄を除去する．なお，本稿では映像の初期フレームに字幕があることを前提とした．

3.3　物体領域検出

3.2.2項と 3.2.3項の処理により，フィールド領域が検出される．さらに，3.1節で述べたように，生成した濃淡画像に Log フィルタをかけ，2 値化した画像とフィールド領域を比べることで，フィールド内の物体領域を取得する．

3.4　選手領域検出

3.4.1　Contour-based Labeling

物体検出・追跡においてラベリングは重要な処理である．ここで，既存のアルゴリズムでは物体画素が見つかる度に，周囲 8 方向のラベルの有無を確認せねばならなかったため，計算時間が大きくなる問題があった．

本手法では，フィールド領域，隠れなし選手領域の 2 検出処理においてラベリングを行なわれる．従って，アルゴリズムの改善により，全体の処理における計算時間を大幅に削減することが期待できる．そこで次のアルゴリズムでラベリングを行なったところ，既存のものに比べ，映像にもよるが，5 〜 6 倍程度の処理速度の向上を確認することができた．

1. 入力画像配列と同じ大きさのラベリング用配列を別途用意
2. 物体領域の画素があれば領域周囲を追跡
3. 追跡中に，周囲画素を中心とする周囲 8 方向の画素についてラベルの有無を確認
4. ラベルがあれば，ラベリング配列において周囲画

7. 処理 2 に戻る．入力画像から物体領域が完全に消

3.4.2　隠れなし選手領域の検出

ラベリング後，各領域の幅 w・高さ h を計算し，次の条件を満たすものを選手領域とする．ここでの選手領域とは，隠れのない選手領域のことを指す．

- $h_{min} \leq h \leq h_{max}$
- $w_{min} \leq w \leq w_{max}$
- $w \leq h$

実験例を図 4に示す．図 4（上）は幅・高さの制限に

選手間の隠れがなく，さらに対応付けられていない選手（重なり画素がない）については，ラインなどの他の物体との隠れが発生していると考えられる．そこで，物

法では選手領域をロバストに検出することを想定しており，そのため多少の雑音がテンプレートに含まれることがある．よって，この雑音の影響によりマッチングがずれたと考えられる．

また，表 3は，フレーム単位で検出の精度を測ったものである．すなわち，フレーム内の全選手を検出できた場合に，正解とする．一度追跡が失敗すると，その後の検出率にも影響するため，表 2 に比べて検出精度は下がっている．ただし，隠れなし選手に関しては表 2 同様，高精度で検出できていることから，誤追跡を解消することで全選手の検出精度を向上させられると考えられる．隠れなし選手の誤検出の原因は，フィールド領域の検出時に誤って観客席を含んでしまい，そのため観客席も問題ないと考えられる．

一方，隠れのある選手の検出・追跡については問題が残る．テンプレートマッチングは画像内からの静止物体の検出には有効であるが，選手のように姿勢が激しく変化する移動物体については高精度な追跡が難しい．これについては，動き予測を別途用いることが考えられる．[9] では，テンプレートマッチングと運動予測を行なっている．また，Kalman フィルタを適応的に設定することで，隠れが発生する状況においても適切に選手を追跡する研究例 [6] が報告されており，これらの手法については今後検討する．また，テンプレート内に含まれる雑音のため，誤追跡になることもあるため，テンプレートからの雑音消去を行なうことも課題である．

[redacted]

度濃淡画像を本手法で得られた濃淡画像の代わりに用

[redacted]

研究,” 第 63 回情処全大，No.6L-03，Vol.2，pp.57-58,

た物体検出手法を提案した．またサッカー映像における選手領域検出に応用し，実験を通してその有効性を確認した．

物体検出手法で得られる濃淡画像は，画像の統計的要素から機械的に領域の判別を行なうものである．すなわち，画素成分の物体の形状などの前知識は用いない．

[5] Gerald Kuhne, Stephan Richter, Markus Beier, “Motion-based Segmentation and Contour-based Classification of Video Objects,” ACM Proc. of the 9th ACM Intl. Conf. on MultiMedia, pp.41-50, 2001.

[6] 三須俊彦, 苗村昌秀, 境田慎一, 鄭文涛, 金次保明, “複数情報の融合によるサッカー選手のロバストな追跡法,”

[redacted]

まれるものの，隠れなしの選手に関してはほぼ完全に検出できている．本研究では，選手領域の位置を基にして映像を自動記述し，検索などの応用に活かすことを目指しており，その意味では選手領域に多少の雑音はあって

pp.306-313, 1996.

[9] 松井健太郎, 平野一視, 田中敏光, 大西昇, “任意視点からサッカーシーンを復元する仮想カメラシステム,” 信学技報, PRMU97-253, 1998.