

LI-012

オブジェクト追跡と背番号認識の連携による動画像用スポーツ選手同定手法

Identification of Moving Athletes Based on Cooperation of Object Tracking and Number Recognition

三須 俊彦[†] 苗村 昌秀[†]
Toshihiko Misu Masahide Naemura高橋 正樹[†] 和泉 吉則[†]
Masaki Takahashi Yoshinori Izumi

1. まえがき

スポーツ番組等において映像オブジェクトに関連する情報を提供できる新しいデータ放送サービス [1] や、映像効果を効率的に提供するシステムの開発を行なっている。映像中においてどこに誰が映っているかといったメタデータを自動的に生成するため、オブジェクト追跡 [2][3] と背番号認識を組み合わせた人物同定手法を開発した。追跡結果によって探索領域を限定し、時間方向の統計処理を行うことで背番号認識の効率化・高精度化が可能になる。本稿では、追跡と認識の連携方法、ガウス混合モデルに基づく背番号文字領域の抽出方法について述べ、サッカー映像を用いた実験結果を報告する。

2. 本手法の構成

本手法は、図 1 に示すように追跡部 (Tracking) と認識部 (Recognition) の 2 部で構成されている。追跡部は、オブジェクト (選手) の色、テクスチャ、動きといった特徴量に基づく複数のテンプレートマッチング手法を用い、頑健にオブジェクトの位置を追跡する。

認識部は、追跡部によって推定される選手の位置情報に基づいておおよその背番号領域を切り出し (方形領域)、文字とユニフォームの領域分割、文字認識、多数決処理を行って、各オブジェクトの所属チーム、背番号、名前を認識する。以下では、動作について順を追って説明する。

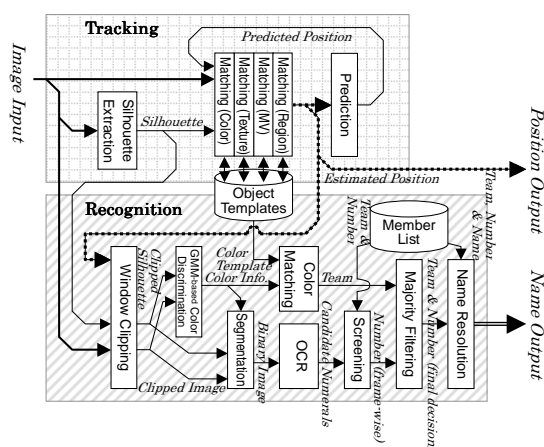


図 1: Block diagram of the system

2.1 追跡部

追跡部は、1 台～複数台のカメラ画像に基づき時々刻々のオブジェクトのシルエット (オブジェクト領域を 1, 背

[†]日本放送協会, NHK

景を 0 とする 2 値画像) 及び位置を抽出する。シルエットはクロマキーもしくは背景差分法により検出する。続いて、オブジェクトの色、テクスチャ、動き場、シルエット領域の各情報に基づきテンプレートマッチングを行い、その結果を拡張 Kalman フィルタをベースとする手法によって情報融合することで、オブジェクトの重心位置を推定する。詳細は文献 [3] を参照されたい。

2.2 背番号候補領域の切り出し

追跡部によって得られる各オブジェクト位置のワールド座標を、カメラ位置、姿勢、焦点距離に基づいてカメラ座標系へ投影する。続いて、投影された座標近傍に存在するシルエットの連結領域を抽出する。その連結領域のバウンディングボックスにおける上下 20% ~ 50% の方形領域内であって、かつシルエット上である領域を背番号候補領域 B とする。

2.3 文字色とユニフォーム色の抽出

本手法では、背番号候補領域が文字色とユニフォーム (文字背景) 色との 2 色で構成される場合を考え、色ベクトル c の分布の確率密度関数 $p(c)$ が基底要素数 2 のガウス混合モデル

$$p(c) = w_1 N(c; \mu_1, \Sigma_1) + w_2 N(c; \mu_2, \Sigma_2) \quad (1)$$

に従うと仮定する。ここに、 μ_m, Σ_m は色の平均及び共分散であって、文字色は $m \in \{1, 2\}$ のいずれか一方、ユニフォーム色はその他方に対応する。また、 $N(c; \mu, \Sigma)$ は平均 μ 、共分散 Σ の多変量ガウス分布、 w_1, w_2 はガウス混合モデルの混合重みである。なお、同一チームの選手は同じユニフォームを着用していると考えられるが、照明条件の違いなどの影響を排除するため、本手法では選手ごとに色分布をモデル化することとした。

EM アルゴリズムにより、第 m の基底要素から色ベクトル c が出力される確率

$$\zeta_m^{(k)}(c) = w_m^{(k)} N(c; \mu_m^{(k)}, \Sigma_m^{(k)}) / \sum_{i=1}^2 w_i^{(k)} N(c; \mu_i^{(k)}, \Sigma_i^{(k)}) \quad (2)$$

を用いて、下記漸化式の繰返計算により混合重み $w_m = \lim_{k \rightarrow \infty} w_m^{(k)}$ 、平均色ベクトル $\mu_m = \lim_{k \rightarrow \infty} \mu_m^{(k)}$ 、共分散行列 $\Sigma_m = \lim_{k \rightarrow \infty} \Sigma_m^{(k)}$ 、確率 $\zeta_m(c) = \lim_{k \rightarrow \infty} \zeta_m^{(k)}(c)$ を求めることができる。

$$w_m^{(k+1)} = \sum_{x \in B} \zeta_m^{(k)}(c(x)) / \sum_{x \in B} 1 \quad (3)$$

$$\mu_m^{(k+1)} = \sum_{x \in B} \zeta_m^{(k)}(c(x)) \cdot c(x) / \sum_{x \in B} \zeta_m^{(k)}(c(x)) \quad (4)$$

$$\Sigma_m^{(k+1)} = \sum_{x \in \mathcal{B}} \zeta_m^{(k)}(c(x)) \cdot \{c(x) - \mu_m^{(k)}\} \{c(x) - \mu_m^{(k)}\}^T / \sum_{x \in \mathcal{B}} \zeta_m^{(k)}(c(x)) \quad (5)$$

続いて, $m \in \{1, 2\}$ のいずれが文字色及びユニフォーム色に対応するかを推定する(その値を m_{num} 及び m_{uni} とおく). 文字色とユニフォーム色の各領域を比較すると, 文字色の領域の方が空間的に拡がり小さいと考えられる. そこで, $\zeta_m(c(x))$ ($m = \{1, 2\}$) の空間分布(水平方向)の2次モーメントを比較し, 小さい方が文字であると見なすこととした.

$$n_m = \sum_{x \in \mathcal{B}} \zeta_m(c(x)), \quad a_m = \frac{1}{n_m} \sum_{x \in \mathcal{B}} x \zeta_m(c(x))$$

$$\begin{bmatrix} W_m & C_m \\ C_m & H_m \end{bmatrix} = \frac{1}{n_m} \sum_{x \in \mathcal{B}} x x^T \zeta_m(c(x)) - a_m a_m^T$$

$$m_{\text{num}} = \arg \min_{m \in \{1, 2\}} W_m, \quad m_{\text{uni}} = 3 - m_{\text{num}} \quad (6)$$

2.4 文字とユニフォームの領域分割

$\zeta_{m_{\text{num}}}(c)$ は, 色ベクトル c が文字色であるときに大きな値をとる. そこで, 下式の閾値処理(閾値 θ)により文字領域が1, その他の領域が0である2値画像 $T'(x)$ を得る.

$$T'(x) = \begin{cases} 1 & (\zeta_{m_{\text{num}}}(c(x)) \geq \theta) \\ 0 & (\text{otherwise}) \end{cases} \quad (7)$$

for $x \in \mathcal{B}$

続いて, 2値画像 $T'(x)$ の文字領域のバウンディングボックスの縦横比から数字が1桁であるか2桁であるかを判別する. 2桁である場合には, 2値画像 $T'(x)$ を垂直投影したヒストグラムの極小点に基づき, 2桁の境界を検出する. 最後に, 各桁の領域に対する面積フィルタリングによってノイズを除去し, 最終結果 $T(x)$ を得る.

2.5 光学文字認識(OCR)

認識エンジンには, エー・アイ・ソフト株式会社様製「読ん del!ココ™ 活字文書 OCR ライブラリ」を使用した(認識対象の文字セットは数字のみとした). このライブラリでは, 各文字につき複数の認識文字候補を信頼度とともに得ることができる.

多くの場合, スポーツにおいては参加選手が予め既知であるため, 出現し得る背番号集合を限定することができる. そこで, 本手法では, 表1に示すようなメンバーリストを用い, 認識精度の向上を図った.

例えば, ある選手に関する2値画像 $T(x)$ のOCRの結果, 表2に示す2桁の数字候補群が得られたとする. まず, 表1のメンバーリストに存在する背番号全てについて, OCR結果の候補に存在するか(桁の組み合わせによりその背番号を作れるか)否かを調べる. この例の場合, 12と15が背番号候補となる.

次に, 全ての背番号候補について信頼度を求める. 2桁の場合, うち最小の信頼度をもって全体の信頼度とする. この例では, 背番号12は信頼度45, 背番号15は信頼度60となる.

表 1: Member List

H: Home (Red)		A: Away (Blue)	
Number	Name	Number	Name
9	松村	3	吉野
12	中村	6	永龍
16	小沢	8	安田
18	真壁	12	芝
22	寺内	15	横山

表 2: Example of Numeral Candidates

Position	Tens		Units	
	Numeral	Confidence	Numeral	Confidence
1 st Candidate	7	65	5	70
2 nd Candidate	1	60	3	55
3 rd Candidate	9	40	2	45

最後に, 背番号候補の中から信頼度最大のものを認識結果として出力する. この例では背番号15が認識結果となる.

2.6 チーム分類

本手法では, ユニフォームの色情報によってチームの分類を行う. 予め, 各チームのユニフォームの色分布の平均(ホーム側チーム μ_H , アウェイ μ_A)と共分散(ホーム Σ_H , アウェイ Σ_A)を求めておく.

上記各統計量と, 2.3で求めたユニフォーム色の平均 $\mu_{m_{\text{uni}}}$ とのマハラノビス距離を比較し, 距離のより近いチームを求めることで分類を行う.

$$(\text{Team}) = \begin{cases} \text{Home} & (d_{\Sigma_H}(\mu_H, \mu_{m_{\text{uni}}}) \leq d_{\Sigma_A}(\mu_A, \mu_{m_{\text{uni}}})) \\ \text{Away} & (\text{otherwise}) \end{cases} \quad (8)$$

$$d_{\Sigma}(\mu, c) = (c - \mu)^T \Sigma^{-1} (c - \mu) \quad (9)$$

2.7 名前の解決

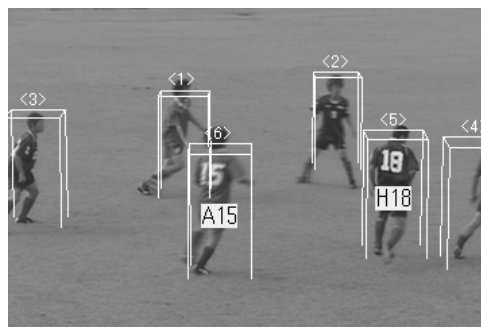
背番号の歪みやオクルージョンなどにより誤認識が生ずる可能性があるため, 時間方向の多数決を行う. 多数決においては, 2.5で述べた信頼度の重み付けを行う. その結果に基づき, 表1に示したメンバーリストを参照し, 背番号とチーム分類から選手名を求める.

3. 実験結果

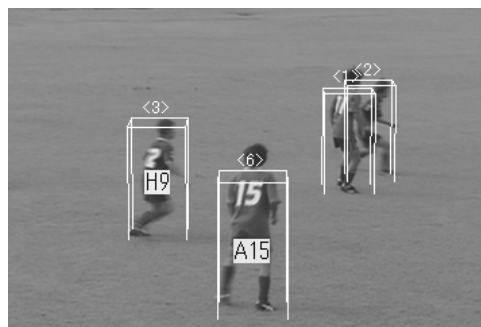
図2に示すサッカーのテスト映像を用い, 本手法による選手同定の実験を行った.

同図(a)は第14フレームにおける選手位置と背番号認識の結果を原画像上に重ねて表示したものである. 計6名の選手のうち, 背を向けているのは<5>及び<6>の2名である. 表3の2値画像に示すように, これら2名の選手に関しては, 背番号領域が正しく抽出できている. 他の選手に関しては, セグメンテーションの段階において文字領域が存在しないと判断され, OCR以降の処理がスキップされている.

これら2値画像に対しOCR及びチーム分類を行った結果を表3のRecognition欄に示す. それぞれホーム側



(a) Frame # 14



(b) Frame # 58

図 2: Tracked positions and recognized numbers

表 3: Result of Number Recognition

Frame #	14			58		
Object #	<1>- <4>	<5>	<6>	<1>, <2>	<3>	<6>
Binarized Image	-			-		
Recognition	-	H18	A15	-	H9	A15
Result	-	True	True	-	False	True
Confidence	0.0	58.0	47.8	0.0	47.1	67.5

18 番とアウェイ側 15 番というように正しい認識結果が得られた。また、表 3 の Confidence 欄を見ると、選手 <5> のように鮮明な 2 値画像が得られた場合には、信頼度が高くなることわかる。

一方、第 58 フレーム(図 2(b))においては、選手 <3> の番号認識に失敗している。この選手の背番号はセルフオクルージョンと歪みによって完全には見えていない。この不完全な画像を認識したため、表 1 のメンバーリスト中にある最も類似した数字 9 を出力している。ただ、表 3 の Confidence 欄の信頼度の値から、この認識結果が信頼できないこともわかる。

図 3 は背番号認識の結果をカラーコードによりタイムライン表示 [4] したものである。時間軸方向に信頼度を重みとする多数決を行い、表 1 のメンバーリストを参照して選手名の自動補完を行った。その結果を図 3 の左端の Names 欄に示す。「真壁」及び「横山」については正解である。

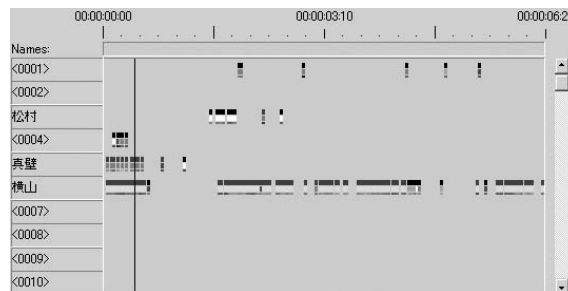


図 3: Time-Line Visualization of Numbers

一方、選手 <3> (図 3 の上から 3 番目) は「松村」となっているがこれは誤りである。表 3 の第 58 フレームの結果にも示したが、この選手は長期間に渡って横を向いており、完全な形で 2 桁の数字が見える機会に乏しかったことが、誤認識の原因であると考えられる。

こうした誤認識を防ぐためには、今回導入したような信頼度を考慮するだけでは未だ不十分である。例えば、選手のシルエット形状に基づいて選手姿勢を検出し、背番号の正しく見える可能性のあるシーンだけを OCR にかけるといった工夫が必要であると考えられる。

4. まとめ

オブジェクト追跡と背番号認識を連携させたスポーツ選手の同定手法を提案した。本構成により、移動する選手の背番号位置を絞り込むことができ、OCR をかけるべき探索範囲を限定することができる。また、ガウス混合モデルによる色分布のモデリングにより、文字色とユニフォーム色のさまざまな組み合わせに対し、無調整で背番号の 2 値画像を得ることができるようになった。また、メンバーリストを利用し、背番号の探索空間を絞り込むことで、背番号認識の頑健化を図っている。

今後は、選手姿勢に応じて背番号の見えないシーンを排除し、誤認識の低減を目指したい。また、背番号認識結果を追跡部へフィードバックすることで、追跡処理の誤り(交差した選手が入り替わって追跡されてしまう誤りなど) を自動的に修正することを考えている。

参考文献

- [1] 苗村, 鄭, 三須, 和泉, 金次: 映像オブジェクトに連動した放送サービス「情報ナビゲーション型未来放送」, 2001 年映像情報メディア学会冬季大会, 4-10, p. 80, 2001.
- [2] T. Misu, M. Naemura, W. Zheng, Y. Izumi, K. Fukui: Robust Tracking of Soccer Players Based on Data Fusion, ICPR 2002, vol. 1, pp. 556-561, 2002.
- [3] 三須, 苗村, 境田, 鄭, 金次: 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, 信学技報 IE2001-47, pp. 23-30, 2001.
- [4] 三須, 鄭, 苗村, 和泉, 金次: 情報ナビゲーション型未来放送のオーサリングツール, 映情メ技報 BCS2002-3, pp. 13-18, 2002.