

複数尤度を用いた 3 次元パーティクルフィルタによる選手の追跡

西濃 拓郎[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院工学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]fantakuro@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

あらまし 本稿では、視点固定サッカー映像において、SVM とパーティクルフィルタを用いて複数の選手を追跡する新しい手法を提案する。従来手法と比較して、この論文には 3 つの特徴がある。1 つ目はパーティクルフィルタの状態量を 2 次元ワールド座標空間ではなく、3 次元ワールド座標空間にとることによって、奥行きを考慮したより現実に近いモデルでの追跡が可能となった。2 つ目は SVM による選手の検出結果をパーティクルフィルタの枠組みに取り入れることで、初期位置の獲得や見失ったときの復帰が可能となった。3 つ目は尤度評価の特徴量としてヒストグラムに基づく尤度と相関値に基づく尤度の重みづけ和を用いることで、長時間正確に追跡できる特徴量となった。キーワード サッカー、選手追跡、パーティクルフィルタ、EMD、SVM

Tracking of Multiple Soccer Players Using 3D Particle Filter Based on Multiple likelihood

Takuro NISHINO[†], Tetsuya TAKIGUCHI^{††}, and Yasuo ARIKI^{††}

[†] Graduate School of Science and Technology, Kobe University

1-1, Rokkodai, Nada, Kobe, Hyogo, 657-8501 Japan

^{††} Organization of Advanced Science and Technology, Kobe University

1-1, Rokkodai, Nada, Kobe, Hyogo, 657-8501 Japan

E-mail: [†]fantakuro@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

Abstract In this paper, a novel multiple soccer players detection and tracking approach based on support vector machine (SVM) and particle filter is proposed to track players in monocular fixed soccer video. The contributions of this paper are focused on three aspects. First, the state \mathbf{x} of particle filter presented in 3D world coordinate space, instead of 2D image coordinate space makes tracking more realistic model because of depth consideration. Second, by integrating SVM into particle filter framework, the initial positions of players can be detected or rediscovered when the tracker loses the players. Third, the feature histogram and correlation are combined when the likelihood for a particle is computed. Thanks to these features, multiple soccer players are tracked accurately for a long duration.

Key words Soccer, Player Tracking, Particle Filter, EMD, SVM

1. はじめに

スポーツの試合を撮影する場合や編集する場合、単に撮影するのではなく、試合の流れやイベントが分かりやすく面白い映像を撮影・編集することが要求される。これをコンピュータで自動的に行う技術として、自動映像生成技術がある。自動映像生成技術の最も簡単なものとして、デジタルカメラワークがある。これは、スポーツの試合において、フィールド全体を一つの高解像度カメラで撮影し、その映像をクリッピングし、連続的につなぎ合わせることによって新しい映像を作成する技術であ

る。この技術によって、試合の内容が分かりやすく面白くかつ、視聴者の嗜好に即した自由な編集が可能となる [7]。

映像をクリッピングするためには、その映像の内容を理解し構造化する必要がある。スポーツ映像におけるイベントを検出し、その種類に基づいて映像を編集する手法が考えられる。単眼サッカー映像においてイベントを検出するためには、ボールと選手の 3 次元位置情報が必要不可欠であるため、それらを正確に追跡する必要がある [8]。

従来研究数多くの追跡手法が提案されている中で、

tracking-by-detection と呼ばれる追跡アルゴリズムの中に、検出の枠組みを取り入れた追跡手法の有用性が示されている [1] [2] [4]。この手法は、もし追跡中に対象を見失っても復帰が可能であることや、初期位置の獲得が容易であることが大きな利点として挙げられる。検出アルゴリズムには大きく分けて SVM [1] [2] と AdaBoost [3] [4] があるが、どちらの手法も多くの学習データを用いれば、正確に検出できることが報告されている。追跡アルゴリズムとしてはそのオクルージョンへの頑健さからパーティクルフィルタが頻繁に用いられている。

本稿では、視点固定サッカー映像において、SVM とパーティクルフィルタを用いて複数の選手を追跡する新しい手法を提案する。従来手法と比較して、この論文には3つの特徴がある。1つ目はパーティクルフィルタの状態量を2次元ワールド座標空間ではなく、3次元ワールド座標空間にとることによって、奥行きを考慮したより現実に近いモデルでの追跡が可能となった。2つ目はSVMによる選手の検出結果をパーティクルフィルタの枠組みに取り入れることで、初期位置の獲得や見失ったときの復帰が可能となった。3つ目は尤度評価の特徴量としてヒストグラムに基づく尤度と相関値に基づく尤度の重みづけ和を用いることで、長時間正確に追跡できる特徴量となった。

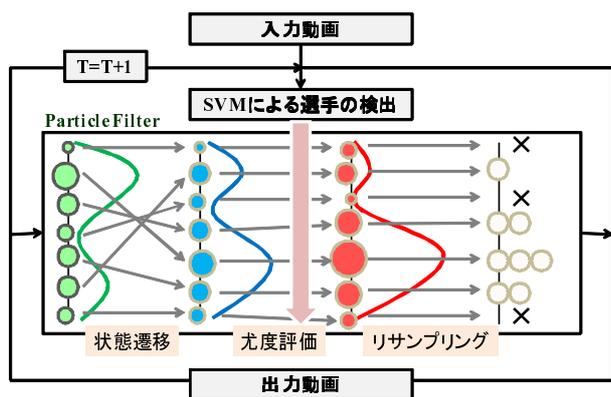


図1 システムの流れ

図1は提案システムの流れを示している。まず毎フレームの始めに選手の検出を行い、パーティクルフィルタによって追跡を行う。パーティクルフィルタは複数の粒子を用いてモデルを離散的に近似しており、毎フレームごとにそれぞれの粒子を状態遷移モデルに基づいて遷移させ、さらに観測モデルに基づいて重みづけすることでその状態を更新していく。この重み付けのフェイズにおいて、SVMによる検出結果を組み込むことで、より正確なモデリングが可能となる。

2. 選手の検出

毎フレームの始めに、SVMによって学習された分類器を用いて選手の検出を行う。今回は図2の上段のようなポジティブサンプル数3000枚、下段のようなネガティブ



図2 学習データサンプル

サンプル数1000枚を用いて、グレースケール画像の画素値を特徴量として学習した。窓探索を窓の大きさを変えながら選手を検出し、検出結果 det の画面座標と SVM スコア $c(det)$ を獲得する。SVM スコアは分類器の識別境界からの距離で定義される。



図3 検出結果

図3は検出結果を表している。SVM スコアの値によっていくつかのグループ分けを行い、選手をそのグループの色の矩形で囲った。最も SVM スコアの高いグループは赤い矩形で表されている。ほぼ正しく検出できているが、選手同士が近付いている場合には、1人の選手として検出してしまっている。

3. 複数選手の追跡

3.1 3次元パーティクルフィルタ

追跡手法は各選手に対して1つのパーティクルフィルタを用いた。これは、状態量と尤度を持つ多数の粒子によって確率分布を離散的に近似し、近似確率モデルを伝搬させることによって、安定した精度で追跡できる手法である。サッカー場にワールド座標系を固定し、その3次元座標を用いて時刻 t における状態量 x は式(1)で定義される。

$$x = [p_x, p_y, v_x, v_y, a_x, a_y]^T \quad (1)$$

x は3次元ワールド座標 (p_x, p_y) と速度 (v_x, v_y) 、加速度 (a_x, a_y) から構成されている。状態量を2次元ワールド座標空間ではなく、3次元ワールド座標空間にとることによって、奥行きを考慮したより現実に近いモデルとなるため、オクルージョンに対してより頑健な追跡ができるようになる。ここでは選手の高さを考慮せず、 (p_x, p_y) は重心の位置を表している。

また、時刻 $t - \alpha$ から t における状態遷移モデルは式(2)のように加速度運動を仮定した。

$$\mathbf{x}_t = C\mathbf{x}_{t-\alpha} + \Upsilon \quad (2)$$

$$C = \begin{bmatrix} I_{2*2} & \alpha I_{2*2} & (\alpha^2/2)I_{2*2} \\ O_{2*2} & I_{2*2} & \alpha I_{2*2} \\ O_{2*2} & O_{2*2} & I_{2*2} \end{bmatrix}$$

$$\Upsilon = [\epsilon_p I_{1*2}, \epsilon_v I_{1*2}, \epsilon_a I_{1*2}]^T$$

Υ はノイズを表しており、変数 $\epsilon_p, \epsilon_v, \epsilon_a$ はいずれも平均 0 の正規分布に従う。位置ノイズの分散 σ_p^2 は追跡対象の大きさによって変化し、速度ノイズの分散 σ_v^2 と加速度ノイズの分散 σ_a^2 は連続追跡成功フレーム数に反比例する。このノイズ処理によって、対象を長時間追跡できればできるほど、パーティクルフィルタの粒子が集束していき、より安定な追跡が可能となる。

3.2 トラックと検出結果の連結

SVM による検出結果を追跡の枠組みの中へ取り入れるためには、どの検出結果がどのトラックに当てはまるのかを決定しなければならない。そのために本研究では [1] で用いられている手法を用いた。このアルゴリズムは、まず検出結果 det とトラック tr のそれぞれの組 (tr, det) について、マッチングスコア $s(tr, det)$ を式 (3) に基づいて計算する。このスコアが高ければ高いほど検出結果とトラックがより似ていることを表している。

$$s(tr, det) = g(tr, det) \cdot (c(det) + \alpha \sum_{p \in tr} p_N(det - p)), \quad (3)$$

$$g(tr, det) = p(size_{det}|tr) = p_N\left(\frac{size_{tr} - size_{det}}{size_{tr}}\right), \quad (4)$$

$p_N(det - p) \sim N(det - p; 0, \sigma_{det}^2)$ は、トラック tr の粒子 p と検出結果 det の位置間距離に基づく正規分布を表している。 $g(tr, det)$ はゲート関数を表しており、 $size_{tr}$ は粒子 p の y 座標 p_y によって決定されるトラックの窓サイズ、 $size_{det}$ は検出結果 det の窓サイズである。 α と σ_{det}^2 は実験的に求めた。そして、検出結果の数を行数、トラックの数を列数としたマッチングスコア $s(tr, det)$ の行列 S を生成する。

最後に S の中で最大のスコアの組 (tr^*, det^*) を見つけ、 S から検出結果 det^* とトラック tr^* に属する行と列を削除する。これを繰り返してトラックと検出結果の連結を行う。

3.3 尤度計算

トラック tr^* の粒子 p の重み $\omega_{tr,p}$ を計算するために、パーティクルフィルタは状態遷移によって得られた新しい観測ベクトル y_t の条件付き尤度を推定する。本研究では、複数の尤度を組み合わせることで、オクルージョン時に頑健な尤度とした。今回用いたのは、検出結果に基づく尤度とヒストグラム間距離に基づく尤度、さらに相

関値に基づく尤度である。

$$\begin{aligned} \omega_{tr^*,p} &= p(y_t|\mathbf{x}_t) \\ &= p_D + \beta \cdot p_H + (1 - \beta) \cdot p_C \end{aligned} \quad (5)$$

p_D は検出結果に基づく尤度を表している。

$$p_D = I(tr^*)(\phi \cdot c(p) + \psi \cdot p_N(p - det^*)), \quad (6)$$

$$I(tr^*) = \begin{cases} 1 & s(tr^*, det^*) > threshold \\ 0 & other \end{cases}$$

$c(p)$ は粒子 p の SVM スコアを表しており、 $p_N(p - det^*)$ は粒子 p と 3.2 で求めた det^* の位置間距離に基づく正規分布である。 $I(tr^*)$ は検出結果がトラックと連結している、すなわち、マッチングスコア $s(tr^*, d^*)$ がある閾値 $threshold$ より大きいときに 1 を返し、それ以外るときに 0 を返す関数である。この関数は、誤検出した結果を尤度に影響させないために用いた。パラメータ ϕ と ψ 、 $threshold$ は実験的に求めた。

式 (5) の 2 番目の項 p_H と 3 番目の項 p_C は、それぞれヒストグラム間距離に基づく尤度と相関値に基づく尤度を表している。前者はオクルージョンや姿勢変化に頑健な特徴量であるが、それ以外るときは粒子の分散が必要以上に大きくなってしまいう問題がある。逆に後者はオクルージョンに弱い粒子の分散が小さいという利点がある。それぞれの利点を生かすために選手間の距離が近く、オクルージョンが起こりそうなときは前者の重みを大きくし、選手間の距離が離れている時は後者の重みを大きくすることで、長時間の頑健な追跡が可能となる。本稿ではヒストグラム間距離として Earth Mover's Distance [6] を、相関値として正規化相互相関値を用いた。それぞれは以下の式ようになる。

$$p_H = p_N(dist_{EMD}), \quad (7)$$

$$p_C = \frac{\sum_{i,j} \{I(i,j) - \bar{I}\} \cdot \{T(i,j) - \bar{T}\}}{\sqrt{\sum_{i,j} \{I(i,j) - \bar{I}\}^2 \cdot \sum_{i,j} \{T(i,j) - \bar{T}\}^2}}, \quad (8)$$

$$\beta = p_N(\min_{i \in P} dist_i), \quad (9)$$

式 (7) における $dist_{EMD}$ は Earth Mover' Distance であり、その距離に基づく正規分布を p_H とする。

正規化相互相関値 p_C は式 (8) で定義される。ここで、 $I(x, y)$ と $T(x, y)$ はそれぞれ対象領域内の座標 (x, y) における輝度値とテンプレート画像内の座標 (x, y) における輝度値を表している。また、 \bar{I} と \bar{T} はそれぞれ対象領域とテンプレート画像の平均輝度値である。

式 (9) は p_H と p_C の重みを表している。 $dist_i$ は追跡対象の選手とその他の選手 i との距離を表している。 P は全選手数である。すなわち、 β は最も近い選手との距離の正規分布 p_N で定義される。

3.4 キャリブレーション

本研究では、状態量 x の定義されているワールド座標系 (X_w, Y_w, Z_w) から、画面座標系 (x_f, y_f) への座標投影モデルとして、ピンホールカメラによる透視投影変換を用いた、このモデルは、式 (10) で定義され、 A は内部パラメータ、 R は回転行列、 t は平行移動ベクトルである。

$$\begin{bmatrix} x_f \\ y_f \\ 1 \end{bmatrix} = A \begin{bmatrix} R & | & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (10)$$

サッカーのフィールドは、一般的な大きさが規定されており、フィールド内のペナルティエリアやゴールエリアなどのラインの長さも正確に定められているため、ワールド座標が既知な点が得やすいという利点を持っている。事前に複数の点を手動で獲得し、Tsai の手法 [5] を用いてキャリブレーションを行った。図 4 にキャリブレーションに用いた点を示す。



図 4 ワールド座標の既知な点

4. 実験

4.1 実験環境

視点固定映像は、1280 * 270 画素、フレームレート 30 f p s のサッカー映像を利用した。内容は第 83 回高校サッカー選手権大会京都府大会決勝、桂高等学校（白チーム）対京都朝鮮高級学校（赤チーム）の試合を撮影したものである。図 5 のように、フィールドの左半分が画面の下半分に映っている映像となっている。この映像から、平均 330 フレームのクリッピング映像を 12 本抽出して、実験を行った。



図 5 実験に使った映像

評価は、次式で定義される追跡精度により行った。

$$\text{追跡精度} = \frac{1}{P} \sum_{i=1}^P \frac{\text{選手 } i \text{ の追跡成功フレーム数}}{\text{選手 } i \text{ の映っている全フレーム数}}$$

あるフレームにおいて、手動で与えた正解 2 次元画面座標データと、パーティクルフィルタによる追跡結果を式 (10) を用いて変換した座標の距離が、5 ピクセル以内であれば追跡成功とした。P は全選手数である。

また、キャリブレーションによって得られたカメラパラメータは、以下の表 1 の通りである。

表 1 カメラパラメータ

外部パラメータ			内部パラメータ	
R	R_x	93.291679[deg]	f	3.155433[m]
	R_y	5.392349[deg]	k	3.609726e-02
	R_z	-0.296688[deg]	s_x	1.0
t	t_x	-28.381686[m]	s_y	1.0
	t_y	12.246007[m]	C_x	640[pixel]
	t_z	71.529384[m]	C_y	360[pixel]

4.2 実験結果

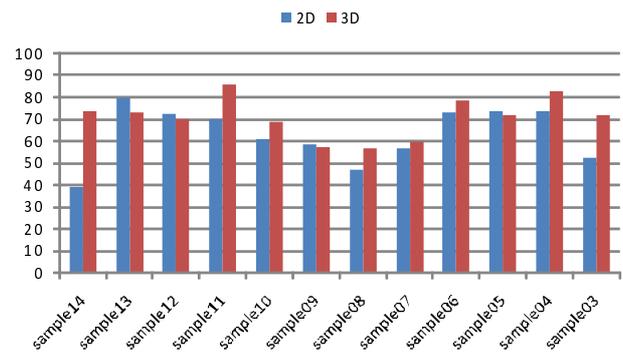


図 6 2次元追跡と3次元追跡の比較結果

図 6 に、従来の 2 次元パーティクルフィルタと提案手法である 3 次元パーティクルフィルタで追跡した結果を示す。尤度評価は共に正規化相互相関値のみ用いた。平均追跡精度は 63.1 % から 70.3 % へ飛躍的に向上した。3 次元ワールド座標系に状態量をとることで、より現実に近いモデル化が可能となり追跡精度が向上しているのが示された。クリッピング映像ごとに考察すると、sample13, sample12, sample09, sample05 はわずかに精度が下がっている。これらの映像に共通している点は、選手にほとんど動きがなく、あったとしても単調な線形の動きが多いという点である。単純な動きを 3 次元で複雑にモデル化したことが精度低下の原因だと考えられる。

表 2 に、3 つの尤度を組み合わせた場合の実験結果を示す。式 (5) における p_D を用いた場合は、用いなかった場合は \times で表している。 p_H は用いなかった場合と Bhattacharyya 距離、もしくは EMD を用いた場合と比較した。 p_C は正規化相互相関値 (NCC) のみを用い

表 2 追跡精度

p_D	p_H	p_C	追跡精度 (%)
×	×	NCC	70.30
×	Bhattacharyya	NCC	72.41
×	EMD	NCC	73.64
	Bhattacharyya	NCC	75.22

た．2行目と3行目を比較すると，同じヒストグラム間距離でも EMD と Bhattacharyya 距離では前者の方の精度が向上している．これは，EMD が後者に比べてよりオクルージョンに強い特徴量であることを示している．また，2行目と4行目を比較すると，検出結果を尤度評価に組み込むことで，飛躍的に追跡精度が向上しているのがわかる．

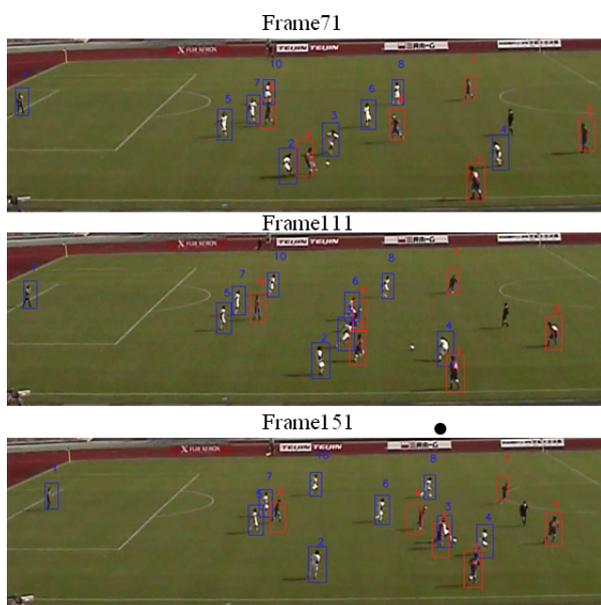


図 7 追跡結果 (広いアングル)

図 7 は実験結果の一部を広いアングルで長いフレーム単位でクリッピングしたものを表している．オクルージョンのあった選手もなかった選手も両方ともに正確に追跡できていることがわかる．図 8 は，実験結果の一部

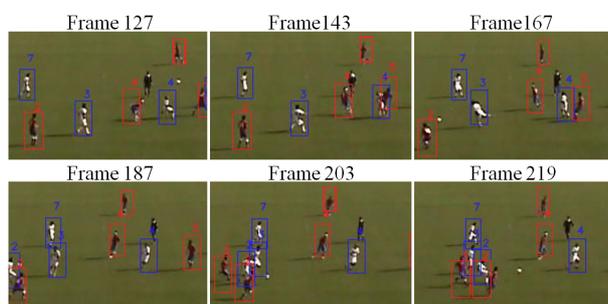


図 8 追跡結果 (近いアングル)

を近いアングルで少ないフレーム単位でクリッピングしたものを表している．短い時間に複数のオクルージョン

が起こって複数の選手がいくつも重なっているが，問題なく追跡できている．

図 9, 図 10, 図 11, 図 12 にクリッピング映像 sample10 と sample13 の最初のフレームと約 5 秒後のフレームを示す．左の図が入力画像における選手の軌跡を表しており，右の図は上面図における選手の軌跡を表している．対応している点が追跡できているのが確認できる．

5. ま と め

本稿では，サッカーの試合における自動映像生成を目的として，単眼サッカー映像において複数の選手を追跡する手法を提案した．2次元画面座標系ではなく3次元ワールド座標系に基づく状態量をパーティクルフィルタに用いることで，より現実に近い奥行きを考慮したモデルによる追跡が可能となった．また，パーティクルフィルタの尤度評価に，SVM による選手検出の結果を組み込み，ヒストグラム間距離に基づく尤度として Earth Mover's Distance を用いることで，よりオクルージョンに頑健な追跡を実現した．

今後は，獲得したボールと選手の3次元位置情報を用いて複雑なイベント（オフサイド，ファウルなど）の検出を行う．

文 献

- [1] M. Breitenstein, F. Reichin, B. Leibe, E. Koller-Meier and L. V. Gool, "Robust Tracking-by-Detection using a Detector Confidence Particle Filter", The 12th IEEE International Conference on Computer Vision (ICCV), pp. 1515–1522, Kyoto, Japan, Sept. 2009.
- [2] G. Zhu, C. Xu, Q. Huang and W. Gao, "Automatic Multi Player Detection and Tracking in Broadcast Sports Video Using Support Vector Machine and Particle Filter", IEEE International Conference on Multimedia & Expo (ICME), pp. 1629–1632, Toronto, Canada, July. 2006.
- [3] Y. Wu, X. Tong, Y. Zhang and H. Lu, "Boosted Interactively Distributed Particle Filter for Automatic Multi-Object Tracking", IEEE International Conference on Image Processing (ICIP), pp. 1844–1847, San Diego, U.S.A, Oct. 2008.
- [4] K. Okuma, A. Taleghani, N. D. Freitas, J. J. Little and D. G. Lowe, "A Boosted Particle Filter: Multi target Detection and Tracking", The 8th European Conference on Computer Vision (ECCV), pp. 28–39, Prague, Czech, May 2004.
- [5] R. Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D-Machine Vision", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 364–374, Miami, U.S.A, Jun. 1986.
- [6] Y. Rubner, C. Tomasi and L. J. Guibas, "The Earth Mover's Distance a Metric for Image Retrieval", International Journal of Computer Vision (IJCV), Vol.40, Number2, pp. 99–121, Nov. 2000.
- [7] 窪田進太郎, "TV サッカー映像のオブジェクト情報に基づいた個人視聴向き内容解説", 神戸大学大学院 修士論文, 2007.
- [8] Xinguo Yu, Xin yan, Yiqun Li, "Tool-Aided Semantics Acquisition for Live Soccer Video", IEEE International Conference on Multimedia & Expo, pp. 893–896, Hannover, Germany, Jun. 2008.

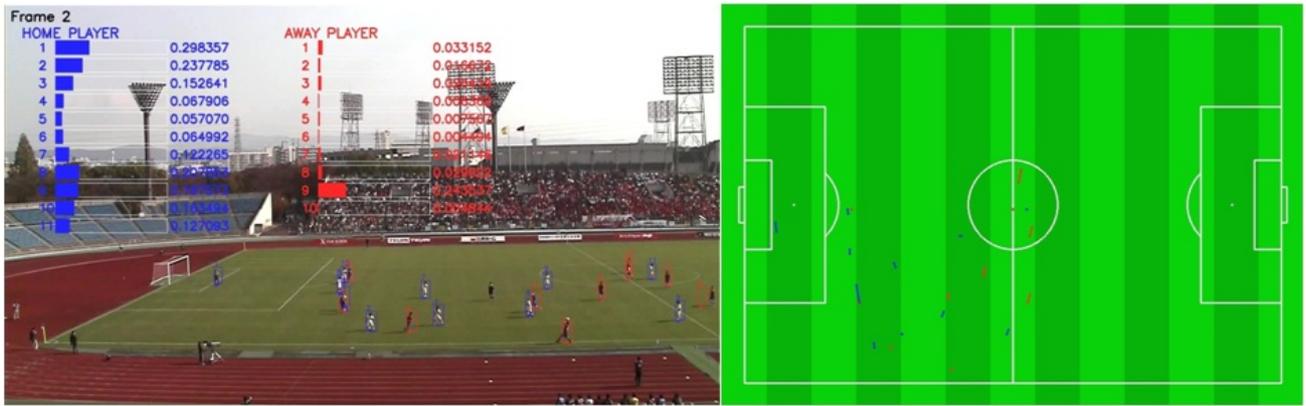


図 9 sample10 における選手の軌跡と上面図（初期フレーム）

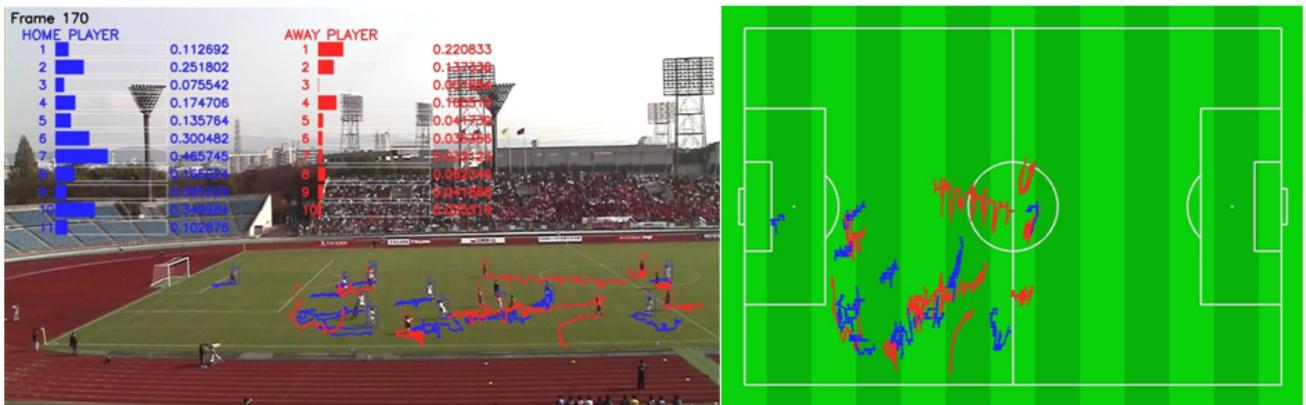


図 10 sample10 における選手の軌跡と上面図（約 5 秒後のフレーム）

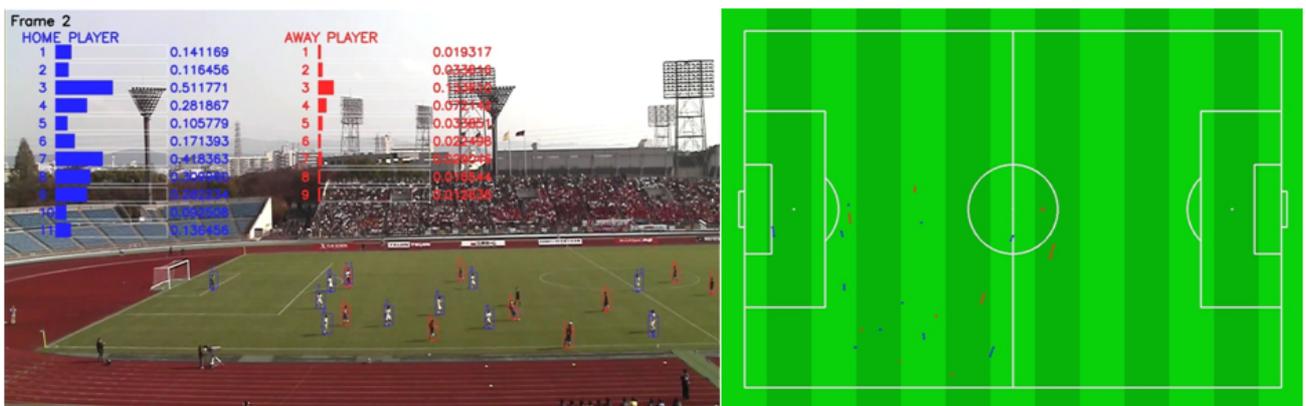


図 11 sample10 における選手の軌跡と上面図（初期フレーム）

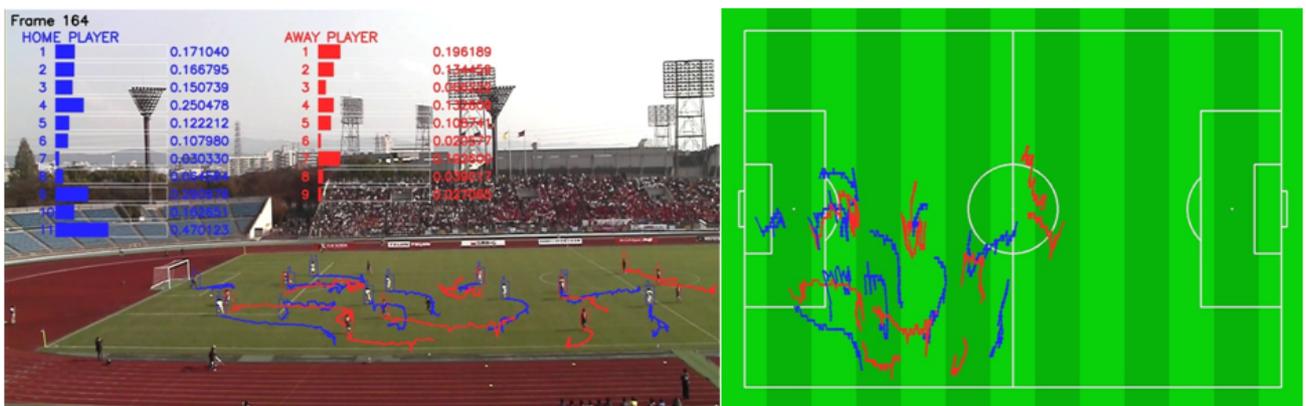


図 12 sample10 における選手の軌跡と上面図（約 5 秒後のフレーム）