

単眼サッカー映像における時間状況グラフを用いた選手追跡

伊藤 拓紀[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院システム情報学研究科 〒 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: [†]itoh@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

あらまし 本研究では、サッカー映像においてオクルージョンにロバストな選手追跡を行うために、時間状況グラフにガイドされたパーティクルフィルタによる新しい選手追跡手法を提案する。従来のパーティクルフィルタによる選手追跡では、映像のフレーム間で複数選手の位置情報を用いていないため、一度対象を見失うと再度発見するのが困難であるという欠点があった。そこで、複数選手の位置情報を時間状況グラフとして表現しておき、これにガイドされる形でパーティクルフィルタを実行すれば、オクルージョンが起こっても選手の誤検出を大幅に減らすことが期待できる。評価実験では、実際の視点固定単眼サッカー映像に対して追跡実験を行い、時間状況グラフを用いないパーティクルフィルタによる選手追跡（従来手法）と、提案手法の時間状況グラフを用いたパーティクルフィルタによる選手追跡を比較した。その結果、従来手法に比べて提案手法の追跡精度は 7.35 ポイント向上した。

キーワード 選手追跡, 時間状況グラフ, パーティクルフィルタ, SVM, “tracking-by-detection”

1. はじめに

近年、個人撮影機器の高画質化やメモリの大容量化など、アマチュアのための撮影環境が整い、映像が誰でも簡単に撮影可能となった。これに伴って、映像編集のニーズが高まり、映像を個人の嗜好に基づいて編集することを支援する自動映像生成技術が注目を集めている。

この技術の一つに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、その映像を切り取り、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という技術である。この技術によって、視聴者の嗜好に即した自由な編集が可能となる。デジタルカメラワークを実現するには、その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必要がある。しかしながら、映像から直接イベント等の高レベルな特徴を抽出することは困難であり、自動映像生成における問題点の一つとなっている。これに対して、より低レベルな特徴を中間に用いることによって、その問題点を克服しようとする研究が盛んに行われている [2], [3]。サッカー映像における低レベルな特徴として、選手やボールの位置が挙げられる。つまり選手やボールの追跡を頑健に行うことは、イベント検出の精度向上につながるため、自動映像生成技術にとって重要である。

従来、映像においてある物体の位置を推定し、追跡する手法としては、mean-shift [4]、カルマンフィルタ [5]、共分散トラッカ [6]、仮説検証 [7] など、多くの手法が提案されている。その中でも注目すべき手法は、パーティクルフィルタと呼ばれる手法 [8] である。この手法を用いてサッカー映像中の選手を追跡する研究は数多く行わ

れている。しかしながら、映像のフレーム間で複数選手の位置情報を用いていないため、オクルージョンなどにより一度追跡対象を見失うと、再度発見するのが困難であり、長時間の追跡が実現できないという問題があった。

また、オクルージョンが生じる下での追跡手法も提案されている。例えば、Grabner ら [9] は、Implicit Shape Model (ISM) によって励起されたモデルを学習することで、オクルージョンによる隠れた部分にも頑健な検出を可能とした。しかしながら、Grabner らの手法では、各選手が異なる動きを見せるサッカーの選手追跡において適用することは困難である。

そこで、本研究では、複数選手の位置情報を時間状況グラフとして表現しておき、これにガイドされる形でパーティクルフィルタを実行することにより、オクルージョンが起こっても選手の誤検出を大幅に減らすことが可能な手法を提案する。

本論文は、全 6 章から構成されている。第 2 章では、提案手法の流れについて述べる。第 3 章では、時間状況グラフの詳細を述べる。第 4 章では、提案手法である時間状況グラフを用いたパーティクルフィルタによる選手追跡の詳細を述べる。第 5 章では、実映像を用いた実験を行い、その結果を示し、考察を行う。最後に第 6 章では、本研究の成果と今後の課題についてまとめる。

2. 提案手法の流れ

提案手法の流れを図 1 に示す。提案手法はグラフ構築部と選手追跡部に分けることができる。グラフ構築部では、まず入力フレームから背景差分法により選手領域のみを抽出する。その後、抽出した選手領域情報をノードに格納して時間状況グラフを構築する。時間状況グラフ

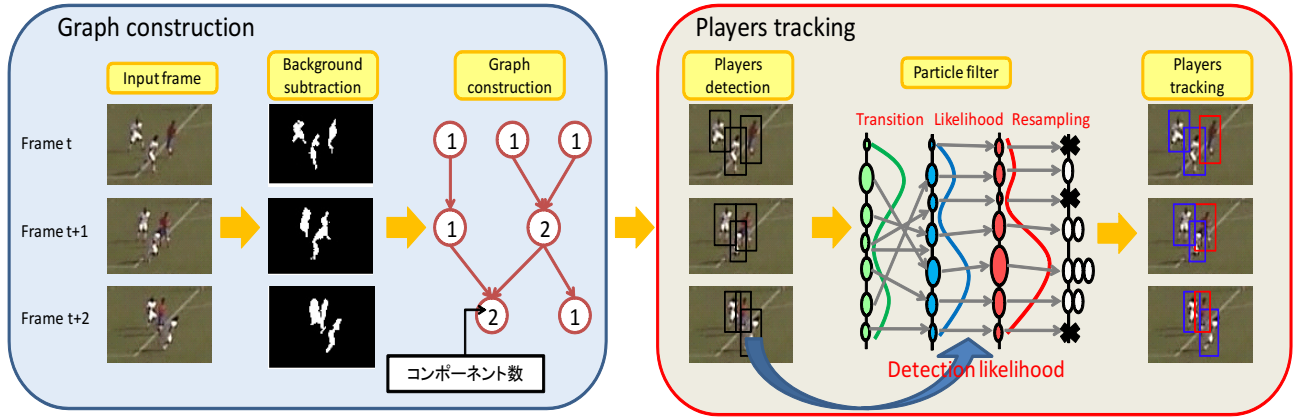


図 1 提案手法の流れ

とは、フレーム毎に背景差分により取得した選手領域内に存在する選手数（以後、コンポーネント数）を遷移させていくグラフである。選手追跡部では、まず時間状況グラフのノード情報により選手検出を行う。そして、時間状況グラフによる検出結果を尤度評価に組み込んだパーティクルフィルタによって、選手の追跡を行う。パーティクルフィルタは、状態量を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を使用している。

3. 時間状況グラフの構築

パーティクルフィルタには、映像のフレーム間で複数の選手の位置情報を用いていないため、一度見失うと再度発見することが困難であるという欠点がある。この欠点を補完するために本研究では、[1] の手法を用いて、複数選手の位置情報を時間状況グラフとして表現しておき、これにガイドされる形でパーティクルフィルタを実行する。こうすることにより、オクルージョンが起こっている選手に対しても、ロバストに追跡が行なえるようになる。

3.1 時間状況グラフ

時間状況グラフは、図 1 左のグラフ構築部のように、映像において毎フレーム、背景差分法により抽出した選手領域の集合を基に構築される。背景差分法は、[10] の手法を用いている。この手法は、背景画像の時間的変化を考慮し、動的に背景を更新するため、照明変化に頑健な手法となっている。時間状況グラフの各ノードは、1 つの選手領域の空間情報を格納し、エッジは選手領域間の距離に基づいて、コンポーネント数を変化させていく。

ノード情報とエッジ情報に関してまとめたものを、表 1 に示す。ノード情報としては、各ノードを識別するためのラベル、選手領域面積、選手領域サイズ（幅、高さ）、選手領域の中心位置座標、コンポーネント数が含まれており、エッジ情報としては、各エッジを識別するためのラベル、フレーム間での選手領域間距離が含まれている。[1] の手法ではグラフで追跡を行なっているため、

エッジ情報に色情報を含んでいるが、本研究で用いた時間状況グラフは、コンポーネント数に基づいて選手を検出するためのものであり、コンポーネント数のみを遷移させればよい。この点から、エッジ情報に色情報は含めていない。

表 1 時間状況グラフの情報

ノード情報	エッジ情報
ラベル	ラベル
選手領域面積	選手領域間距離
選手領域サイズ（幅、高さ）	
選手領域の中心位置座標	
コンポーネント数	

3.2 時間状況グラフの構築

G を有向グラフである時間状況グラフとする。 $n_i(t)$ はフレーム t でのノードであり、 i はフレーム t でのノードのラベル（識別番号）である。 $d_{i,j}$ は、ノード $n_i(t)$ と $n_j(t+1)$ の間のノード間距離であり、 $d_{max(i,j)}$ は、選手領域が 1 フレームで動く推定最大距離で、選手領域のワールド座標から見積もられる。ノード間距離は、フレーム t における選手領域の中心座標と、フレーム $t+1$ における選手領域の中心座標との間のユークリッド距離で求めることができる。また、図 2 では、ノード $n_1(t)$ に関するノード間距離を求めている。フレーム $t+1$ における黄色の点は、フレーム t におけるノード $n_1(t)$ の位置であり、そこからフレーム $t+1$ における各ノードの中心点（青色の点）に延びている赤色の線分がノード間距離となっている。 $e_{i,j}$ は、ノード $n_i(t)$ と $n_j(t+1)$ の間のエッジである。

時間状況グラフを構築するアルゴリズムは、以下のステップのようになる。

1. 初期フレーム ($t = 1$) で、各選手領域 i に対するノード $n_i(t)$ を生成し、グラフ G へ挿入する。
2. フレーム $t+1$ で、各選手領域 j に対するノード $n_j(t+1)$ を生成し、グラフ G へ挿入する。

3. 各ノード $n_i(t)$ と $n_j(t+1)$ の中心座標間のユークリッド距離を計算する.
4. 条件 $d_{i,j} < d_{max(i,j)}$ を満足するエッジ $e_{i,j}$ を生成する.
5. ノードのグループ化を行い, 選手領域面積を基にコンポーネント数を決定する. (詳細は 3.3 で述べる.)
6. 最終フレームになるまでステップ 2~5 を繰り返す.

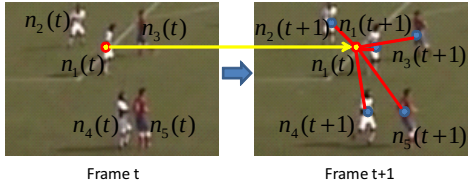


図 2 ノード間距離

3.3 コンポーネント数の決定

コンポーネント数は, 1つの選手領域内に存在する選手数を表す. このコンポーネント数を正確に決定することは, 選手検出を行う上で重要である. コンポーネント数は, まず時間状況グラフにおいてエッジを用いたノードのグループ化を行い, その後, ノード情報の1つである選手領域面積を用いてコンポーネント数を決定する.

3.3.1 ノードのグループ化

図 3 に示されているように, ノード間のエッジによって, ノードをグループ化する方法について述べる. フレーム $t+1$ で, 時間状況グラフ中に, エッジ e_{v_3, w_4} と e_{v_4, w_4} が存在するようなノード $n_{w_4}(t+1)$ がある場合, フレーム t で2つのノード $n_{v_3}(t)$ と $n_{v_4}(t)$ は同じグループに属する. つまり, フレーム t , および $t+1$ において, エッジでつながっているノードは全て同じグループにまとめる. このグループ決定方法は, 時間状況グラフのフレーム t , および $t+1$ のノードとエッジによって定義されるグラフにおいて, 深さ優先探索を行うことと同じである.

グラフ上の新しい探索ごとに, 新しいグループナンバーが定義される. この探索中にたどった各ノードは, 定義されたグループナンバーを受け取る. 従って, 図 3 の例では, ノード $n_{v_1}(t)$ から始まる探索は, Group1 と定義され, ノード $n_{v_3}(t)$ から始まる探索は, Group2 と定義される.

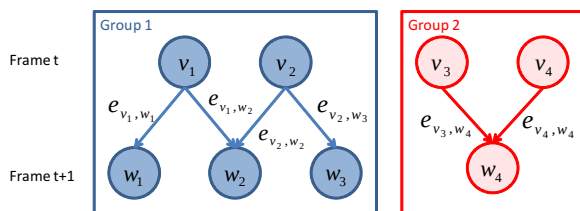


図 3 ノードのグループ化

3.3.2 コンポーネント数の決定

コンポーネント数の決定は, コンポーネント数に基づいて選手検出を行う本研究では重要である. 図 4 のように 3 人の選手が密集してプレーしている状況を例に, コンポーネント数を決定する流れを図 5 に示し, これらのステップを要約したアルゴリズムを以下に述べる. なお, サッカーでは, キックオフの時にオクルージョンが発生している選手はいないと考えられるので, 初期フレームにおけるコンポーネント数は全て 1 としている.

アルゴリズム

[ステップ 1] フレーム t で, Group X に属する各ノード $n_v(t)$ のコンポーネント数 num_v を総和することで, Group X に属するコンポーネント数 num_X を以下のよう

$$num_X = \sum_{v \in X} (num_v)$$

図 5 では, 初期状態 (a) において, フレーム t で Group X に属するノードは $n_v(t)$ のみなので, Group X に属するコンポーネント数 num_X は 3 と決定される.

[ステップ 2] フレーム $t+1$ で, Group X に属する各ノード $n_{w_i}(t+1)$ において, 選手 1 名分の大きさである選手領域面積 A_p を推定する. これは, フィールド上の選手のワールド座標を基に推定できる. この推定選手領域面積 A_p を用いて, ノード $n_{w_i}(t+1)$ の選手領域面積 A_w を次のように更新する.

$$A_w \leftarrow A_w - A_p$$

図 5 では, 初期状態 (a) において, ノード $n_{w_1}(t+1)$, $n_{w_2}(t+1)$ の選手領域面積 A_{w_1} , A_{w_2} は, それぞれ 242, 139 であったが, 状態 (b) においては, 選手領域の位置より推定された選手 1 名分の面積 A_p の 130 を引いて, それぞれ 112, 9 となっている.

[ステップ 3] フレーム t において, Group X に属するコンポーネント数 num_X を 1 減らし, ノード $n_w(t+1)$ のコンポーネント数を 1 に初期化する. この作業を Group X に属するフレーム $t+1$ の全ノードに対して繰り返す.

$$num_X \leftarrow num_X - 1$$

$$num_w = 1$$

図 5 では, 状態 (c) において, ノード $n_{w_1}(t+1)$, $n_{w_2}(t+1)$ のコンポーネント数 num_{w_1} , num_{w_2} が, それぞれ 0 から 1 となっている. そのため, Group X に属するコンポーネント数 num_X は 1 となる.

[ステップ4] フレーム t で、Group X に属するコンポーネント数 num_X が 0 になるまで、フレーム $t+1$ において、最も大きい選手領域面積 A_w をもつノード $n_w(t+1)$ を見つけ、以下の作業を繰り返す。

$$A_w \leftarrow A_w - A_p$$

$$num_X \leftarrow num_X - 1$$

$$num_w \leftarrow num_w + 1$$

図 5 では、状態 (c) において、ノード $n_{w_1}(t+1)$, $n_{w_2}(t+1)$ の選手領域面積 A_{w_1} , A_{w_2} は、それぞれ 112, 9 であり、ノード $n_{w_1}(t+1)$ の選手領域面積 A_{w_1} が最も大きいので、最終状態 (d) では、 A_{w_1} から選手領域位置より推定された面積 A_p の 130 を引いて、-18 となっている。また、ノード $n_{w_1}(t+1)$ のコンポーネント数 num_{w_1} は、1 増えて 2 となっている。そのため、Group X に属するコンポーネント数 num_X は 0 となり、Group X は、最終状態 (d) のコンポーネント数で決定する。

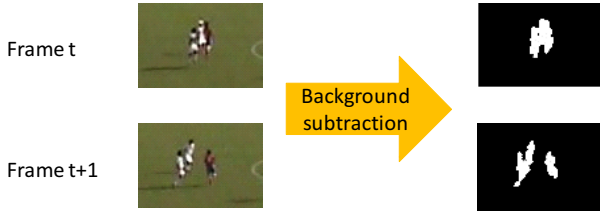


図 4 入力フレーム例

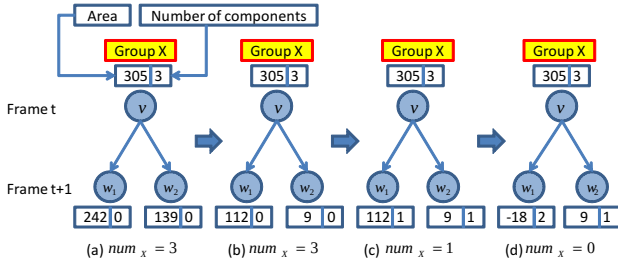


図 5 コンポーネント数決定の流れ

4. 時間状況グラフを用いたパーティクルフィルタによる選手追跡

4.1 ノード情報に基づく選手検出

本研究では、選手検出と追跡を組み合わせた“tracking-by-detection”手法を用いているので、選手検出の精度が追跡精度に及ぼす影響は大きい。前章で述べたように、時間状況グラフの各ノードの主要な情報はコンポーネント数、つまり、背景差分法によって抽出された選手領域内に何人の選手が存在しているのかという情報である。このコンポーネント数により、1 選手領域内で検出する選手数が把握できるので、複雑なオクルージョンが起こっている選手領域でも頑健な選手検出が可能となる。また、選手検出手法は、時間状況グラフのノードに含まれるコンポーネント数によって異なる。

4.1.1 コンポーネント数が 1 名もしくは 2 名の場合

ノードのコンポーネント数が 1 名の場合は、ノードの選手領域のサイズから容易に検出できる。ノードのコンポーネント数が 2 名の場合は、図 6 (b) のように、ノードの選手領域情報を用いて、選手を検出する。すなわち、選手領域の y 座標が最小となる座標 $\{Min = (x_{y_{min}}, y_{min})\}$ と、選手領域の y 座標が最大となる座標 $\{Max = (x_{y_{max}}, y_{max})\}$ 、ノードに含まれる選手領域位置において、ワールド座標により推定された選手のウィンドウサイズを用いて、図 6 (c) のように 2 名の選手を検出する。但し、選手の各水平位置は、 Min , Max それぞれの x 座標から決定している。図 6 (c) における選手枠の上の数字は、検出する際に用いたノードのコンポーネント数である。

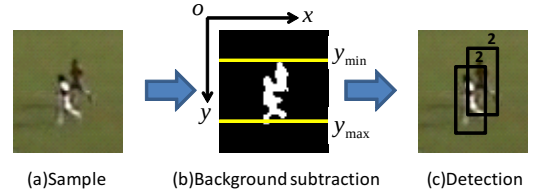


図 6 コンポーネント数 2 の選手検出

4.1.2 コンポーネント数が 3 名以上の場合

ノードのコンポーネント数が 3 名以上の場合は、2 名の場合と同様に、図 7 (b) のように、ノードの選手領域情報を用いて選手を検出する。すなわち、選手領域の y 座標が最小となる座標 $\{Min = (x_{y_{min}}, y_{min})\}$ と、選手領域の y 座標が最大となる座標 $\{Max = (x_{y_{max}}, y_{max})\}$ 、ノードに含まれる選手領域の中心において、ワールド座標により推定された選手のウィンドウサイズを用いて、コンポーネント数が 1 名もしくは 2 名の場合と同等に、上下 2 名の選手を検出する。残りの選手に関しては、様々なアプリケーションで広く使われ、認識性能が高いことで知られる SVM [12] を用いて、選手領域の範囲を絞ってスライディングウィンドウにより検出する。

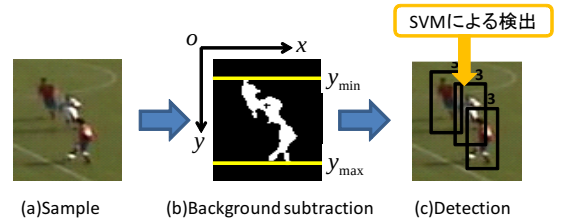


図 7 コンポーネント数 3 の選手検出

本研究では、図 8 のようなポジティブサンプル数 3000 枚、図 9 のようなネガティブサンプル数 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 SVM を学習した。検出は、ノードの選手領域をある程度縮小した範囲で、フィールド上のワールド座標から適切なウィンドウサイズを推定し、検出尤度と検出位置に基づ

いて行う。同一選手を重複して検出することを避けるため、各検出位置の距離において閾値処理を行っている。

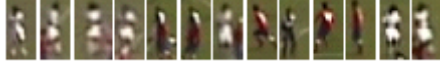


図 8 ポジティブサンプルの例

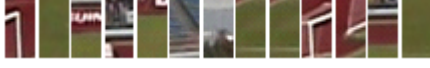


図 9 ネガティブサンプルの例

図 10 は、実際の映像フレームで選手検出を行ったものである。時間状況グラフのコンポーネント数より、オクルージョンが発生している選手でも、正確に検出できていることがわかる。

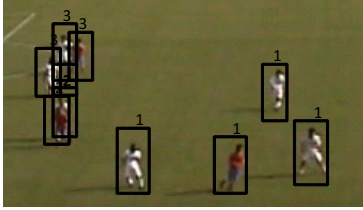


図 10 選手検出結果の例

4.2 パーティクルフィルタの状態空間と状態遷移モデル

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 [11] を用いた。時刻 t における状態量 $\vec{x}_p$ は式 (1) で定義される。

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

$\vec{x}_p$ はワールド座標 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。状態量を画像座標ではなく、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 は選手の重心の座標位置を表している。また、時刻 $t - \beta$ から t に至る状態遷移モデルは、式 (2) のように等加速度運動を仮定している。

$$\vec{x}_p(t) = C\vec{x}_p(t - \beta) + \Upsilon \quad (2)$$

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

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

$I_{2 \times 2}$ は 2×2 の単位行列であり、 $J_{1 \times 2}$ は全ての要素が 1 である 1×2 の行列を表す。 Υ はノイズを表しており、変数 ϵ_p (位置ノイズ)、 ϵ_v (速度ノイズ)、 ϵ_a (加速度ノイズ)

はいずれも平均 0 の正規分布に従う。位置ノイズ ϵ_p の分散は追跡対象の大きさによって変化し、速度ノイズ ϵ_v の分散と加速度ノイズ ϵ_a の分散は、連続して追跡に成功した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だけ、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。

4.3 時間状況グラフを考慮した尤度評価

4.3.1 トラックと検出器の結果統合

検出器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、どの検出器による結果がどのトラック（追跡器）に対応するのかを決定しなければならない。検出器とは、4.1 で述べたように、時間状況グラフのノード情報と SVM を用いたものである。また、トラックとは、選手 1 人 1 人に割り当てられたパーティクルフィルタであり、1 トラック当たり 200 個の粒子で構成されている。本研究では [6] で用いられている手法を用いている。

このアルゴリズムは、図 11 のような流れとなっている。まず、各選手の検出器による検出結果 $det^k (k = 0, \dots, K)$ と、各選手のトラックによる追跡結果 $tr^r (r = 0, \dots, K)$ のそれぞれの組 (tr^r, det^k) について、マッチングスコア $s(tr^r, det^k)$ 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、検出結果と追跡結果がより似ていることを表している。ただし、 K は、選手数を表しており、 k と r は、それぞれ検出器とトラックの識別番号を表している。

$$s(tr^r, det^k) = g(tr^r, det^k) \cdot \left(\frac{\alpha_p}{N_p} \sum_{\vec{p} \in tr^r} (p_N(det^k - \vec{p}) + p_C^{\vec{p}}) \right) \quad (3)$$

$$g(tr^r, det^k) = p(size_{det^k} | tr^r) = \frac{1}{N_p} \sum_{\vec{p} \in tr^r} p_N\left(\frac{size_{\vec{p}} - size_{det^k}}{size_{\vec{p}}}\right) \quad (4)$$

$$p_C^{\vec{p}} = \frac{\Sigma_{i,j}\{I(i,j) - \bar{I}\} \cdot \{T(i,j) - \bar{T}\}}{\sqrt{\Sigma_{i,j}\{I(i,j) - \bar{I}\}^2 \cdot \Sigma_{i,j}\{T(i,j) - \bar{T}\}^2}} \quad (5)$$

$p_N(det^k - \vec{p}) \sim N(det^k - \vec{p}; \vec{0}, \sigma_{det^k}^2 I_{2 \times 2})$ は、トラック tr^r の粒子の位置 $\vec{p} = [p_x, p_y]$ と検出結果 det^k の位置間距離を、正規分布により確率に変換したものである。 $g(tr^r, det^k)$ はゲート関数を表しており、 $size_{\vec{p}}$ は粒子 $\vec{p}$ の y 座標 p_y によって決定されるトラックのウィンドウサイズ、 $size_{det^k}$ は検出結果 det^k のウィンドウサイズ、 N_p はトラックの粒子数、 $I_{2 \times 2}$ は 2×2 の単位行列である。 $p_N(\frac{size_{\vec{p}} - size_{det^k}}{size_{\vec{p}}})$ は、トラック tr^r の位置 $\vec{p} = [p_x, p_y]$ における粒子のサイズと検出結果 det^k のサイズの差を、正規分布により確率に変換したものである。 α_p と $\sigma_{det^k}^2$ は実験的に求めた。

また、 $p_C^{\vec{p}}$ は、式 (5) で定義される正規化相互相関値

である．式 (5) において，領域内の座標 (i, j) における輝度値を $I(i, j)$ と表し，同様にテンプレート画像内の座標 (i, j) における輝度値を $T(i, j)$ と表す．ただし， $\bar{I}$ と $\bar{T}$ はそれぞれ領域画像とテンプレート画像の平均輝度値とする．以上より，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 $s(tr^r, det^k)$ の行列 S が生成できる．(図 11(a)) そして，Greedy アルゴリズムを用いて， S の中で最大のスコアの組 (tr^*, det^*) を見つけ (図 11(b))， 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．(図 11(c)) これを繰り返してトラックによる追跡結果と検出器による検出結果を統合する．(図 11(d))

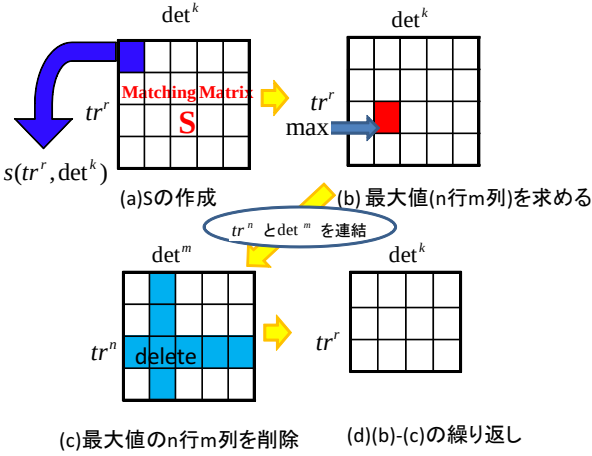


図 11 トラックと検出器の結果統合の方法

4.3.2 尤度計算

トラック tr^* の粒子 $\vec{p}$ の尤度 $\omega_{tr, \vec{p}}$ 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，時間状況グラフのノードに含まれるコンポーネント数の変化から計算されるオクルージョン情報を用いて，式 (6) のように重み付けした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を求めている．用いた複数の尤度は，検出結果に基づく尤度 $p_D^{\vec{p}}$ とヒストグラム間距離に基づく尤度 $p_H^{\vec{p}}$ ，さらに相関値に基づく尤度 $p_C^{\vec{p}}$ である．

$$\omega_{tr, \vec{p}} = p(y_t | \vec{x}_t) = p_D^{\vec{p}} + \beta^{\vec{p}} \cdot p_H^{\vec{p}} + (1 - \beta^{\vec{p}}) \cdot p_C^{\vec{p}} \quad (6)$$

$$\beta^{\vec{p}} = p_N(\min_{i \in P} dist_i) \quad (7)$$

検出結果に基づく尤度 $p_D^{\vec{p}}$ は，次のようにして求められる．

$$p_D^{\vec{p}} = \alpha \cdot p_N((\vec{p} - det^*)^2) \quad (8)$$

$$\alpha = \begin{cases} 1 & num_{comp} = 1 \\ \frac{A_{estimate}}{A_{real}} & num_{comp} > 1 \end{cases}$$

$p_N((\vec{p} - det^*)^2)$ は，粒子の位置 $\vec{p}$ と検出結果の位置 det^* の位置間距離の二乗を，正規分布により確率に変換

したものである．平均は 0，分散は実験的に求めた値である．オクルージョンが発生している場合は，尤度が低くなる選手がいると考えられるので，検出結果の尤度を高める必要がある．そこで，本研究では，オクルージョンの度合いを表すパラメータ α を用いて，オクルージョン時の尤度低下を抑えている．

α の計算で用いる $A_{estimate}$ は，検出に用いたノードのワールド座標におけるオクルージョンを考えない場合の推定選手領域面積， A_{real} は，検出に用いたノードの実際の選手領域面積， num_{comp} は，検出に用いたノードのコンポーネント数である． $I(tr^*)$ は，ノードのコンポーネント数が 1 ならば，オクルージョンが発生していないので， α の値を 1 とする関数である．オクルージョンが発生し，その度合いが大きくなればなるほど，実際の面積が小さくなるので，パラメータ α の値は大きくなる．つまり，パラメータ α の値を大きくすることで，検出結果の尤度 $p_D^{\vec{p}}$ を高くし，オクルージョン発生時の全体の尤度低下を抑えている．

式 (6) の 2 番目の項 $p_H^{\vec{p}}$ と 3 番目の項 $p_C^{\vec{p}}$ 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，それ以外の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きているときは前者の重みを大きくし，逆に選手間の距離が離れていて，オクルージョンが起きていないときは後者の重みを大きくする．こうすることで，長時間にわたる頑健な追跡が可能となる．式 (7) の $\beta^{\vec{p}}$ は $p_H^{\vec{p}}$ と $p_C^{\vec{p}}$ の重みを表しており，最も近い選手との距離を正規分布 p_N により確率に変換したものである．ただし， $dist_i$ は追跡対象の選手とその他の選手 i との距離， P は追跡対象の選手を除く全選手数である．

本研究ではヒストグラム間距離として Earth Mover's Distance [13], [14] を，相関値として正規化相互相関値を用いた．

$$p_H^{\vec{p}} = p_N(EMD(\vec{p}, \vec{T})) \quad (9)$$

式 (9) における $EMD(\vec{p}, \vec{T})$ は粒子の位置 $\vec{p}$ における入力画像とテンプレート画像 $\vec{T}$ との Earth Mover's Distance であり，その距離を正規分布により確率 $p_H^{\vec{p}}$ に変換している．正規化相互相関値 $p_C^{\vec{p}}$ は式 (5) で定義される．Earth Mover's Distance (EMD) は近年注目されている尺度であり，他の尺度より人間の感覚に近いことが知られている．

5. 評価実験

実際の視点固定単眼サッカー映像に対して，追跡実験を行った．時間状況グラフを用いないパーティクルフィルタによる手法と，時間状況グラフを組み合わせたパーティクルフィルタによる提案手法を比較する．

5.1 実験環境

視点固定映像は、1280 × 720 画素、フレームレート 30fps のサッカー映像を用いた、内容は第 83 回全国高校サッカー選手権大会京都府大会決勝、桂高等学校（以下白チーム）対京都朝鮮高級学校（以下赤チーム）の試合を撮影したものである。なお、追跡範囲は、図 12 のように、フィールドの左半分のみとなっており、選手が右フィールドに出てしまうと追跡できないため、選手が左フィールドでプレーしている平均 350 フレームの映像を 10 本抽出して、実験を行なった。



図 12 実験フレーム

5.2 実験結果と考察

追跡精度評価は、式 (10) で定義される選手の平均追跡精度 A_p で行った。

$$A_p = \frac{1}{SP} \sum_{i=1}^S \sum_{j=1}^P \frac{W^{i,j}}{N^{i,j}} \quad (10)$$

$N^{i,j}$ はサンプル映像 i において、選手 j が映っているフレーム数、 $W^{i,j}$ はサンプル映像 i において、選手 j の追跡成功フレーム数を表している。あるフレームにおいて、手動で与えた 2 次元画像座標正解データと、ワールド座標空間で追跡した結果を 2 次元画像座標に変換した座標データとの距離が、10 画素以内であれば追跡成功とした。ただし、 S はサンプル映像数、 P は全選手数である。図 13 に、サンプル別の平均追跡精度の結果を示す。サンプル平均の追跡精度は、表 2 より従来手法の 72.15 % から 79.50 % へ 7.35 ポイント向上した。詳細に述べると、本研究で用いた時間状況グラフを選手検出に用いることで、従来手法の 72.15 % から 78.31 % へ 6.16 ポイント向上した。また、パーティクルフィルタの尤度評価において、重み α を用いなかった場合と用いた場合で比較すると、重み α を用いなかった場合の 78.31 % から 79.50 % へ 1.19 ポイント向上した。

サンプル映像毎に考察すると、sample07 を除く全てのサンプル映像で提案手法である時間状況グラフを組み合わせたパーティクルフィルタによる選手追跡が、従来手法を上回った。これは、オクルージョンが発生している時でも、時間状況グラフによって背景差分から得た選手領域のコンポーネント数が分かり、選手検出精度が向

上したことで、オクルージョンによって生じるパーティクルフィルタの尤度低下を抑えたことが要因に挙げられる。特に sample06 では、約 20 % の精度向上が見られた。sample06 では、オクルージョンの発生回数が多く、複雑なオクルージョンの発生も多かったために、時間状況グラフの効果が高かったと考えられる。また、sample07 では、オクルージョンの発生回数が少なかったことで、同じチームの選手同士のオクルージョン発生後、互いに異なる選手を追跡してしまったために、追跡精度が下がったと考えられる。しかし、同じチームの選手同士の追跡器が逆転しても、その後に尤度低下が起こらないため、一旦逆転すると正しい追跡器に戻すことは困難な問題である。

表 2 平均追跡精度

Method	$A_p(\%)$
Conventional method	72.15
Proposed method without occlusion likelihood α	78.31
Proposed method with occlusion likelihood α	79.50

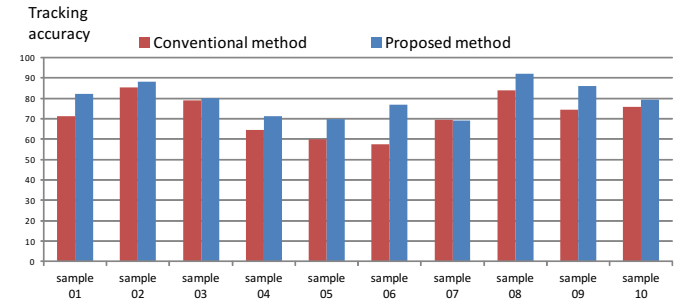


図 13 サンプル動画毎の従来手法と提案手法の追跡精度の比較

図 14 は、実験結果の一部を広いアングルで比較的長時間追跡したものを表している。オクルージョンが発生していた選手も、発生していなかった選手もほぼ正確に追跡できていることがわかる。

6. 結 論

本研究では、サッカー映像においてオクルージョンにロバストな選手追跡を行うために、時間状況グラフにガイドされたパーティクルフィルタによる新しい選手追跡手法を提案した。従来のパーティクルフィルタによる追跡では、映像のフレーム間で複数選手の位置情報を用いていないため、一度見失うと再度発見することが困難であるという問題があった。そこで本研究では、複数選手の位置情報を時間状況グラフとして構築した。この結果、評価実験においては、従来手法と比較してオクルージョンに対してロバストとなり、7.35 ポイントの精度向上が見られた。

追跡に失敗する例としては、同一チームの選手同士がすれ違うことなどにより追跡器が逆転することや、1 人

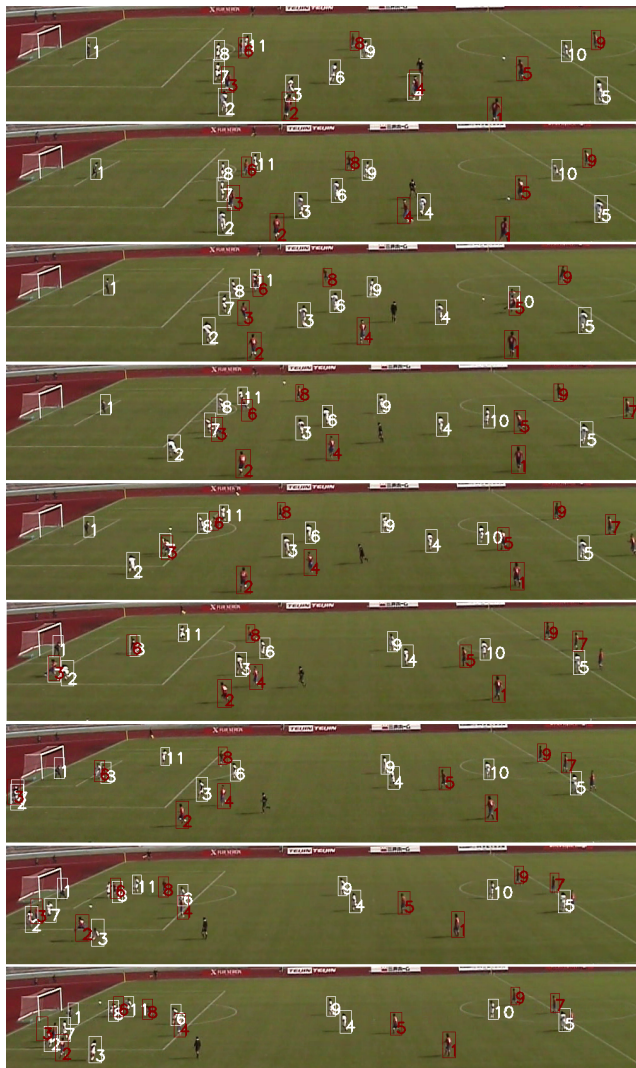


図 14 追跡結果

の選手を重複して追跡してしまうことが挙げられる。前者に関しては同一チームのためユニフォームの色が同じであることにより、追跡が逆転してもパーティクルフィルタの尤度は低下しないことが要因と考えられる。後者に関しては、時間状況グラフのコンポーネント数の誤りが要因と考えられる。

今後の課題として、ゴール前などの選手密集地域における追跡精度の向上が挙げられる。そのためには、まず背景差分の精度向上が必要となる。背景差分によって正確に選手領域を抽出することは、時間状況グラフのコンポーネント数決定の精度向上につながり、選手の検出精度にも関係してくるため重要である。また、コンポーネント数の訂正アルゴリズムを導入する必要があると考えられる。コンポーネント数の誤りは、検出精度と直接関係していて、追跡精度低下の要因となるためである。また、パーティクルフィルタの尤度評価においては、現在はヒューリスティックに求めているパラメータがあるため、これらを動的に最適化すれば精度が向上すると考えられる。また、状態空間を、様々な特徴量を主成分分析 (PCA) して得られる部分空間とすることで、より追跡対

象らしさを表すことができるため、追跡精度の向上が期待できる。

文 献

- [1] Pascual J. Figueroa a, Neucimar J. Leite , Ricardo M.L. Barros, "Tracking soccer players aiming their kinematical motion analysis", Computer Vision and Image Understanding (CVIU), pp. 122-135, 2005.
- [2] M. Zheng and D. Kudenko, "Automated Event Recognition for Football Commentary Generation", International Journal of Gaming and Computer-Mediated Simulations (IJGCMs), Vol. 2, pp. 67-84, 2010.
- [3] V. Tovinkere and R. J. Qian, "Detecting Semantic Events in Soccer Games: Towards A Complete Solution", IEEE International Conference on Multimedia & Expo (ICME), pp. 1040-1043, 2001.
- [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] 片山 徹, "応用カルマンフィルタ", 朝倉書店, 2000.
- [6] 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, 2009-9.
- [7] 澤田裕介, 尺長健, "仮説検証に基づく自然環境下での複数歩行者追跡", 画像の認識・理解シンポジウム (MIRU), pp. 718-725, 2011-7.
- [8] Takuro Nishino, Yasuo Ariki, Tetsuya Takiguchi, "Tracking of Multiple Soccer Players Using a 3D Particle Filter Based on Detector Confidence", Advances in Computer Science and Engineering (ACSE), pp. 93-104, 2011.
- [9] H. Grabner, J. Matas, L. Van Gool, P. Cattin, "Tracking the Invisible: Learning Where the Object Might be", Computer Vision and Pattern Recognition (CVPR), pp. 1285-1292, 2010.
- [10] 森田 真司, 山澤 一誠, 寺沢 征彦, 横矢 直和, "全方位画像センサを用いたネットワーク対応型遠隔監視システム", 電子情報通信学会論文誌 (D-II), Vol. J88-D-II, No. 5, pp. 864-875, 2005-05.
- [11] 樋口知之, "粒子フィルタ", 電子情報通信学会誌, Vol. 88, No. 12, pp. 989-994, 2005-12.
- [12] M.J. Vapnik, "The Nature of Statistical Learning Theory", Springer, Heidelberg, 2001.
- [13] 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, No. 2, pp. 99-121, Nov. 2000.
- [14] X. Wan and Y. Peng, "The Earth Mover's Distance as a Semantic Measure for Document Similarity", Proc. of the 14th ACM International Conference on Information and Knowledge Management, pp. 301-302, 2005.