

移動ロボット制御のための 人物シルエットの重なりを考慮した複数人物追跡

佐竹 純二^{†1} 三浦 純^{†1}

本研究では移動ロボットにカメラを取り付け、撮影した動画画像から複数の歩行者を検出・追跡することを目的とする。移動ロボットからの人物検出にはレーザーレンジセンサやステレオカメラが広く用いられているが、複数物体のオクルージョンはあまり取り扱われていなかった。そこで、本研究ではステレオカメラを用い、複数人物の重なりを考慮したシルエット形状を利用することで、それぞれの人物を正確かつ安定に追跡する方法を提案する。そして、実際に移動ロボットによる特定人物の追従実験を行い、提案手法の有効性を示す。また、周囲環境に設置されたカメラなど、複数台のカメラを併用した場合の拡張方法についても検討する。

Multi-Person Tracking for a Mobile Robot using Overlapping Silhouette Templates

JUNJI SATAKE^{†1} and JUN MIURA^{†1}

This paper describes a stereo-based person tracking method for a person following robot. Many previous works on person tracking use laser range finders which can provide very accurate range measurements. Stereo-based systems have also been popular, but most of them are not used for controlling a real robot. In this paper, we propose an accurate, stable tracking method using overlapping silhouette templates which consider how persons overlap in the image. Experimental results show the effectiveness of the proposed method. In addition, we examine the possibility of extending the method to multi-camera cases.

^{†1} 豊橋技術科学大学
Toyohashi University of Technology

1. はじめに

人物追従は移動ロボットにおける重要なタスクの一つである。パブリックスペースでの人物追従を実現するためには、移動カメラで撮影された動画画像から複数の人物を検出・追跡する必要がある。動画画像を用いた人物検出・追跡の研究^{1)–4)}はこれまでも数多く行われているが、その多くは固定カメラを使用しており、カメラが移動する場合には背景と移動物体の分離が重要な問題となる。

本研究では、移動ロボットの人物追従制御を行うための複数人物の検出・追跡方法を述べる。移動ロボットにおける障害物の検出にはレーザーレンジセンサが広く用いられており、人物の検出・追跡に利用した例もある^{5),6)}。しかし、複数人物の分離や、見え情報（色・テクスチャなど）を用いた識別、人物の向きや状態の認識などを行うためには画像情報が必要である。また、全方位カメラを用いた研究^{7),8)}も提案されており、交差を取り扱った実験結果も示されている。しかし、全方位カメラは解像度が低いため、重なった人物の分離や人物と人物以外の物体の識別など、複雑な状況の認識は難しいと考えられる。ステレオカメラもまた移動物体の検出・追跡によく用いられている^{9)–11)}。しかし、これらの研究では距離情報を探索領域の限定に用いるだけで、複数人物のオクルージョンは取り扱われていない。

Essら^{12),13)}は様々な手がかり（見えベースの物体検出、距離の推定、ビジュアルオドメトリ、地表面の検出など）を統合して歩行者を追跡する方法を提案している。この方法は複雑な状況に対しても高い追跡性能を示しているが、実際にロボットの追従制御を行うにはまだ計算コストが高いと考えられる。

パーティクルフィルタを用いて複数物体を追跡する手法も提案されている^{14)–16)}。これらの手法では、probabilistic exclusion principleを加えることで複数物体の追跡を実現している。Khanら¹⁵⁾は交錯する複数の昆虫の追跡する手法を提案しているが、2次元平面上を動き回る対象を真上から撮影しているため、複雑なオクルージョンはあまり起こっていない。TweedとCalway¹⁶⁾は重なりあった対象の間に連結情報を追加することで、飛行する多数の鳥の追跡を実現している。ただし、翼の部分的な重なりはしばしば生じているが、鳥のほぼ全身が常に見えている。本研究では、移動ロボットに取り付けたカメラで撮影した動画画像を用いるため、歩行者が他の歩行者に完全に隠されてしまうような複雑なオクルージョンが発生する。そこで、本研究ではステレオカメラを用い、距離情報を利用して複数人物の分離や追跡を行う。特に本稿では、人物シルエットの重なりを考慮したテンプレートを用い、複数人物を正確かつ安定に追跡する方法を提案する。

2. ステレオカメラを用いた複数人物追跡

2.1 距離情報を利用した人物検出

移動カメラで撮影した画像を用いて安定な人物追跡を実現するため、本研究では距離画像テンプレート¹⁷⁾ (図1)を用いる。このテンプレートはステレオカメラから得られた距離画像をもとに背景領域と人物領域を2値とするシルエット画像であり、人物がカメラからの距離 2[m] の時の距離画像から作成している。

画像探索時にはカメラから対象までの距離に応じてテンプレートの大きさを変更することで、画像上でのスケールの変化に対応する。また、テンプレート中の人物領域の画素値はカメラから人物までの距離とする。人物追跡処理の詳細については2.2節で述べる。背景領域の画素値については、カメラが移動するため、各フレームの入力画像から背景の距離を推定する(距離のヒストグラムを求め、累積90%となる距離を背景の距離とした)。

サイズ $H \times W$ の距離画像テンプレートを $T(x, y)$ 、入力された距離画像を $I_D(x, y)$ とした時、テンプレートマッチングの相違度 d は次のように求められる。

$$d = \frac{1}{HW} \sqrt{\sum_p \sum_q [T(p, q) - I_D(x+p, y+q)]^2} \quad (1)$$

なお、体の向きの変化に対応するため、図1のような4方向3種類(正面と後ろ向きは共通)のテンプレートを併用し、最も相違度が低いものをマッチングの結果として使用する。

距離画像テンプレートを用いた人物検出の例を図2に示す。3種類のテンプレートの探索結果の内、最も評価値の良いもの(相違度の低いもの)を太線で表示しており、複数のテンプレートを用いることで体の向きの変化に対応できていることが確認できる。

2.2 パーティクルフィルタを用いた3次元位置推定

本研究で使用するロボットやカメラと、各座標系の関係を図3に示す。時刻 t における人物の3次元位置を (X_t, Y_t, Z_t) とし、状態変数 \mathbf{x}_t を次のように定義する。

$$\mathbf{x}_t = \begin{bmatrix} X_t & Y_t & Z_t & \dot{X}_t & \dot{Y}_t \end{bmatrix}^T \quad (2)$$

ただし、 \dot{X}_t, \dot{Y}_t は X - Y 平面上の移動速度である。この時、状態方程式は次のようになる。

$$\mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{w}_t \quad (3)$$

各人物の3次元位置をパーティクルフィルタを用いて推定する。式(1)で定義した相違度 d を用い、各パーティクルの尤度 L を次のように計算する。

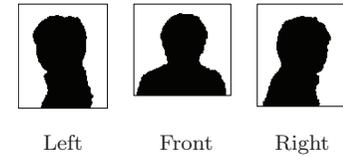
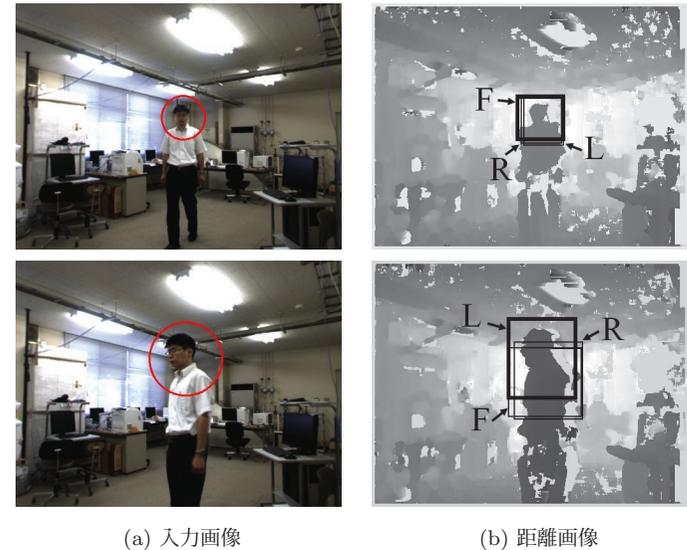


図1 距離画像テンプレート
Fig. 1 Depth templates



(a) 入力画像 (b) 距離画像
図2 距離画像テンプレートを用いた人物検出の例
Fig. 2 Detection examples using depth templates

$$L = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{d^2}{2\sigma^2}\right) \quad (4)$$

人物位置は各パーティクルの尤度を重みとした重み付き平均で決定する。

2.3 シルエットの重なりを考慮した複数人物追跡

接近した複数人物を正確かつ安定に追跡するため、本研究ではシルエットの重なりを考慮した複数人テンプレートを作成する。まず、追跡対象人物が他の人物と離れている場合には、 N 個のパーティクルを用いて独立に追跡する。2人の人物 A, B が接近し、領域の重な

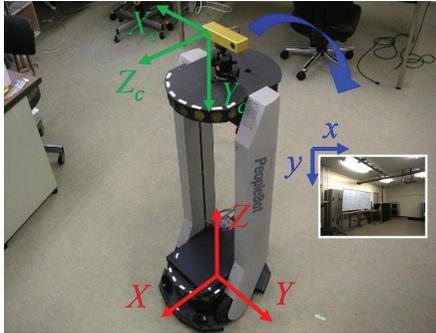


図 3 各座標系の定義
Fig. 3 Definition of coordinate systems

りが生じた時、それぞれの人物の状態変数 \mathbf{x}_t^A と \mathbf{x}_t^B を結合し、新たな状態変数 \mathbf{x}_t^{AB} を次のように作成する。

$$\mathbf{x}_t^{AB} = \begin{bmatrix} \mathbf{x}_t^A \\ \mathbf{x}_t^B \end{bmatrix} \quad (5)$$

ここで、各人物のパーティクル数が N 個の時、組合せの総数は $N \times N$ となる。しかし、全ての組合せについて評価（マッチング）を行うのは計算コストが大きいため、本研究ではそれぞれの人物のパーティクルの中で尤度の高いもののみを結合に用いる。（例えば実験では、孤立した人物をそれぞれ $N = 100$ 個のパーティクルで独立に追跡し、2人の人物が接近した時にはそれぞれ上位 25 個を結合して $N^{AB} = 25 \times 25$ 個のパーティクルを作成した。）結合したパーティクルの初期の尤度 L^{AB} は結合に用いたパーティクルの尤度 L^A, L^B の積 $L^{AB} = L^A L^B$ とする。また、状態方程式は次のようになる。

$$\mathbf{x}_{t+1}^{AB} = \begin{bmatrix} \mathbf{F}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_t \end{bmatrix} \mathbf{x}_t^{AB} + \begin{bmatrix} \mathbf{w}_t \\ \mathbf{w}_t \end{bmatrix} \quad (6)$$

複数人テンプレートを用いた追跡の手順を図 4 に示す。まず、結合したパーティクルそれぞれについて、2人のシルエットが重なりあったテンプレートを作成する。2人の3次元位置から画像上での位置関係を計算し、カメラから遠い人物のテンプレートの上に、近い人物のテンプレートを上書きする。それぞれの人物領域の画素値と背景領域の画素値は 2.1

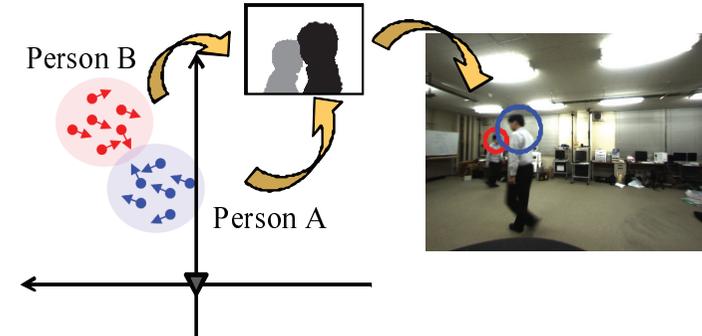


図 4 複数人テンプレートを用いた追跡の手順
Fig. 4 Procedure of tracking using an overlapping silhouette template

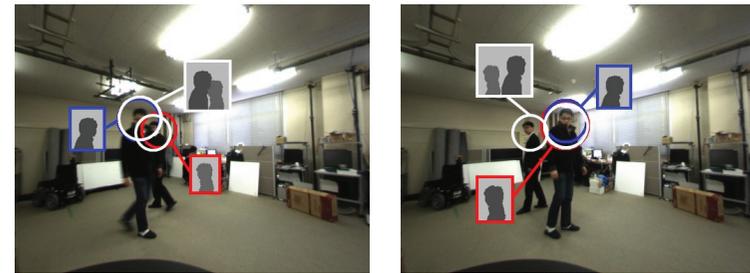


図 5 複数人テンプレート（白枠）と個別テンプレート（赤枠・青枠）の追跡結果の比較
Fig. 5 Comparison of tracking results with overlapping silhouette templates (drawn by white frame) and with individual templates (drawn by red or blue frame)

節と同じ方法で決定する。なお、ここでは計算コストを抑えるため、図 1 の 3 方向のテンプレートの内、各人物の移動方向に対応するテンプレート 1 つのみを用いる。そして、2.1 節で述べた個別テンプレートの場合と同様に複数人テンプレートと距離画像とのマッチングを行い、尤度 L^{AB} を計算する。

接近していた人物 A, B が離れ、距離が閾値を超えた時、状態変数 \mathbf{x}_t^{AB} を \mathbf{x}_t^A と \mathbf{x}_t^B に分離する。分離時には、結合された N^{AB} 個のパーティクルの内、尤度の高い N 個のパーティクルを用いる。

複数人テンプレートを用いた場合と、それぞれ個別のテンプレートを用いた場合の追跡結

果の比較を図5に示す。複数人テンプレートを用いた追跡結果を白色、個別テンプレートを用いた追跡結果を赤色と青色の円で描画している。個別のテンプレートを用いた追跡では奥の人物（赤色）の推定位置がずれており、これはシルエットの一部が手前の人物によって隠されてしまったためと考えられる（図5左）。また、奥の人物が完全に隠されてしまった場合には、対象人物を見失い、誤って別の人物を追跡してしまった（図5右）。一方、複数人テンプレートを用いた場合にはそれぞれの人物を正しく追跡することができており、シルエットの重なりを考慮することで他の人物に隠された場合でも安定な追跡が実現できていることが確認できた。

3. 複数台カメラへの拡張方法の検討

パブリックスペースでの特定人物追従を実現する際、人物は移動ロボットの前方だけではなく、ロボットの周囲を自由に移動する。そのため、安全にロボットを制御するためには、複数のカメラをロボットに取り付け、視野を拡張する必要がある。また、静止障害物が多く存在し、ロボットカメラだけで人物追跡を行うことが困難な状況では、周囲環境に設置されたカメラ（以下、環境カメラと呼ぶ）を併用する必要がある。複数のカメラやレーザレンジセンサを用いた研究¹⁸⁾も数多く行われているが、本研究では特に移動カメラによる人物追跡を主とし、本章では提案手法を複数台カメラで併用する場合の拡張方法について検討する。ロボットカメラ1台と環境カメラ2台を併用した場合の例を図6に示す。このシステムは各カメラ画像を処理する追跡モジュールとそれらの結果を統合する統合モジュールで構成され、カメラ1台は移動ロボットを想定しているため、各追跡モジュールは別PCで実行するものとする。以下では、図7に示した2種類の統合方法について述べる。

3.1 各カメラで個別に追跡する方法

まず、各カメラで個別に追跡を行う方法を述べる（図7(a)）。この方法では、それぞれの追跡モジュールは各人物について N 個のパーティクルを保持し、2章で述べた方法を用いて追跡を行う。なお、図6(c)のようにカメラ上で複数の人物が接近した場合には、提案手法を用いて人物シルエットの重なりを考慮した追跡を行う。そして、推定した人物の3次元位置を統合モジュールへ送信する。統合モジュールでは、各追跡モジュールで推定された人物位置の情報を統合する。この時、各カメラ間の人物の対応付けを行い、あるカメラでの追跡に誤りがある場合には、その情報を該当の追跡モジュールにフィードバックする。

3.2 複数カメラを統合して追跡する方法

次に、複数カメラを統合して追跡を行う方法を述べる（図7(b)）。この方法では、統合モ

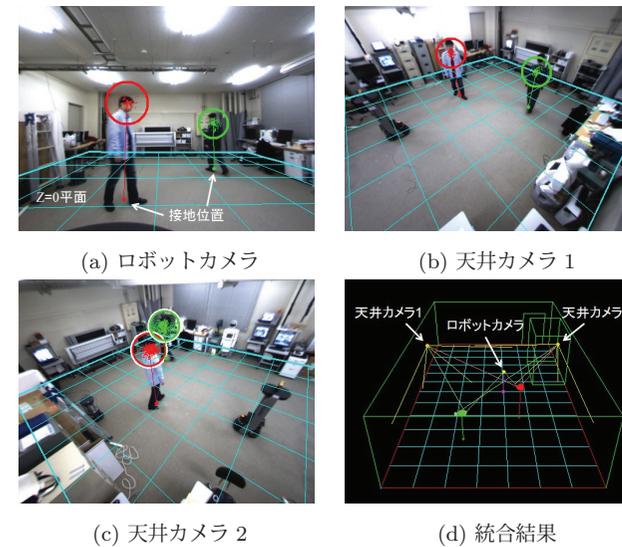


図6 複数台カメラを用いた人物追跡
Fig.6 Tracking using multiple cameras

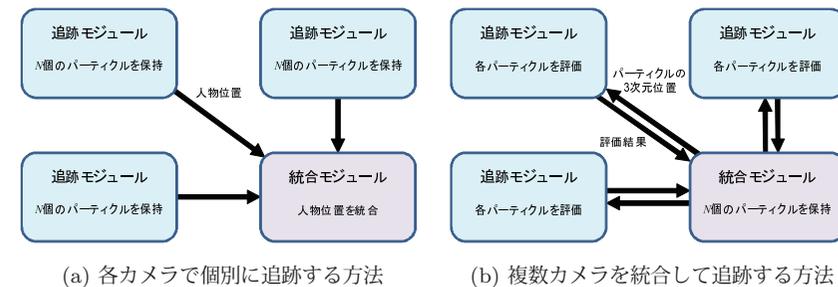


図7 複数台カメラを用いた追跡の統合方法
Fig.7 Integration of tracking with multiple cameras

ジュールで各人物について N 個のパーティクルを保持する。各パーティクルの3次元位置をそれぞれ全ての追跡モジュールに送信し、追跡モジュールは各パーティクルの評価のみを行う。そして、その評価結果を統合モジュールに送信し、統合モジュールで統合する。この時、カメラの視界の外に居る人物や、評価値が低く、他の人物や静止物体に隠れていると判

断される人物については、そのカメラの評価結果を破棄する。このように複数台のカメラを用いて多方向からの評価を同時に行うことで、より安定な人物追跡が実現できると考えられる。ただし、2.3節で述べた接近した人物の結合をどのように実装するかは今後の課題である。

図6のようにロボットカメラと環境カメラを併用する場合、視野が重複した領域が大きいため、この統合方法が有効だと考えられる。一方、ロボットの前方と後方にカメラを設置するなど、視野の重複が小さく、それぞれのカメラで個別に追跡を行う場合には3.1節で述べた個別に追跡を行う方法が有効だと考えられる。ただし、統合して追跡を行う方法では、 N 個全てのパーティクルの情報をやり取りするために通信コストが大きくなり、カメラ間の同期も必要である。また、1台のPCで処理を行う場合にはその計算コストも問題となる。現状では通信量はあまり問題とならないため、4.2節の実験では統合して追跡を行う方法を採用した。

なお、単に移動ロボットの見えていないところの情報を環境カメラから取得するだけでなく、オクルージョンの状況に応じてロボットをうまく移動させることで、固定カメラだけでは不十分な情報を補うことができると考えられる。このように、視覚移動ロボットと環境カメラとの相互協調によって安定な人物追跡を実現する方法について今後検討する。

4. 実験結果

4.1 ロボットカメラ1台を用いた人物追跡

図3に示した車輪移動ロボットとステレオカメラを用い、複数人物の追跡実験を行った。ロボットカメラは天井カメラに比べて低い位置にあるため、歩行者が他の歩行者に完全に隠されてしまうことがある。移動ロボットはMobileRobots社製のPeopleBotを用い、その上部にPointGreyResearch社製のステレオカメラBumblebee2 (XGA, 20fps, $f=2.5$ [mm], ベースライン12[cm])を設置した。1台のノートPC (Core2Duo, 3.06 [GHz])により、ステレオ処理、人物検出・追跡、ロボット制御の全ての処理を行っている。処理に用いた画像サイズは 512×384 である。

ロボットの動き情報の誤差(滑りや遅延)の影響を排除するため、まずはロボットを移動させず、10[fps]で撮影したオフライン画像を用いて追跡を行った結果を図8に示す。パーティクル数は $N = 100$, $N^{AB} = 25 \times 25$ とした。白緑の付いた人物はパーティクルを結合していることを表している。#160~#202の人物Bや#310~#324の人物Aのように、他の人物に隠された状態で移動方向が変化した場合でも正しく追跡が行われている。これは生

表1 パーティクル数を変化させた場合の追跡結果の比較

Table 1 Comparison of tracking results concerning the number of particles
 (a) using only individual templates

number of particles	success rate	positional error	processing time
$N = 100$	73.3 [%]	9.56 [pixel]	248 [ms]
$N = 200$	75.0 [%]	9.95 [pixel]	383 [ms]

(b) using overlapping silhouette templates

number of particles	success rate	positional error	processing time
$N^{AB} = 15 \times 15$	86.7 [%]	7.09 [pixel]	226 [ms]
$N^{AB} = 20 \times 20$	90.0 [%]	6.74 [pixel]	314 [ms]
$N^{AB} = 25 \times 25$	93.3 [%]	6.64 [pixel]	436 [ms]
$N^{AB} = 30 \times 30$	90.0 [%]	6.53 [pixel]	568 [ms]
$N^{AB} = 100 \times 100$	91.7 [%]	6.43 [pixel]	4993 [ms]

成されたシルエットの重なり方の仮説が、十分な多様性を持っていたためだと考えられる。

60回のオクルージョン(12ケースの画像セットをそれぞれ5回テスト)について追跡の比較を行った結果を表1に示す。それぞれの画像セットは2人の人物が近づき、交差して離れるまでのオフライン動画であり、各人物の移動方向や移動速度が様々に変化している。追跡の成功率は、のべ60回の交差の内、最後までそれぞれの人物を正しく追跡できたものの割合とした。また、2次元画像上の追跡誤差と1フレーム当たりの平均処理時間の計算には追跡に成功した画像セットのみを用いた。なお、各フレームにおけるそれぞれの人物の正解位置は手動で与えた。以上の結果から、シルエットの重なりを考慮した複数人テンプレートを用いた追跡は、個別のテンプレートを用いた場合よりも正確な人物位置を推定できていることが確認できた。

最後に、実際にロボットを制御し、特定人物を追従した結果を図9に示す。この実験では $N = 100$, $N^{AB} = 20 \times 20$ とし、ロボットは最初に検出した人物Aを追従対象とするようにした。別の人物Bがロボットと人物Aの間を横切り、オクルージョンが発生した場合でも正しく追従が行われていることが確認できた。

4.2 ロボットカメラと環境カメラを用いた人物追跡

図6に示したロボットカメラ1台と環境カメラ2台を併用して人物追跡実験を行った。3つの追跡モジュールと統合モジュールはそれぞれ別のPCで実行し、4台のPC間でTCP/IPを用いて情報を通信した。なお、天井カメラでは画面端に近づくにつれて人物の傾きが大きくなるが、ある程度の追跡性能が得られたため、今回はロボットカメラと同じ直立した人物

テンプレートを用いている。より正確・安定な追跡を実現するためにはこの傾きを考慮する必要がある。

3.2節で述べた統合方法を用いて追跡を行った結果を図10に示す。ロボットカメラの画像中で人物位置を示す円が黒くなっているものは、全パーティクルの評価値が小さく、オクルージョンが発生していると判断されたことを示している。これらの結果から、他の人物や静止障害物によってオクルージョンが発生した場合でも、正しく追跡できていることが確認できた。

5. まとめ

本研究では移動ロボットにステレオカメラを取り付け、距離情報を利用して複数人物の追跡を行う方法を提案した。シルエットの重なりを考慮した距離画像テンプレートを用いることで、正確かつ安定な追跡を実現した。そして、実際に移動ロボットによる特定人物の追従実験を行い、提案手法の有効性を確認した。また、複数台のカメラへ拡張する方法を検討し、1台のロボットカメラと2台の環境カメラを併用した場合の追跡実験を行った。

今後の課題としては、まず、人物の素早い動きに対応するために処理速度を向上させる必要がある。また、複数人テンプレートを人物の検出に拡張し、既知の人物2人が接近した場合だけでなく、追跡中の人物の後ろに重なっている未知の人物を検出できるようにすることで、より複雑な状況に対応できるようになると考えられる。

謝辞 本研究の一部は立石科学技術振興財団、倉田記念日立科学技術財団の研究助成を受けたものであり、記して謝意を表す。

参考文献

- 1) P. Viola, M.J. Jones, and D. Snow: Detecting pedestrians using patterns of motion and appearance, *Int. J. Computer Vision*, Vol.63, No.2, pp.153–161 (2005).
- 2) N. Dalal and B. Briggs: Histograms of oriented gradients for human detection, *IEEE Conf. Computer Vision and Pattern Recognition*, pp.886–893 (2005).
- 3) B. Han, S. W. Joo, and L. S. Davis: Probabilistic fusion tracking using mixture kernel-based Bayesian filtering, *Int. Conf. Computer Vision* (2007).
- 4) S. Munder, C. Schnorr, and D. M. Gavrilu: Pedestrian detection and tracking using a mixture of view-based shape-texture models, *IEEE Trans. Intelligent Transportation Systems*, Vol.9, No.2, pp.333–343 (2008).
- 5) D. Schulz, W. Burgard, D. Fox, and A. B. Cremers: People tracking with a mobile robot using sample-based joint probabilistic data association filters, *Int. J. Robotics Research*, Vol.22, No.2, pp.99–116 (2003).
- 6) N. Bellotto and H. Hu: Multisensor data fusion for joint people tracking and identification with a service robot, *IEEE Int. Conf. Robotics and Biomimetics*, pp.1494–1499 (2007).
- 7) H. Koyasu, J. Miura, and Y. Shirai: Realtime omnidirectional stereo for obstacle detection and tracking in dynamic environments, *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp.31–36 (2001).
- 8) M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia: People tracking and following with mobile robot using omnidirectional camera and a laser, *IEEE Int. Conf. Robotics and Automation*, pp.557–562 (2006).
- 9) D. Beymer and K. Konolige: Real-time tracking of multiple people using continuous detection, *Int. Conf. Computer Vision* (1999).
- 10) A. Howard, L. H. Matthies, A. Huertas, M. Bajracharya, and A. Rankin: Detecting pedestrians with stereo vision: safe operation of autonomous ground vehicles in dynamic environments, *Int. Symp. Robotics Research* (2007).
- 11) D. Calisi, L. Iocchi, and R. Leone: Person following through appearance models and stereo vision using a mobile robot, *VISAPP 2007 Workshop on Robot Vision*, pp.46–56 (2007).
- 12) A. Ess, B. Leibe, and L. V. Cool: Depth and appearance for mobile scene analysis, *Int. Conf. Computer Vision* (2007).
- 13) A. Ess, B. Leibe, K. Schindler, and L. V. Cool: A mobile vision system for robust multi-person tracking, *IEEE Conf. Computer Vision and Pattern Recognition* (2008).
- 14) J. MacCormick and A. Blake: A probabilistic exclusion principle for tracking multiple objects, *Int. Conf. Computer Vision*, pp.572–578 (1999).
- 15) Z. Khan, T. Balch, and F. Dellaert: An MCMC-based particle filter for tracking multiple interacting targets, *European Conf. Computer Vision*, pp.279–290 (2004).
- 16) D. Tweed and A. Calway: Tracking many objects using subordinated Condensation, *British Machine Vision Conf.*, pp.283–292 (2002).
- 17) J. Satake and J. Miura: Robust stereo-based person detection and tracking for a person following robot, *IEEE ICRA 2009 Workshop on People Detection and Tracking* (2009).
- 18) 倉爪亮, 山田弘幸, 曾我部光司, 村上剛司, 岩下友美, 長谷川勉: SIR/MCMC パーティクルフィルタを用いた分散カメラとレーザによる複数移動体の同時追跡, *日本ロボット学会誌*, Vol.28, No.1, pp.65–76 (2008).



図 8 オフライン画像を用いた人物追跡の結果
Fig. 8 Experimental result of tracking using off-line images



図 9 特定人物追従の結果
Fig. 9 Result on a person following control

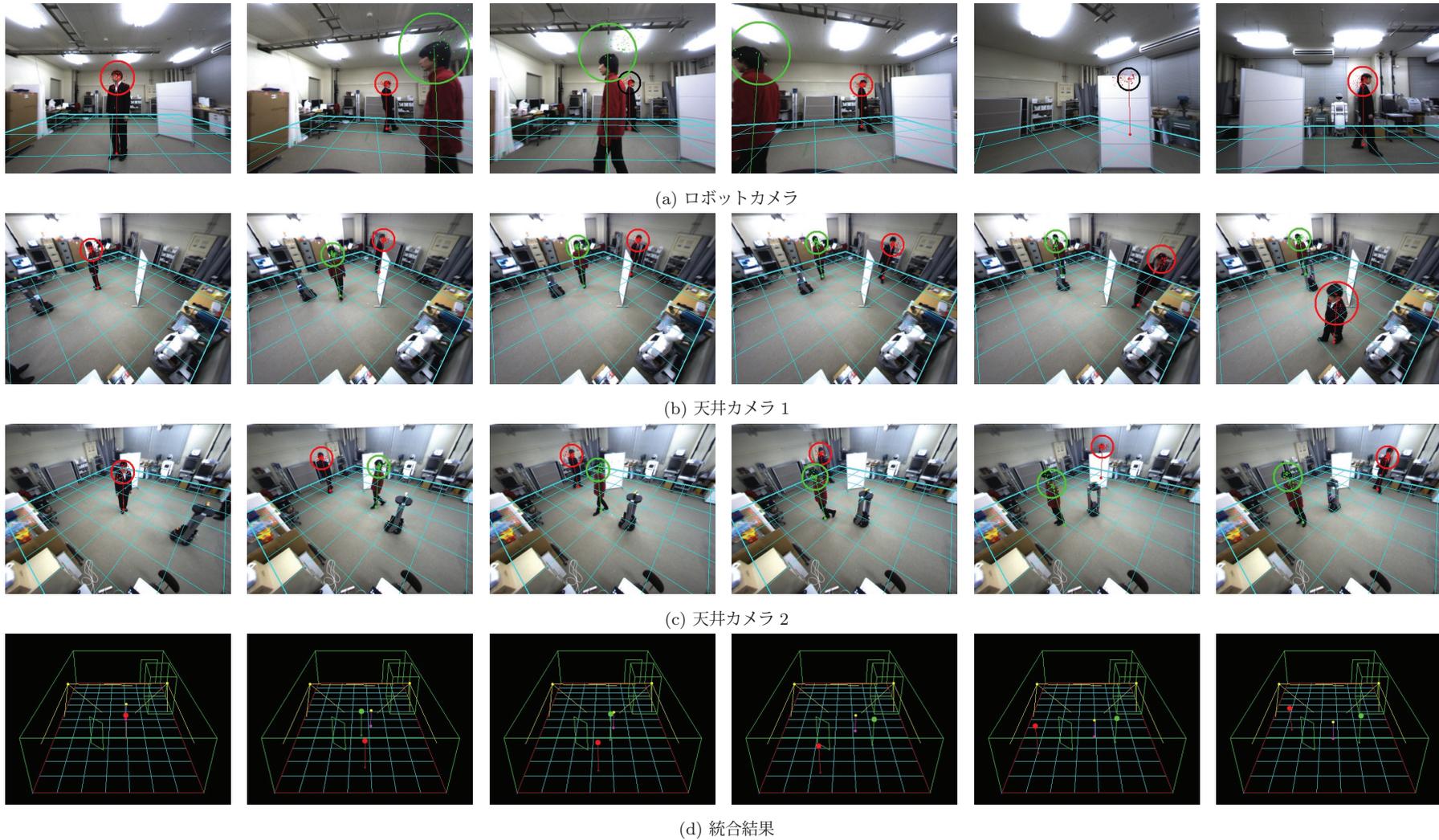


図 10 複数台カメラを用いた人物追跡結果
Fig. 10 Tracking result using multiple cameras