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たステレオビジョン複数人物追跡

佐 竹 純 二* 三 浦 純*

Stereo-Based Multi-Person Tracking using Overlapping Silhouette Templates

Junji Satake* and Jun Miura*

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. We previously developed a tracking method which uses depth templates of person shape applied to a dense depth image. The method, however, sometimes failed when complex occlusions occurred. In this paper, we propose an accurate, stable tracking method using overlapping silhouette templates which consider how persons overlap in the image. Moreover, we decrease the number of image comparison by using an approximated likelihood map based on kernel density estimation. Experimental results show that the proposed method is able to reduce the processing time greatly without dropping the tracking performance.

Key Words: Person Tracking, Mobile Robot, Particle Filter, Stereo Camera, Silhouette Templates

1. は じ め に

人物追跡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跡を実現するためには、移動カメラで撮影された動画像から複数の人物を検出・追跡する必要がある。動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、特に移動ロボットではオンラインでの実時間処理、移動する背景と人物の分離、複数人物の重なりなどが重要な課題である。

現在、HOG特徴[2]を用いた人物検出の研究が広く行われている[5][6]。また、Microsoft社のKinectのようなRGB-Dカメラから得られた距離情報とHOGを組み合わせた人物追跡手法も提案されている[7][8]。しかし、これらの手法では人物の重なりを明示的には扱っていない。移動ロボットに取り付けたカメラで撮影した動画像では歩行者が他の歩行者に完全に隠されてしまうオクルージョンも発生するため、本研究では人物同士の重なりを考慮した追跡を行う。また、屋外環境での使用も想定し、赤外線を利用したアクティブセンサなどはいずれにステレオカメラのみを用いた方法を開発する。

Essら[9][10]は様々な手がかり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、画像1枚当たり約6秒の処理時間がかかっており[10]、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。

池村と藤吉[11]は距離画像中の局所特徴（首から肩のラインなど）を使った人検出の手法を提案している。TOFカメラから得られた詳細な距離情報を用いて他の人物に隠されている領域を抽出し、局所領域の隠されている割合に応じて人識別器の評価値を下げることでオクルージョンに対応している。これに対し、ステレオカメラのように大まかな輪郭形状しか得られない場合にも有効な手法として、本研究では人物上半身のシルエットを用い、複数人物の重なりあったシルエット形状を直接利用した追跡方法を提案する。

また、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[12]~[14]。これらの手法では、複数物体が空間上で重ならないように確率的排他規則を加えることで複数物体の追跡を実現している。Khanら[13]は交錯する複数の昆虫を追跡する手法を提案しているが、二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Calway[14]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。そのため、これらの手法ではパーティクルの評価は追跡対象それぞれに対して行われている。これに対し、本研究では同様の排他規則に加え、複数人物の重なったシルエット形状のテンプレートを作成して距

原 2011 年 12 月 9 日

*豊橋技術科学大学

*Toyohashi University of Technology

■ 本論文は有用性で評価されました。

離画像と直接比較し、複数の追跡対象が重なった状況そのものを評価することで正確な位置推定を行う。

我々は移動ロボットに取り付けたカメラから撮影された画像を用い、人物のシルエット形状を距離画像テンプレートとして人物の検出を行う方法 [15] を開発してきた。本論文では、交錯する人物をより安定かつ正確に追跡するため、画像上でのシルエットの重なりを考慮したテンプレートをを用いた追跡方法を提案する。さらに、カーネル密度推定により尤度マップを近似し、テンプレートマッチングの回数を削減することで処理速度を向上させる。そして、実画像を用いた実験的検証により提案手法の有効性を示す。

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 [15]

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 (Fig. 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作成した、人物領域と背景領域の2領域からなるシルエット画像である。基準となるテンプレートは人物がカメラからの距離 2 [m] のときの距離画像から作成した。

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。具体的には、まず距離 2 [m] の距離画像から作成したテンプレートのサイズを $H_0 \times W_0$ とし、透視投影を考慮して距離 X [mm] のテンプレートサイズ $H \times W$ を次のように求める[†]。

$$H = \frac{2000}{X} H_0, \quad W = \frac{2000}{X} W_0$$

なお、テンプレート中の人物領域の画素値は実際の距離である。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)。距離に応じて作成した距離画像テンプレートを $T(x, y)$ 、入力された距離画像を $I_D(x, y)$ としたとき、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。

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

ここで、距離画像やテンプレートの画素値はカメラからの距離を表す 32 [bit] の実数値であり、画像として表示する際にのみ 8 [bit] の整数値に変換している。また、体の向きの変化に対応するため、Fig. 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、最も評価値の良いもの (相違度の低いもの) をマッチングの結果として使用する。なお、斜めを追加した 8 方向 5 種類 (左右それぞれの斜め前と後ろは共通) のテンプレートをを用いた場合でも追跡結果はほとんど変化しなかったため、3 種類のテンプレートで十分だと考えられる。また、ステレオで得られる距離情報にはノイズが含まれており、三次元形状を持つテンプレートをを用いても追跡結果がほとんど変

Fig. 1 Depth templates

Fig. 2 Examples of tracking result using depth template

Fig. 3 Definition of each coordinate system

化しなかったため、人物領域の画素値を単一の距離値とした平坦な形状のテンプレートをを用いた。

入力された距離画像の各画素に対してテンプレートと比較し、画面内に新しく出現した人物を検出する。ただし、探索範囲を三次元空間における高さ 0.7~2.0 [m]、カメラからの距離 0.5~5.5 [m] に限定している。また、シルエット形状が人と似ている物体を誤検出してしまうことがあるため、そのような物体をあらかじめ学習しておき、SVM を用いて人物かどうかを判定する [15]。

距離画像テンプレートをを用いた人物検出の例を Fig. 2 に示す。3 種類のテンプレートの探索結果のうち、最も相違度の小さいものを太線で表示しており、複数のテンプレートを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三次元位置推定

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 Fig. 3 に示す。移動ロボットによる人物の追従・回避を実現するためには特に各人物の XY 平面座標が必要であり、また、二次元画像上で人物上半身の位置を予測するためには高さ情報も必要である。そこで、本研究では人物頭部の三次元位置を追跡し、それを XY 平面に投影してロボットの行動計画に用いる。

[†]高速化のため、実際には pixel 値で離散化したすべての大きさのテンプレート形状をあらかじめ用意している。

時刻 t での, ロボット座標系における人物頭部の三次元位置を (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.

$$\mathbf{x}_t = [X_t \ Y_t \ Z_t \ \dot{X}_t \ \dot{Y}_t]^T$$

ただし, $\dot{X}_t, \dot{Y}_t$ は XY 平面上の移動速度である. このとき, 状態方程式は次のようになる.

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{w}_t$$

ただし,

$$\mathbf{F}_t = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$\mathbf{w}_t$ はシステムノイズを表すベクトルである. なお, ロボットの動きを考慮するため, オドメトリ情報 (前時刻との位置の差分) を用いて人物位置を現時刻のロボット座標系へと変換する. 人物位置をワールド座標系ではなくロボット座標系で扱っているのは制御量が計算しやすいからである.

自由に移動する人物を安定に追跡するため, 本研究ではパーティクルフィルタを用いて各人物の三次元位置を推定する. 2.1 節で述べた相違度 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.

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

また, 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. すべてのパーティクルの尤度がしきい値よりも小さい状態が一定時間 (実験では 10 フレームとした) 続いた場合にはその人物の追跡を中断し, パーティクルを破棄する.

3. シルエットの重なりを用いた複数人物追跡

3.1 2 人の重なり

接近した複数人物を正確かつ安定に追跡するため, 複数人物の重なったシルエット形状のテンプレートを作成し, 距離画像と直接比較する.

まず, 追跡対象人物が画像上で他の人物と離れている場合には, それぞれの人物を独立に追跡する. 2 人の人物 A, B が接近し, 画像上のテンプレートの重なりが生じた時, 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.

$$\mathbf{x}_t^{AB} = \begin{bmatrix} \mathbf{x}_t^A \\ \mathbf{x}_t^B \end{bmatrix}$$

なお, この人物の重なり判定は各パーティクルの尤度を求めた後, リサンプリングの前に行う. パーティクルの尤度を重みとした各人物位置の重み付き平均 $(\bar{X}^A, \bar{Y}^A, \bar{Z}^A)$, $(\bar{X}^B, \bar{Y}^B, \bar{Z}^B)$ とその画像投影座標 $(\bar{u}^A, \bar{v}^A)$, $(\bar{u}^B, \bar{v}^B)$ をもとに, 以下の場合に重なりが生じたと判定する.

$$|\bar{u}^A - \bar{u}^B| < \frac{\bar{W}^A + \bar{W}^B}{2}$$

Fig. 4 Procedure of tracking using overlapping silhouette templates

ここで, $\bar{W}^A, \bar{W}^B$ はそれぞれの平均距離 $\bar{X}^A, \bar{X}^B$ でのテンプレート幅であり, 以下のように求める.

$$\bar{W}^A = \frac{2000}{\bar{X}^A} W_0, \quad \bar{W}^B = \frac{2000}{\bar{X}^B} W_0$$

すべての組み合わせについて評価 (マッチング) を行うのは計算コストが大きいため, 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. ここで, 他の人物と離れている 1 人の人物の追跡に用いるパーティクルの数を $N_{(1)}$, 重なった 2 人を結合して追跡する場合に用いるパーティクルの数を $N_{(2)}$ とおく. 実験では他の人物と離れている人物をそれぞれ $N_{(1)}=100$ 個のパーティクルで独立に追跡し, 2 人が接近したときにはそれぞれ上位 20 個を結合して $N_{(2)}=20 \times 20$ 個のパーティクルを作成した. なお, 追跡に使用するパーティクル数は, 屋内で自然に歩行する場合に十分な数を経験的に設定した. 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.

$$\mathbf{x}_t^{AB} = \begin{bmatrix} \mathbf{F}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_t \end{bmatrix} \mathbf{x}_{t-1}^{AB} + \begin{bmatrix} \mathbf{w}_t^A \\ \mathbf{w}_t^B \end{bmatrix}$$

重なりを考慮したテンプレートを用いた追跡の手順を Fig. 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 (以下, 重なりテンプレートと呼ぶ) を作成する. 具体的には状態変数 $\mathbf{x}^{AB}$ 中の 2 人の三次元位置 (X^A, Y^A, Z^A) , (X^B, Y^B, Z^B) からそれぞれの画像投影座標 (u^A, v^A) , (u^B, v^B) とテンプレートサイズ $H^A \times W^A$, $H^B \times W^B$ を求め, カメラから遠い人物のテンプレートの上に, 近い人物の人物領域を上書きする. 人物領域の画素値はそれぞれの距離 X^A, X^B であり, 背景領域の画素値は 2.1 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, Fig. 1 の 3 方向のテンプレートのうち, 各人物の移動方向に対応するテンプレート一つのみを用いる. 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重なり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, 画像上のテンプレートが重ならない距離になったとき (すなわち $|\bar{u}^A - \bar{u}^B| > (\bar{W}^A + \bar{W}^B)/2$ となったとき), 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 $N_{(2)}$ 個のパーティクルのうち, 尤度の高い $N_{(1)}$ 個のパーティクルを用いる.

2 人の人物の重なりに対し, それぞれ個別のテンプレートを用いた場合と, 重なりテンプレートを用いた場合の追跡結果の比

Fig. 5 Comparison of tracking results

較を Fig. 5 に示す。個別テンプレートを用いた追跡結果を灰色と黒色、重なりテンプレートを用いた追跡結果を白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（灰色）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（Fig. 5 (a) 上）。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（Fig. 5 (a) 下）。一方、重なりテンプレートを用いた場合には、2 人の重なったシルエット形状のテンプレートを用いて距離画像と直接比較することで、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（Fig. 5 (b)）。

3.2 3 人の重なり

3 人の重なりが生じた場合にも、2 人の場合と同様に状態変数を結合し、3 人の重なりテンプレートを作成する。本研究では移動ロボットによる特定人物追従を目的としており、ロボットと追従対象人物の間を 1~2 人の人物が横切るような状況を想定しているため、3 人までの重なりテンプレートでほぼ対応できると考えられる。

3 人の人物 A, B, C が接近し、画像上のテンプレートの重なりが生じたとき、それぞれの人物の状態変数を結合し、新たな状態変数 $\mathbf{x}_t^{ABC}$ を次のように作成する。

$$\mathbf{x}_t^{ABC} = \begin{bmatrix} \mathbf{x}_t^A \\ \mathbf{x}_t^B \\ \mathbf{x}_t^C \end{bmatrix}$$

結合した 3 人の追跡に用いるパーティクル数を、実験では $N_{(3)} = 729 (= 9^3)$ とした。ここで、3 人の重なりが生じるケースには次の二つが考えられる。一つは 3 人の独立した人物が現在のフレームで同時に重なるケースであり、この場合は各人物の上位 $N_{(3)}^{1/3}$ 個のパーティクルを結合することで $N_{(3)}$ 個のパーティクルを作成する。もう一つは 1 人の独立した人物と 2 人の重なった（結合された）人物が重なるケースであり、この場合は 1 人の人物から $N_{(3)}^{1/3}$ 個、2 人のグループから $N_{(3)}^{2/3}$ 個のパーティクルを取り出して結合する。

重なっていた 3 人の人物が離れ、画像上のテンプレートが重ならない距離になったときも同様に状態変数を分離する。1 人

Fig. 6 Estimation of each particle's likelihood

ずつに分かれた場合には $N_{(3)}$ 個のパーティクルの中から尤度の大きい $N_{(1)}$ 個、1 人と 2 人に分かれた場合には尤度の大きい $N_{(1)}$ 個と $N_{(2)}$ 個を用いて該当人物の状態変数を抽出する。

4. 尤度マップの近似による追跡処理の高速化

3 章で述べた追跡方法では、すべてのパーティクルについてテンプレートマッチングを行うために画像比較のコストが大きいという問題がある（Fig. 6 (a)）。そこで、Fig. 6 (b) のように、カーネル密度推定 [16] [17] に基づき、少ないサンプルから尤度マップを近似する方法を提案する。具体的には、ある一定数のパーティクルでのみテンプレートマッチングを行い、近似した尤度マップから残りのパーティクルの尤度を推定することで高速な追跡を実現する。

4.1 1 人の場合の尤度マップ近似

まず、1 人の人物の場合の尤度マップ近似方法について述べる。他の人物と離れた 1 人の人物の追跡には $N_{(1)}$ 個のパーティクルを用いる。最初に、前フレームでの尤度が大きかったパーティクルをある一定数選び、現フレームの予測位置でテンプレートと画像を比較してその評価値を求める。ここで、選択された k 番目のパーティクルの尤度を L_k とする。次に、カーネル密度推定に基づき、XYZ 空間における混合正規分布モデルにより尤度マップを近似する。実際の尤度マップの形状を参考に、本研究では以下のように混合正規分布を用いて近似する。

$$L(\mathbf{x}) = \sum_k L_k |\Sigma_k|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \Delta_k(\mathbf{x})^T \Sigma_k^{-1} \Delta_k(\mathbf{x}) \right\},$$

(a) Likelihood distribution considering projection

(b) Actual likelihood map

(c) Evaluation value of 50 particles

(d) Approximated likelihood map

Fig. 7 Approximation of likelihood map based on kernel density estimation

$$\Delta_k(\mathbf{x}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} (\mathbf{x} - \mathbf{x}_k) = \begin{bmatrix} X - X_k \\ Y - Y_k \\ Z - Z_k \end{bmatrix},$$

$$\Sigma_k = \mathbf{R}_Z(\theta_k) \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} \mathbf{R}_Z(\theta_k)^T$$

ここで、 θ_k は Z 軸まわりの回転角度であり、人物はカメラ中心と画像平面上の位置を結ぶ直線状に存在するため、透視投影を考慮して、分布を Z 軸まわりに回転する (Fig. 7(a)). θ_k の大きさは以下のように求める。

$$\theta_k = \tan^{-1} \frac{Y_k}{X_k}$$

なお、人物がカメラ正面にいる場合 ($\theta_k = 0$) の分布の X, Y, Z 軸方向の標準偏差 $\sigma_X, \sigma_Y, \sigma_Z$ は実際の尤度マップの形状に基づいて経験的に設定した。これらは実際の人物サイズに依存するものであるため、シーンが変わっても有効である。また、距離画像とテンプレートのマッチングにおいて、位置のずれによる尤度 L の変化の仕方は画像上の位置に依存しない。さら

Fig. 8 Sigmoid function used for exclusion principle

に、透視投影を仮定しているため、画像の横方向のずれと三次元中の Y 軸方向のずれは比例し、またその大きさは Y 軸上の位置には依存しない。そのため、分布の中央を Y 軸方向に切った場合の断面の標準偏差 s_Y が一定になるように (Fig. 7(a) を参照)、回転された分布の主軸方向の標準偏差 σ_1, σ_2 を回転角度 θ_k から計算する。

$$\sigma_1 = \frac{\sigma_X}{\cos \theta_k}, \quad \sigma_2 = s_Y \cos \theta_k, \quad \sigma_3 = \sigma_Z$$

選択された 50 個のパーティクルの尤度 (Fig. 7(c)) をもとに、混合正規分布モデルにより近似した尤度マップの例を Fig. 7(d) に示す。それぞれの位置についてテンプレートマッチングを行って求めた実際の尤度マップ (Fig. 7(b)) とよく似た形状が得られていることが確認できる。なお、近似に使用するパーティクルの数は、十分な近似形状を得るために必要な数を経験的に設定した。最後に、近似された尤度マップから残りのパーティクルの尤度を推定することで高速な追跡を実現する。

4.2 2 人と 3 人の場合の尤度マップ近似

まず、2 人の重なりの場合には、特徴空間を六次元に拡張して尤度マップ近似を行う。テンプレートのずれの影響は各人物において独立と仮定し、以下のように近似する。

$$L_{(2)}(\mathbf{x}^{AB}) = \varsigma_{(2)}(\mathbf{x}^{AB}) \times \sum_k L_k |\Sigma_k^2|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \Delta_k^2(\mathbf{x}^{AB})^T (\Sigma_k^2)^{-1} \Delta_k^2(\mathbf{x}^{AB}) \right\},$$

ただし、

$$\Delta_k^2(\mathbf{x}^{AB}) = \begin{bmatrix} \begin{bmatrix} X^A - X_k^A \\ Y^A - Y_k^A \\ Z^A - Z_k^A \end{bmatrix}^T \begin{bmatrix} X^B - X_k^B \\ Y^B - Y_k^B \\ Z^B - Z_k^B \end{bmatrix}^T \end{bmatrix}^T,$$

$$\Sigma_k^2 = \begin{bmatrix} \Sigma_k & 0 \\ 0 & \Sigma_k \end{bmatrix}$$

ここで、 $\varsigma_{(2)}(\mathbf{x}^{AB})$ は 2 人の人物間の距離 δ に基づくシグモイド関数である。

$$\varsigma_{(2)}(\mathbf{x}^{AB}) = \frac{1}{1 + e^{-a\{\delta - b\}}}$$

$$\delta = \sqrt{(X^A - X^B)^2 + (Y^A - Y^B)^2}$$

パーティクルフィルタでは各人物の動きを独立に予測しているが、実際には複数の人物が実空間上で同じ位置に重なることはない

Fig. 9 Examples of approximation in the case of two and three overlapping persons

ため、本研究では人物間の距離に基づく係数を排他規則として尤度マップに乗じる。実際に使用したシグモイド関数 ($a = 0.02$, $b = 500$) を **Fig. 8** に示す。このパラメータも一般的な人物の大きさに依存したものである。

同様に、3 人の重なりの場合には、以下のように特徴空間を九次元に拡張して尤度マップ近似を行う。

$$L_{(3)}(\mathbf{x}^{ABC}) = \zeta_{(3)}(\mathbf{x}^{ABC}) \times \sum_k L_k |\Sigma_k^3|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \Delta_k^3(\mathbf{x}^{ABC})^T (\Sigma_k^3)^{-1} \Delta_k^3(\mathbf{x}^{ABC}) \right\}$$

2 人と 3 人の重なりにおいて近似された尤度マップの例をそれぞれ **Fig. 9** (d) に示す[†]。複数人の重なりが生じた場合でも、各位置について求めた実際の尤度マップ (**Fig. 9** (b)) とよく似た形状の尤度マップが近似できていることが確認できる^{††}。

5. 実験

提案手法の有効性を確認するため、**Fig. 3** のような移動ロボットとステレオカメラを用い、実際に撮影した動画像を用いて人物の追跡実験を行った。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, 20 [fps], $f = 2.5$ [mm] (水平画角 100°), ベースライン 12 [cm]) を設置した。カメラの高さは約 125 [cm] である。1 台のノート PC (Core i7, 2.7 [GHz]) により、ステレオ処理、人物検出・追跡、ロボット制御のすべての処理を行う。処理に用いた画像サイズは 512×384 である。今回は最大で 3 人の重なりが生じる状況について検証を行った。

5.1 重なりテンプレートを用いた複数人物追跡の検証

10 [fps] で撮影された動画像に対し、重なりテンプレートを用いて追跡を行った結果を **Fig. 10** に示す。ここではパーティクル数を $N_{(1)} = 100$, $N_{(2)} = 25 \times 25$ とした。白縁の付いた人物の組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 **Table 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動画像であり、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、延べ 60 回の交差のうち、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、二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、重なり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。

5.2 尤度マップの近似による追跡処理の高速化の検証

10 [fps] で撮影された動画像を用い、尤度マップ近似を用いて追跡を行った結果を **Fig. 11** に示す。前節と同様に白縁の付いた複数人物は結合した状態変数を用いた追跡結果であることを表している。3 人の重なりが生じた場合でもそれぞれの人物位置を正しく推定できていることが確認できる。

動画像の中からオクルージョンが発生する画像セットを抽出し、360 回のオクルージョン (12 ケースの画像セットをそれぞれ 30 回テスト、合計 15,830 フレーム) について追跡の比較を行った結果を **Table 2** に示す。追跡の成功率や追跡誤差、平均処理時間の求め方は前節と同様である。まず、(1) 個別テンプレートを用いた追跡では、従来手法と同様に排他規則を用いているが、重なりを考慮した評価を行っていないために隠れてしまった人物を見失い、追跡成功率が低くなっている。また、(3) 重なりテンプレートを用いた追跡では追跡成功率が大きく向上しているが、すべてのパーティクルについてテンプレートマッチングを行うために処理時間が大きい。これに対し、(2) 個別テンプレートを用いた追跡に尤度マップ近似による高速化のみを行った場合や、(4) 重なりテンプレートを用いた追跡に尤度マップ近似による高速化を行った場合 (提案手法) では、追

[†] 図では Y^A , Y^B についてのみ変化させ、それ以外はパーティクルの重み付き平均の値について表示している。

^{††} **Fig. 9** (b) 下の左側の二つのピークは人物 B が人物 A の後ろに隠れた場合のものである。今回の表示範囲内には小さなピークが存在するが、実際の追跡の上では距離が離れているためにパーティクルが存在せず、(d) の近似した尤度マップで復元されていなくても影響はない。

Fig. 10 Experimental result of tracking using overlapping silhouette templates

Fig. 11 Experimental result using overlapping silhouette templates with approximated likelihood map

Table 1 Comparison of tracking results concerning the number of particles

(a) using only individual templates			
number of particles	processing time	positional error	success rate
$N_{(1)} = 100$	248 [ms/frame]	9.56 [pixel]	73.3 [%]
$N_{(1)} = 200$	383 [ms/frame]	9.95 [pixel]	75.0 [%]
(b) using overlapping silhouette templates			
number of particles	processing time	positional error	success rate
$N_{(2)} = 15 \times 15$	197 [ms/frame]	7.09 [pixel]	86.7 [%]
$N_{(2)} = 20 \times 20$	275 [ms/frame]	6.74 [pixel]	90.0 [%]
$N_{(2)} = 25 \times 25$	367 [ms/frame]	6.64 [pixel]	93.3 [%]
$N_{(2)} = 30 \times 30$	498 [ms/frame]	6.53 [pixel]	90.0 [%]
$N_{(2)} = 100 \times 100$	4684 [ms/frame]	6.43 [pixel]	91.7 [%]

Table 2 Comparison of tracking performances

(a) two overlapping persons (240 occlusion cases)

tracking method	template matching	processing time	positional error	success rate
(1) individual templates	200 [times/frame]	146.6 [ms/frame]	6.24 [pixel]	80.8 [%]
(2) individual templates using approximation	100 [times/frame]	80.6 [ms/frame]	6.52 [pixel]	78.8 [%]
(3) overlapping templates	400 [times/frame]	195.0 [ms/frame]	5.30 [pixel]	95.0 [%]
(4) overlapping templates using approximation (proposed)	100 [times/frame]	65.9 [ms/frame]	5.40 [pixel]	94.6 [%]

(b) three overlapping persons (120 occlusion cases)

tracking method	template matching	processing time	positional error	success rate
(1) individual templates	300 [times/frame]	202.5 [ms/frame]	4.75 [pixel]	73.3 [%]
(2) individual templates using approximation	150 [times/frame]	134.4 [ms/frame]	4.86 [pixel]	71.7 [%]
(3) overlapping templates	729 [times/frame]	363.3 [ms/frame]	4.38 [pixel]	86.7 [%]
(4) overlapping templates using approximation (proposed)	200 [times/frame]	128.7 [ms/frame]	4.54 [pixel]	85.2 [%]

Fig. 12 Experimental result on a person following control

追跡成功率を維持したまま、処理時間を大きく削減することができており、その有効性を確認することができた。

5.3 移動ロボットによる特定人物追従実験

実際に移動ロボットを制御し、特定人物を追従した結果を Fig. 12 に示す。今回は追従対象人物の識別に色情報を用い、赤色の服を着た人物 A を追従するようにした。なお、ロボットの最大移動速度は 900 [mm/s] であるが、動き情報の誤差の影響を抑えるために今回は 350 [mm/s] を上限とした。

カメラから入力された画像をオンラインで処理し、ロボットの移動によって背景が変動した場合でも正しく追跡が行われている。また、別の人物 B がロボットと人物 A の間を横切り、1 秒程度のオクルージョンが発生した場合でもそれぞれの人物を正しく追跡できている。以上の実験結果から、移動ロボットによる特定人物追従が実際にオンラインで実現できていることを確認した。

6. おわりに

本論文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検出・追跡を行う方法を開発した。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。また、カーネル密度推定に基づき、少ない画像比較回数で尤度マップを近似することにより、高速な追跡処理を実現した。そして、最大 3 人の重なりを含む動画画像を用いて実際に人物追跡実験を行い、重なりテンプレートを用いることで個別のテンプレートを用いた場合に比べて追跡成功率が約 10% 向上することを確認した。なお、本論文では 3 人までの重なりについて述べたが、同様に n 人の重なりにも拡張することが可能である。ただし、人数が多くなると追跡に必要なパーティクル数も多くなるため、並列処理などを実装する必要がある。本研究では移動ロボットと追従対象人物の間を 1~2 人の人物が横切るような状況を想定しており、3 人までの重なりテンプレートではほぼ対応できると考えられる。

今後の課題としては、より多人数が交錯する複雑な状況での検証や、追従対象人物やロボットの進路上の人物を優先して処理するための注視制御などが挙げられる。

謝辞 本研究の一部は NEDO「次世代ロボット知能化技術開発プロジェクト」、科研費 (23700203) の助成によるものである。

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. Gavrila: "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] M. Enzweiler and D. Gavrila: "Monocular Pedestrian Detection: Survey and Experiments," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, no.12, pp.2179–2195, 2009.
- [6] P. Dollar, C. Wojek, B. Schiele and P. Perona: "Pedestrian Detection: A Benchmark," *Proc. 25th Conf. on Computer Vision and Pattern Recognition*, pp.304–311, 2009.
- [7] L. Spinello and K.O. Arras: "People Detection in RGB-D Data," *Int. Conf. on Intelligent Robots and Systems*, pp.3838–3843, 2011.
- [8] J. Salas and C. Tomasi: "People Detection Using Color and Depth Images," *Mexican Conf. on Pattern Recognition*, pp.127–135, 2011.
- [9] A. Ess, B. Leibe, K. Schindler and L.V. Gool: "Moving Obstacle Detection in Highly Dynamic Scene," *Int. Conf. on Robotics and Automation*, pp.56–63, 2009.
- [10] A. Ess, B. Leibe, K. Schindler and L. V. Gool: "Object Detection and Tracking for Autonomous Navigation in Dynamic Environments," *Int. J. of Robotics Research*, vol.29, no.14, pp.1707–1725, 2010.
- [11] 池村翔, 藤吉弘亘: "距離情報に基づく局所特徴量によるリアルタイム人検出", *電子情報通信学会論文誌*, vol.J93-D, no.3, pp.355–364, 2010.
- [12] J. MacCormick and A. Blake: "A probabilistic exclusion principle for tracking multiple objects," *Int. Conf. Computer Vision*, pp.572–578, 1999.
- [13] 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.
- [14] D. Tweed and A. Calway: "Tracking many objects using subordinated Condensation," *British Machine Vision Conf.*, pp.283–292, 2002.
- [15] 佐竹純二, 三浦純: "ステレオビジョンを用いた移動ロボットの人物追従制御", *日本ロボット学会誌*, vol.28, no.9, pp.1091–1099, 2010.

- [16] E. Parzen: "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol.33, no.3, pp.1065–1076, 1962.

- [17] R.O. Duda and P.E. Hart: *Pattern Classification and Scene Analysis*. pp.44–84, John Wiley & Sons, 1973.

佐竹純二 (Junji Satake)

1998 年岡山大学工学部情報工学科卒業。2003 年同大学大学院自然科学研究科博士後期課程修了, 博士 (工学)。独立行政法人情報通信研究機構専攻研究員, 豊橋技術科学大学情報工学系助手を経て, 2010 年同大学大学院工学研究科助教。コンピュータビジョン, パターン認識, ヒューマンインタフェースの研究に従事。電子情報通信学会, 情報処理学会各会員。

(日本ロボット学会正会員)

三浦 純 (Jun Miura)

1984 年東京大学工学部機械工学科卒業。1989 年同大学大学院工学系研究科情報工学専攻博士課程修了, 工学博士。同年大阪大学助手。同大学大学院工学研究科機械工学専攻助教授を経て, 2007 年豊橋技術科学大学情報工学系教授。知能ロボット, 人工知能, コンピュータビジョンの研究に従事。1994～1995 年 CMU 客員研究員。1997 年ロボット学会論文賞受賞。人工知能学会, 電子情報通信学会, 情報処理学会, システム制御情報学会, 日本機械学会, IEEE, AAAI 各会員。 (日本ロボット学会正会員)