応用測量論文集〔論文〕
JAST Vol.20, 2009, pp15-22

ステレオビデオカメラによる複数人物追跡の可能性

A Study on Multiple Human Tracking by Stereo Video Camera

布施孝志*・母里明陽**
By Takashi Fuse* and Akihiro Mori**

Abstract: Recently, interest for observations of human movements is increasing. Such kinds of observations require individual human tracking under the circumstance where the multiple human exist, and the observations are still challenging. Though video cameras can acquire color data, these have difficulty of tracking objects having occlusion. On the other hand, range data make the tracking more stable against the occlusion. Stereo video cameras can acquire color and range data simultaneously. This study proposes a multiple human tracking technique by using the stereo video camera. To be specific, human tracking is accomplished by combining color and range data. Generalized state-space model is adopted as a basic method, and color and range data are combined into observation model. The human positions are estimated by using particle filter. The proposed technique is applied to real sequential images, and the effectiveness of the technique is confirmed.

Keywords: multiple human tracking, stereo video cameras, color and range data, particle filter

1．はじめに

近年のセンシング技術の発展の中、人物の動きの観測に対して注目が集まっており、その自動観測手法の構築への要請が高まっている。個々の人物の動きが観測できれば、監視・防犯システム、商業施設における顧客動線の取得によるマーケティング分析、駅構内や横断歩道といった公共空間での人物の動きの解析、その結果に基づく施設設計や流動制御への応用等、その応用分野も多岐にわたる。これらの応用対象は、いずれも複数人物が存在している状況において、個人の動きを認識しなくてはならない。例えば、横断歩道における人物の流れの解析を例に取ると、群集における個人の動きを把握することにより、群集の密度と個人の動きや群集化との関係性を議論することが可能となり、より効率的に人物の流れを制御するための方策への貢献が期待される。

人物の動きを観測するために、従来は一般的に、ビデオカメラによる観測が行われてきた。ビデオカメラによる観測では、低コスト、多数の設置可能性、目視判定の容易さが利点である一方、人物相互の遮蔽（オクルージョン）が発生した場合には、自動追跡が非常に困難となる。オクルージョンの問題に対する画像認識手法は、現在もチャレンジングな課題として取り組まれている状況である[10)-13)]。オクルージョンの問題に対処するために、レーザスキャナを用いた距離情報による人物抽出手法も開発されている[5)]。レーザスキャナによれば、混雑時においても安定的に人物の抽出が可能であるが、人物の外観を把握することはできないため、同一人物を同定することが困難となる。

一方で、ステレオビデオカメラは、人物の外観が把握可能である色情報と、人物抽出に有効である距離情報を同時に取得することが可能なセンサである。ステレオビデオカメラを用いれば、距離情報を用いることにより、オクルージョンに対して頑健であり、かつ色情報を用いることにより、同一人物の追跡を可能とし、その結果、複数人物の追跡が容易になると考えられる。既に、距離情報を自動取得可能なステレオビデオカメラも製品化されており、今後の実用化に向けても期待が高まっていくことが予想される。

* 国土技術政策総合研究所　National Institute for Land and Infrastructure Management
** 三菱商事株式会社　Mitsubishi Corporation

以上の背景の下、本研究は、ステレオビデオカメラによる複数人物追跡の可能性を検討することを目的とする。特に、従来から行われている色情報のみを用いた画像認識手法に対し、距離情報を統合することにより、複数人物追跡の基礎手法の適用可能性を検討する。

2．複数人物追跡のための基礎手法

動物体の追跡においては、動きを対象としていることから、時系列データを扱うことになる。また、本研究では、色情報と距離情報の複数種類の観測値を用いるため、それらの情報の統合に適した手法が必要となる。本要件に合致した手法の一つとして、本研究では、一般状態空間モデル [1), 6)]に基づき、基礎手法を構築するものとする。一般状態空間モデルでは、人物位置（状態量）の推定を、過去の観測値に基づいた予測値と現在の観測値から最適な推定値を求める時系列フィルタリング [2)]により実現する。これは、時系列解析において一般的に知られるモデルであり、多様な分野において用いられている。また、時系列フィルタリングは、画像解析の分野においても、動物体認識のための手法として、近年、特に着目されているものである。

（1）一般状態空間モデル

一般状態空間モデルは、時刻 t における観測可能な観測ベクトル $\mathbf{z}_t$（例えば、色や距離等）と観測不可能な（観測できないものとする）状態ベクトル $\mathbf{x}_t$（例えば、人物の位置等）から構成される。これらは確率変数であり、時刻 t-1 から時刻 t の状態ベクトルの関係を表現するシステムモデル、観測ベクトルと状態ベクトルとの関係を表現する観測モデルに従う（図１）。時刻 t における観測ベクトルが得られる毎に、事後確率最大化基準の下、逐次、最適な状態ベクトルを推定する。

ここで、時刻 1 から t までの観測量を $\mathbf{z}_{1:t}=\{\mathbf{z}_1, \cdots, \mathbf{z}_t\}$ と表すと、ベイズの法則より、事後確率 $p(\mathbf{x}_t|\mathbf{z}_{1:t})$は、下記の通り展開される [7)]。

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) \propto p(\mathbf{z}_t \mid \mathbf{x}_t)\, p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1}) = p(\mathbf{z}_t \mid \mathbf{x}_t) \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1} \mid \mathbf{z}_{1:t-1})\, d\mathbf{x}_{t-1} \tag{1}$$

右辺における、$p(\mathbf{z}_t|\mathbf{x}_t)$、$p(\mathbf{x}_t|\mathbf{x}_{t-1})$がそれぞれ観測モデルとシステムモデルに対応する。

本研究では、観測ベクトルは、ステレオビデオカメラから得られる各画素 (i, j) の画素値 (r_{ij}, g_{ij}, b_{ij}) と対応する 3 次元座標 (X_{ij}, Y_{ij}, Z_{ij}) とする。すなわち、時刻 t における観測ベクトルは

$$\mathbf{z}_t = (r_{ijt}, g_{ijt}, b_{ijt}, X_{ijt}, Y_{ijt}, Z_{ijt}) \tag{2}$$

である。状態ベクトルを定義するために、追跡対象とする人物を楕円体モデルで表現する（図２）。楕円体の

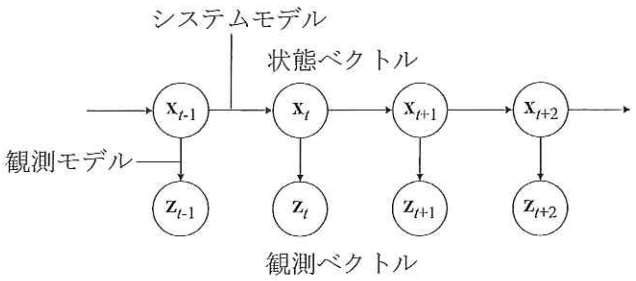


図１　一般状態空間モデル

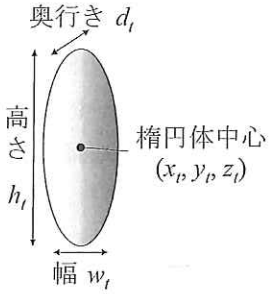


図２　楕円体モデル

中心を(x, y, z)、各軸の長さを (w, h, d)、更に、人物の移動速度を (v_x, v_y, v_z) とすれば、時刻 t における状態ベクトルは、

$$\mathbf{x}_t = (x_t, y_t, z_t, w_t, h_t, d_t, v_{xt}, v_{yt}, v_{zt}) \tag{3}$$

で表される。

（2）システムモデル

システムモデル $p(\mathbf{x}_t|\mathbf{x}_{t-1})$は、時刻 t の状態ベクトル $\mathbf{x}_t$ と 1 期前の状態ベクトル $\mathbf{x}_{t-1}$ との関係を確率的に表現したものである。このシステムモデルは、

$$\mathbf{x}_t = \mathbf{A}\,\mathbf{x}_{t-1} + \mathbf{w} \tag{4}$$

に従うものとする。ただし、

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

であり、Δt はビデオカメラの時間分解能、$\mathbf{w}$ は平均 0、分散 Σ のホワイトノイズである。すなわち、システムモデル $p(\mathbf{x}_t|\mathbf{x}_{t-1})$に従って、時刻 t-1 の状態ベクトル（人物の楕円体中心、楕円体形状、移動速度）から、時刻 t のそれらを予測することとなる。

（3）観測モデル

観測モデル $p(\mathbf{z}_t|\mathbf{x}_t)$は、観測ベクトルと状態ベクトルの関係を確率的に表現したものであり、尤度に相当する。本研究では、色情報を用いた観測モデルと距離情報を用いた観測モデルをそれぞれ定義し、それらを統合することにより、観測モデルを決定する。

① 色情報を用いた観測モデル

色情報を用いた観測モデルでは、時刻 t-1 における人物位置周辺の色分布と時刻 t において予測された人物位置周辺の色分布との類似度に従って確率を与えるように設定する。ステレオビデオカメラにより、各画素に対応する 3 次元座標が得られているため、楕円体モデル内部に存在する地点に対応する画素を判別することが可能である。楕円体モデルを画像に投影した場合の領域は楕円となり、この領域を Ω で表すこととする。

時刻 t-1 における楕円 Ω_{t-1} の内部の色分布と時刻 t において予測された楕円 Ω_t の内部の色分布を比較することにより、色情報を用いた観測モデルを計算する。追跡対象人物には、オクルージョンが発生していることから、色分布の比較には、カラーヒストグラムの相関係数である Bhattacharyya 係数[4), 13)]を採用する。いま、時刻 t-1 におけるヒストグラム分布を p_i（i は画素値に相当）、時刻 t の分布を q_i とし、画素値の取りうる値

が 1 から m とすると，Bhattacharyya 係数 ρ は，

$$\rho = \sum_{i=1}^{m} p_i q_i \tag{6}$$

となる。ここで、p_i、q_i は $\sum_{i=1}^{m} p_i = 1$、$\sum_{i=1}^{m} q_i = 1$ と正規化されているのものとする。この係数は、2 つの色ヒストグラム分布間の類似度が高いほど大きな値をとり、両者が完全に一致すると 1 を返す。よって、この係数を最大化する位置を探索することで追跡を実現するものである。この Bhattacharyya 係数を、各色 r、g、b に関して計算し、それらの値を乗じたものを、色情報を用いた観測モデル $p\left(\mathbf{z}_t^{color} \mid \mathbf{x}_t\right)$ とする。

② 距離情報を用いた観測モデル

距離情報を用いた観測モデルを計算するために、楕円体モデル内部に存在する地点の 3 次元座標と楕円体モデル表面の 3 次元座標を比較する。ステレオビデオカメラより得られた距離画像より、時刻 t における楕円体モデルの内部に存在する地点 P を抽出し、その 3 次元座標を$(X_t(P), Y_t(P), Z_t(P))$とする。この 3 次元座標と楕円体中心座標 (x_t, y_t, z_t) との距離を $d_t(P)$とする。一方、楕円体中心から点 P の方向への線分を延長し、楕円体モデル表面と交わる地点を求め、楕円体モデル中心からの距離を $\hat{d}_t(P)$ とする。この $d_t(P)$と $\hat{d}_t(P)$ との差分の総和が小さいほど、追跡対象の存在確率が高いといえる。そのため、楕円体内部に対応する点数を I とし、距離情報を用いた観測モデルを、

$$p\left(\mathbf{z}_t^{\text{range}} \mid \mathbf{x}_t\right) = a - b\left(\frac{1}{I}\sum_{P}\left(d_t(P) - \hat{d}_t(P)\right)^2\right) \tag{7}$$

と定義することとした。ここで、a、b は正の定数であり、これらの値は実験的に与えることとする。

③ 色情報と距離情報を用いた観測モデルの統合

色情報を用いた観測モデル $p\left(\mathbf{z}_t^{color} \mid \mathbf{x}_t\right)$ と距離情報を用いた観測モデル $p\left(\mathbf{z}_t^{\text{range}} \mid \mathbf{x}_t\right)$ が得られたならば、両者を統合することにより、最終的な観測モデル $p(\mathbf{z}_t|\mathbf{x}_t)$とする。統合の際には、確率分布であることより、下記の通りとする。

$$p\left(\mathbf{z}_t \mid \mathbf{x}_t\right) = p\left(\mathbf{z}_t^{color} \mid \mathbf{x}_t\right) \cdot p\left(\mathbf{z}_t^{\text{range}} \mid \mathbf{x}_t\right) \tag{8}$$

3．パーティクルフィルタ

本研究では、画像解析の分野においても、近年、特に着目されている手法の一つであるパーティクルフィルタ [8)-10)]により、時系列フィルタリングを実現する。

パーティクルフィルタは、確率分布をその分布からの多数の仮説（パーティクル）によって近似表現するものである [3),6)]。すなわち、時刻 t における事後確率分布 $p(\mathbf{x}_t|\mathbf{z}_t)$を、状態 $\mathbf{x}_t$ のパーティクル群 $\mathbf{s}_t^{(n)}$ と各パーテ

ィクルの重み $\pi_t^{(n)}$ の組によって、離散的に表現する（図３）。図中、パーティクルの大きさが重みに対応する。ここで、各パーティクル $\mathbf{s}_t^{(n)}$ の重み $\pi_t^{(n)}$ は、観測モデルから、$\pi_t^{(n)} = p\left(\mathbf{z}_t \mid \mathbf{x}_t = \mathbf{s}_t^{(n)}\right)$ として計算されるものである。時刻 t-1 での確率分布 $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$から、システムモデルと観測モデルに従って、時刻 t での確率分布を $p(\mathbf{x}_t|\mathbf{z}_{1:t})$を近似することにより、逐次、最適状態量を推定していくことになる。以下にその流れを示す（図４）。

① 時刻 t-1 における確率分布 $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$が、重み（観測モデル）を伴った N 個のパーティクル群 $\left\{\left(\mathbf{s}_{t-1}^{(n)}, \pi_{t-1}^{(n)}\right),\ n=1,\cdots,N\right\}$ によって近似されているとする。各パーティクルの重みの比に従ってリサンプリングすることにより、新しいパーティクル群 $\mathbf{s}'^{(n)}_{t-1}$ を生成する。

② 生成されたパーティクル群 $\mathbf{s}'^{(n)}_{t-1}$ を、システムモデル $p\left(\mathbf{x}_t \mid \mathbf{x}_{t-1} = \mathbf{s}'^{(n)}_{t-1}\right)$ に従って伝播させ、$p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$に相当する、時刻 t におけるパーティクル群 $\mathbf{s}_t^{(n)}$ を生成する。

③ 各パーティクルの重みを、観測モデル $\pi_t^{(n)} = p\left(\mathbf{z}_t \mid \mathbf{x}_t = \mathbf{s}_t^{(n)}\right)$ に従い計算する。ここで、$\sum_{n=1}^{N} \pi_t^{(n)} = 1$ となるよう正規化する。この結果、$p(\mathbf{x}_t|\mathbf{z}_{1:t})$を近似的に表現したパーティクル群 $\left\{\left(\mathbf{s}_t^{(n)}, \pi_t^{(n)}\right),\ n=1,\cdots,N\right\}$ が得られる。このパーティクル群の期待値を時刻 t における推定状態ベクトル $\mathbf{x}_t$ とする。

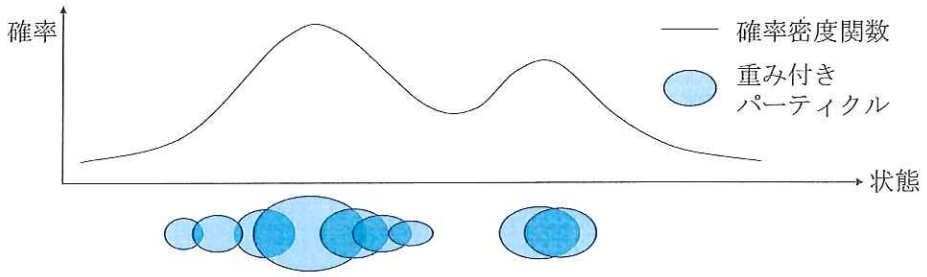


図３　重み付きパーティクル群による離散的な確率分布の表現

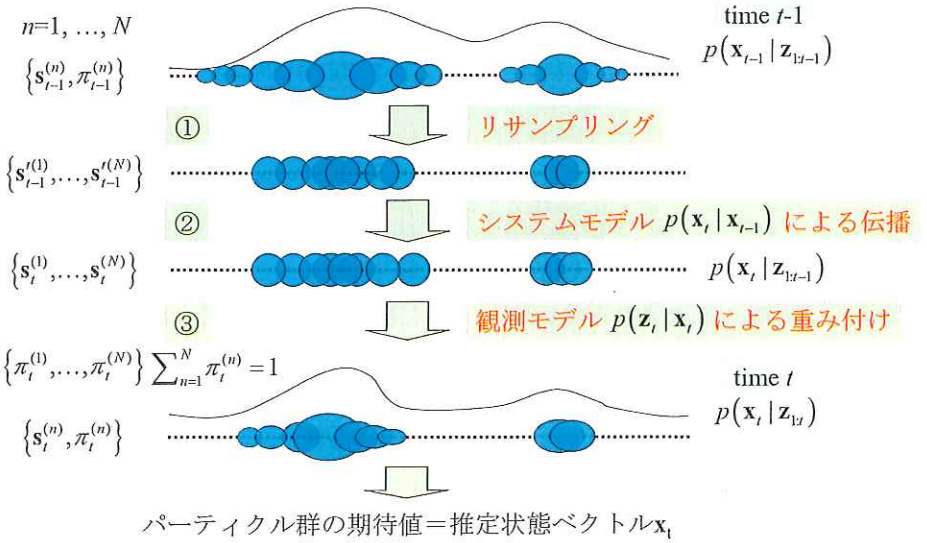


図４　パーティクルフィルタの流れ

4．適用

（1）撮影条件

提案した基礎手法を、ステレオビデオカメラで撮影した実動画像へ適用する。ステレオビデオカメラ（カメラ：SONY DFW-SX900（145 万画素））は事前にキャリブレーションを行い（表 1）、フレームレート 3.75 frames/s での撮影（約 30 秒間）を行った。ただし、カメラ座標は、レンズ中心からの左右方向に X 軸（撮影方向に向かって左が正）、上下方向に Y 軸（下が正）、奥行方向に Z 軸（奥が正）をとっている。本ステレオビデオカメラで撮影したカメラ 1 の画像例を図５に示す。画像中の楕円で囲んだ人物を追跡対象とする。画像内の人物は、画面奥から手前に、一般的な徒歩速度で移動している。

表 1　ステレオビデオカメラのキャリブレーション結果

	カメラ 1	カメラ 2
焦点距離 [mm]	8.261089	8.282733
1 次レンズ歪み係数 2 次レンズ歪み係数 [1/mm^2]	0.001691 -4.2E-05	0.00164 -5.3E-05
回転行列	$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$	$\begin{bmatrix}0.999495 & -0.00559 & 0.031263\\ 0.005142 & 0.999885 & 0.014326\\ -0.03134 & -0.01416 & 0.999408\end{bmatrix}$
並進ベクトル [mm]	$[0 \quad 0 \quad 0]^T$	$[-661.238 \quad 9.19235 \quad -1.41488]^T$
画像中心座標 [pix]	(676.8637, 489.5823)	(700.4433 491.2373)

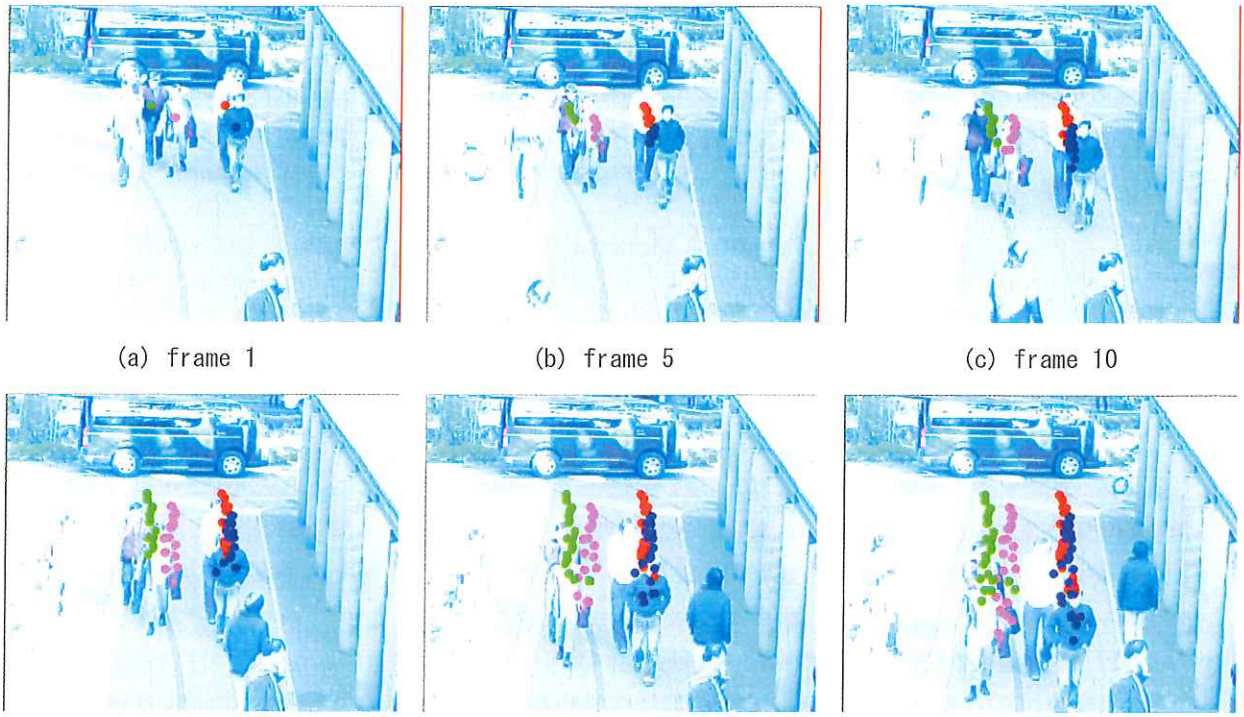
図５　ステレオビデオカメラによる取得画像例

（2）適用結果

提案手法において事前に設定する必要があるパラメータおよび初期値は、パーティクルフィルタにおけるパーティクル数 N、距離情報を用いた観測モデルにおけるパラメータ a、b、状態ベクトルの初期値である。パーティクル数は過去の事例に倣い[8]、N=100 と設定した。なお、パーティクル数を 100 から 1000 まで変化させても結果に大きな差異は見られなかったため、以後は、N=100 の場合のみを示すものとする。距離情報を用いた観測モデルにおいては、試行実験より a=b=1 と設定した。また、状態ベクトルの初期値として、各人物の位置は、マニュアルにより取得した。楕円体形状と移動速度に関しては、一般的な人物の大きさや徒

歩速度を考慮し、(w_0, h_0, d_0, v_{x0}, v_{y0}, v_{z0})=(0.5(m), 1.7(m), 0.3(m), 0(m/frame), 0.2(m/frame), 0.2(m/frame))とした。

提案手法を、ステレオビデオカメラにより撮影した実動画像へ適用した結果を図６に示す。図中の点は追跡された各人に対応する楕円体モデルの中心を表し、追跡結果を示すために、移動履歴を点列として表現している。人物の半分程度にオクルージョンが生じている場合にも、安定して追跡可能であることを確認した。距離情報により、追跡の頑健性が向上し、更には、3 次元空間における人物追跡を実現している。また、色情報を用いていることにより、軌跡が錯綜する場合にも、同一人物の追跡が可能となっており、提案手法が有効であることを確認した。なお、位置精度に関しては、目視によって追跡した結果と比較して、最大でも30cm 程度の差となった。今回は、個人の動きの変化がある程度の大きさまでならば、追跡が可能であるという結果が得られた。これは、今回使用したステレオビデオカメラのフレームレートが低いことによるものであり、高フレームレートのステレオビデオカメラを用いれば、大きな問題とはならないことが予想される。



(a) frame 1　(b) frame 5　(c) frame 10

(d) frame 15　(e) frame 20　(f) frame 25

図６　適用結果

５．おわりに

本研究では、人物相互のオクルージョンを含む複数人物の追跡に対し、ステレオビデオカメラを用いて、色情報と距離情報を統合した人物追跡手法を提案した。提案手法を実動画像へ適用し、人物の半分程度にオクルージョンが生じている場合にも、安定して追跡可能であることを確認した。本手法は、その他のセンサ（例えば、レーザスキャナ等）によるデータとの融合も可能である。以上の結果より、本研究で提案した手法は、複数人物追跡の基礎手法ではあるものの、今後の適用可能性を示したといえる。

本研究で提案した手法では、一定の成果を得たものの、あくまで基礎手法に過ぎない。今後、実際の駅構内や横断歩道といった、更に多数の人物が存在している公共空間における適用を考えた場合、例えば、以下の課題が挙げられる。

・人物の動きモデル（システムモデル）の拡張

提案手法では、人物の動きを単純なモデルとして表現している。より安定した追跡を実現するために、例えばセルオートマトンによる群集の動きのシミュレーションモデルを導入することによるシステムモデルの拡張が考えられる。

・人物の出現・消失を考慮した手法への発展

本研究では、初期フレームの人物の位置は手動で与えている。今後の自動観測を考えた場合には、画像における人物の出現・消失を考慮しなくてはならない。人物の出現・消失には、画像周縁部における距離情報の変化を計算することにより、その判定が可能であることが予想される。更には、人数をも状態量として含めた手法への発展も考えられる。

今後は、上記の手法拡張を行うことにより、より多数かつ高密度な複数人物追跡への対応が重要となる。

謝辞

本研究のアプローチに関して、清水英範教授（東京大学）より貴重なご意見を頂いた。また、ステレオビデオカメラによる撮影に関して、山口博義氏（コマツエンジニアリング（株））より多大なご協力を頂いた。ここに記して感謝の意を表する。

参考文献

1) 伊庭幸人, 種村正美, 大森裕浩, 和合肇, 佐藤整尚, 高橋明彦: 計算統計 II: マルコフ連鎖モンテカルロ法とその周辺, 岩波書店, 2005.
2) 北川源四郎: 時系列解析入門, 岩波書店, 2005.
3) 佐藤達也, 岩崎慎介, 小林貴訓, 佐藤洋一, 杉本晃宏: 環境モデルの導入による人物追跡の安定化, 電子情報通信学会論文誌, Vol.J88-D-II, No.8, pp.1592-1600, 2005.
4) 出口光一朗, 岡谷貴之, 中島平: 能動視覚による動的な空間知覚と立体形状認識機構の解明とその応用システムの構築, 情報学平成 13 年度成果報告会, A03-02, 2002.
5) 中村克行, 趙卉菁, 柴崎亮介, 坂本圭司, 大鋸朋生, 鈴川尚毅: 複数のレーザレンジスキャナを用いた歩行者トラッキングとその信頼性評価, 電子情報通信学会論文誌, Vol.J88-D-II, No.7, pp.1143-1152, 2005.
6) 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.
7) 渡部洋: ベイズ統計学入門, 福村出版, 1999.
8) Isard, M. and Blake, A.: Condensation – Conditional density propagation for visual tracking, International Journal of Computer Vision, Vol.29, No.1, pp.5-28, 1998.
9) Isard, M. and MacCormick, J.: BraMBLe: A bayesian multiple-blob tracker, Proceedings of International Conference on Computer Vision, pp.34-41, 2001.
10) Lu, W.-L., Okuma, K. and Littlea, J. J.: Tracking and recognizing actions of multiple hockey players using the boosted particle filter, Image and Vision Computing, Vol.27, Issues.1-2, pp.189-205, 2009.
11) Tao, H., Sawhney, H.S. and Kumar, R.: A sampling algorithm for tracking multiple objects, Proceedings of Workshop of Vision Algorithms with ICCV99, pp.53-68, 1999.
12) Zhao, T. and Nevatia, R.: Tracking multiple humans in complex situations, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.26, No.9, pp. 1208-1221, 2004.
13) Zhao, T. and Nevatia, R.: Tracking multiple humans in crowded environment. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 406-413, 2004.