

環境モデルの導入による人物追跡の安定化

鈴木 達也[†] 岩崎 慎介[†] 佐藤 洋一[†] 杉本 晃宏^{††}

[†] 東京大学生産技術研究所 〒 153-8505 東京都目黒区駒場 4-6-1

^{††} 国立情報学研究所 〒 101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: [†]{ttyszk,iwasaki,ysato}@iis.u-tokyo.ac.jp, ^{††}sugimoto@nii.ac.jp

あらし 分散配置されたセンサ群からの入力情報の統合により、3次元空間における実時間追跡を行う。具体的には、パーティクルフィルタによる人物追跡において、多視点画像の統合に加え、レンジセンサから得られる室内形状情報を利用し、室内空間における人物頭部の存在可能性の偏りを考慮した人物頭部の位置及び向きを推定を行う。提案手法により、室内空間の環境モデルを考慮した人物頭部の安定した実時間追跡が可能になる。実環境下での実験によって本手法の有効性を確認した。

キーワード 人物追跡, パーティクルフィルタ, 情報統合, 環境モデル

Stability Improvement in Tracking People by Introducing an Environment Model

Tatsuya SUZUKI[†], Shinsuke IWASAKI[†], Yoichi SATO[†], and Akihiro SUGIMOTO^{††}

[†] Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, JAPAN

^{††} National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN

E-mail: [†]{ttyszk,iwasaki,ysato}@iis.u-tokyo.ac.jp, ^{††}sugimoto@nii.ac.jp

Abstract We propose a method for tracking people in a cluttered indoor environment by integrating information from distributed sensors. Our method estimates the location and direction of a user's head by fusing multiple visual cues from multiple sensors based on particle filtering. Our method also utilizes a prior knowledge about the environment obtained from a laser range sensor for taking into account unevenly distributed probability of a user's position. The use of the environment model as well as integration of multiple visual cues from distributed sensors contributes to increasing tracking performance. We show the results of preliminary experiments that demonstrate the effectiveness of our proposed method.

Key words tracking people, particle filter, sensor fusion, environment model

1. はじめに

移動する対象、特に人物を検出し追跡する技術は、コンピュータビジョンの分野で最も重要な課題の1つである。人物を検出し、その動きを視覚的に認識することは、監視システムや遠隔講義などの自動撮影システムをはじめとする多くの応用システムにおいて重要な課題となっているからである。

人物追跡についてはこれまで多くの研究が行われてきた。追跡にあたり、不完全な観測値しか得られない状況で目標物体を追跡する場合、時系列フィルタリングが有効であることが知られている。時系列フィルタリングは、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手法である。その中でも、複雑背景の下、観測値が非ガウス型になる状況においても頑健な人物追跡を実現する時系列フィルタリングとして、パーティクルフィルタ [3], [4] が提案され、その有効性が報告されている。パーティクルフィルタは、ベイジアンフィルタ、逐次モンテカルロフィルタ、Condensation と呼ばれ、追跡対象を状態量と尤度を持つ多数の仮説群によって離散的な確率密度として表現し、確率モデルを用いて伝播させることで、ノイズや環境変動に対してロバストな追跡を実現する手法である。

一方、このような追跡の枠組みの中で性能の向上を図るには、観測による人物の検出の精度の向上が求められ、様々な手法が

提案されてきた。まず、単一の固定カメラを用いた人物頭部の評価について、人物頭部を楕円と仮定し、人物頭部のカラーヒストグラム [7]、人物頭部輪郭の輝度変化を利用する手法が提案されてきた [1], [13]。また、照明変化の影響を受けにくい距離画像を用いた人物検出手法も提案されている [2], [8]。さらに、これらの複数の観測情報の統合により観測の精度を向上させる手法も提案されている [5]。複数の視点の観測情報を用い、観測不能な領域を削減し、オクルージョンに対してロバストな追跡を実現する手法としては、多視点画像から構成される視体積交差 [6] に基づいたアプローチ [11]、多視点画像の情報を 2 次元平面で融合する手法 [12], [14] が提案されている。

しかし、実際の室内空間での 3 次元空間における人物頭部の追跡は、これまでの単純な手法はそのままでは必ずしも有効ではない。複数カメラを用いた 3 次元空間追跡を実現するうえで、人物頭部が必ずしもカメラ方向を向いていない場合も考慮に入れ人物頭部らしさの評価を行う必要がある。また、実際の環境において、照明変化が大きく、人物の肌色や髪色が床面や壁面と区別できないような複雑背景下でも安定した追跡を行うためには、環境に関する情報の積極的な利用も有効であると考えられる。

そこで、本稿では、ステレオビジョンカメラ及びレンジセンサを用い、多視点画像の統合に加え、レンジセンサから得られる室内形状情報を利用し、室内空間における人物頭部の存在可能性の偏りを考慮した 3 次元空間実時間追跡手法を提案する。

ステレオビジョンカメラからは色情報及び距離情報 (図 1) が得られる。色情報は、各画素がその画素に投影された 3 次元空間内の点の RGB 値 (3 次元ベクトルとして表現される) を持つ。距離情報は、各画素がその画素に投影された 3 次元空間内の点のカメラ座標系における 3 次元座標値を持つ。一方、レンジセンサからは壁や机といった環境の 3 次元形状情報 (図 2) が得られる。



図 1 ステレオビジョンカメラから得られる色情報及び距離情報



図 2 レンジセンサから得られる室内形状

提案手法は、具体的には次のようになる。追跡対象を人物頭部とし、向きをもった楕円体としてモデル化する。追跡に用いる

時系列フィルタリングとしては、パーティクルフィルタを用いる。観測にあたっては、室内の天井四隅にそれぞれ室内中心方向を向いたステレオビジョンカメラを配置し (図 3)、色情報と距離情報を取得し、それぞれで人物頭部らしさを評価する。そして、これら多視点画像から得られる情報の統合により、人物頭部の 3 次元空間位置、及び、人物頭部の向きの評価を行う。そのうえで、レンジセンサによって得られた室内形状から、壁や机の配置による室内空間中の人物頭部の存在可能性を考慮した環境モデルを評価に加えることにより、より安定した追跡を実現する。以上の手法により、人物の追跡を行い、実環境下での実験によって本手法の有効性を確認する。



図 3 センサ配置

2. パーティクルフィルタ

本稿では、文献 [3] で述べられているパーティクルフィルタを用いる。ここではパーティクルフィルタのアルゴリズムについて説明する。

2.1 時系列フィルタリング

ある時刻 t における追跡対象の状態量を \mathbf{x}_t 、画像から得られる観測結果を \mathbf{z}_t とおく。また、時刻 t までに得られる観測結果を $\mathbf{Z}_t = (\mathbf{z}_1, \dots, \mathbf{z}_t)$ とする。このとき、時刻 t までの観測結果が得られた時の状態 \mathbf{x}_t の確率密度関数 $P(\mathbf{x}_t | \mathbf{Z}_t)$ を推定する。

追跡対象の、時刻 $t-1$ における確率密度関数 $P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$ と、時刻 $t-1$ から t への動きモデル $P(\mathbf{x}_t | \mathbf{x}_{t-1})$ が与えられ、時刻 t における事前確率 $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$ は、マルコフ過程を仮定することにより次のようになる。

$$P(\mathbf{x}_t | \mathbf{Z}_{t-1}) = \int P(\mathbf{x}_t | \mathbf{x}_{t-1})P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})d\mathbf{x}_{t-1}. \quad (1)$$

このとき、時刻 t における尤度 $P(\mathbf{z}_t | \mathbf{x}_t)$ を画像から推定すると、時刻 t における確率密度関数 $P(\mathbf{x}_t | \mathbf{Z}_t)$ は、ベイズの法則に従い、次のようになる。

$$P(\mathbf{x}_t | \mathbf{Z}_t) \propto P(\mathbf{z}_t | \mathbf{x}_t)P(\mathbf{x}_t | \mathbf{Z}_{t-1}). \quad (2)$$

2.2 重みつきサンプリング

パーティクルフィルタでは、ある時刻 t における確率密度関数 $P(\mathbf{x}_t | \mathbf{Z}_t)$ を、状態 \mathbf{x}_t の仮説群 $\{s_t^{(1)}, \dots, s_t^{(N)}\}$ と各仮説の重み $\{\pi_t^{(1)}, \dots, \pi_t^{(N)}\}$ の組によって離散的に表現する。ここで、時刻 t における n 番目の仮説の状態量を $s_t^{(n)}$ とし、重みは $\pi_t^{(n)} = P(\mathbf{z}_t | \mathbf{x}_t = s_t^{(n)})$ により評価する。

2.3 パーティクルフィルタによる追跡

仮説群に適用されるプロセスは、次の3つの部分から構成される。このプロセスの繰り返しによって追跡が実現される。

(1) 時刻 $t-1$ において、観測 \mathbf{Z}_{t-1} が得られたときの状態量 \mathbf{x}_{t-1} の分布 $P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$ が、 N 個の重みつき仮説群 $\{(s_{t-1}^{(n)}, \pi_{t-1}^{(n)}), n = 1, \dots, N\}$ で表されているとき、各仮説の重み $\{\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}\}$ の比に従い、仮説群 $\{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ を選択する。

(2) 選択された仮説群を、動きモデル $P(\mathbf{x}_t | \mathbf{x}_{t-1} = s_{t-1}^{(n)})$ に従い伝播し、 $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$ に相当する時刻 t における N 個の仮説群 $s_t^{(n)}$ を生成する。

(3) 重み $\pi_t^{(n)}$ を画像から推定することで、新しいサンプル $s_t^{(n)}$ の重み $\pi_t^{(n)} = P(\mathbf{z}_t | \mathbf{x}_t = s_t^{(n)})$ を求める。ただし、重み $\pi_t^{(n)}$ が $\sum_{n=1}^N \pi_t^{(n)} = 1$ となるように正規化を行う。その結果、時刻 t における $P(\mathbf{x}_t | \mathbf{Z}_t)$ の近似表現である $\{(s_t^{(n)}, \pi_t^{(n)}), n = 1, \dots, N\}$ を得る。また、追跡対象の最適な状態量推定として、仮説群の期待値を用いる。

3. 人物頭部の位置及び向き の推定

3.1 人物頭部モデル

人物頭部のモデルとして楕円体を仮定する。まず、室内空間に3次元世界座標 XYZ をとる。このとき、床面が XY 平面と一致し、高さ方向が Z 軸となるようにする。人物頭部の形状は不変とし、位置を楕円体の中心座標 (x, y, z) で表す。人物頭部の向きは、上下方向にはあまり動かないと仮定すると、 X 軸を基準とした Z 軸回りの回転 θ のみで表される。この様子を図4に示す。

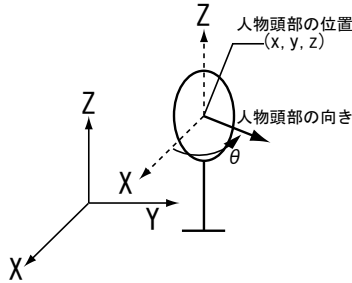


図4 人物頭部モデル

以上により、人物頭部は4次元の状態量 $\mathbf{s} = (x, y, z, \theta)$ で表される。なお、人物頭部の状態量について、時刻 t における n 番目の仮説を、 $s_t^{(n)} = (x_t^{(n)}, y_t^{(n)}, z_t^{(n)}, \theta_t^{(n)})$ と表すことにする。

ここで、時刻 t における n 番目の仮説を表す楕円体 $\Gamma_t^{(n)}$ を i 番目のカメラの画像に投影したときに得られる領域を $\Omega_{i,t}^{(n)}$ とする。投影関数を F_i とすると、次のようになる。

$$\Omega_{i,t}^{(n)} = F_i(\Gamma_t^{(n)}). \quad (3)$$

人物頭部を楕円体でモデル化しているため、 $\Omega_{i,t}^{(n)}$ は楕円となる。なお、領域 $\Omega_{i,t}^{(n)}$ 内の総画素数を $|\Omega_{i,t}^{(n)}|$ で表す。

3.2 色情報の利用による人物頭部らしさの評価

i 番目のカメラのカラー画像から仮説 $s_t^{(n)}$ の色情報に基づく重み $\pi_{i,t}^{color,(n)}$ を求める。ここで、本稿での重みとは、人物頭部らしさをいう。

背景や髪色の個人差の影響を減らすため、評価には肌色のみ

を用いる。色空間としては、HSV 色空間を用いる。これは、多くの研究でこの色空間が肌色領域の抽出に最も適した色空間のひとつであることが示されているためである [9], [10]。しかし、ここでは、照明変化の影響を減らすため V を無視し、肌色領域の判定に H と S からなる2次元の色空間を用いる。そのうえで、人間の肌色の学習データに基づき、あらかじめ肌色領域を定義する。そして、画素の HS 色空間での色度とその領域に含まれれば、その画素は肌色であると判断する。

重みの評価には、楕円 $\Omega_{i,t}^{(n)}$ 内部の肌色画素の割合とあらかじめ得ておいた人物頭部の肌色画素の割合との類似度を用いる。観測によって得られる i 番目のカラー画像において、楕円 $\Omega_{i,t}^{(n)}$ 内部の肌色領域に含まれる画素数を $c_{i,t}^{(n)}$ 、その割合を $\bar{c}_{i,t}^{(n)}$ とすると、次式が成立する。

$$\bar{c}_{i,t}^{(n)} = \frac{1}{|\Omega_{i,t}^{(n)}|} c_{i,t}^{(n)}. \quad (4)$$

また、追跡対象の人物に対して、あらかじめ得ておいた人物頭部の肌色画素の割合を $\hat{c}_{i,t}^{(n)}$ とする。人物頭部がカメラ中心を向いているほど肌色領域は大きく、カメラ中心に対して背を向いているほど肌色領域は小さい。このとき、重み $\pi_{i,t}^{color,(n)}$ は、これらの差分が少ないほど高く、次のような評価関数により与えることができる。

$$\pi_{i,t}^{color,(n)} = a_i^{color} - b_i^{color} \left| \bar{c}_{i,t}^{(n)} - \hat{c}_{i,t}^{(n)} \right|. \quad (5)$$

ここで、 $a_i^{color} (> 0)$ 、 $b_i^{color} (> 0)$ は定数である。なお、実験では、これらの値を経験的に与え、仮説による予測と観測の類似度が高いほど高い重みを与えた。

3.3 距離情報の利用による人物頭部らしさの評価

i 番目のカメラの距離画像から仮説 $s_t^{(n)}$ の距離情報に基づく重み $\pi_{i,t}^{depth,(n)}$ を求める。

観測によって得られる、 i 番目のカメラの距離画像に対して、楕円 $\Omega_{i,t}^{(n)}$ 内部の画素 p のカメラ座標を $\tilde{\mathbf{w}}_{i,t}^{(n)}(p)$ 、楕円体の中心 $\mathbf{v}_t^{(n)} = (x_t^{(n)}, y_t^{(n)}, z_t^{(n)})$ を i 番目のカメラのカメラ座標系へ投影した結果の座標を $\tilde{\mathbf{v}}_{i,t}^{(n)}$ 、 $\tilde{\mathbf{w}}_{i,t}^{(n)}(p)$ と $\tilde{\mathbf{v}}_{i,t}^{(n)}$ との距離を $d_{i,t}^{(n)}(p)$ とすると、次式が成立する。

$$d_{i,t}^{(n)}(p) = \left| \tilde{\mathbf{w}}_{i,t}^{(n)}(p) - \tilde{\mathbf{v}}_{i,t}^{(n)} \right|. \quad (6)$$

画素 p に対応する人物頭部楕円体上の点の楕円体中心からの距離は既知であるので、これを $\hat{d}_{i,t}^{(n)}(p)$ とする。このとき、重み $\pi_{i,t}^{depth,(n)}$ は、各画素におけるこれらの差分の総和が少ないほど高く、次のような評価関数により与えることができる。

$$\pi_{i,t}^{depth,(n)} = a_i^{depth} - b_i^{depth} \left(\frac{1}{|\Omega_{i,t}^{(n)}|} \sum_{p \in \Omega_{i,t}^{(n)}} \left| d_{i,t}^{(n)}(p) - \hat{d}_{i,t}^{(n)}(p) \right| \right). \quad (7)$$

ここで、 $a_i^{depth} (> 0)$ 、 $b_i^{depth} (> 0)$ は定数である。なお、実験では、色情報の場合と同様、これらの値を経験的に与え、仮説による予測と観測の類似度が高いほど高い重みを与えた。

3.4 環境モデルの導入

本稿ではさらに、環境モデルの導入により追跡の安定化を図る。

レンジセンサからあらかじめ得ておいた室内の3次元形状に基づき、壁や机の配置による室内空間における人物頭部の存在可能性を考える。こうして求められた空間中の人物頭部の存在可能性を環境モデルと呼ぶことにする。環境モデルを考慮することにより、パーティクルフィルタの仮説群の評価において、背景中に存在する物体内部のような人物頭部が存在し得ない領域、あるいは机や棚などの上部といった人物頭部の存在可能性の低い領域に生成される仮説の重みを抑えることが可能となる。その結果、人物頭部の存在可能性の低い領域からの仮説の生成が抑えられ、より安定した追跡が期待される。

環境モデルとして、次のような3つの領域からなるモデルを考える。

- 領域A：壁の外側及び机や棚など静的物体の内部。人物頭部は存在しない。
- 領域B：机や棚など静的物体の垂直方向。その他の空間に比べて人物頭部の存在可能性は低い。
- 領域C：低い領域、あるいは身長より上の領域。その他の空間に比べて人物頭部の存在可能性は低い。

また、このような環境モデルの例を図5に示す。

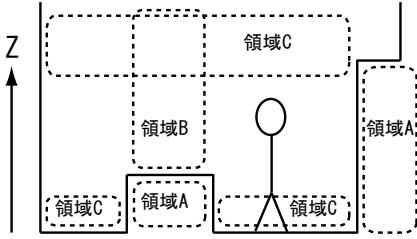


図5 環境モデル

これに基づき、状態量 $s_t^{(n)}$ における人物頭部の存在可能性に基づく重み $e_t^{(n)}$ を求める。各領域における人物頭部の存在可能性をそれぞれ $e_t^{A,(n)}$, $e_t^{B,(n)}$, $e_t^{C,(n)}$ 、どの領域にも属さない場合の人物頭部の存在可能性を α とし、領域が重なった場合はその最小値をとる。すなわち、

$$e_t^{(n)} = \min(e_t^{A,(n)}, e_t^{B,(n)}, e_t^{C,(n)}, \alpha). \quad (8)$$

なお、実験では、 $e_t^{A,(n)} = 0$, $e_t^{B,(n)} = 0.5$, $e_t^{C,(n)} = 0.2$, $\alpha = 1.0$ とした。

3.5 情報の統合

複数カメラからの色情報及び距離情報と環境モデルとの情報の統合を行う。

ここで、色情報は、人物頭部の対称性により単一のカメラでは左右対称な状態を判別できず、距離情報からは向きの判別はできないが追跡対象と背景を区別することができる。また、環境モデルを導入することにより、室内空間中の人物頭部の存在可能性の高い仮説に高い重みを与えることができる。従って、3次元での追跡を実現し、かつ、安定した追跡を実現するためには、これらの情報の統合が必要となる。

統合手法としては、実時間性を損なわないよう単純化し、仮説 $s_t^{(n)}$ における各カメラの色情報に基づく重み $\pi_{i,t}^{color,(n)}$ 及び距離情報に基づく重み $\pi_{i,t}^{depth,(n)}$ と環境モデルによって定義される人物頭部の存在可能性 $e_t^{(n)}$ との積をとる。すなわち、

$$\pi_t^{(n)} = e_t^{(n)} \prod_i \pi_{i,t}^{color,(n)} \prod_i \pi_{i,t}^{depth,(n)}. \quad (9)$$

このようにして、環境モデルを考慮した仮説の重みが得られる。

4. 実験結果

4.1 全体の処理の流れ

システムは1台のサーバPCと4台のクライアントPC (CPU Intel Pentium4 2.0GHz, Memory 1.0GByte) からなり、これらを通信速度が1GbpsのGigabit Ethernetで接続した。また、ステレオビジョンカメラとしてPoint Grey製のDigiclopsを、レンジセンサとしてSICK製のLMS200を用いた。これらのセンサはあらかじめキャリブレーションしておいた。これにより、各センサの観測をあらかじめ設定しておいた世界座標を介して連携することが可能である。クライアントPCにはそれぞれ1つのDigiclopsカメラが取り付けられており、色情報及び距離情報を取得できる。また、あらかじめレンジセンサから得られた室内形状から環境モデルを作成しておく。

処理の全体の流れは次のようになる。

- (1) サーバ、仮説群 $s_t^{(n)} = (x_t^{(n)}, y_t^{(n)}, z_t^{(n)}, \theta_t^{(n)})$ 生成。
- (2) サーバ、仮説群 $s_t^{(n)}$ 送信。
- (3) クライアント i 、仮説群 $s_t^{(n)}$ 受信。
- (4) クライアント i 、カラー情報及び距離情報取得。
- (5) クライアント i 、仮説群 $s_t^{(n)}$ の重みを評価。
 - 色情報から $s_t^{(n)}$ の重み $\pi_{i,t}^{color,(n)}$ を評価。
 - 距離情報から $s_t^{(n)}$ の重み $\pi_{i,t}^{depth,(n)}$ を評価。
- (6) クライアント i 、重み $(\pi_{i,t}^{color,(n)}, \pi_{i,t}^{depth,(n)})$ を送信。
- (7) サーバ、重み $(\pi_{i,t}^{color,(n)}, \pi_{i,t}^{depth,(n)})$ を受信。
- (8) サーバ、環境モデル $\pi_t^{e,(n)}$ を統合し、重み $\pi_t^{(n)}$ を評価。
- (9) サーバ、 $(s_t^{(n)}, \pi_t^{(n)})$ 取得。
 - 仮説群 $s_{t+1}^{(n)}$ 生成、(1)に戻る。
 - 期待値による状態量推定。

なお、追跡にあたり人物頭部の状態量の初期値は既知とし、仮説数 $N = 512$ とし人物頭部の追跡を行った。追跡時のフレームレートは約4fpsであった。

4.2 追跡結果

以上の準備の下、本稿で提案した手法を用いて、実環境下で人物頭部追跡実験を行った。

座っている状態から立ち上がり、室内を頭部の向きを変化させて歩いた場合について追跡実験を行った。具体的には、座った状態から(#30)、立ち上がり(#40)、向きを変え(#50)、まっすぐ歩く(#60)。そして再び方向を変え(#70)、歩きながら徐々に方向を変え(#80)、反転して元の方向に向かい(#90)、静止する(#100)というシーケンスに対して追跡実験を行った。実験で得られた画像の一部を図6に示す。各フレームには、情報統合後の人物頭部の状態量の期待値による推定結果を、カメラの方向を向いているほど輝度が高くなるような円として、カラー画像に重ねて表示した。なお、仮説の重みの評価には、色情報、距離情報に加え環境モデルを用いた。

図6から、仮説群の期待値によって、ほぼ人物頭部中心を推定できていることがわかる。これは、情報統合後の重みが高い仮説が、全体として人物頭部付近に集まっていることによるためであると考えられる。また、色情報による評価において、カメラの向きに応じた人物頭部の色モデルが有効であり、複数カメラでの観測結果の情報統合により、人物頭部の位置及び向きの

推定が有効に働いていると考えられる。

また、提案手法による追跡の精度を定量的に調べるために、画像中の人物頭部を手動で求め、複数画像から逆投影して求めた3次元座標値を真の位置と見なし、推定結果と比較した。図7に、推定結果とそれに対応する人物頭部の真の位置の3次元空間上及びXY平面上での軌跡を示す。また、Z軸方向及びXY平面上での誤差の平均及び標準偏差を表1に示す。

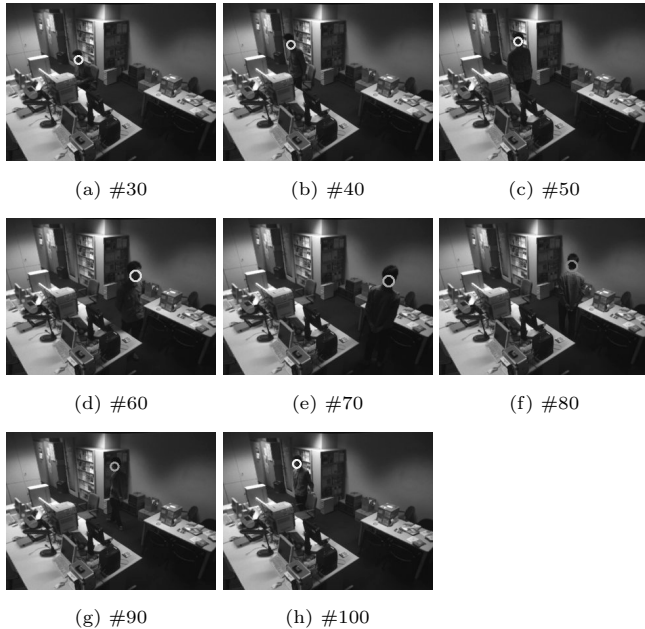


図6 人物頭部の追跡結果

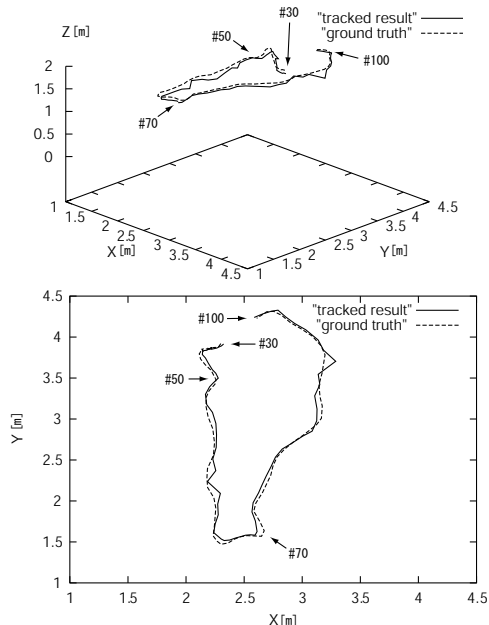


図7 人物頭部の追跡結果の軌跡

ここで、Z軸方向の検出誤差は6cm程度であり、下方に検出された。これは、色情報による評価が首も検出していることと、距離情報による評価が服も検出しているためと推定される。一方、XY平面上での検出誤差は4cm程度であり、十分正確に追跡できているといえる。

表1 人物頭部の追跡における検出誤差

	平均 [cm]	標準偏差 [cm]
Z軸方向	6.02	2.62
XY平面	4.34	2.70

4.3 色情報及び距離情報の統合の有効性

同じ画像シーケンスに対して、色情報、距離情報を統合することの有効性を確認する実験を行った。ただし、パーティクルフィルタにおける重みの計算には環境モデルを統合した。

フレーム(#40)での色情報と距離情報の観測結果を図8に示す。各フレームには、パーティクルフィルタによって生成された仮説を各カメラの画像へ投影し、それらの色情報及び距離情報からの評価結果に対応する点を、取得したカラー画像に重ねて表示した。輝度は観測結果から求められた重みを表し、輝度が高いほど重みが大きいことを表す。



図8 色情報と距離情報の観測結果

色情報の評価については、人物頭部の向きがカメラの方向に近い場合は、色情報モデルとして肌色領域が大きいモデルが適用されるため、人物頭部の中心に近いほど高い重みをもつことがわかる。しかし、背景に肌色に近い色をもつ物体があると、その部分に投影された仮説の重みが大きくなっていることがわかる。

一方、距離情報の評価については、人物頭部だけでなく、人物の体の部分などカメラからみて人物頭部と近い位置にある仮説は大きな重みをもつが、背後の物体に対しては重みは低くなっている。

このとき、真値に近い仮説ほど、色情報及び距離情報を用いた評価による重みが共に高く、その結果、統合後の重みが大きいといえる。

4.4 環境モデルの有効性

同じ画像シーケンスに対して、環境モデルの導入の有効性を確認するために、環境モデルを利用しない場合と利用した場合についての比較を行った。

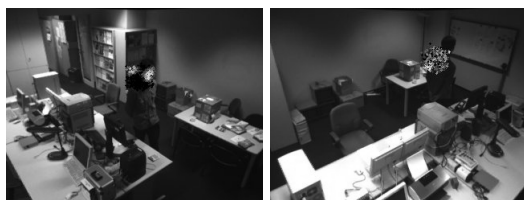
環境モデルを利用しなかった場合、(#60)の時点で、追跡対象が4つのカメラ全ての視野範囲内にあるにもかかわらず仮説が複数のカメラの視野範囲外に発散し、状態量推定に失敗した。そこで、この直前のフレームにおけるパーティクルフィルタの状態を色情報の評価結果と共に図9に表示した。また、比較のため、環境モデルを利用した場合の結果を図10に表示した。

環境モデルを利用しなかった場合、図9から、カメラ1(#58)において、背後の肌色に近い物体について色評価による誤検出が発生しており、この結果、仮説が人物頭部以外の領域に引き寄せられると共に、仮説の向きについても誤検出が起きていることがわかる。この様子を別のカメラ2(#58)から見ると、

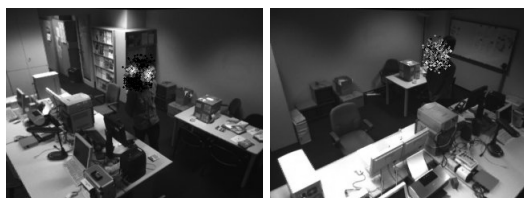
人物頭部には、仮説群がほとんど発生していないことがわかる。そのため、環境モデルを利用しない場合では追跡に失敗しているといえる。

一方、環境モデルを利用した場合、図 10 から、カメラ 2 (#58) において、人物頭部にも仮説が発生していることがわかる。机などの物体や高さの考慮という環境モデルの効果により、誤検出による仮説の生成が抑制された結果、人物頭部の存在可能性の高い部分に効率的に仮説が発生し、追跡を続けられたといえる。

以上のように、環境モデルの導入により、人物追跡の安定性が向上したといえる。



(a) #58 (カメラ 1) (b) #58 (カメラ 2)
図 9 環境モデルを利用しない場合の色情報の観測



(a) #58 (カメラ 1) (b) #58 (カメラ 2)
図 10 環境モデルを利用した場合の色情報の観測

5. おわりに

本稿では、複数センサからの入力情報の統合による、人物頭部の向きを考慮した 3 次元空間実時間追跡手法を提案した。

パーティクルフィルタを用いた人物頭部の追跡において、複数のステレオビジョンカメラから得られる色情報及び距離情報による人物頭部らしさの評価に加え、レンジセンサから得られる環境モデルを統合した。これにより、室内空間における人物頭部の存在可能性の偏りを考慮したより安定した人物頭部の位置及び向きの推定を実現することができた。

実環境下での人物頭部の追跡実験により、複雑背景の室内空間において、立ち上がる動作や任意の方向への歩行など、人物の自然な動作について、人物頭部の各カメラへの向きが変わっても人物頭部の向きを推定できており、追跡に成功していることが示された。これは、色情報と距離情報を統合することで、背景に肌色に近い色をもつ物体がある場合でも、各評価での誤検出の影響が少なくなり、真値に近い仮説ほど全体として大きい重みを得ることができたことによる。また、人物頭部の存在可能性を利用した環境モデルを統合することにより、机上の物体などの誤検出の影響を減らし、より安定した追跡を実現できることが示された。

今後は、オクルージョンによって追跡対象が遮蔽された場合を区別するなど情報の信頼性を考慮した情報統合、また、長期の観測から人物の行動履歴を取得し、人物の通りやすい経路や通

ることが少ない経路などを考慮し、環境モデルに反映させる手法を検討していく予定である。

謝辞 本研究の一部は、文部科学省科学研究費補助金特定領域 (C)(2) 「人間の意図・行動理解に基づく柔軟なヒューマン・マシン・インタラクションの実現」(課題番号 13224051) の助成により行った。

文 献

- [1] S. Birchfield: Elliptical Head Tracking Using Intensity Gradients and Color Histograms, Proc. the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '98), pp.232-237, June 1998.
- [2] I. Haritaoglu, D. Harwood and L. S. Davis: A Real-Time System for Detecting and Tracking People in 2 1/2 D, Proc. 5th European Conference on Computer Vision (ECCV '98), pp.877-892, 1998.
- [3] M. Isard and A. Blake: Condensation - Conditional Density Propagation for Visual Tracking, International Journal of Computer Vision, vol.29, no.1, pp.5-28, 1998.
- [4] M. Isard and A. Blake: ICondensation: Unifying Low-Level and High-Level Tracking in a Stochastic Framework, Proc. 5th European Conference on Computer Vision (ECCV '98), vol.1, pp.893-908, 1998.
- [5] G. Loy, L. Fletcher, N. Apostoloff and A. Zelinsky: An Adaptive Fusion Architecture for Target Tracking, Proc. 5th IEEE International Conference on Automatic Face and Gesture Recognition (FG '02), pp.261-265, May 2002.
- [6] W. N. Martin and J. K. Aggarwal: Volumetric Descriptions of Objects from Multiple Views, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-5, no.2, pp.150-158, March 1983.
- [7] M. Swain and D. Ballard: Color Indexing, International Journal of Computer Vision, vol.7, no.1, pp.11-32, 1991.
- [8] L. Taycher and T. Darrell: Range Segmentation Using Visibility Constraints, International Journal of Computer Vision, 47(1-3), pp.89-98, 2002.
- [9] J. Terrillon, A. Pilpre, Y. Niwa and K. Yamamoto: Analysis of Human Skin Color Images for a Large Set of Color Space and for Different Camera Systems, Proc. IAPR Workshop on Machine Vision Applications (MVA '02) pp.20-25, 2002.
- [10] B. D. Zarit, D. J. Super and F. K. H. Quek: Comparison of Five Color Models in Skin Pixel Classification, Proc. International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pp.58-63, September 1999.
- [11] 浮田宗伯, 松山隆司: 能動視覚エージェント群による複数対象の実時間協調追跡, 情報処理学会 CVIM 研究会論文誌, vol.43, no.SIG11, pp.64-79, December 2002.
- [12] 大塚和弘, 武川直樹: 多視点観測に基づく複数物体の相互オクルージョン解析と逐次状態推定, 情報処理学会 CVIM 研究会論文誌, vol.44, no.SIG17, 2003.
- [13] 杉本晃宏, 谷内清剛, 松山隆司: 確信度付き仮説群の相互作用に基づく複数対象追跡, 情報処理学会論文誌: コンピュータビジョンとイメージメディア研究会論文誌, vol.43 no.SIG04, June 2002.
- [14] 中島平, 浜崎浩二, 岡谷貴之, 出口光一郎: CONDENSATION を用いた多視点画像の融合による複数人物の追跡, MIRU 2002, vol.2, pp.317-322, 2002.