

## 環境モデルの導入による人物追跡の安定化

鈴木 達也<sup>†</sup>      岩崎 慎介<sup>†</sup>      小林 貴訓<sup>†a)</sup>      佐藤 洋一<sup>†b)</sup>  
杉本 晃宏<sup>†c)</sup>

Incorporating Environment Models for Improving Vision-Based Tracking of People  
Tatsuya SUZUKI<sup>†</sup>, Shinsuke IWASAKI<sup>†</sup>, Yoshinori KOBAYASHI<sup>†a)</sup>, Yoichi SATO<sup>†b)</sup>,  
and Akihiro SUGIMOTO<sup>†c)</sup>

あらまし 分散配置されたセンサ群からの入力情報の統合により、三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

## 1. ま え が き

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一つである。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、複雑背景のもと、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ゆが度をも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搬させることで、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]。更に、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二次元平面で融合する手法 [12], [14] が提案されている。

しかし、実際の室内などの三次元空間における人物頭部の追跡では、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三次元空間追跡を実現する上で、人物頭部が必ずしもカメラ

<sup>†</sup> 東京大学生産技術研究所, 東京都  
Institute of Industrial Science, The University of Tokyo,  
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

<sup>††</sup> 国立情報学研究所, 東京都  
National Institute of Informatics, 2-1-2 Hitotsubashi,  
Chiyoda-ku, Tokyo, 101-8430 Japan

a) E-mail: yosinori@iis.u-tokyo.ac.jp

b) E-mail: ysato@iis.u-tokyo.ac.jp

c) E-mail: sugimoto@nii.ac.jp

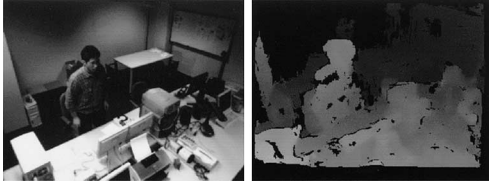


図 1 ステレオビジョンカメラから得られる色情報及び距離情報

Fig.1 Color and depth images captured from a stereo-vision camera system.



図 2 レンジセンサから得られる室内形状

Fig.2 Three dimensional shape of an environment measured by using a range finder.

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、本論文では、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三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三次元空間内の点の RGB 値(三次元ベクトルとして表現される)をもつ。距離情報は、各画素がその画素に投影された三次元空間内の点のカメラ座標系における三次元座標値をもつ。一方、レンジセンサからは壁や机といった環境の三次元形状情報(図 2) が得られる。

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



図 3 センサ配置

Fig.3 The placement of the sensors used in our system.

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、人物頭部の三次元空間位置、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。

## 2. パーティクルフィルタ

本論文では、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  $x_t$ 、画像から得られる観測結果を  $z_t$  とおく。また、時刻  $t$  までに得られる観測結果を  $Z_t = (z_1, \dots, z_t)$  とする。このとき、時刻  $t$  までの観測結果が得られたときの状態  $x_t$  の確率密度関数  $P(x_t | Z_t)$  を推定する。

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

$$P(x_t | Z_{t-1}) = \int P(x_t | x_{t-1})P(x_{t-1} | Z_{t-1})dx_{t-1}. \quad (1)$$

このとき、時刻  $t$  におけるゆう度  $P(\mathbf{z}_t | \mathbf{x}_t)$  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 重み付きサンプリング

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、次の三つの部分から構成される。このプロセスの繰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、 $N$  個の重み付き仮説群  $\{(s_{t-1}^{(n)}, \pi_{t-1}^{(n)}), n = 1, \dots, N\}$  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  $P(\mathbf{x}_t | \mathbf{x}_{t-1} = s_{t-1}^{(n)})$  に従い伝搬し、 $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$  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、室内空間に三次元世界座標  $XYZ$  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  $Z$  軸周りの回転  $\theta$  のみで表される。この様子を図 4 に示す。

以上により、人物頭部は四次元の状態量  $\mathbf{s} = (x, y, z, \theta)$  で表される。なお、人物頭部の状態量に

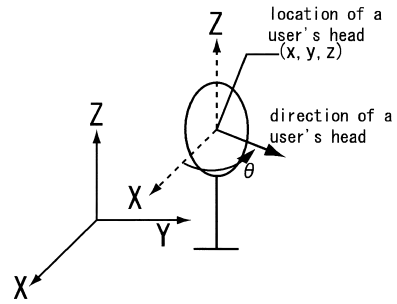


図 4 人物頭部モデル  
Fig. 4 The model of a user's head.

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ので、 $\Omega_{i,t}^{(n)}$  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、本論文での重みとは、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一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  $S$  からなる二次元の色空間を用いる。その上で、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  $HS$  色空間での色度が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、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けているほど肌色領域は小さいことから、人物が正面を向いた場合に最大をとり、正反対を向いた場合に最小となる状態量  $\theta$  に応じて決定される値  $\hat{c}_{i,t}^{(n)}$  をあらかじめ得ておく。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  $\tilde{w}_{i,t}^{(n)}(p)$ 、楕円体の中心  $v_t^{(n)} = (x_t^{(n)}, y_t^{(n)}, z_t^{(n)})$  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  $\tilde{v}_{i,t}^{(n)}$ 、 $\tilde{w}_{i,t}^{(n)}(p)$  と  $\tilde{v}_{i,t}^{(n)}$  との距離を  $d_{i,t}^{(n)}(p)$  とすると、次式が成立する。

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

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

本論文では更に、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三次元形状に基づき、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、次のような三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。

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  $\alpha$  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。

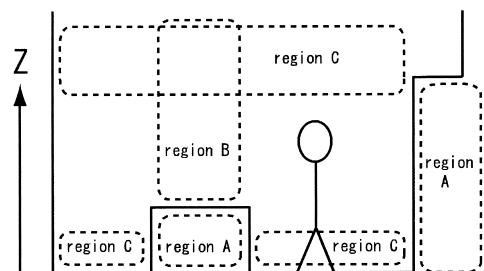


図 5 環境モデル

Fig. 5 The model of an environment.

### 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、色情報では、人物頭部の向きの判別が可能であるが、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区別できない。また、距離情報では、追跡対象と背景を区別することができるが、距離情報による評価では、仮説が人物頭部位置にある場合だけでなく、例えば壁面近くに仮説が存在する場合でも高い評価値をとる場合がある。更に、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。したがって、三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、これらの情報を統合し、最終的に正しい位置にある仮説が高い評価値となる必要がある。

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 2.0 GHz, Memory 1.0 GByte) からなり、これらを通信速度が 1 Gbps 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 にはそれぞれ一つの 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  $\pi_t^{(n)}$  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  $N = 512$  として人物頭部の追跡を行った。追跡時のフレームレートは約 4 fps であった。

### 4.2 追跡結果

以上の準備のもと、本論文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。

座った状態から (#30), 立上り (#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), 反転してもとの方向に向かい (#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分かる。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三次元座標値を真の位置とみなし、

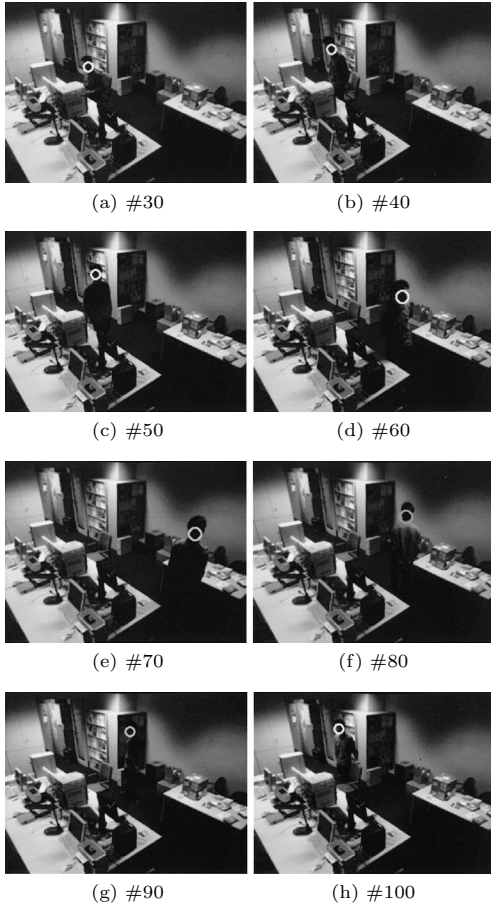


図 6 人物頭部の追跡結果 1  
Fig. 6 Tracking results 1.

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三次元空間上及び  $XY$  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。

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。

また、本論文で提案した手法の頑健性の確認のため、全く同じシステムにおいて先の実験とは異なる人物による頭部追跡実験を行った。人物は室内に置かれたテーブルの周りに沿って頭部の向きを変化させながら歩いている。先の実験と同様に、得られた 60 フレー

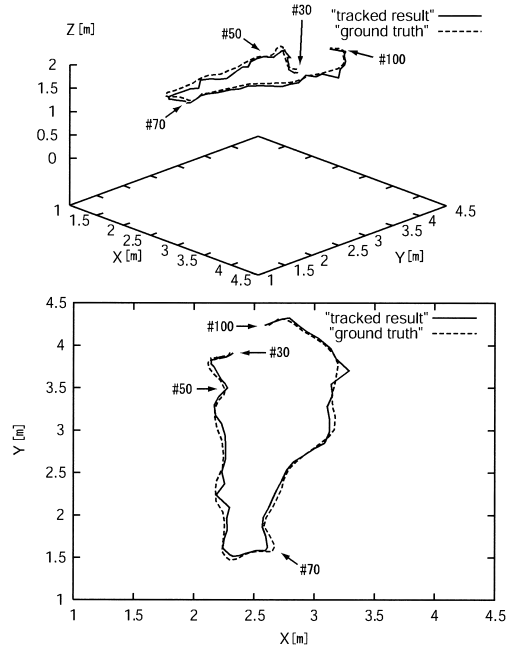


図 7 人物頭部の追跡結果の軌跡 1  
Fig. 7 Trajectory of a user's head position 1.

表 1 人物頭部の追跡における検出誤差 1  
Table 1 Tracking error 1.

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

ムの画像の一部を図 8 に示す。

図 9 に、推定結果とそれに対応する人物頭部の真の位置の  $XY$  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 2 に示す。 $XY$  平面上及び  $Z$  軸方向での検出誤差はそれぞれ 2cm, 3cm 程度であり、先の実験と同様に正確な追跡結果を得ることができた。

更に、パーティクルフィルタが仮説群により確率密度を近似していることに起因する推定位置のばらつきの評価のため、あらかじめ保存した入力画像列について検出アルゴリズムを 30 回適用した。各フレームに対して推定位置の標準偏差を求め、その標準偏差を全フレームに対して平均した値を表 3 に示す。また、標準偏差が最大となったフレームにおける値も併記する。

ばらつきが最も大きかった  $XY$  平面におけるフレーム (#37) においても標準偏差は 0.7cm 程度であり、検出誤差に比べて十分小さいといえる。よって、十分な数の仮説により安定に人物位置が推定されているこ

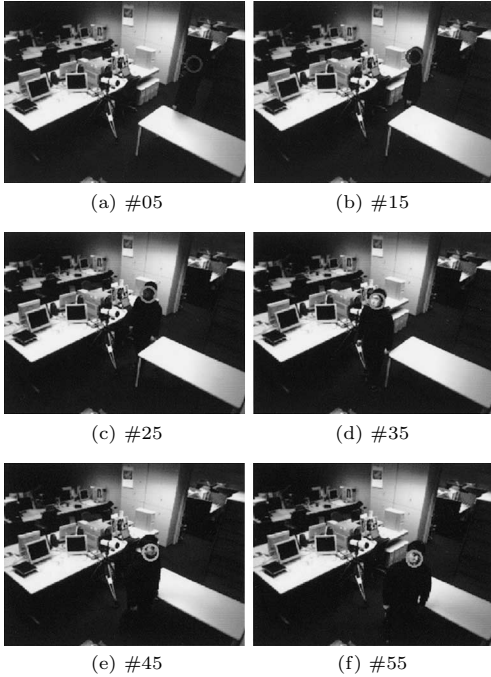


図 8 人物頭部の追跡結果 2  
Fig. 8 Tracking results 2.

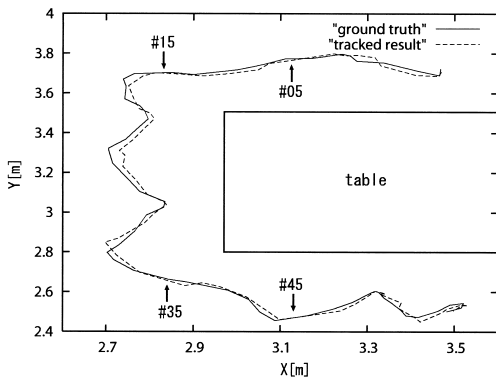


図 9 人物頭部の追跡結果の軌跡 2  
Fig. 9 Trajectory of a user's head position 2.

表 2 人物頭部の追跡における検出誤差 2  
Table 2 Tracking error 2.

	平均 [cm]	標準偏差 [cm]
Z 軸方向	3.25	1.77
XY 平面	2.17	1.20

とが分かる。

#### 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

表 3 同一入力画像列における各フレームの推定頭部位置のばらつき

Table 3 Variation of a user's head position in every frame by using same image sequences.

	平均標準偏差 [cm]	最大標準偏差 [cm] (フレーム)
Z 軸方向	0.27	0.50 (#36)
XY 平面	0.36	0.68 (#37)

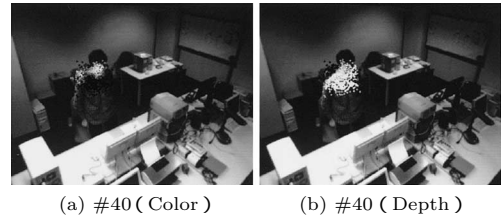


図 10 色情報と距離情報の観測結果  
Fig. 10 Evaluation result based on both color and depth information.

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 10 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。

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分かる。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分かる。

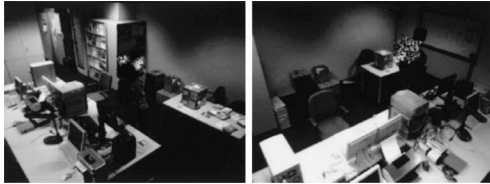
一方、距離情報の評価については、人物頭部だけでなく、人物の体の部分などカメラから見て人物頭部と近い位置にある仮説は大きな重みをもつが、背後の物体に対しては重みは低くなっている。

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ともに高く、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、



(a) #58 (Camera 1) (b) #58 (Camera 2)

図 11 環境モデルを利用しない場合の色情報の観測

Fig. 11 Evaluation result without an environment model



(a) #58 (Camera 1) (b) #58 (Camera 2)

図 12 環境モデルを利用した場合の色情報の観測

Fig. 12 Evaluation result with an environment model

追跡対象が四つのカメラすべ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ともに図 11 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 12 に表示した。

環境モデルを利用しなかった場合、図 11 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ともに、仮説の向きについても誤検出が起きていることが分かる。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分かる。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 12 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分かる。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。

## 5. む す び

本論文では、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三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へい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「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. 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," Int. J. Comput. Vis., 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 Trans. Pattern Anal. Mach. Intell., vol.PAMI-5, no.2, pp.150-158, March 1983.
- [7] M. Swain and D. Ballard, "Color indexing," Int. J. Comput. Vis., vol.7, no.1, pp.11-32, 1991.
- [8] L. Taycher and T. Darrell, "Range segmentation using visibility constraints," Int. J. Comput. Vis., vol.47, no.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, Sept. 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," 情処学 CVIM 論文誌, vol.43, no.SIG11, pp.64-79, Dec. 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," 情処学 CVIM 論文誌, vol.44, no.SIG17, pp.109-125, 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," 情処学コンピュータビジョンとイメージメディア論文誌, vol.43, no.SIG04, pp.69-84, 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.

(平成 16 年 10 月 7 日受付, 17 年 2 月 7 日再受付)



鈴木 達也

2002 東大・工・電子情報卒。2004 同大学院情報理工学系研究科電子情報学専攻修士課程了。室内環境における人物追跡に関する研究に従事。同年(株)日立製作所に入社。



岩崎 慎介

2005 東京大学大学院情報理工学系研究科電子情報学専攻修士課程了。室内環境における人物追跡に関する研究に従事。



小林 貴訓

2000 電気通信大学大学院情報システム学研究科情報システム運用学専攻修士課程了。同年三菱電機設計システム技術センターに入社。現在、東京大学大学院情報理工学系研究科電子情報学専攻博士課程在学中。



佐藤 洋一 (正員)

1997 カーネギーメロン大学計算機科学部ロボティクス学科博士課程了。同年より東京大学生産技術研究所研究機関研究員, 講師を経て, 現在同研究所助教授。Ph.D.in Robotics。コンピュータビジョン, ヒューマン・コンピュータ・インタラクション, コンピュータグラフィックスに関する研究に従事。平 11 山下記念研究賞, MIRU2000 最優秀論文賞, 平 11 年度日本バーチャルリアリティ学会論文誌論文賞, IEEE VR2001 Honorable Mention for the Outstanding Paper Award, MIRU2004 優秀論文賞等を受賞。情報処理学会, 日本バーチャルリアリティ学会, ACM, IEEE 各会員。



杉本 晃宏 (正員)

1987 東大・工・計数工卒。1989 同大学院工学系研究科修士課程了(数理工学専攻)。同年, 日立製作所基礎研究所に入社。1991~1995 ATR に出向。1999 京都大学講師。2002 国立情報学研究所助教授, 現在に至る。2005 総合研究大学院大学複合科学研究科助教授併任。博士(工学)。視覚情報処理や離散システム・アルゴリズムなどに興味をもち, 数理的手法に基づいたコンピュータビジョンの研究に従事。2001 情報処理学会論文賞受賞。