

楕円体モデルを用いたリアルタイム人物追跡

加藤 博一[†] 中澤 篤志^{††} 井口 征士^{††}

本論文では、屋内既知環境におけるリアルタイムの人物追跡の手法を提案する。カメラパラメータと環境情報を用いて人物の見え方をシミュレーションし、入力画像とマッチングを行い、類似度から人物の移動位置を求める。人物を1つの楕円体で表現することにより、計算コストを低減させ、汎用パーソナルコンピュータでのリアルタイム処理を可能とした。また、他のシステムとの統合を容易にするために、3次元実環境中に固定された世界座標系で人物位置の計測・管理を行っている。静止人物に対する精度評価実験では、頭部と足下がともに隠れることがない限り、誤差の平均値で約0.2m以下という結果が得られた。2人の人物の追跡実験においては、人物同士の見かけ上の重なりが発生する場合でも、良好な結果が得られた。しかし、影や写り込みの影響で追跡に失敗する場合もあった。今後は複数システムの統合問題に関して検討する予定である。

Realtime Human Tracking Using Ellipsoid Model

HIROKAZU KATO,[†] ATSUSHI NAKAZAWA^{††} and SEIJI INOKUCHI^{††}

This paper describes a real-time human tracking method using ellipsoid model. Several images that simulate real image are generated by using a camera parameter and environmental information. These images are compared with an input image and human positions are determined by those similarities. A human is simply represented by an ellipsoid to achieve a real-time operation on a standard personal computer. To integrate some tracking systems using this method can be achieved easily because human positions are measured and managed on a world coordinates. In accuracy evaluations for a standing human, the average errors were within about 0.2 meter except the case that both head and foot were occluded. In the tracking experiment for 2 humans, the algorithm was able to work successfully. However, tracking sometimes failed because of shadows or reflections. We will consider the integration of several systems.

1. はじめに

人物の位置情報検出は、防犯システム、店舗内人流調査などのほかに、最近ではヒューマンインタフェース応用においても注目されている。位置情報を利用することによりユーザに適切な情報提供を行うことを目指した情報サービスシステムの研究¹⁾や、福祉目的では、徘徊老人の介護問題対策においても、人物追跡システムに対する期待が高まっている²⁾。このようなシステムにおける位置検出の方法として、アクティブバッジという装着型のセンサを用いる方法³⁾や、GPSやPHSを利用した位置検出システム⁴⁾などが提案さ

れている。しかし、屋内環境での利用を考えると、器具装着の必要性の点や、精度の点から、視覚システムによる受動的で高精度な位置検出システムの登場も期待されている。

視覚システムによって人物追跡を実現する研究はこれまでも数多く提案されている。差分画像から人物領域の特徴点を抽出し、フレームごとに対応づけを行うことによってトラッキングを実現するもの⁵⁾や、オプティカルフローによる画像領域のクラスタリングを用いるもの⁶⁾、両眼視差画像と運動視差画像から生成されるヒストグラム上でクラスタリングを行うもの⁷⁾などがあげられる。しかし画像特徴の対応づけやクラスタリングは、対応づけの困難さや計算量の増加という問題を持っている。またオプティカルフローや視差画像の計算をリアルタイムで実行するためには、高速なハードウェアが要求される。

一方、Hogg⁸⁾は計算機内の人物モデルを画像と比較することでトラッキングを行う方法を提案している。

[†] 広島市立大学情報科学部情報機械システム工学科
Department of Information Machines and Interfaces,
Faculty of Information Sciences, Hiroshima City
University

^{††} 大阪大学大学院基礎工学研究科システム人間系専攻
Department of Systems and Human Science, Graduate
School of Engineering Science, Osaka University

このようなモデルベーストラッキングのアプローチを採用した研究も多くあり、2次元画像領域としてモデルを表現し画像との類似度の比較を行う方法⁹⁾や、人物領域の輪郭と人物モデルを比較することで人物姿勢を求めるもの¹⁰⁾などがある。しかし、前者では人物位置計測は2次元画像平面内で行われており、実世界の3次元座標系での位置計測は行われていない。後者は3次元座標系での人物位置の計測を行うものであるが、複数人物の追跡には対応していない。ただし、これらの研究は人物の位置検出のみではなく、姿勢やジェスチャーの認識を狙っているものである。我々が検討しているのは、屋内既知環境における人物位置検出のための視覚システムである。その中で、

- (1) 低い計算コスト。
- (2) 複数人の追跡。
- (3) 複数の視覚システムや異なるセンサシステムとの情報交換・統合の容易性。

を主な課題としている。そこで、モデルに基づくトップダウン的手法を用いることで計算コストの抑制とロバスト性の向上を図る。具体的には、屋内既知環境を想定し環境の構造的知識を積極的に用いる。また、人物の姿勢情報を必要としないため、人物全体を1つの楕円体で表現し、処理を簡素化する。さらにこの方法を拡張することで複数人の追跡を可能とする。また、計測された3次元位置情報をカメラに固定されたカメラ座標系ではなく床面に固定された世界座標系で表現することにより、他のセンサシステムとの情報交換・統合を容易にする。

2. 処理の概要

図1に処理の概要を示す。カメラから入力された画像に対し、事前に取得しておいた背景画像の差分処理、および固定閾値による2値化処理を行う。これにより生成された画像を実差分画像と呼ぶ。実差分画像には背景画像と閾値以上の濃度差のある領域が抽出される。理想的にはこの領域が人物領域となるが、実際には影や他の移動物などの領域の混入や人物と背景の濃度値の差が少ない部分における領域の欠損がノイズとして生じる。

人物の位置情報は、世界座標系で表現される人物位置管理マップに登録されている。その情報と事前に準備した環境モデル・カメラパラメータを用いて、実際のカメラから観測され背景差分・2値化処理が施された実差分画像、つまり人物領域のみが抽出された画像のシミュレーションを行うことができる。これにより生成された画像をシミュレーション差分画像と呼ぶ。そ

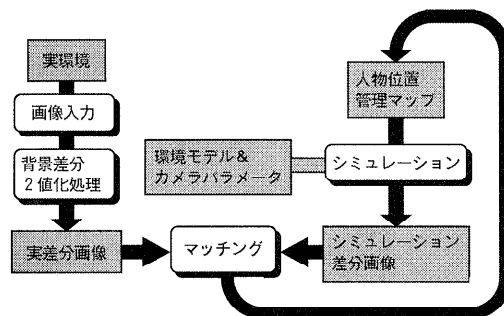


図1 処理の概要

Fig. 1 Our method is based on the matching of an input image and simulated images.

ここで、人物位置管理マップから読み出した人物位置をいろいろな方向に微小移動させ、複数枚のシミュレーション差分画像を生成する。これら複数のシミュレーション差分画像と実差分画像を比較し、実差分画像に最も類似したシミュレーション差分画像における人物位置を現在の人物位置として、人物位置管理マップの情報を更新する。この処理を繰り返す。

2.1 モデル表現

2.1.1 環境モデル

環境モデルは世界座標系で表現する。世界座標系は、 (X, Y, Z) の3次元直交座標系とし、 Z 軸を鉛直方向上向き、 $Z=0$ 平面を床面としている。環境モデルは2つの目的で利用され、処理の高速化を図るために、2つに分けて記述している。

(1) 人物移動可能領域の記述

壁や棚、パーティションなどにより、人物移動が制限される。そこで、人物の移動可能領域を明示することで、現実には移動不可能な位置に対するシミュレーションを防止する。

(2) 構造物の記述

シミュレーション差分画像を生成するとき、オクルージョン問題も考慮する必要がある。そこで、机、パーティションなど人物を隠蔽する可能性のある物体を3次元モデルとして記述しておき、オクルージョンを考慮したシミュレーションを行う。ただし、シミュレーション差分画像には人物領域のみを描画すればよく、これら構造物を描画する必要はない。つまり、人物領域を隠蔽しているかどうかという判定にのみ利用されるので、構造物の色やテクスチャー情報も必要なく、3次元形状情報のみが表現されていけばよい。そこで、物体表面形状を表現するサーフェイスモデルでこれら構造物の記述を行う。また形状の記述精度や記述すべき物体の範囲に関しては、人物自体も1つの楕

円体という近似を行っていることを考えると、物体の細かな凹凸を表現する必要はないと考え、1辺が数十cm程度の長方形パッチで構造物の概形を表現するにとどめている。

2.1.2 人物モデル（楕円体モデル）

シミュレーション差分画像の生成では、描画すべき人物像のモデルによって、その精度と計算コストが変化する。本手法では実差分画像との類似度を評価に用いるが、複数のシミュレーション差分画像間での相対的評価のために、生成画像の絶対的な精度は要求されない。しかし、計算コストに関してはリアルタイム処理を目的としているために可能な限り低く抑える必要がある。そこで人物を1つの楕円体で近似表現し描画することにした。

2.1.3 人物位置管理マップ

追跡中のすべての人物の位置情報を記述する。現在は世界座標系の $Z = 0$ 平面を床面としているので、世界座標系の残りの2成分である X, Y 座標値を記述している。またシミュレーション差分画像の生成で必要とされる楕円体モデルの高さと幅情報も各人物ごとに与える。この値は他のセンサから獲得するか事前知識として準備しておく必要があるが、現在は成人を仮定して、高さ165[cm]、幅40[cm]の固定値を用いている。

2.1.4 カメラパラメータ

カメラは、環境内の高い位置から人物が視野内に入るように固定する。環境モデルと同一の世界座標系でキャリブレーションされたカメラパラメータを用いる。このカメラパラメータは、20点以上の特徴点の世界座標系の3次元座標値とそれに対応する画像平面内の2次元座標値から、事前に求めておく。このデータは3行4列の行列として表現され、世界座標系の座標値 (X, Y, Z) とカメラ画像平面内の座標値 (x_c, y_c) の間に次式の関係を与える¹¹⁾。また、画像サイズもカメラパラメータとともに記述しておく。

$$\begin{bmatrix} hx_c \\ hy_c \\ h \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (1)$$

3. シミュレーション差分画像の生成

3.1 視野領域内人物の選択

本手法は複数の視覚システムや他の位置管理システムとの統合を前提としているために、人物位置管理マップには、カメラに写らない人物も登録されていると仮定する。そこでまずカメラに写る人物を選択する。

これは実際に画像内に描画される人物を事前に選択することで処理コストを削減するためである。

人物位置管理マップに登録されている人物 P の位置を (X_p, Y_p) 、その人物の高さを H_p とすると、式(1)に人物の最高点 (X_p, Y_p, H_p) と最低点 $(X_p, Y_p, 0)$ を代入することで、それらの画像平面内での座標値を求めることができる。この座標値のどちらかが画像サイズ内に含まれる場合にその人物を描画するものとする。

3.2 楕円パラメータの計算

シミュレーション差分画像は人物領域が画素値1で塗りつぶされた2値画像である。人物を3次元空間内の楕円体で近似しているため、その楕円体がカメラに写った場合の見え方をシミュレーションすることになる。今、人物の世界座標系での位置を (X_p, Y_p) 、その人物の幅と高さを $2R_a, 2R_b$ とすると人物を表現する楕円体の方程式は次のようになる。

$$\frac{(X - X_p)^2}{R_a^2} + \frac{(Y - Y_p)^2}{R_b^2} + \frac{(Z - R_b)^2}{R_b^2} = 1. \quad (2)$$

また、カメラパラメータの基本式(式(1))に対して形式的に4行目を追加し正方形列とすることで、その逆行列を求め、次の式を得ることができる。

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} hx_c \\ hy_c \\ h \\ 1 \end{bmatrix}. \quad (3)$$

この式の X, Y, Z を式(2)に代入することで、 h に関する2次方程式を得る。この解が重解の場合、カメラ画像平面内の座標値 (x_c, y_c) を通過する視線方程式が式(2)の楕円体と接することを意味する。そこで判別式 $D = 0$ とする方程式をつくることで、その式の x_c, y_c は、楕円体がカメラに写ったときの輪郭線を表現する。この式は最終的には、次の形式の楕円方程式にまとめることができる。

$$K_1 x_c^2 + K_2 y_c^2 + K_3 x_c y_c + K_4 x_c + K_5 y_c + K_6 = 0. \quad (4)$$

詳細な式は付録に示す。

3.3 奥行きマップによるオクルージョンへの対処

式(4)で表される楕円方程式は画像平面内での人物領域の輪郭線を表現しているため、この領域内部を画素値1で塗りつぶすことでシミュレーション差分画像が生成できる。しかし、環境中の構造物などによるオクルージョンを考慮する必要がある。これにはコンピュータグラフィックスの分野で隠面消去法として一般的な Z バッファ法を利用する。

環境モデルとして構造物の記述を行っているため、

その記述をもとに画像内の各画素ごとに構造物までの距離を計算し、奥行きマップ、いわゆる Z バッファを事前に生成しておく。この距離値と描画される楕円体モデルの距離値を各画素ごとに比較し、楕円体モデルの方がカメラに近い画素のみ描画を行う。

Z バッファの作成は事前に行うことができるためにその計算速度は問題にはならないが、塗りつぶされる楕円内の各画素に対して距離値を求めるための計算コストが問題となる。そこで、楕円内全領域の奥行き値を楕円体中心の奥行き値で代表させることで計算コストを削減した。また、Z バッファの生成も含めた距離計算には、3次元距離ではなく、Z = 0 平面に射影した2次元平面内での距離値を利用することにした。

4. 比較処理

4.1 マッチングによる人物位置の決定

人物位置管理マップに記述されている人物位置を微小移動させ、3章で述べた方法により複数枚のシミュレーション差分画像を生成し、実画像との類似度から現時点での人物位置を決定する。カメラ視線方向を床面に投影し、その方向とそれに垂直な方向の2軸を考え、その各軸方向に、-50 [mm], 0 [mm], +50 [mm] の3段階に人物位置を変化させ、合計9枚のシミュレーション差分画像を生成する。1回の画像入力に対して、この9枚のシミュレーション差分画像とのマッチングによる位置更新を3回繰り返し、最後に求められた人物位置を人物位置管理マップに記録する。つまり1回の画像入力で、最大150 [mm] の移動に追従でき、10 frame/sec の処理速度が達成できれば、1500 mm/sec の移動速度まで対応可能になる。

4.2 3値画像によるマッチング

マッチングでは、以下の式で評価を行う。

$$E = \frac{\sum_{x,y} \{\text{sim}(x,y) \cdot \text{real}(x,y)\}}{\sum_{x,y} |\text{sim}(x,y)|} \quad (5)$$

$\text{sim}(x,y)$, $\text{real}(x,y)$ はそれぞれシミュレーション差分画像、実差分画像の座標値 (x,y) での画素値である。両画像とも画像サイズは同一であり、基本的には0, 1を値とする2値画像である。しかし、図2に示すようにシミュレーション画像が人物領域に完全に包含されるような場合、評価値は最大となりマッチングを誤る。また、影などの影響により2値画像に多くのノイズが発生した場合も、ノイズ領域を包含する楕円領域では評価値は高くなり、マッチングを誤る可能性が高くなる。このような影響を抑制するために輪郭の

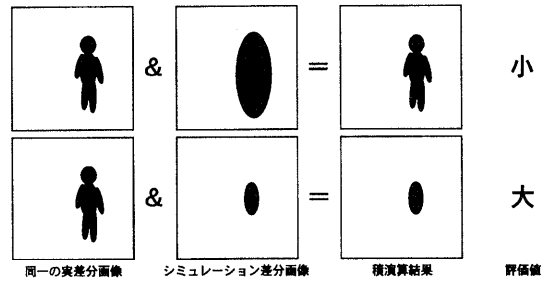


図2 類似度計算において問題となる状況

Fig. 2 These two cases result in unfavorable similarity values, when eq. 5 applies to binary images.

マッチング度合いを評価値に反映させることとした。

つまり、シミュレーション差分画像、実差分画像のそれぞれにおいて、1の値の領域に対して膨張処理を施し、膨張領域には-1の値を割り当てる。これらの画像で式(5)の計算を行うと、ともに-1という値を持つ画素では評価値を高めるが、1と-1とくいちがう画素では評価値を下げる効果を生み、輪郭のマッチング度合いを評価値に反映させることができる。6章の実験においては膨張幅を5画素としている。

また、画素値のいずれかが0の部分は評価式分子部の計算に不要なので、シミュレーション差分画像の描画領域情報から評価値計算領域を限定でき、ここでも計算コストを削減できる。

5. 拡張

以上に述べたことが基本的な処理方法であるが、ここでは複数人の追跡のための処理、および、画像視野外から進入してくる人物の追跡に関する処理について説明する。

5.1 複数人追跡への拡張

複数人の追跡において問題となるのは、人物同士が見かけ上で重なる場合である。本手法においても、各人物ごとに独立に処理を行っては問題にうまく対応できない。そこで、すべての人物に対する処理を一括して行う。図3にその手順を示す。1人に対して9通りの位置変化を与えるとすると、N人だと9^N通りの組合せが考えられる。しかし、計算コストの点で問題が生じるので、1人を変化させるときには他の人物の位置は固定しマッチングを行い、その人物についてのみ位置を更新する。すべての人物について、順次この処理を繰り返す。この方法だと、N人だと9Nの組み合わせとなる。

5.2 追跡モードと捕捉モード

複数システムや他のセンサ情報との統合を考えると、

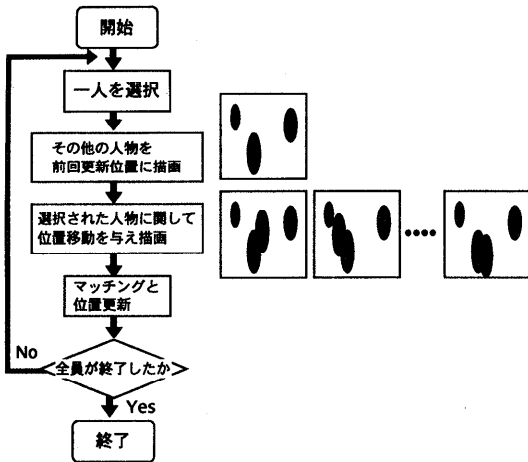


図3 複数人の場合の処理手順

Fig. 3 When plural persons appear in a image, tracking of each person is repeated in sequence.

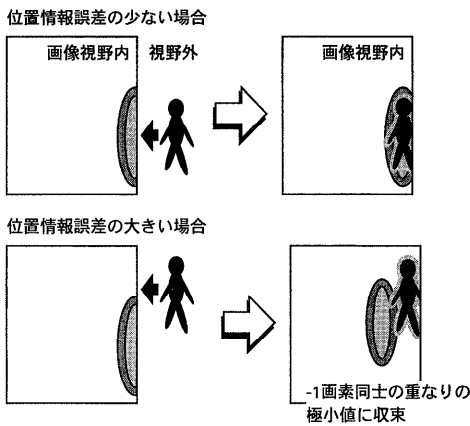


図4 収束が困難な状況

Fig. 4 When position information does not have high accuracy, the ellipse might stay away from a human region.

カメラ視野外から視野内に人物が出現する場合や視野内から視野外に消滅する場合に対応する必要がある。本手法においては、視野外から視野内に出現する人物の追跡を開始する場合に、観測データの誤差に起因する問題が生じる。これは、4章で示した3値画像のマッチングに関連し、図4に示すように、-1という画素値の効果で実画像上の人物領域にシミュレーション画像の人物領域が収束しない状況が生じるからである。そこで、視野外から視野内に出現する人物の初期補足の段階では、初期捕捉モードとして-1領域を設定しないマッチングを行う。そして式(5)の評価値が閾値0.7を超えた時点で通常の追跡モードに復帰するようにした。

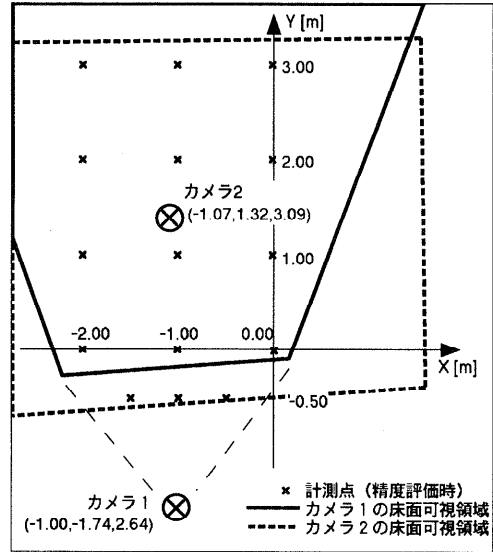


図5 カメラ配置

Fig. 5 2 cameras installed with different angle were used for experiments.

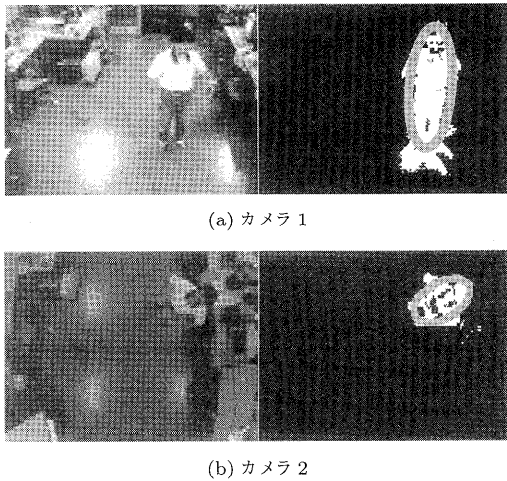
6. 実験と検討

6.1 実験環境

提案手法の人物静止状態での位置検出精度の評価、および人物移動時の通過軌跡の評価実験を行った。カメラ配置を図5に示す。水平床面上の適当な場所に世界座標原点(Z軸が鉛直方向上向き)を置き、その座標系内において、人物を斜め上方から見下ろす配置でカメラ1、真上から見下ろす配置でカメラ2を設置した。この2種類の配置のカメラを用いてそれぞれ実験を行い、カメラ配置による影響も考察する。カメラからのビデオ信号は、パーソナルコンピュータ(Intel PentiumII-333MHz)で縦横各120×160画素の大きさで取り込まれ処理される。

6.2 カメラのキャリブレーション

縦横各300mm、高さ1100mmの基準直方体を作成し、これを用いてカメラのキャリブレーションを行った。基準直方体を世界座標系での定点に置き、画像中の基準直方体のマーカ位置をマウスで指示するとともに、対応する世界座標系での座標値を入力する。基準直方体を移動させながら、この操作を繰り返し、30点のデータを用いて式(1)のカメラパラメータを最小2乗近似により求めた。得られたカメラパラメータより計算された各カメラの位置は、それぞれ(-1.00m, -1.74m, 2.64m), (-1.07mm, 1.32m, 3.09m)であった。また、カメラ2では広角レンズ使用による樽型歪みが観測されたために、その影響も調べた。キャリブレーションに用い



(a) カメラ 1

(b) カメラ 2

図 6 2つのカメラ画像(左は入力濃淡画像, 右は背景差分画像の画素値 1 の領域と選択されたシミュレーション画像の画素値 -1 の領域を重ね合わせたもの)

Fig. 6 Input gray scale images are left. Right images represent both regions where pixel value is 1 in a background subtraction image and where pixel value is -1 in a simulated image.

たデータの画像中の座標値に対応する世界座標系での床面位置をカメラパラメータを用いて逆算し, 真値と比較したところ, 画像中央部で 0.02 [m] 程度, 画像周辺部で 0.10 [m] 程度の誤差が確認できた。つまり, カメラ 2 による計測ではつねにこの程度の誤差が生じると考えられる。

6.3 計測精度評価

人物静止状態での位置検出精度を評価するため, 人物が定点に立ち止まっているときの計測誤差を評価した。図 5 に示すように計測点を取り, カメラ 1 では $(x, y) = (0\text{m}, -0.5\text{m}) \sim (-2\text{m}, 3\text{m})$ の範囲内にある計 15 点, カメラ 2 では $(x, y) = (0\text{m}, 0\text{m}) \sim (-2\text{m}, 3\text{m})$ の範囲内にある計 12 点で計測を行った。カメラ 1, カメラ 2 による画像を図 6 に示す。背景差分処理により多少の領域欠損をともなって人物領域の抽出は行われている。また, 影や写り込み部分も同時に抽出されていることが確認できる。また, 背景の単調さ・複雑さによる人物領域抽出性能の変化はなかったが, 単調な背景の上に同一の濃淡レベルの人物像が写る場合には大きな領域欠損が生じる場合もあった。

カメラ 2 では, 人物を上方向から垂直に見下ろす視線となるのでほとんどの場合画面中には人物の全身が写るが, カメラ 1 では人物を斜めから見下ろす画像が得られるため, 画像境界によって人物の上半身が写らない状態(図 7(a))や下半身が写らない状態(図 7(b))が生じる。そこでカメラ 1 では,



(a) 上半身が不可視

(b) 下半身が不可視

図 7 画像境界で部分的なオクルージョンが生じた画像

Fig. 7 Example images in which a part of human is masked by the border of the image.

表 1 計測誤差

Table 1 Average errors of measured human position.

	計測点数	誤差平均 [m]	X軸方向 [m]	Y軸方向 [m]
カメラ 1 (a)	9 点	0.11	0.05	0.09
(b)	3 点	0.12	0.04	0.11
(c)	3 点	0.06	0.05	0.03
カメラ 2	12 点	0.21	0.11	0.17

- (a) 人物の全身が写った場合(計測範囲 $(0\text{m}, 0\text{m}) \sim (-2\text{m}, 2\text{m})$): 計 9 点)
 (b) 人物の下半身のみが写った場合(計測範囲 $(0\text{m}, 3\text{m}) \sim (-2\text{m}, 3\text{m})$): 計 3 点)
 (c) 人物の上半身のみが写った場合(計測範囲 $(0.5\text{m}, -0.5\text{m}) \sim (-1.5\text{m}, -0.5\text{m})$): 計 3 点)
 の 3 タイプに分割して評価した。なお被験者は身長が 166 [cm], 胴体横幅最大値は 42 [cm] であった。楕円体モデルには高さ 165 [cm], 幅 40 [cm] という値を用いている。表 1 に計測誤差を示す。計測値と真値の距離の平均を誤差平均とした。また, XY 各軸方向に関する誤差平均も示す。

計測値の量子化誤差が 0.025 [m] ということを考慮すると, 全体的には良好な結果が得られたと思われる。カメラ 2 の結果の誤差が大きいのは, 楕円体モデルの影響のほかに, ある計測点における人物の影による比較的大きな誤差が原因となった。これは真上から見下ろすというカメラ配置固有の問題ではなく, 斜めからの場合でも照明との位置関係により起こりうる問題である。この点を考慮すると両カメラにおける計測精度の比較において画像の楕円体モデルによる違い以外の差は見られ

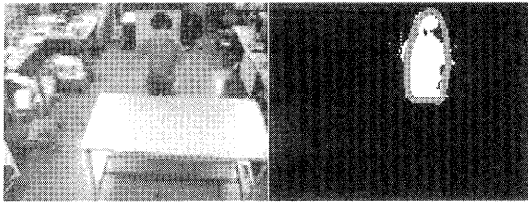


図8 構造物によるオクルージョンが生じた画像

Fig. 8 In this image, a part of the human is occluded by the desk.

表2 計測誤差 (構造物によるオクルージョンの影響)

Table 2 Average errors in case that a part of the human is occluded by the desk.

	計測点数	誤差平均 [m]	X軸方向 [m]	Y軸方向 [m]
カメラ1 (a)	9点	0.18	0.06	0.15
(b)	3点	1.31	0.18	1.29

なかった。また、カメラ1の視線は世界座標系においてはほぼY軸に沿ったものなので、X軸が画像面内方向、Y軸が奥行き方向に対応する。表中カメラ1(a)、(b)の結果から奥行き方向の誤差がわずかに大きなことが確認できるが、個別のデータからも奥行き方向に対する誤差が大きいたことが確認できた。これは斜めから見下ろすカメラ配置において、奥行き方向の人物移動に対する画像変化が面内方向に比べて小さいことに起因する。

次にカメラ1において障害物によるオクルージョンが生じた場合の計測精度を評価する。計測範囲は(0m, 0m)~(-2m, 3m)の12点とし、すべての計測点において人物の足下が隠れるように、高さ0.73mの机を移動させながら計測した(図8)。なお、この机の配置はそのつど環境モデルに記述した。計測結果を表2に示す。足下は隠れるが上半身が写っている(a)の場合では、Y軸方向の誤差が多少増加したのが分かる。(b)の場合では、足元と頭部の両方が隠れてしまい胴体の一部が写る画像が得られる。この場合、人物のY軸方向の変化に対しシミュレーション差分画像はほとんど変化がなく、表からもY軸方向に大きな誤差が生じているのが分かる。頭部と足元が同時に隠れる状況に対しては、本手法では計測不能領域として取り扱わなければならない。

6.4 複数人を対象とした追跡実験

複数人物を対象にした追跡実験として、6.3節と同様の2台のカメラによって、2人の人物が同時にカメラ視野内を移動する際の人物の通過軌跡を評価した。人物の歩行経路を図9に示す。人物1の被験者の身長は173[cm]、胴体横幅最大値は48[cm]で、人物2は6.3節の被験者と同一人物である。楕円体モデルの

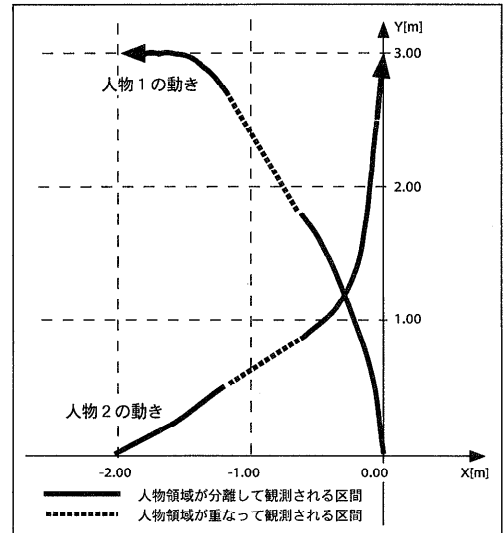


図9 人物の歩行経路

Fig. 9 In this experiment, 2 persons walked as these trajectories show.



図10 人物の重なり合い

Fig. 10 Tracking was done successfully for overlapped persons.

高さ・幅値は各160[cm]・40[cm]と変更していない。図中の破線で示した位置では、カメラ1の画像中で人物が重なって観測される(図10)。カメラ2では人物は重なって観測される箇所はない。計測された人物の歩行軌跡を図11に示す。

全体的評価として歩行追跡は両方のカメラで可能だった。カメラ1においては人物が重なって観測される部分においても良好な結果が得られた。カメラ2の結果には計測誤差が目立つ。これはカメラ2において、画像中の人物領域が比較的小さいにもかかわらず、床に映る人物の影が大きく写ったことが影響した。

6.5 検 討

本実験は特別な環境を設置することなく通常の屋内環境で行われており、影や写り込みが生じる環境にもかかわらず安定した追跡結果が得られた。処理速度は、1人の追跡の場合20~30 frame/sec、2人の場合5.6~10 frame/sec、3人の場合2.7~5.5 frame/secであった。1人の場合は小走りする人物の追跡も可能だった

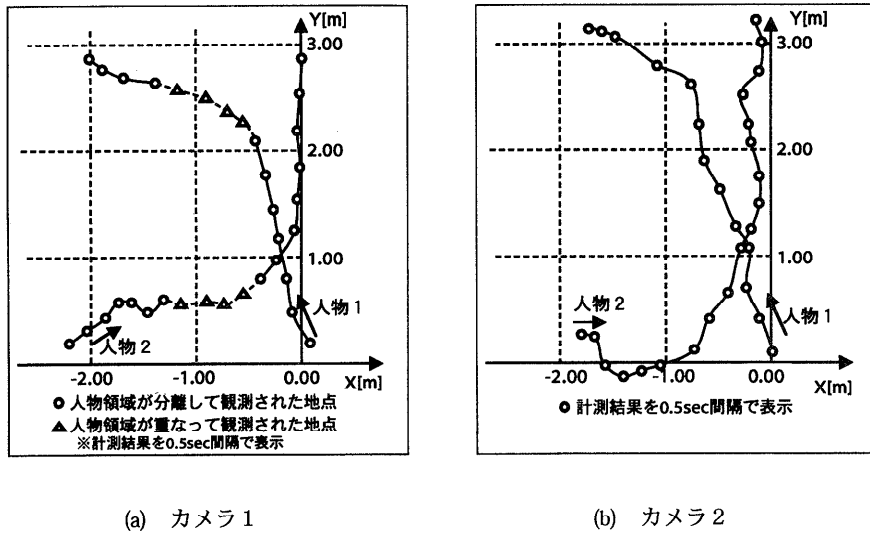


図 11 歩行経路の計測結果

Fig. 11 Measurement results from the use of each camera.

が、3人の場合ではゆっくりとした歩行でない場合には追跡に失敗する場合も見られた。本手法では、ノイズなどの影響による処理速度の変化はないが、シミュレーション画像の生成枚数を起因とする、追跡人数にはほぼ比例した処理時間を必要とする。このため、処理速度の観点からは、現状の処理装置では3人の追跡が限界となっている。しかし、視野の重なる複数のシステムに処理を分担させることにより、この問題も解決可能と思われる。

複数人物が画像上で重なりあう状況に関しても、本手法では良好な追跡結果を得ることができた。しかし、人物同士が見かけ上ではなく実際に非常に接近して交差するような場合には、追跡人物が入れ替わるという問題が生じる。現在は二値の背景差分画像のみを情報として用いているが、このような人物の入れ替わりが予想される場合には、色情報などを用いた人物識別を検証処理として追加し、人物の入れ替わりを修正する方法が考えられる。また、人物の密度が高い環境においては、真上から見下ろすカメラ配置を採用することで、人物の見かけ上の重なりを低減することも可能である。

計測精度に関しては、位置情報に基づく情報提供サービスなどへの応用を考えた場合には、十分な精度が得られたと考えられる。計測精度に影響する主な要因は、

- (1) 人物モデルと実際の人物の大きさの差異
- (2) 影・写り込み領域の発生・人物領域の欠損
- (3) 画像内の人物領域の大きさ

(4) 人物の大きな姿勢変化

の4点であった。(1)に関してはモデルの高さ情報の値が実際の人物の身長と異なる場合に定常的な誤差が観測されたが、5%程度の差であれば、その影響により追跡不能に陥ることはなかった。(2)に関しては、影や写り込みが人物領域以上に大きな領域として観測されるような場合や人物像の50%以上の領域が抽出できなかった場合に、大きな誤差が発生し、追跡不能に陥ることもあった。影・写り込みをシミュレーションし画像に反映させる方法も考えられるが、処理速度の増加が予想される。領域欠損に関してはカラー画像による背景差分処理の利用も考えられる。さらに、実応用においては、影や写り込みの発生を抑えるような環境に対する工夫や、それらの影響を受けにくいカメラ配置の検討が必要になるとと思われる。また、照明環境の変化や人物以外の物体移動に対処するためには、背景画像の更新処理も導入しなければならない。(3)に関しては、今回の実験において人物全身が写る場合の人物領域の画像面積に対する割合は3~8%程度であったが、その範囲における計測精度は6.3節で示したとおりである。しかし、廊下などに俯角を浅くカメラを設置した場合、遠くの人物まで観測することは可能である。この場合は人物移動に対する画像変化の減少に加え、画像内の人物領域が小さくなることによるノイズ領域による影響が大きくなる。画像面積に対する人物領域の割合が1%以下の場合では、事実上、追跡は不可能であった。(4)に関しては、座り込むといった姿勢変化によって大きな位置誤差が生じたが、追跡に

は大きな影響を与えなかった。

人物追跡という実応用を考えた場合、複数台のシステムを統合した広い環境での追跡処理が要求される。本手法では人物管理マップ上に人物位置を世界座標系で表現しているの、これを複数システムで共有することにより、統合は容易に行えると考えられる。しかし、複数のカメラに同一人物が写るような場合の効果的な処理の分担方法などを検討する必要もある。また、現在は水平な床面を仮定しているが、追跡処理の中からその制約を取り除くことは容易であり、環境モデルの表現法を工夫することで、階段やスロープといった環境に対しても本手法は適応可能である。これらの点に関して今後さらに検討を加える予定である。

7. ま と め

本論文では、屋内既知環境におけるリアルタイム人物追跡の手法を提案した。本手法は、人物移動に対するカメラによる見え方を複数シミュレーションし、それら画像と入力画像のマッチングを行うことを特徴としている。人物を1つの楕円体で表現することにより、計算コストを低減し、汎用型パーソナルコンピュータでリアルタイム処理が可能となった。また、影や写り込みによるノイズにも強く、安定な人物追跡が可能となった。今後は、複数の視覚システムや他のセンサシステムとの統合による、広範囲な領域での人物追跡手法を検討する予定である。

参 考 文 献

- 1) 丹 康雄, 日比野靖: 位置情報ブローカの提案とそのアーキテクチャ, 情報処理学会研究報告, 97-MBL-2, pp.7-12 (1997).
- 2) 清水孝一: 徘徊者定位システムの開発—バイオテレメトリによる在宅医療支援, 日本 ME 学会雑誌, Vol.10, No.5, pp.3-10 (1996).
- 3) Want, R., Hopper, A., Falcao, V. and Gibbons, J.: The Active Badge Location System, *ACM Trans. Information Systems*, Vol.10, No.1, pp.91-102 (1992).
- 4) 河村邦明, 清水孝一, 山本克之: 痴呆性徘徊対処システムの開発—移動体通信と GPS による徘徊者定位の試み, 信学技報, MBE94-96, pp.93-98 (1995).
- 5) Segen, J. and Pingali, S.G.: A Camera-Based System for Tracking People in Real Time, *Proc. IEEE ICPR '96*, pp.63-67 (1996).
- 6) 岡田隆三, 白井良明, 三浦 純, 久野義徳: オプティカルフローと距離情報に基づく動物体追跡, 電子情報通信学会論文誌 (D-II), Vol.J80-D-II, No.6, pp.1530-1538 (1997).
- 7) Nordlund, P. and Eklundh, J.-O.: Towards a Seeing Agent, *Proc. 1st International Workshop on Cooperative Distributed Vision*, pp.93-120 (1997).
- 8) Hogg, D.: Model-based vision: a program to see a walking person, *Image and Vision Computing*, Vol.1, No.1, pp.5-20 (1983).
- 9) Wren, C.R., Azarbayejani, A., Darrell, T. and Pentland, A.P.: Pfunder: Real-Time Tracking of the Human Body, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.780-785 (1997).
- 10) Rohr, K.: Towards Model-Based Recognition of Human Movements in Image Sequences, *CVGIP: Image Understanding*, Vol.59, No.1, pp.94-115 (1994).
- 11) 井口征士, 佐藤宏介: 三次元画像計測, 昭晃堂 (1990).

付 録

式 (3) において,

$$\begin{bmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}$$

と置き換え, X, Y, Z を式 (2) に代入すると, 以下の h に関する 2 次方程式が得られる.

$$(a_1x_c^2 + a_2y_c^2 + a_3x_cy_c + a_4x_c + a_5y_c + a_6)h^2 + 2(b_1x_c + b_2y_c + b_3)h + c_1 = 0$$

$$a_1 = R_b^2(D_{11}^2 + D_{21}^2) + R_a^2D_{31}^2$$

$$a_2 = R_b^2(D_{12}^2 + D_{22}^2) + R_a^2D_{32}^2$$

$$a_3 = 2R_b^2(D_{11}D_{12} + D_{21}D_{22}) + 2R_a^2D_{31}D_{32}$$

$$a_4 = 2R_b^2(D_{11}D_{13} + D_{21}D_{23}) + 2R_a^2D_{31}D_{33}$$

$$a_5 = 2R_b^2(D_{12}D_{13} + D_{22}D_{23}) + 2R_a^2D_{32}D_{33}$$

$$a_6 = R_b^2(D_{13}^2 + D_{23}^2) + R_a^2D_{33}^2$$

$$b_1 = R_b^2D_{11}(D_{14} - X_p) + R_b^2D_{21}(D_{24} - Y_p) + R_a^2D_{31}(D_{34} - R_b)$$

$$b_2 = R_b^2D_{12}(D_{14} - X_p) + R_b^2D_{22}(D_{24} - Y_p) + R_a^2D_{32}(D_{34} - R_b)$$

$$b_3 = R_b^2D_{13}(D_{14} - X_p) + R_b^2D_{23}(D_{24} - Y_p) + R_a^2D_{33}(D_{34} - R_b)$$

$$c_1 = R_b^2(D_{14} - X_p)^2 + R_b^2(D_{24} - Y_p)^2 + R_a^2(D_{34} - R_b)^2 - R_a^2R_b^2$$

この解が重解となる条件は,

$$\frac{1}{4}D = (b_1^2 - a_1c_1)x_c^2 + (b_2^2 - a_2c_1)y_c^2 \\ + (2b_1b_2 - a_3c_1)x_cy_c + (2b_1b_3 - a_4c_1)x_c \\ + (2b_2b_3 - a_5c_1)y_c + (b_3^2 - a_6c_1) = 0$$

であり、この式が楕円体の画像平面内での輪郭線となる。

(平成 10 年 8 月 21 日受付)

(平成 11 年 9 月 2 日採録)



加藤 博一 (正会員)

1963 年生。1986 年大阪大学基礎工学部制御工学科卒業。1988 年同大学院修士課程修了。1989 年同大学基礎工学部助手、1996 年講師。1998 年ワシントン大学客員研究員。1999 年より広島市立大学情報科学部助教授。工学博士。ヒューマンインタフェース、画像計測の研究に従事。ヒューマンインタフェース学会、電子情報通信学会、計測自動制御学会、ACM 等各会員。



中澤 篤志 (学生会員)

1974 年生。1997 年大阪大学基礎工学部システム工学科卒業、1999 年同大学院基礎工学研究科修士課程修了。現在、同博士後期課程在学中。画像計測、分散エージェントシステムの研究に従事。システム制御情報学会、ヒューマンインタフェース学会各会員。



井口 征士 (正会員)

1940 年生。1962 年大阪大学工学部電気工学科卒業。1964 年同大学院修士課程修了。1965 年同大学基礎工学部助手。1969 年助教授。1984 年教授 (現大学院基礎工学研究科システム人間系)。(財)イメージ情報科学研究所関西研究センター所長兼任。1966~67 年フランス・サクレ原子力研究所、1979~80 年南カリフォルニア大学画像処理研究所客員。工学博士。専門はパターン計測、特に三次元画像・音響信号等の計測。感性情報処理。著書「三次元画像計測」(昭晃堂)、「感性情報処理」(オーム社)等。電子情報通信学会、計測自動制御学会、システム制御情報学会、人工知能学会、日本ロボット学会、IEEE 等会員。