

複数カメラを用いた CONDENSATION のための 隠れに対して頑健な重み統合法

松元 郁 佑[†] 加藤 丈 和[†] 和田 俊 和[†]

本論文では、多視点画像を用いて人物頭部の 3 次元位置を追跡する問題を取り扱う。従来、3 次元空間中で追跡対象に関する仮説を生成し、個々の仮説を各画像に投影して複数の 2 次元の重みを求め、それらを統合して 3 次元の重みを推定する方法が提案されている。しかし、この方法では各画像で求めた重みを積によって統合するため、オクルージョンが発生した場合にはある画像での重みが 0 に近くなるため、統合された重みが 0 に近い値になり、追跡が行えなくなる。この問題を解決するため、我々はオクルージョンロバストな追跡のための重みの統合法を提案する。あるカメラでオクルージョンが発生していることを検出できる場合、このカメラ以外の重みの積で統合するべきであるが、高速にオクルージョンを検出する方法は存在しない。そこで、我々は各カメラで得られる重みの統合法について詳細な考察を行い、オクルージョンに対してロバストな統合法を提案する。提案手法によるシステムを、カメラ 6 台とネットワーク接続された計算機 7 台を用いて試作し、1 人の人物頭部の追跡を約 4ms 間隔で実行できること、およびオクルージョンに対する頑健さ、同時に 6 人までの追跡が実行できることを実証した。

An Occlusion Robust Weight Integration Method for Multi-camera CONDENSATION

YUSUKE MATSUMOTO,[†] TAKEKAZU KATO[†] and TOSHIKAZU WADA[†]

This paper presents a novel method for human head tracking using multiple cameras. Most existing methods estimate 3D target position according to 2D tracking results at different viewpoints. This framework can be easily affected by the inconsistent tracking results on 2D images, which leads 3D tracking failure. For solving this problem, an extension of CONDENSATION using multiple images has been proposed. The method generates many hypotheses on a target (human head) in 3D space and estimates the weight of each hypothesis by integrating viewpoint dependent weight values of 2D hypotheses projected on image planes. In theory, viewpoint dependent weight values should be integrated by multiplication, however, it is easily affected by occlusions. Thus we investigate this problem and propose a novel integration method in this paper and implemented a prototype system consisting of six set of PCs and cameras. We confirmed the robustness against occlusions and the efficiency of our method.

1. はじめに

3 次元空間中の人物追跡は、セキュリティやヒューマンインタフェース、人物の行動解析などに応用可能な技術である。特に、多視点画像を用いた人物追跡技術は対象を最もよく観測できる視点を選択し、オクルージョンなどの影響を回避できる可能性があることから、数多くの研究例がある。これらは、追跡を行う空間によって、以下に述べる 2 種類に大別される。

2 次元追跡統合型：これは、各カメラで撮影された画

像平面上で追跡を行い、この 2 次元の追跡結果を統合することで 3 次元の位置を復元するという手法¹⁾⁻³⁾である。この手法には、1) 通信量が少なく、PC の台数を容易に増やすことができる、2) 各カメラによる独立した追跡処理による処理の並列化が可能である、という 2 つの利点がある。しかしながら、3 次元の追跡結果と 2 次元の追跡結果との整合性を評価しないため、3 次元の追跡が不安定になるという問題がある。この問題は、2 次元追跡統合型の手法に不可避な問題である。

3 次元仮説生成型：これは、各画像平面上で追跡を行うのではなく、3 次元空間中对象モデルを仮定し、そのモデルと各画像との整合性を評価することで追跡するという手法⁴⁾⁻¹⁰⁾である。

[†] 和歌山大学システム工学研究科
Graduate School of Systems Engineering, Wakayama
University

表 1 従来手法との比較

Table 1 Comparison of conventional methods with our method.

手法	追跡次元	オクルージョン
従来手法 (1) (浮田らの手法)	2 次元 (画像平面)	対応 (環境マップ)
従来手法 (2) (鈴木らの手法)	3 次元 (実空間)	未対応
提案手法	3 次元 (実空間)	対応 (重みの統合法)

たとえば、森らの手法⁴⁾では、3次元空間中の人物モデルと、背景差分によって得られた複数のシルエット画像との整合性を評価することで人物の追跡を実現している。このような手法では、2次元追跡統合型と異なり、各画像はモデルの妥当性評価に用いられるだけであるので、3次元空間中の対象位置と画像上での対象位置に食い違いは生じないという特長がある。しかし、各カメラの撮影時刻のずれを吸収できるように、カルマンフィルタを用いて対象の運動予測を行っている^{4),5)}ため、単峰性の運動パラメータの予測しかできないという制限がある。

この問題点を解決するために、CONDENSATION¹¹⁾を用いて3次元空間中の多峰性の確率分布推定による追跡を行う手法が提案されている⁶⁾⁻¹⁰⁾。このような手法では、3次元空間中の仮説を各画像に投影した2次元の仮説の重みを求め、それらの積によって元の3次元空間の仮説の重みを計算しているため、遮蔽物などによって一部のカメラから対象が撮影できない場合には追跡することができない。

本論文ではCONDENSATIONを用いた3次元空間中の多峰性の確率分布を推定する手法に基づき、重みの統合法を変更することにより、オクルージョンに対してロバストな追跡を実現する方法を提案する。さらに、人物の検出、複数人物の同時追跡機能を有する実時間人物検出・追跡システムの構築法を明らかにする。

従来の多視点画像を用いた追跡システムと提案手法に基づくシステムとの違いを示したのが表1である。従来法ではオクルージョンへの対応に問題があり(この問題点については3章で示す)、我々の追跡システムでは、重みの統合法によって従来法の問題を解決している。

2. CONDENSATION による 3 次元人物追跡

本章ではCONDENSATIONアルゴリズムと、これを用いた3次元人物追跡の方法を説明する(図1)。

2.1 CONDENSATION アルゴリズム

CONDENSATIONでは、追跡対象の位置や姿勢、形状を表すパラメータ空間中の確率分布(事後分布)を各時刻で推定することによって追跡を実現する。時刻 t における対象の状態を確率変数 \mathbf{X}_t で表し、同時刻

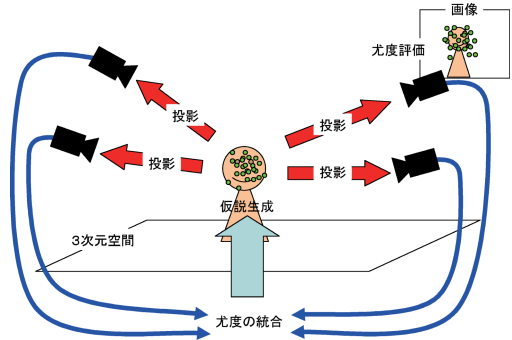


図 1 3次元人物追跡の概要

Fig.1 Outline of 3D human tracking.

の画像特徴を \mathbf{Z}_t とする。また、時刻 t までに得られる画像特徴系列を $\mathcal{Z}_t = (\mathbf{Z}_1, \dots, \mathbf{Z}_t)$ とおく。このとき、 \mathbf{X}_t の事後分布 $p(\mathbf{X}_t | \mathcal{Z}_t)$ はベイズの定理により、次のように変形できる。

$$p(\mathbf{X}_t | \mathcal{Z}_t) = k_t p(\mathbf{Z}_t | \mathbf{X}_t) p(\mathbf{X}_t | \mathcal{Z}_{t-1}) \quad (1)$$

ただし、 k_t は正規化項であり、 $p(\mathbf{Z}_t | \mathbf{X}_t)$ は状態 \mathbf{X}_t の尤度、 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ は事前分布である。ここで、 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ は前の時刻 $t-1$ の事後分布 $p(\mathbf{X}_{t-1} | \mathcal{Z}_{t-1})$ と、時間が経過するときの分布の推移確率 $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ をもとに、次式のように求められる。

$$p(\mathbf{X}_t | \mathcal{Z}_{t-1}) = \int_{\mathbf{X}_{t-1}} p(\mathbf{X}_t | \mathbf{X}_{t-1}) p(\mathbf{X}_{t-1} | \mathcal{Z}_{t-1}) d\mathbf{X}_{t-1} \quad (2)$$

CONDENSATIONでは、事後確率を各時刻 t における、 N 個の仮説 $\mathbf{s}_i^{(n)}$ ($n = 1, \dots, N$)と、その整合性評価によって得られる重み $\pi(\mathbf{s}_i^{(n)})$ を用いて、離散的に効率良く推定する。つまり、重み $\pi(\mathbf{s}_i^{(n)})$ によって尤度 $p(\mathbf{Z}_t | \mathbf{X}_t)$ を近似している。この離散化によって、式(1)は次式のように表される。

$$p(\mathbf{X}_t | \mathcal{Z}_t) \approx \sum_{n=1}^N \pi(\mathbf{s}_i^{(n)}) \delta(\mathbf{X}_t - \mathbf{s}_i^{(n)}) \quad (3)$$

ただし、 δ はディラックデルタ関数である。

2.2 3次元人物追跡への拡張

CONDENSATIONでは、ある2次元画像平面における仮説 $\mathbf{s}_i^{(n)}$ とその重み $\pi(\mathbf{s}_i^{(n)})$ によって対象追跡を

実現した．これを次のようにして，複数視点を用いた 3 次元の対象追跡に拡張することができる．

3 次元の対象の仮説を 2 次元の場合と同様に $s_t^{(n)}$ としたとき，これらが 3 次元空間中の対象とどの程度一致しているのかを評価しなければならない．このとき，単純に画像を撮影するだけの計算機内では，3 次元物体の情報を持っていないため，直接 3 次元空間中で重みを計算することは不可能である．そこで，各カメラの画像 c に対して 2 次元の重み $\pi_c(s_t^{(n)})$, ($c = 1, \dots, C$) を計算し，これらを統合することで 3 次元空間中の重み $\pi(s_t^{(n)})$ を推定する．この概念を示したのが図 2 である．図 2 においてカメラから延びる白い帯は各カメラから推定される対象位置，つまり重みが大きい部分を表しており，交わりの部分は，3 次元空間中での対象人物の位置を表している．

このような統合法の 1 つに，以下のように積によって統合する方法がある．

$$\pi(s_t^{(n)}) \equiv \prod_{c=1}^C \pi_c(s_t^{(n)}) \quad (4)$$

このように統合することによって，対象位置の 3 次元の重みを計算することができる．しかし，あるカメラ c_0 でオクルージョンが発生し， $\pi_{c_0}(s_t^{(n)})$ が小さくなったとき，3 次元の重み $\pi(s_t^{(n)})$ の値も小さくなってしまい，追跡に失敗してしまうという問題がある．我々はこのような問題に対して，重みの統合法の詳細な考察を行い，オクルージョンに対して敏感に反応しない重みの統合法を提案する．

以下，我々の人物追跡システムでは，人物のモデルとして 3 次元の楕円体モデルを用い，このモデルを画像平面上に投影した結果と，画像のエッジとの一致度を各画像から得られる重みとしている．この処理の詳細については，4.1，4.2 節で示す．また，動的モデルとして CONDENSATION の推移確率 $p(X_t|X_{t-1})$ には，正規分布を用いる．

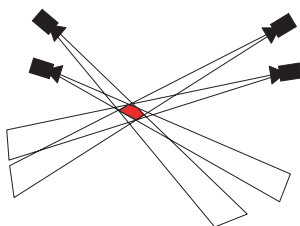


図 2 重みの統合
Fig.2 Weight integration.

3. オクルージョンロバストな追跡のための重み統合

本研究においてオクルージョンとは，次の 2 つを示す．

- (1) 他の追跡対象によって隠される場合
 - (2) 追跡対象でない人物・物体によって隠される場合
- それぞれ，(1) の場合に対しては，すべての追跡情報を統合し，これを用いて他の追跡対象によるオクルージョンを計算する方法¹²⁾，(2) の場合に対しては，オクルージョンを動的に計算する方法¹³⁾ や，環境モデルを構築する方法^{3),6)} などが提案されてきた．しかし，(1) に対処する方法だけでは(2)のオクルージョンの発生に対処できない．また，(2)に対処する方法では，高速にオクルージョンの判定を行える手法は今のところ明らかにされておらず，追跡をリアルタイム (30 fps) で行えないという問題がある．あるいは，環境モデルの構築を行うと，動的な環境の変化に対応できないという問題がある．本章では，このような処理を行わず，2 次元の重み $\pi_c(s_t^{(n)})$ の統合方法を工夫することで，オクルージョンに敏感に反応せず 3 次元の重み $\pi(s_t^{(n)})$ を求める方法を提案する．

まず，我々は式 (4) を用いた重みの統合実験を行った．この実験に用いたカメラ配置は図 3 であり，図 4 のような人物頭部について，床面と平行な人物頭部の高さと同じ平面における重みのマップを作成した．図 5 にその結果を示す．(a) にすべてのカメラから人

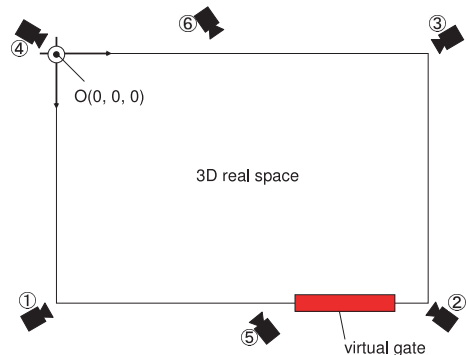


図 3 カメラ 6 台の配置
Fig.3 Placement of 6 cameras.

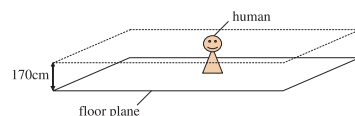


図 4 人物頭部が存在する平面
Fig.4 A plane where a human head exists.

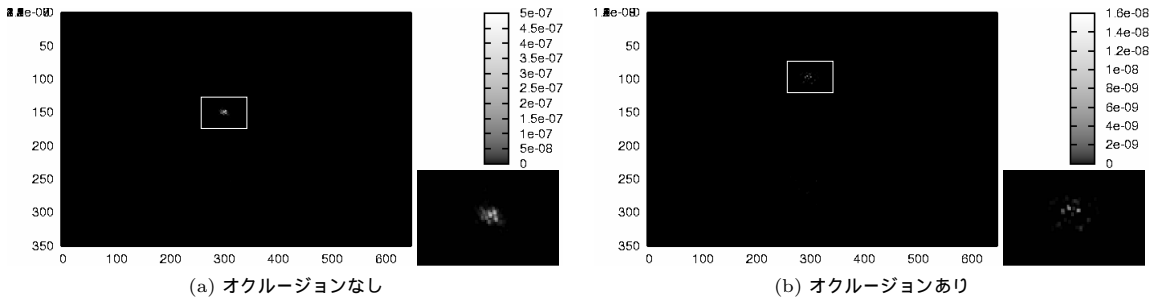


図 5 積を用いた重みマップ: (a) オクルージョンなし, 人物頭部の位置 $(X, Y, Z) = (300, 150, 170)$, (b) オクルージョンあり, 人物頭部の位置 $(X, Y, Z) = (300, 100, 170)$

Fig. 5 Weight map using integration by product: (a) occlusion-free, object position $(X, Y, Z) = (300, 150, 170)$, (b) occluded, object position $(X, Y, Z) = (300, 100, 170)$.

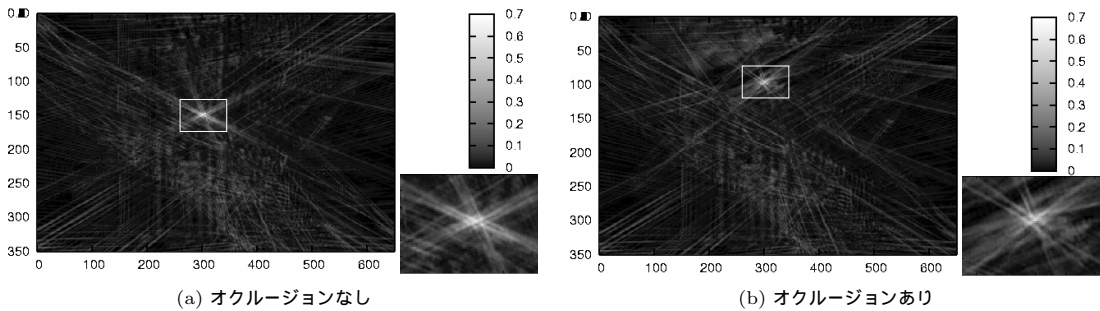


図 6 和を用いた重みマップ

Fig. 6 Weight map using integration by summation.

物頭部が観測可能な場合, (b) にカメラ 4 と 6 でオクルージョンが発生している場合を示す.

この図から, 積による重みの統合は, オクルージョンがない場合は人物頭部周辺の 3 次元の重みを推定することができるが, オクルージョンが発生したときに, 敏感に反応してしまい, 重みが小さくなるということが分かる. このため, 対象位置以外の重みとの差が得られず, 追跡の失敗を引き起こすと考えられる.

そこで, オクルージョンに対して敏感に反応してしまわないような統合法を考える必要がある. 我々は, まず次のような和による重みの統合を考える.

$$\pi(\mathbf{s}_t^{(n)}) \equiv \sum_{c=1}^C \pi_c(\mathbf{s}_t^{(n)}) \quad (5)$$

この統合法では, あるカメラにオクルージョンが発生したときに, そのカメラの重みは小さくなるが, 他のカメラの重みが大きければ, それが 3 次元の重みをサポートし, 3 次元の重みは小さくなりにくい性質がある. その様子を示したのが図 6 であり, 図 5 と同様のデータを用いた結果を示す. 図 6 から, 1) 積の場合と比べ, 対象位置の重みの分布がブロードである,

2) 対象位置以外の場所に重みが大きい場所が多数存在する, 3) オクルージョンに敏感に反応しない, という 3 つの性質があることが分かる.

これら 3 つの性質のうち, 3) のオクルージョンに敏感に反応しないことが積による統合に対する最大のアドバンテージである. しかし, 図 6 から, 対象位置以外で重みが大きくなるのが分かる. これは, 他の場所を追跡しやすくなっていることと等価であるため, この性質は排除しておきたい.

これに対し, 積による統合の性質を考えてみると, 1) 対象位置の重み分布がブロードでない, 2) 対象位置以外の場所の重みがほぼ 0 である, 3) オクルージョンに敏感に反応する, という 3 つの性質があることが分かる.

ここで, 我々が想定している追跡システムに必要な性質は, 1) 対象位置の重みの分布がブロードである (追跡対象を見失いにくい), 2) 対象位置以外での重みが小さい (対象位置以外に引き寄せられにくい), 3) オクルージョンに敏感に反応しない (オクルージョンロバスタな追跡ができる), という 3 点である. この性質を得るためには, 積による統合の性質と和による統合

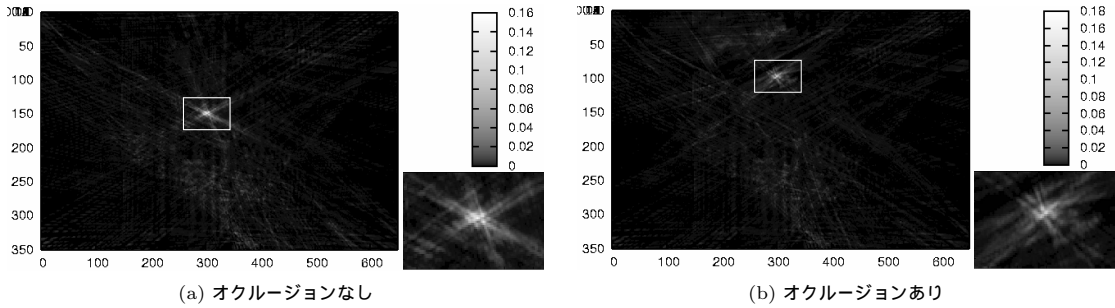


図 7 式 (6) を用いた重みマップ
Fig. 7 Weight map using integration by Eq. (6).

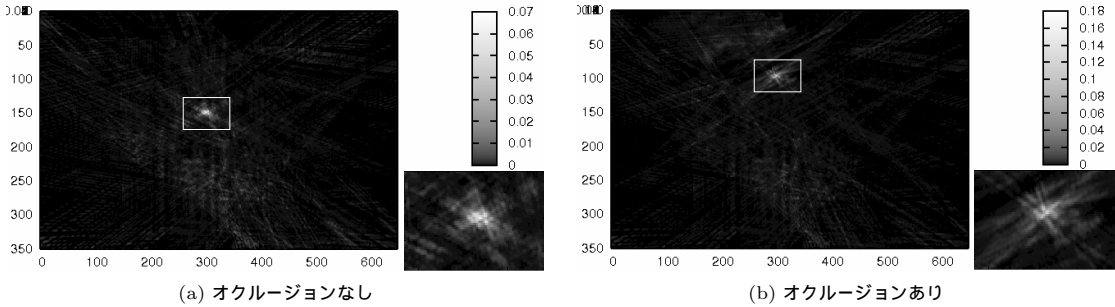


図 8 提案手法による重みマップ
Fig. 8 Weight map using our method.

の性質を組み合わせることが必要ではないかと考える。

そこで、複数の 2 次元の重みのうち、これらの組合せの積を計算し、その和を 3 次元の重みとして用いる方法を考える。このように統合することで、i) 積の性質から、組合せに用いる重みがそれぞれ大きくないと、この積の結果が大きくならないため、対象位置以外の重みを小さくでき、ii) 和の性質から、ある 1 つの積の結果の値が小さくなったとき、他の積の結果が大きくなれば和の結果も大きくなるため、オクルージョンに対して敏感に反応しない 3 次元の重みが得られる。

ここで、 C 個の 2 次元の重みの中から n 個を取り出す組合せを用いる場合、最低限 n 個の重みが大きい必要がある。つまり、組合せに用いる重みの数が多ければ多いほど、最低限大きくないとならない重みの数も増える。よって、我々は組合せに必要な最小限の 2 つの組合せを用いることにした。

以上の考察から、我々は 3 次元の重みを、2 次元の重みのうち 2 つの組合せについて積を計算し、その積の和を計算することで推定する。この統合方法を表したのが次の式である。

$$\pi(\mathbf{s}_t^{(n)}) \equiv \sum_{c_1, c_2, c_1 \neq c_2} \pi_{c_1}(\mathbf{s}_t^{(n)}) \pi_{c_2}(\mathbf{s}_t^{(n)}) \quad (6)$$

この式で重みを統合した結果が図 7 であり、積の場合と比べ、オクルージョンに敏感に反応せず、かつ

和の場合と比べ、対象位置以外の場所では値が小さくなっていることが分かる。

しかし、図 7 (a) では、対象位置周辺に放射状に伸びる重み値の高い部分が現れている。式 (5) によって統合した場合にも示したが、対象位置の周辺に重み値の高い部分が存在すると、追跡が不安定になってしまう。この部分は人物とカメラ 2 台が直線状に並んだ状態で積を計算した結果生じると考えられる。したがって、このような人物位置におけるカメラ 2 台の組合せは 3 次元位置を特定するためには有効に利用できない。この問題を改善するために、積の計算を行う際にこのような組合せの結果を排除して重み統合を行うことにする。

以上の考察に基づき、本研究で追跡に用いる重みの統合を次式のように定義する。

$$T(c_i, c_j, \mathbf{X}_t) = \begin{cases} 1, & \frac{(\mathbf{X}_t - \mathbf{O}_{c_i}) \cdot (\mathbf{X}_t - \mathbf{O}_{c_j})}{\|\mathbf{X}_t - \mathbf{O}_{c_i}\| \|\mathbf{X}_t - \mathbf{O}_{c_j}\|} < \beta \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$\begin{aligned} \pi(\mathbf{s}_t^{(n)}) &\equiv \sum_{c_1, c_2, c_1 \neq c_2} T(c_1, c_2, \mathbf{X}_t) \pi_{c_1}(\mathbf{s}_t^{(n)}) \pi_{c_2}(\mathbf{s}_t^{(n)}) \end{aligned}$$

ここで、 \mathbf{O}_{c_i} はカメラ c_i の焦点位置を表している。式 (7) で重みを統合した結果が図 8 である。実際に

図 8 と図 7 との差を計算したところ、(a) では最大で 0.015 (同じ場所での図 7 の 52.9%) の不要な重みを除去できていた。(b) ではほとんど違いが見られなかったが、これは対象位置周辺においてすべてのカメラの組合せが T が 0 となったためである。

4. 人物頭部追跡システム

提案手法は、1 章で示したように様々な応用に適用することができる追跡手法のための重み統合法である。この応用例のうち、我々は人物行動解析ができるような、複数の人物が行き交う場所での適用を考える。本章では、これを実現するシステムとして、人物の頭部追跡システム構築例を説明する。

4.1 座標系とモデル

本研究では、カメラは人物を上から観測できるように図 3 のように設置した 6 台の固定カメラを用いる。このときの座標系は図 9 のようになる。ここで、ワールド座標系と各画像座標系の関係についてはあらかじめ校正済みである。この校正は、12 点の対応点を用いた最小二乗法によって行い、3 次元空間から 2 次元画像への投影行列を計算している。

このとき、同次座標で表現したワールド座標系上のある点 $\mathbf{a} = (X, Y, Z, 1)^T$ に対応するカメラ c での画像座標系の点 $\mathbf{b}_c = (x_c, y_c, w_c)^T$ は、キャリブレーションで得られた 3 次元空間と画像平面との関係を表す投影行列 P_c によって次のように計算することができる。

$$\mathbf{b}_c = P_c \mathbf{a} \tag{8}$$

また、本研究では人物頭部を表現するモデルとして、図 10 のような回転楕円体モデルを用いる。このモデルは中心位置 (X, Y, Z) のみを持つ回転楕円体モデル

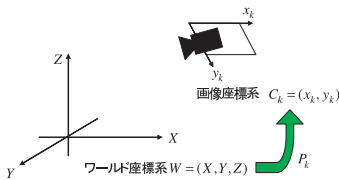


図 9 ワールド座標系と画像座標系

Fig. 9 World coordinate system and image plane coordinate system.

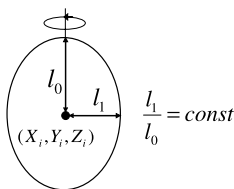


図 10 回転楕円体モデル
Fig. 10 Spheroid model.

で、これを 1 つの仮説として追跡する。

4.2 仮説の 2 次元重みの計算

ある仮説 $s_t^{(n)}$ がカメラ c の画像座標系に投影されたとき、この投影座標 $(x_{n,c}, y_{n,c}, w_{n,c})$ は 4.1 節で示した式 (8) によって得られる。このとき、仮説を表す図 10 の回転楕円体は、図 11 のような楕円として投影される。これを用いて、各カメラ画像における各仮説の 2 次元の重みを計算する。

本研究では、図 11 で示したような楕円上の点 (これを輪郭点 μ とする) における法線 \mathbf{N} と輝度勾配ベクトル \mathbf{E} (Sobel オペレータを用いて x, y それぞれの方向について計算) の内積を計算し、その和によって仮説を評価する。つまり、カメラ c における仮説 $s_t^{(n)}$ の重み $\pi_c(s_t^{(n)})$ は、仮説 $s_t^{(n)}$ を画像平面に投影した楕円とエッジとの一致度を表す。これは、次式のようになる。

$$\pi_c(s_t^{(n)}) \equiv \sum_{\mu} \mathbf{N}_{\mu} \cdot \mathbf{E}_{\mu} \tag{9}$$

ただし、 \mathbf{N}_{μ} は正規化されており、 \mathbf{E}_{μ} は正規化されていない。これは、エッジの強さも一致度に含まれるとして考えているためである。つまり、図 12 に示すように、 \mathbf{E} と \mathbf{N} が同一方向を示し、かつ \mathbf{E} の長さが長いときに、一致度が大きくなる。本研究で用いた輪郭点は図 11 のように、人物頭部の上部分だけを考慮した。これは、人物頭部の上部周辺の輪郭がよく現れるためである。このように一致度を計算し、重みとして用いることにより、頭部の追跡を安定に行うことができる。

4.3 新規人物の検出・追跡の終了

本研究では、人物の出入りを考慮した環境を考える。

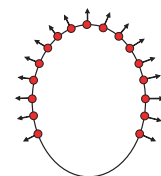


図 11 楕円上の法線

Fig. 11 Normals of ellipse.

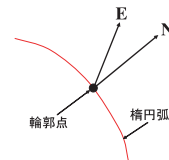


図 12 楕円上の法線 \mathbf{N} とエッジ勾配の方向ベクトル \mathbf{E} の関係
Fig. 12 Relationship of normal of ellipse \mathbf{N} and edge gradient vector \mathbf{E} .

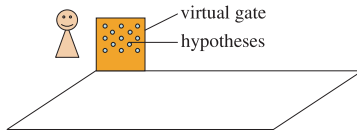


図 13 人物の検出

Fig. 13 Human detection.

このとき、人物の出入りはある決まった場所で起こるとする（この場所を図 13 のような仮想的なゲートとする）。つまり、このゲート以外では人物が急に現れたり、いなくなったりすることはないという前提条件を加える。

まず、人物の初期検出について説明する。これは図 13 のように、仮想ゲートで仮説をできるだけ広い範囲に存在するように発生させておき、この仮説の形状と同等のエッジが画像上で現れたときに追跡を開始する。つまり、発生させた仮説に対し追跡時と同様の 3 次元の重み計算をしておき、この重みが大きくなったときが人物が近づいたときだと考え、このときに追跡を開始する。

次に、追跡の終了について説明する。これは前述のゲートを通過したときや、追跡範囲から出たことによる追跡の失敗を検出すればよい。このゲートの通過は追跡位置によって計算できる。

以上の新規人物の検出と追跡の終了を自動的に行うことで、より実用的なシステムにすることができる。

4.4 複数人物の追跡

複数人物の追跡を行うためには、それぞれの人物に対し CONDENSATION を適用し、事後分布を推定する。この CONDENSATION では推定される分布の多峰性と複数人物を区別できないため、これらの分布の仮説群が混ざり合うという問題が生じる。

このような問題に対して、仮説が混ざることを出し、これにより分布間の重なりを回避している手法^{8),14)}が提案されている。ほかに、Khan らによる手法¹⁵⁾では、MCMC の枠組みに MRF (Markov Random Field) を導入し、2 対象間の相互作用を表現することで複数対象の追跡を成功している。我々は、直感的に理解しやすい前者の考え方をヒントに、次のような方法でそれぞれの分布が交わらないようにした。

まず、各人物に対する分布が交わる状態とは、ある人物 m_0 に対する分布 $p_{m_0}(\mathbf{X}_t | \mathcal{Z}_t)$ と、その他の人物 m_1 に対する分布 $p_{m_1}(\mathbf{X}_t | \mathcal{Z}_t)$ が図 14 のように重なっている状態である。このとき、それぞれの人物 m_0 と m_1 の仮説集合 $\{s_{m_0,t}^{(n)}\}$ と $\{s_{m_1,t}^{(n)}\}$ も混ざ

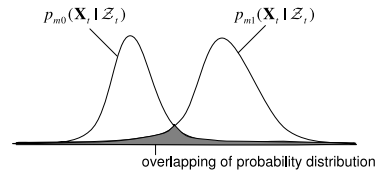


図 14 分布の重なり

Fig. 14 Overlapping of probability distributions.

りあってしまう。

この問題を解決するため、分布の重なりを判定し、重なっている部分の仮説の重みを下げようする方法について検討する。これは、結果的に仮説群間の斥力を生み出すことになる。

本来、人物 m_0 , m_1 の仮説 $s_{m_0,t}^{(n)}$, $s_{m_1,t}^{(n)}$ は

$$\begin{aligned} p_{m_0}(s_{m_0,t}^{(n)} | \mathcal{Z}_t) &> p_{m_1}(s_{m_0,t}^{(n)} | \mathcal{Z}_t) \\ p_{m_1}(s_{m_1,t}^{(n)} | \mathcal{Z}_t) &> p_{m_0}(s_{m_1,t}^{(n)} | \mathcal{Z}_t) \end{aligned} \quad (10)$$

を満足するものが大半であるが、仮説の分布が近接した場合、分布の交わり部分で

$$\begin{aligned} p_{m_0}(s_{m_0,t}^{(n)} | \mathcal{Z}_t) &< p_{m_1}(s_{m_0,t}^{(n)} | \mathcal{Z}_t) \\ p_{m_1}(s_{m_1,t}^{(n)} | \mathcal{Z}_t) &< p_{m_0}(s_{m_1,t}^{(n)} | \mathcal{Z}_t) \end{aligned} \quad (11)$$

となってしまう。このような仮説に対する重みを強制的に 0 とすることで、推定される分布が混ざり合わない処理ができる。

この判定を行うには、 $p_{m_0}(\mathbf{X}_t | \mathcal{Z}_t)$ と $p_{m_1}(\mathbf{X}_t | \mathcal{Z}_t)$ の値を求めなければならないが、CONDENSATION では分布を明示的に確率分布の形で求めていないため、この計算は高速に実行することができない。

本来、CONDENSATION で推定される多峰性の分布間の重なりを求めるためには、かなりの計算時間を要する。我々は、簡単化のため、この分布間の重なりを求める際に限り、単一の人物頭部の分布が単峰性の形をしていると考え、Gaussian で十分近似できるとする。また、この近似を行った結果、各人物の分布は多峰性の形をしていれば問題はないと考えている。このように近似することで、各人物の分布間距離を求めることができるため、仮説間の重なりが判定できる。

ある人物 m に対する分布の平均を $\overline{s_{m,t}}$ 、共分散行列を $\Sigma_{m,t}$ とすると、これらは仮説の重み ($\sum_n \pi(s_{m,t}^{(n)}) = 1$ とする) による重み付け平均、および重み付け共分散行列として表現できる。

$$\overline{s_{m,t}} = \sum_n \pi(s_{m,t}^{(n)}) s_{m,t}^{(n)} \quad (12)$$

$$\Sigma_{m,t} = \sum_n \pi(s_{m,t}^{(n)}) (s_{m,t}^{(n)} - \overline{s_{m,t}}) (s_{m,t}^{(n)} - \overline{s_{m,t}})^T$$

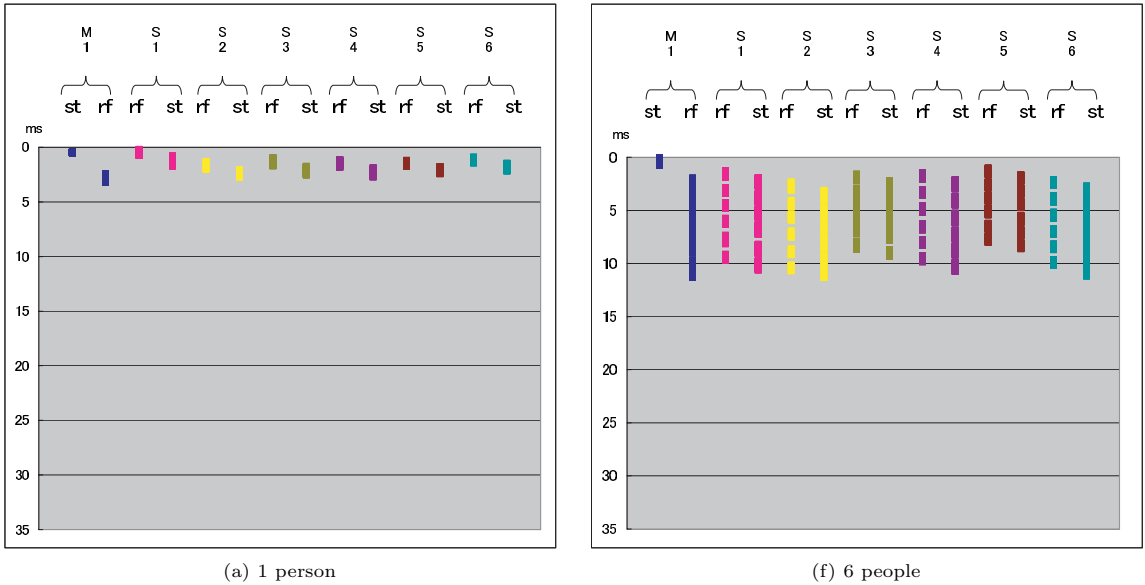


図 15 1 フレームの通信のタイムチャート
Fig. 15 Time chart of 1 frame communication.

このとき、人物 m の分布に対する、ある人物 m_0 の仮説 $s_{m_0,t}^{(n)}$ のマハラノビス距離 $D(m_0, \Sigma_{m,t})$ は、次のように表される。

$$D(m_0, \Sigma_{m,t}) = \left(s_{m_0,t}^{(n)} - \bar{s}_m \right)^T \Sigma_{m,t}^{-1} \left(s_{m_0,t}^{(n)} - \bar{s}_m \right) \quad (13)$$

各人物に関する確率分布の共分散行列は等しいと仮定しているので、確率密度の大小関係はマハラノビス距離の大小関係で評価することができる。具体的には、この $D(m_0, \Sigma_{m,t})$ が、本来属している人物 m_0 の分布とのマハラノビス距離 $D(m_0, \Sigma_{m_0,t})$ よりも小さければ重なっていると判定する。

$$\pi(s_{m_0,t}^{(n)}) = \begin{cases} \pi(s_{m_0,t}^{(n)}) & D(m_0, \Sigma_{m_0,t}) \leq D(m_0, \Sigma_{m,t}) \\ 0 & otherwise \end{cases} \quad (14)$$

また、複数の人物が存在する場合は、全人物についての分布とのマハラノビス距離を計算して判定する。

5. 実験結果

本研究では、前章までに説明した人物頭部追跡の試作システムを6台のカメラとそれぞれを接続したPCを用いて実装した。本章では、これを用いて人物頭部の追跡実験を行った結果を示す。用いたシステムは、CPUがPentium 4 3.6 GHz、メモリが1 GBytesのスレーブPCにカメラを接続し、このスレーブPCが

仮説の評価を行う。さらに、仮説の生成と重みの統合を行うマスタPCとしてPentium D 3.0 GHz、メモリ512 MBytesのPCを用いた。カメラの配置は図3のようになる。

このシステムを用いた、1フレームあたりの処理タイミングは図15に示すようになる。本研究で用いるシステムでは、前述のようなマスタPCとスレーブPCによる通信が行われる。通信の順番として、マスタPCからスレーブPCに対して仮説の送信、スレーブPCで仮説の受信、スレーブPCからマスタPCに対して重みの送信、マスタPCで重みの受信となる。図15では、縦方向に時間軸をとり、横方向にそれぞれのタイミングの種類を表す。また、タイミングの種類は左からM1がマスタPCを表し、“st”は仮説の送信、“rf”はスレーブPCから返信される重みの受信、S1からS6までは各スレーブPCを表し、“rf”はマスタPCからの仮説の受信、“st”はマスタPCへの重み送信のタイミングをそれぞれ表す。このグラフでの時間軸の単位はmsであり、各計算機において絶対的な時間はそろっていると仮定し、マスタPCが最初の仮説を送信した時間を基準とした相対時間を用いた。また、今回は1人の追跡を安定に行える仮説の個数を、経験的に150個と決定し、実験を行った。

これらによれば、1フレームあたりの通信が1人の追跡の場合で3ms強、6人の場合で13msの時間を要することが分かった。つまり、6人の人物を追跡する場合でも提案手法の十分な実時間性を確認できる。

表 2 追跡人数と処理時間 (ms)

Table 2 Relationship between number of tracking and processing time (ms).

追跡人数	仮説生成 (ランダムサンプリング)	画像処理	重み統合	1 フレームの処理時間
1 人	0.129	1.531	0.136	3.635
2 人	0.245	2.995	0.351	5.010
3 人	0.363	4.454	0.423	6.522
4 人	0.476	5.776	0.487	7.899
5 人	0.606	7.049	0.584	9.000
6 人	0.715	8.586	0.392	11.563

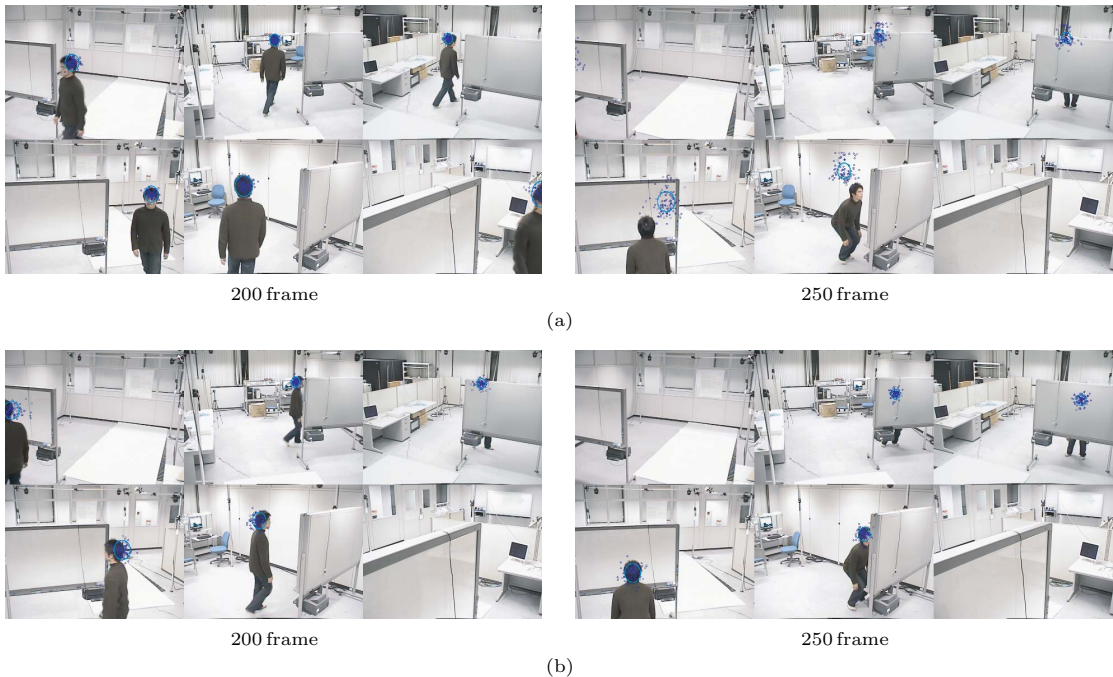


図 16 オクルージョンを含む場面での追跡結果：(a) 積で統合した場合、(b) 提案手法

Fig. 16 Tracking result when image includes occlusion: (a) product integration, (b) our method.

また、追跡人数の増減による処理時間の変化と、仮説生成、重み計算の画像処理、重み統合、それぞれの時間を表 2 に示す。この表から、提案手法である重み統合の時間がほとんど必要ないということが分かる。

以下では、まず、オクルージョンに対する頑健さを実証するための実験として、環境内にオクルージョンを発生させて追跡を行う様子を示す。次に、人物の検出と追跡の終了を示す結果を示す。最後に複数人物を追跡した結果を示す。

5.1 オクルージョンを含む追跡実験

オクルージョンによる影響を評価するために、実際にオクルージョンが発生するシーンでの追跡実験を行った。図 16 にその結果を示す。これらはそれぞれ各フレームで撮影された画像を示し、1 つの画像に 6 台のカメラ (図 3 に示した、1~6 までのカメラ) から撮影された画像 (画像中の左上の画像から右方向に順番に

カメラ 1~6 で撮影した画像を示す) を合成して示した。図 16 で、各カメラで撮影された画像に写っている人物の頭部にある多数の点が、仮説 $s_{m,t}^{(n)}$ を画像平面上に投影したときの楕円モデルの中心位置を表す。また、大きな楕円は仮説のそれぞれの画像に対する期待値であり、式 (12) の $\overline{s_{m,t}}$ を画像上に投影したものを示す。

今回の実験では、オクルージョンによる追跡の失敗判定は目視で行っている。具体的には、各カメラ画像を見て、追跡結果が明らかに人物頭部から離れている場合を失敗、頭部にある場合を成功としている。このとき、オクルージョンが発生していない 200 フレームで、追跡に成功していることが確認できている。

同図では、(a) の系列が各カメラで得られる重みの積で統合を行った場合、(b) の系列が提案手法で、どちらも 250 frame でオクルージョンが発生しているときを示す。ただし、オクルージョンが発生しているカメラ

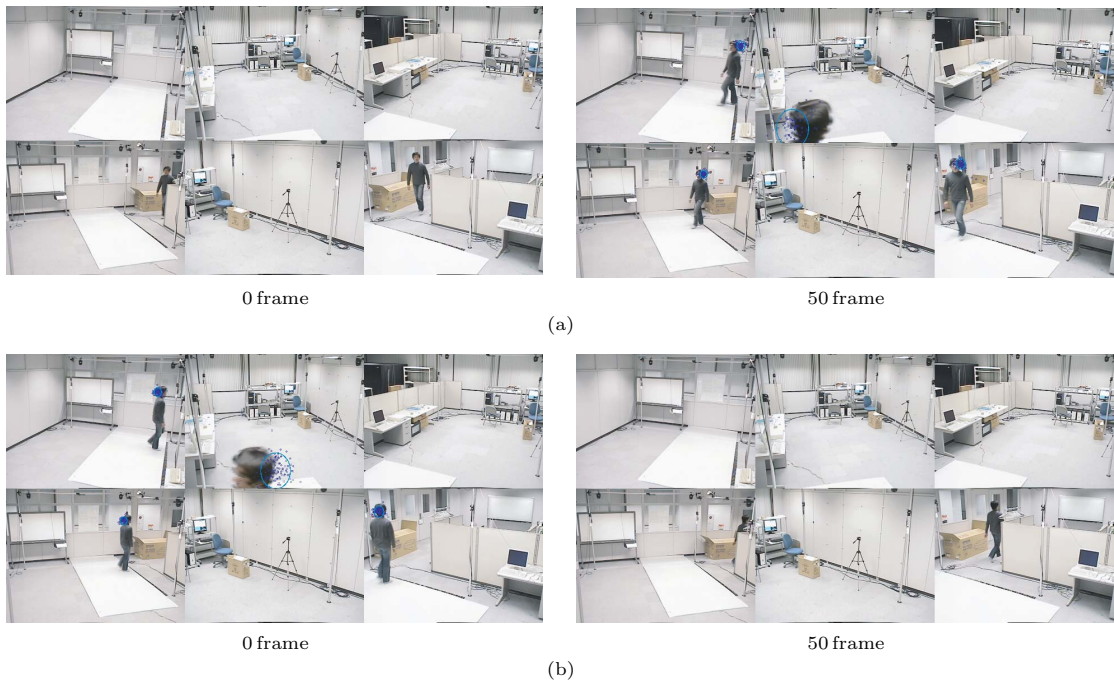


図 17 人物の検出と追跡の終了 (左: 処理前, 右: 処理後): (a) 人物の検出の様子, (b) 追跡終了の様子

Fig. 17 Left side: before, right side: after, (a) detection (b) end of tracking.

で撮影される画像はそれぞれの画像の上段中央, 右であり, 観測可能なのは下段左, 中央の画像である. そのほかはカメラの視野に入らない位置に人物がいるので無視できる. この結果から, 積の場合ではオクルージョンが発生したときに追跡できないのに対し, 提案手法では正しく追跡できていることが分かる. このように, 重み統合の違いで追跡の性能が大きく変化するということがいえ, 提案手法がオクルージョンロバストな追跡を実現する重み統合法だということがいえる. ただし, 提案手法では 2 つの組合せを用いているため, 最低 2 つのカメラから観測可能である必要がある.

人物の個人差による追跡性能の変化を示すため, この実験を 10 人の人物に対して行った. その結果, いずれの人物に対しても, オクルージョンが発生するとき, 積による統合では追跡に失敗し, 提案手法では追跡に成功した.

5.2 人物の検出・追跡の終了

4.3 節で示したように, 本研究では人物の検出と追跡の終了を自動的に行う. この自動化によって, より実用的なシステムになると考えられる.

図 17 に, 人物の検出と追跡の終了における実験結果を示す. 同図 (a) は人物の検出の様子, (b) は追跡の終了の様子を表し, 仮想ゲートは画像の 6 カメラのうち, 上段は左, 下段は左, 右の 3 カメラから観測可能

な位置に存在する (図 3 に仮想ゲートの位置を示す).

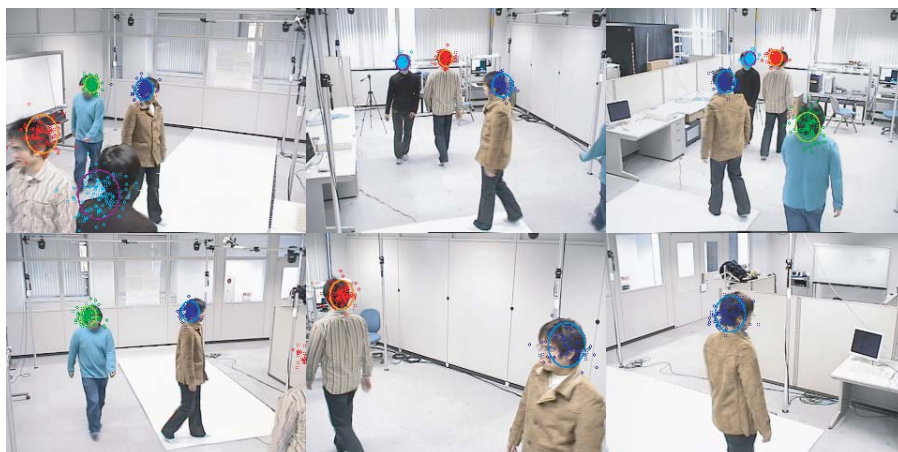
はじめに, 人物の検出の様子について説明する. 図 17 (a) で示すように, 前述のようにゲート付近に人物が近づいたときに検出できていることが分かる. ただし, 本システムの処理の関係上, 1 フレームで 1 人ずつしか検出できない. この理由として, 複数人が入室するとき, 1 人が入室するときでは, 撮影される画像のエッジが変化してしまうため安定な重みが得られず, 設定した閾値よりも低い重み得られる場合があるためである. この検出率を計算するために, 10 人の人物に対して同様の実験を行った. その結果, 8 人の検出に成功した. 失敗の原因は, 個人差によって頭部周辺のエッジが異なるため, 閾値以上の重み得られなかったことがあげられる. そのほかにも原因があると考えられるが, この考察は今後の課題とする.

次に, 追跡の終了の様子について説明する. 同図 (b) に示すように, 検出と同様に, 追跡の終了も行えていることが分かる. これは, ゲートを通過したことを検出していることを示し, 安定に追跡が行えていることを示す結果でもある.

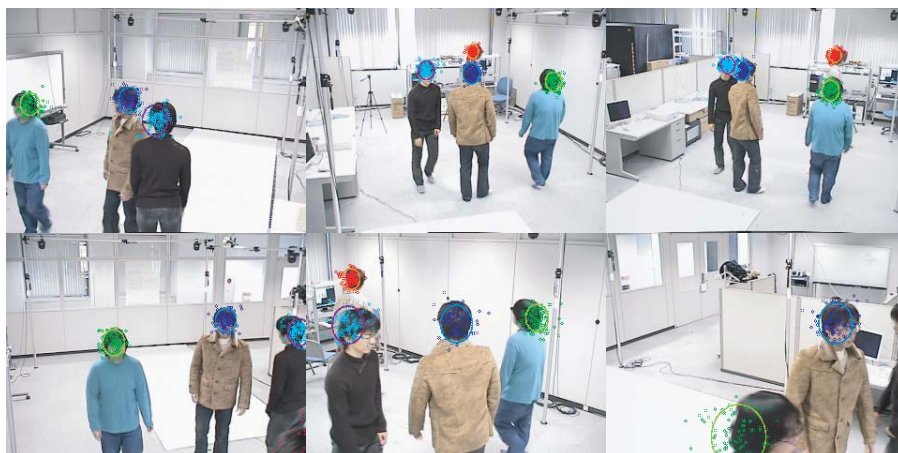
5.3 複数人物の追跡

本節では 4.4 節で示した方法で複数人物の追跡を行った結果について示す.

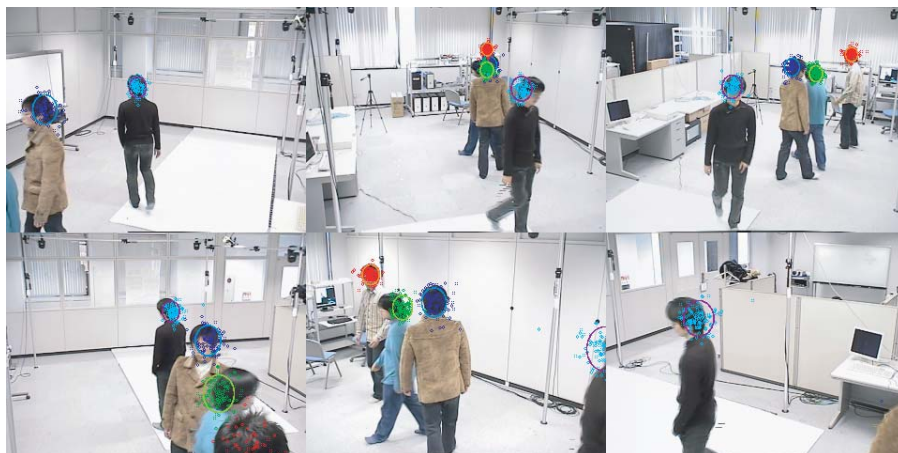
図 18, 図 19 に追跡結果を示す. このように, 複数



(a) 0 frame



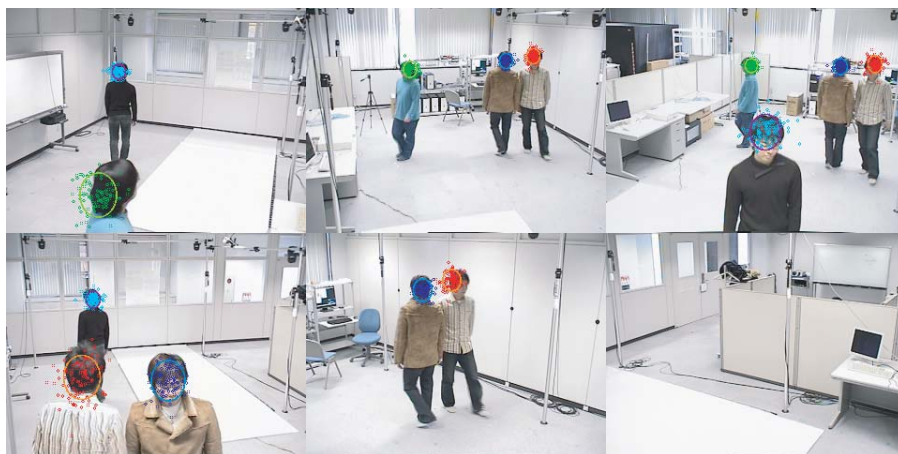
(b) 50 frame



(c) 100 frame

図 18 複数人物の追跡

Fig. 18 Tracking of multiple people.



(d) 150 frame



(e) 200 frame



(f) 250 frame

図 19 複数人物の追跡

Fig. 19 Tracking of multiple people.

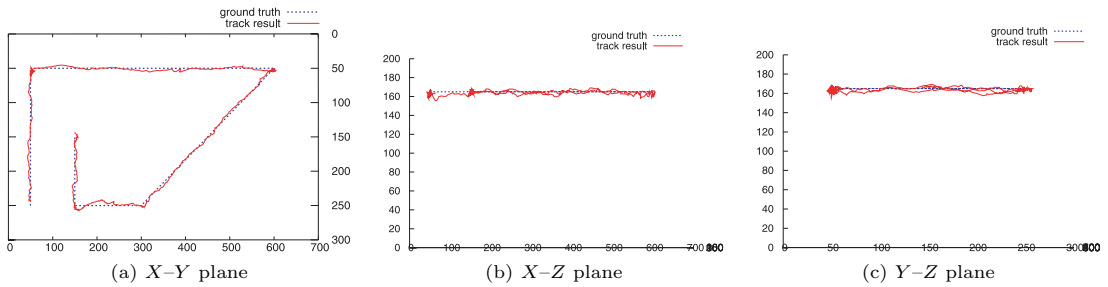


図 21 真値と追跡結果の比較 (cm)

Fig. 21 Comparison of tracking result with ground truth (cm).

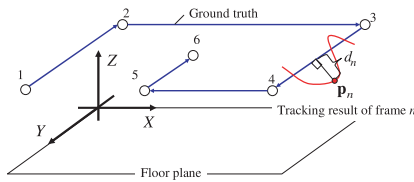


図 20 誤差の計算

Fig. 20 Calculate tracking error.

の人物それぞれの分布を推定することによって正しく追跡できていることが分かる。さらに、100 frame の画像では上段中央のカメラ画像で 1 人の人物が他の人物にほとんど隠された状態になっているが、追跡に失敗していないことが分かる。これは、提案手法によってオクルージョンに対してロバストな追跡が行えている結果である。このように、我々が提案した重み統合法が人物同士のオクルージョンにも対応できたことが分かった。

また、人物同士が近づいた状態になってもそれぞれが干渉することなく追跡できていることも分かる。これは 150 frame の画像の 2 人の人物が近寄っていることからいえる。この後もそれぞれの人物の分布が重なったり入れ替わったりすることなく、追跡できている。このように、本研究で提案した分布の重なりを除去方法は有効であるといえる。

5.4 追跡精度

追跡誤差を求める方法として、従来研究では次の 2 つが用いられていることが確認できた。

- (1) 画像上の対象位置から 3 次元復元を行うことでそれを真値とする方法
- (2) あらかじめ決められた場所を人物が歩くことで、その場所を真値とする方法

前者の方法は鈴木らの手法⁽⁵⁾で用いられ、同時刻の画像が必要であり、カメラの撮影タイミングの同期が必要である。これに対し、後者の方法は森らの手法⁽⁴⁾で用いられ、実際の 3 次元位置との差があることが問

表 3 追跡誤差 (cm, 有効桁数小数点以下 3 桁)

Table 3 Tracking error (cm, number of significant figures is three places of decimals).

	平均誤差	最小誤差	最大誤差
全体	4.413	0.019	34.727
人物 1	4.263	0.019	14.832

題であるが、画像の撮影に同期が必要でない。我々のシステムでは画像の同期をとっていないため、後者の方法を用いる。

図 20 の ground truth で示すような場所を人物がたどるように移動し、このとき得られたあるフレーム n での追跡結果 p_n に対して、これらに対応する真値とする直線との距離 d_n を全フレームに対して計算し、その平均を誤差とした。また、高さ方向 Z に対する真値は人物の身長から、本研究で用いる仮説の精円モデルの長軸の長さ l_0 を差し引いた値を用いた。

以下では、まず 1 人を追跡した結果についての誤差を示し、次に複数人物を同時に追跡したときの追跡誤差を示す。

5.4.1 人物 1 人の追跡誤差

1 人の人物を追跡した結果と真値を示したのが図 21 である。同図では、直線が真値であり、その周囲にある線が追跡結果である。図 21 から、 X, Y 方向に対しての精度は良いが、 Z 方向の精度がそれほど良くないように感じられる。

今回は人物の違いによって誤差が変化するかを調べるために、この追跡実験を 10 人の人物それぞれ 1 人ずつに対して行った。これらを用いて計測誤差を計算した結果の一部を表 3 に示す。表 3 から、最小で 0.019 cm, 平均 4.632 cm という非常に良い精度であることが分かった。

5.4.2 複数人物の追跡誤差

人物 1 人の追跡誤差の計算と同様にして、複数人物を同時に追跡するときの誤差を計測した。ただし、人物がその場所を繰り返し移動できるように、図 20 の

表 4 6 人を追跡したときの追跡誤差 (単位 cm)

Table 4 Error of tracking 6 peoples.

	平均誤差	最小誤差	最大誤差
全体	4.225	0.021	18.491
人物 1	4.993	0.194	18.491

1→2→3→4→1 という経路を設定した。今回は 6 人の同時追跡実験を行い、この結果に対して誤差を計算した。

追跡実験においては、1 人の人物のときの追跡と同様の結果がそれぞれの人物に対して得られ、この追跡結果をもとに、追跡誤差を計算した結果の一部を表 4 に示す。この表から、追跡の精度は人物の増加による隠れの発生などに影響されないことが分かる。

6. まとめ

本研究では、複数視点を用いた 3 次元追跡法として CONDENSATION の拡張法を用いた。この拡張法では 3 次元の仮説とその重みによって事後分布を近似することで追跡を行う。しかし、3 次元の仮説の重みは、複数の 2 次元画像から 2 次元の重みを計算し、これらを統合することで推定する必要がある。従来法では、この統合が積であったため、オクルージョンの影響によって追跡に失敗するという問題があった。我々は、この問題に対して 2 次元の重み統合の詳細な検討を行い、それによる追跡実験を行った。これにより、提案した統合方法がオクルージョンに対して有効であることを示した。さらに、実用的なシステムを構築するために、自動的に人物の検出・追跡の終了を行うように実装し、その有効性を示した。最後に、複数人物への適用を行い、他の人物によるオクルージョンや、人物間での追跡の入れ替わりや重なりがないことを確認した。

提案手法による実験結果から、検出率の向上が今後の課題にあげられる。また、追跡の成功率などが不十分であると考えられるため、これらを計算するための大規模な実地実験を行っていきたい。

今後の予定として、システムの拡張性を考慮する必要がある。本論文で示した追跡手法と、他のセンサを用いた追跡手法を組み合わせた方法を考えることで、それぞれの追跡を補い合い、追跡を安定に行えと考える。この場合、システム全体のスケーラビリティを持たせるためのネットワークの構成や通信方法について検討する。

参考文献

- 1) Matsuyama, T.: Cooperative Distributed Vision—Dynamic integration of visual perception, action, and communication, *Image Understanding Workshop*, Monterey, CA, pp.365–384 (1998).
- 2) Yachi, K., Wada, T. and Matsuyama, T.: Human head tracking using adaptive appearance models with a fixed-viewpoint pan-tilt-zoom camera, *4th International Conference of Automatic Face and Gesture Recognition*, pp.150–155 (2000).
- 3) Ukita, N. and Matsuyama, T.: Incremental observable-area modeling for cooperative tracking, *15th International Conference on Pattern Recognition (ICPR2000)*, pp.192–196 (2000).
- 4) 森 大樹, 内海 章, 大谷 淳, 谷内田正彦, 中津良平: 非同期多視点画像による人物追跡システムの構築, 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.1, pp.102–110 (2001).
- 5) Chiuso, A., Favaro, P., Jin, H. and Soatto, S.: Structure from motion causally integrated over time, *IEEE Trans. PAMI*, No.39, pp.523–535 (2002).
- 6) 鈴木達也, 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 D-II, Vol.J88-D-II, No.8 (2005).
- 7) Oka, K., Sato, Y., Nakanishi, Y. and Koike, H.: Head pose estimation system based on particle filtering with adaptive diffusion control, *IAPR Conf. Machine Vision Applications*, pp.586–589 (2005).
- 8) Hamasaki, K., Nakajima, T., Okatani, T. and Deguchi, K.: Tracking multiple three-dimensional motions by using modified condensation algorithm and multiple images, *International Conference on Intelligent Robots and Systems*, Las Vegas, Nevada, pp.236–241 (2003).
- 9) Qian, G., Chellappa, R. and Zheng, Q.: A Bayesian approach to simultaneous motion estimation of multiple independently moving objects, *16th International Conference on Pattern Recognition (ICPR2002)* (2002).
- 10) Qian, G. and Chellappa, R.: Structure from motion using sequential monte carlo methods, *International Journal of Computer Vision*, pp.5–31 (2004).
- 11) Isard, M. and Blake, A.: Condensation—conditional density propagation for visual tracking, *International Journal of Computer*

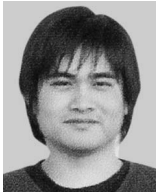
Vision, Vol.29, No.1, pp.5–28 (1998).

- 12) Lanz, O.: Occlusion robust tracking of multiple objects, *Proc. International Conference on Computer Vision and Graphics* (2004).
- 13) Nguyen, H.T., Worring, M. and van den Boomgaard, R.: Occlusion robust adaptive template tracking, *Proc. IEEE Conf. on Computer Vision, ICCV2001*, pp.678–683 (2001).
- 14) MacCormick, J. and Blake, A.: A probabilistic exclusion principle for tracking multiple objects, *International Journal of Computer Vision*, No.39, pp.57–71 (2000).
- 15) Khan, Z., Balch, T. and Dellaert, F.: Mcmc-based particle filtering for tracking a variable number of interacting targets, *IEEE Trans. PAMI*, No.11, pp.1805–1819 (2005).

(平成 18 年 9 月 10 日受付)

(平成 19 年 3 月 20 日採録)

(担当編集委員 村上 和人)



松元 郁佑

1982 年生 . 2006 年和歌山大学大学院システム工学研究科システム工学専攻博士前期課程修了 . 同年同大学院後期課程入学 . コンピュータビジョンの研究に従事 .



加藤 文和

1974 年生 . 1997 年岡山大学工学部情報工学科卒業 . 2001 年同大学院博士課程修了 . 2001 年 4 月から 2002 年 12 月まで産業技術総合研究所 . 2003 年 1 月より和歌山大学に勤務 . コンピュータビジョン , パターン認識等の研究に従事 . 博士 (工学) . 電子情報通信学会 , IEEE 各会員 .



和田 俊和 (正会員)

1959 年生 . 1987 年東京工業大学大学院修士課程修了 . 1990 年同大学院博士課程修了 . 同年岡山大学工学部助手 . 1994 年同大学大学院自然科学研究科助手 . 1995 年同大学工学部講師 . 1997 年京都大学大学院工学研究科助教授 . 1998 年同大学院情報科学研究科助教授 . 2002 年から和歌山大学システム工学部教授 . 工学博士 . 画像理解 , パターン認識の研究に従事 . 1995 年 David Marr 賞 , 1997 年情報処理学会山下記念研究賞 , 1999 年電子情報通信学会論文賞各受賞 . 人工知能学会 , 電子情報通信学会各会員 .