

## 複数カメラを用いた CONDENSATION による複数人物頭部の実時間検出・追跡

松元 郁佑<sup>†</sup> 加藤 丈和<sup>†</sup>  
和田 俊和<sup>†</sup> 上田 博唯<sup>††</sup>

本論文では、複数の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、各2次元画像上で追跡を行い、これを統合することで3次元位置を追跡する方法が多かったが、このような方法では各画像上で求められる対象位置が3次元空間中の同一物体に対応しないという食い違いが生じる。本論文では、この問題を解決するために、3次元空間中で追跡対象に関する仮説を生成し、それらを各画像に投影して尤度を求め、それらを統合するというCONDENSATIONの自然な拡張法について検討する。この手法では、各画像で求める尤度の統合を積によって行うべきであるが、オクルージョンが発生した場合には追跡ができない。このため、本論文では特に、尤度の統合法について詳しく検討し、新たな尤度統合方法を提案する。提案手法によるシステムを、カメラを接続した計算機6台を用いて試作し、オクルージョンに対する頑健さ、複数人物の追跡に対する安定さを実験的に確認した。

### Real time tracking and detection multiple peoples using multiple cameras based on CONDENSATION

YUSUKE MATSUMOTO,<sup>†</sup> TAKEKAZU KATO,<sup>†</sup> TOSHIKAZU WADA<sup>†</sup>  
and HIROTADA UEDA<sup>††</sup>

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, we investigate a natural extension of CONDENSATION to multi-viewpoint images. Our method generates many hypotheses on a target (human head) in 3D space and estimates the likelihood of each hypothesis by integrating viewpoint dependent likelihood values of 2D hypotheses projected on image planes. In theory, view point dependent likelihood 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次元空間中の人物の実時間追跡は、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としてその実現が待ち望まれている。特に、複数カメラを用いた人物追跡技術は対象を最も良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。

(1) 各カメラで撮影された画像平面上で追跡を行い、その情報を統合することで対象の3次元位置を推定するという手法<sup>1)</sup>がある。この手法では、各カメラによる追跡結果(2次元位置など)を統合するため、カメラ間で交換する情報が少なく、カメラ台数を容易に増やすことが可能である。一方、各カメラでの追跡が失敗した場合、異なるカメラで撮影された画像上の対象位置が単一の3次元位置に対応しなくなるため、推定される対象の3次元位置が不安定になる。これは、複数の画像平面上での追跡結果を統合する手法にとって不可避的な問題である。

(2) 各画像平面上で追跡を行うのではなく、3次元空間中に対象モデルを仮定し、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がある。

<sup>†</sup> 和歌山大学システム工学研究科  
Graduate School of Systems Engineering  
<sup>††</sup> NiCT

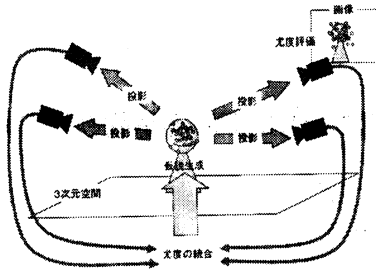


図1 NAMA-CONの概要  
Fig.1 outline of NAMA-CON

例えば、森らの手法<sup>2)</sup>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、(1)の手法と異なり、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。

この問題点を解決したのが、鈴木らによる手法<sup>3)</sup>や岡らによる手法<sup>4)</sup>である。これらの手法はCONDENSATION<sup>5)</sup>を3次元空間中で適用することで、多峰性の確率分布推定による追跡を行う。この手法では、確率分布を表す仮説の評価を各画像で行い、それらの積によって尤度を推定しているため、一部のカメラでのオクルージョンの発生には対処できない。

本研究では(2)のような、3次元空間中の多峰性の確率分布を推定する手法のアプローチを元に、CONDENSATIONの自然な拡張法である、NAMA-CON (Network Augmented Multisensor Association CONDENSATION)を提案する。この手法は、基本的には前述の手法と同様に、追跡対象の3次元空間における事後分布を推定する方法である(図1参照)。本論文では、特に前述のオクルージョンの問題に対処するために、各画像上で得られる仮説の尤度統合法を詳細に検討し、オクルージョンに対して頑健な尤度統合法を提案する。さらに、人物頭部を対象として、確率分布推定を用いた仮想ゲートでの対象検出ならびに複数対象の実時間追跡を行うことができることを実証した。

## 2. NAMA-CON

本章ではCONDENSATION<sup>5)</sup>アルゴリズムを簡単に

説明し、その拡張法であるNAMA-CONを提案する。

### 2.1 CONDENSATION アルゴリズム

本研究は人物の頭部を追跡することを目的としており、求める頭部の位置を3次元実空間中の確率分布として推定する。この確率分布をロバストに推定する手法としてCONDENSATIONを適用した。本節では、CONDENSATIONの簡単な説明をする。

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を時系列で推定することで追跡を実現する。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 $\mathbf{Z}_t = \{\mathbf{Z}_1, \dots, \mathbf{Z}_t\}$ とおく。このとき、 $\mathbf{X}_t$ の事後分布 $p(\mathbf{X}_t | \mathbf{Z}_t)$ はベイズの定理により、次のように変形できる。

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

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、 $p(\mathbf{X}_t | \mathbf{Z}_{t-1})$ は事前分布である。ここで、 $p(\mathbf{X}_t | \mathbf{Z}_{t-1})$ は前の時刻 $t-1$ の事後分布 $p(\mathbf{X}_{t-1} | \mathbf{Z}_{t-1})$ 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。

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

この手法では、事後分布をランダムサンプリングとそれによって発生されるサンプル点の尤度推定によって離散的に近似することで、事後分布を効率良く推定する。これは、各時間 $t$ における、 $N$ 個の仮説 $\mathbf{s}_t^{(n)}$ とその尤度評価によって得られる重み $\pi_t^{(n)}$ を用いて次式のように表される。

$$p(\mathbf{X}_t | \mathbf{Z}_t) \approx \sum_{n=0}^{N-1} \pi_t^{(n)} \delta(\mathbf{s}_t^{(n)}) \quad (3)$$

ただし、 $\delta$ はディラックデルタ関数である。

### 2.2 NAMA-CONの概要

本節では、前節で示したCONDENSATIONを3次元実空間に適用したNAMA-CONについて説明する。

提案手法は、図1に示すように、3次元空間中の位置を状態 $\mathbf{X}_t$ として、事前分布 $p(\mathbf{X}_t | \mathbf{Z}_{t-1})$ に従ったランダムサンプリングにより、3次元空間中の仮説を生成し、それらを各カメラ座標に投影して尤度推定する(カメラ $k$ での画像特徴を $\mathbf{z}_{t(k)}$ とする)。さらに、この尤度を統合することによって、元の3次元の仮説の尤度 $p(\mathbf{Z}_t | \mathbf{X}_t)$ を推定する。これらの仮説と統合した尤度によって、確率分布 $p(\mathbf{X}_t | \mathbf{Z}_t)$ の推定を行うこ

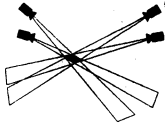


図2 尤度評価の統合  
Fig. 2 likelihood integration

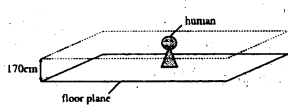


図3 人物頭部が存在する平面  
Fig. 3 a plane which exists human head

とが追跡の目的である。

提案手法では、3次元空間中に仮説を発生させるため、3次元空間から2次元の各カメラ座標への変換により各画像の対応する画素を計算でき、各カメラ間の画像情報が食い違わない。このような計算によって、人物頭部の3次元空間中の位置を正しく追跡できる。さらに、各カメラの情報を有効に扱うことによってオクルージョンや背景などの周囲環境にロバストな追跡が可能である。

### 3. 尤度統合

前章で示したように提案手法では、3次元空間中での頭部位置の仮説生成し、それらを各カメラの画像座標に投影して評価する。これは各カメラから見える情報のみを用いて推定した尤度であり、3次元空間中の仮説の尤度を推定するためには、これらを統合する必要がある。

図2はこの尤度統合の概念を示したものである。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、つまり尤度が高い部分を表しており、全ての帯の共通部分は3次元空間中での対象人物の位置を表わしている。つまり、このような位置で尤度が高くなり、それ以外では低くなるような尤度統合方法を考える必要がある。

以下、まずオクルージョンの無い単純な状況での尤度統合について議論し、そのあとでオクルージョンが生じる場合の尤度統合方法について議論する。

#### 3.1 オクルージョンを考慮しない尤度統合

本節では、オクルージョンがない場合の尤度統合について考察する。カメラ  $k$  において推定される尤度は、前章で示した式を用いて  $p(\mathbf{z}_{t(k)}|\mathbf{X}_t)$  と表わす。このとき、各カメラからの観測が独立であるとする、全てのカメラにおける観測  $\mathbf{Z}_t$  から推定される尤度  $p(\mathbf{Z}_t|\mathbf{X}_t)$  は、同時確率として次のように計算できる。

$$p(\mathbf{Z}_t|\mathbf{X}_t) = \prod_k p(\mathbf{z}_{t(k)}|\mathbf{X}_t) \quad (4)$$

この方法では、各カメラから観測する場所に遮蔽物などでオクルージョンのない場合は安定な値を推定でき

るが、オクルージョンが生じた場合には全体の値として安定したものが得られない。その様子を表したのが図4(a)(b)である。

この図は、図3のような人物頭部が存在する高さにおける、床平面と平行な平面に対して後述する尤度評価を行った時の結果を画像の輝度として表現したものであり、カメラ6台を用いた結果を示す。また、人物頭部が存在する付近(同図の四角形で囲んだ部分)を右下に拡大して表示した。

図4(a)はオクルージョンが生じていない場合であり、同図(b)は2台のカメラでオクルージョンが生じた場合の結果である。(a)の対象位置にはある程度大きなピークが現れているが、(b)では、オクルージョンの影響によって対象位置と他の場所の差がほとんど見られない。あるいは、オクルージョンの影響で値自体も小さいため、追跡が不安定になる。このように積によって統合すると、オクルージョンに弱くなる。

#### 3.2 オクルージョンを考慮した尤度統合

前節で示したように、全てのカメラの積によって統合すると、オクルージョンに非常に弱くなってしまふ。ここで、どのカメラでオクルージョンが起こっているかが分かっている仮定すると、オクルージョンが生じていないカメラから得られる尤度の積が正しい値であるということは明らかであるが、オクルージョンの有無を画像情報だけを用いて動的に推定することは非常に困難である。つまり、どのカメラの情報を用いるべきか分からない状況で尤度を推定しなければいけない。

##### 3.2.1 相互生起性

このような状況に対応するために次のように考える。まず、カメラが4台あるとして、カメラ1から見える  $(C_1)$ 、...、カメラ4から見える  $(C_4)$  の4つが起こりえる事象として考えられる。ここで、求める確率は、どのカメラから対象を観測できるかが分からないので、 $C_1$  から  $C_4$  ままで相互に起こり得る、相互生起確率で表される。これらの事象が独立であるとする、次のように表される。

$$p(C_1 \cup C_2 \cup C_3 \cup C_4) = p(C_1) + p(C_2) + p(C_3) + p(C_4) \quad (5)$$

ただし、1つずつの事象の相互性だけではなく、同時性も考えなければならない。

##### 3.2.2 同時生起性

あるカメラから対象を観測できるかどうかの情報は、その事象が相互に生起する確率として表現できることを前小節で示した。本小節では、あるカメラ同士が同時に見えていることについて考察する。これは、 $C_1$

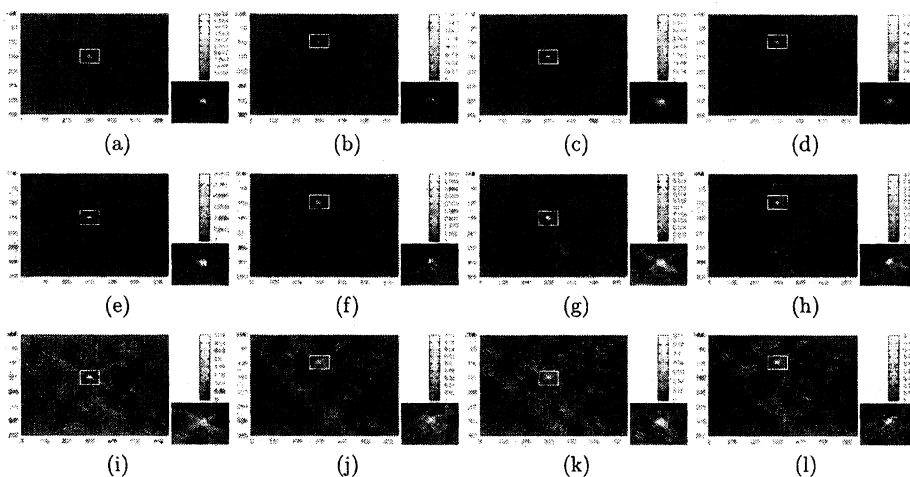


図 4 尤度統合の比較: (a)(c)(e)(g)(i)(k) オクルージョンがないとき, (b)(d)(f)(h)(j)(l) オクルージョンがあるとき, (a)(b) カメラの積によって統合した場合, (c)~(j) 同時に生起するカメラの数を変えた場合, (k)(l) 提案手法

Fig. 4 method of likelihood integration: (a)(c)(e)(g)(i)(k) excluding occlusion, (b)(d)(f)(h)(j)(l) including occlusion, (a)(b) case of product of each camera's likelihood, (c)-(j) case of changed number of simultaneous occurrence camera, (k)(l) our method

から  $C_4$  が同時に起こり得る場合, 全事象の中から 3 つが同時に起こり得る場合, 同様に 2 つの場合, となり, この同時生起確率で表される. 例えば, 2 つの事象が同時に起こり得る場合は,  $p(C_1 \cap C_2)$  のような同時生起確率を考え, それらが独立に起こり得るので, 求めたい確率はこれらが相互に起こり得る確率になる. これを表したのが図 4(c)~(j) である. この図も図 4(a)(b) と同様の図であり, (c)(d) が 5 つのカメラから同時に見えていると仮定した場合, (e)(f) は 4 つ, (g)(h) は 3 つ, (i)(j) 2 つの場合を表す. このように, 同時に見えるカメラの数を変えることで, 全体の評価値が変わることが分かる.

ここで, 対象を同時に観測できるカメラの台数は不明であるため, 本来はこれら全ての可能性の組合せを計算するべきであるが, 計算量が膨大になり, また異なる台数の統合結果を単純に統合できない. そこで, 本研究では最小限の組合せを用いて近似的に推定する.

三角測量の原理から, 3次元位置を測定するのに必要な最小の 2 台のカメラから見えている場合を, 一般的に起こり得る同時生起の事象として用いる. このとき, 最低 2 つのカメラから見えていればこの統合によって尤度推定が可能である. この統合方法は以下のように表すことができる.

$$p(\mathbf{Z}_t | \mathbf{X}_t) = \sum_{i,j,i \neq j} p(\mathbf{z}_{t(i)} | \mathbf{X}_t) p(\mathbf{z}_{t(j)} | \mathbf{X}_t) \quad (6)$$

このように統合した結果が図 4(i)(j) である. オクルージョンが生じても他のカメラの情報を有効に扱うことによって, 対象位置を推定できることが分かる.

ここで, 対象位置周辺を注目したい. 周辺には正しい位置以外の場所でも放射状に値の大きな場所が現れている. これは, ある 2 つカメラを用いたときに, そのカメラ 2 台と対象の位置がほぼ一直線上にある場合に起こる. つまり, この 2 つのカメラの組合せは 3次元位置が計測できない位置関係にあるために起こる. そこで, このような組合せを用いて尤度を統合した. その結果が図 4(k)(l) である. 図 4(i)(j) で現れていた対象以外のピークがほぼ除去できていることが分かる.

以上のように考察を行った尤度の統合によって, 本研究で追跡に用いる尤度を推定する.

#### 4. 人物の検出・複数人物の追跡

ユビキタスホームなどのより一般的な環境下では, 人物の出入りを検出したり, 複数の人物を追跡する必要である. 本章ではこれらについての考察を行う.

##### 4.1 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. 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 5 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, いなくなることはないという前提条件を加える.

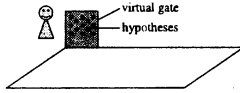


図5 人物の検出  
Fig. 5 human detection

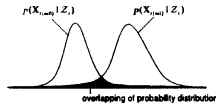


図6 分布の重なり  
Fig. 6 overlapping of probability distribution

まず、人物の初期検出について説明する。これは図5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、追跡時と同様の尤度推定を行う。この尤度が大きくなったときに人物が近付いたときだと考え、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じた時や、追跡範囲から出たことによる追跡の失敗を検出すれば良い。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。

#### 4.2 複数人物の追跡

複数人物の追跡はそれぞれの人物に対し事後分布を推定することで行う。このとき、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。そこで、我々は次のような方法で、それぞれの分布が交わらないようにした。

まず、各人物に対する仮説群が混ざる状態とは、ある人物  $m_0$  に対する分布  $p(\mathbf{X}_{t(m_0)}|Z_t)$  と、その他の人物  $m_1$  に対する分布  $p(\mathbf{X}_{t(m_1)}|Z_t)$  が図6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  $m$  の分布は複数の仮説  $s_{t(m)}^{(n)}$  とその重み  $\pi_{t(m)}^{(n)}$  によって表現され、これら仮説群が混ざりあってしまう。

そこで、多分布に正規分布を当てはめ、ある仮説とそれ以外の分布とのマハラノビス距離を計算することで、その仮説が他の分布に対して重なりあっているかを判定する。ある人物  $m$  に対する分布の平均を  $\bar{s}_m$ 、分散を  $\sigma_m^2$  とすると、これらは仮説と重み ( $\sum_n \pi_{t(m)}^{(n)} = 1$  とする) による重み付け平均、及び重み付け分散として表現できる。

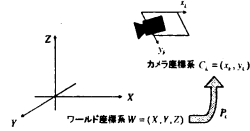


図7 ワールド座標系とカメラ座標系

Fig. 7 world coordinate system and camera coordinate system

$$\begin{aligned} \bar{s}_m &= \sum_n s_{t(m)}^{(n)} \pi_{t(m)}^{(n)} \\ \sigma_m^2 &= \sum_n (s_{t(m)}^{(n)} - \bar{s}_m)^2 \pi_{t(m)}^{(n)} \end{aligned} \quad (7)$$

これを用いた人物  $m$  に対する分散共分散行列を  $\Sigma_{t(m)}$  とすると、人物  $m_0$  の仮説  $n$  がどれだけ人物  $m_1$  の分布に重なっているのかを示す、マハラノビス距離  $d_{m_0 m_1}$  は、次のように表される。

$$d_{m_0 m_1} = \left( s_{t(m_0)}^{(n)} - \bar{s}_{m_1} \right)^T \Sigma_{t(m_1)}^{-1} \left( s_{t(m_0)}^{(n)} - \bar{s}_{m_1} \right) \quad (8)$$

この  $d_{m_0 m_1}$  が、本来重なるべき分布  $m_0$  とのマハラノビス距離  $d_{m_0 m_0}$  よりも小さければ重なっていると判定する。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。この判定で重なっているとされた仮説は尤度を0にする。

### 5. 人物頭部追跡システムの構築

本章では、NAMA-CONによる人物の頭部追跡システムの構築について説明する。

#### 5.1 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7のようになる。ここで、ワールド座標系と各カメラ座標系の関係についてはあらかじめ校正済みである。また、この関係は3次元から2次元への透視投影変換によって表現する。

このとき、同次座標で表現したワールド座標上のある点  $\mathbf{a} = (X, Y, Z, 1)^T$  に対応するカメラ  $C_k$  ( $k$ はカメラ番号) 座標系の点  $\mathbf{b}_k = (x_k, y_k, w_k)^T$  は、キャリブレーションで得られたワールド座標系とカメラ座標系との関係を表す投影行列  $P_k$  によって次のように表現することができる。

$$\mathbf{b}_k = P_k \mathbf{a} \quad (9)$$

また、本研究では人物頭部を表現するモデルとして、図8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で、

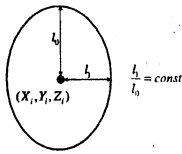


図 8 楕円モデル Fig. 8 ellipse model



図 9 用いる楕円上の法線 Fig. 9 normals of ellipse



図 10 楕円上の法線 N とエッジ勾配 E の関係 Fig. 10 relationship of normal at ellipse N and edge gradient E

これを一つの仮説として追跡する。

### 5.2 仮説の尤度評価

ある仮説  $s_{t(m)}^{(i)}$  がカメラ  $C_k$  座標系に投影されたとき、この仮説のパラメータ  $(x_{k(i)}, y_{k(i)}, w_{k(i)})$  は 5.1 節に示した計算で得られる。これを用いて、各カメラ画像における各仮説の尤度を評価する。

本研究では、図 9 のように、仮説の評価に楕円上の点(これを輪郭点  $\mu$  とする)における法線  $N$  とエッジ勾配  $E$  の内積を用いる。つまり、求めたいカメラ  $k$  における仮説  $s_{t(m)}^{(i)}$  の尤度  $\pi_{t(m,k)}^{(i)}$  は次のようになる。

$$\pi_{t(m,k)}^{(i)} = \sum_{\mu} \frac{N_{\mu} \cdot E_{\mu}}{\|N_{\mu}\| \|E_{\mu}\|} \quad (10)$$

この尤度  $\pi_{t(m,k)}^{(i)}$  は図 10 に示すように、 $E$  と  $N$  が同一方向を示すとき最大となる。つまり、仮説  $s_{t(m)}^{(i)}$  が表す楕円がどれだけその画像にフィットしているかという度合を表す。本研究で用いた輪郭点は図 9 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。このように尤度を評価することにより、頭部の追跡を安定に行うことができる。

## 6. 実験結果

本研究では、前章の NAMA-CON による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、図 11 に示すようにカメラを配置し、それぞれ CPU が Pentium 4 3.6GHz、メモリが 1Gbytes のスレーブ PC に接続されている。このスレーブ PC が仮説の評価を行う。また、図 12 に用いたカメラを示



図 11 実際のカメラの配置 Fig. 11 camera location

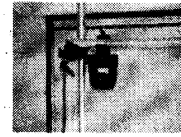
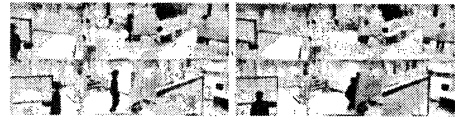


図 12 用いたカメラ Fig. 12 camera which used



(a.0)

(a.1)



(b.0)

(b.1)

図 13 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. 13 tracking result at scene that including occlusion : (a) production integration, (b) our method

す。さらに、仮説の生成と尤度の統合を行うマスター PC として Pentium D 3.0GHz、メモリ 512MBytes の PC を用いた。

以下では、本研究で示した手法について示し、追跡精度の計測結果について示すことで提案手法の有効性を示す。

### 6.1 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 13 にその結果を示す。同図では、人物の頭部付近に存在する点が各仮説を表し、大きな楕円が推定された頭部位置を示す。以下の追跡結果の画像ではこれを示す。

また、(a) の系列が各カメラで得られる尤度の積で統合を行った場合、(b) の系列が提案手法で、どちらも右側でオクルージョンが発生しているときを示す。この結果、積の場合ではオクルージョンに敏感に反応してしまうのに対し、提案手法では他のカメラの情報によって正しく追跡できている。このように、尤度統合の違いで追跡の性能が大きく変わることが分かる。

### 6.2 人物の検出・追跡の終了

4.1 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 14 に、人物の検出と追跡の終了における実験結

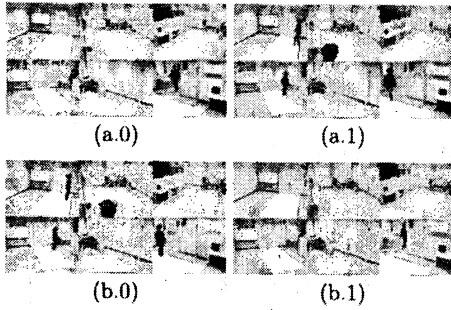


図 14 人物の検出と追跡の終了(左:処理前, 右:処理後): (a)人物の検出の様子, (b) 追跡終了の様子

Fig. 14 left side: before, right side: after, (a) detection (b) end of tracking

果を示す。同図上段は人物の検出の様子, 下段は追跡の終了の様子を表す。

はじめに, 人物の検出の様子について説明する。図 14 に, (a.0) から (a.1) に時間が経過したときの様子を示す。このように, 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わかる。ただし, 1 フレームで 1 人ずつしか検出できないため, 複数人物が同時にゲートを通することはできない。

次に, 追跡の終了の様子について説明する。同図に, (b.0) から (b.1) へ時間が経過したときの様子を示す。このように, 追跡の終了が行えていることが分かる。これは, ゲートの通過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。

このように, 提案手法を用いることで, 安定に人物の検出・追跡の終了が行えることが分かる。

### 6.3 複数人物の追跡

本節では 4.2 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 15 に追跡結果を示す。このように, 複数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一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, 追跡に失敗することなく継続できていることが分かる。これは, 複数のカメラの情報をおクルージョンに強い尤度統合によって得られた結果である。

また, 人物同士が近付いた状態になってもそれぞれが干渉すること無く追跡できていることも分かる。これは 150 frame の画像の二人の人物が近寄っていることから分かる。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, 追跡できていることが分かる。

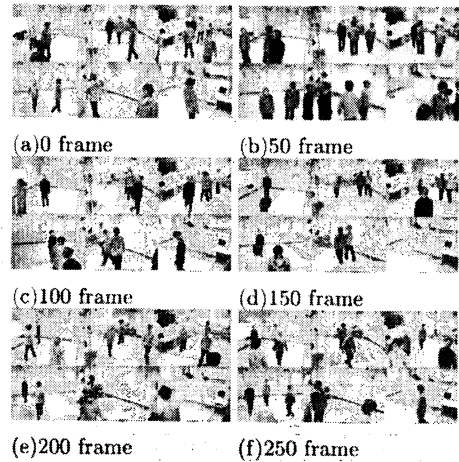


図 15 複数人物の追跡

Fig. 15 tracking of multiple human

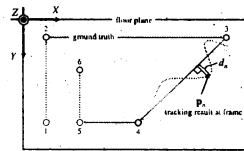


図 16 誤差の計算

Fig. 16 calculate tracking error

表 1 追跡誤差 (単位 cm, 有効桁数小数点以下 3 桁)

Table 1 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
人物 6	4.852	0.105	34.727

### 6.4 追跡精度

提案手法による追跡では, 人物頭部の 3 次元位置が推定できる。本節では, これがどの程度妥当に推定できているかを定量的に評価する。

図 16 に, 誤差の計算方法について示す。この図で示すように, 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, 5.2 節で示した楕円の長軸の長さ  $l_0$  を差し引いた値を用いた。

以下では, まず一人での追跡結果についての誤差を計算し, 次に複数人物を追跡したときの追跡誤差を計算する。

#### 6.4.1 人物 1 人の追跡誤差

まず一人の追跡によって得られた追跡誤差を計算す

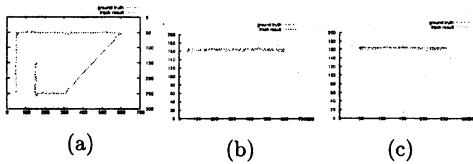


図 17 真値と追跡結果の比較:(a) $X - Y$  平面, (b) $X - Z$  平面, (c) $Y - Z$  平面

Fig. 17 comparison of tracking result with ground truth:(a) $X - Y$  plane, (b) $X - Z$  plane, (c) $Y - Z$  plane

表 2 6 人を追跡したときの追跡誤差  
Table 2 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
人物 5	3.734	0.021	11.495

る。本節のはじめに説明した誤差の計算を行うための追跡結果が図 17 であり、直線が真値であり、その周囲にある線が追跡結果である。見た目としては、 $X, Y$  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どうかを調べるために、この追跡実験を 10 人の人物に対して行った。それらを用いて計測誤差を計算した結果の一部を表 1 に示す。このように、最小では 0.019cm の精度が得られ、また平均 4.632cm の誤差という非常に良い結果が得られた。また、この結果には最大 30cm 以上の誤差が現れていることが分かる。これは、一時的な追跡の失敗による大きな誤差であり、一時的な追跡であればその後の追跡に失敗することはないため、問題ないと考えられる。

#### 6.4.2 複数人物の追跡誤差

人物一人の追跡誤差の計算と同様にして、複数人物を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 16 の  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$  という経路を設定した。今回は 6 人の追跡実験を行い、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、このような追跡結果を元に、追跡誤差を計算した結果の一部を表 2 示す。この表で示すように、人物が増えることによる隠れの発生に影響されないことが分かる。

このように、本研究による追跡では、1 人の追跡を行った場合でも、複数人物の追跡を行った場合でも、追跡誤差は変化すること無く追跡できることが分かった。このような結果は、本研究による追跡がユビキタ

スホームなどで実用可能なことを示している。

## 7. まとめ

本研究では、人物頭部の位置を 3 次元空間中の事後確率としてとらえ、CONDENSATION を 3 次元空間に拡張した NAMA-CON を提案した。また、本論文では、尤度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、およびその追跡性能について実験を行い、本手法の有効性を示した。

今後の展開として、システムの拡張性を考慮する必要がある。大規模システムなどに拡張する場合、通信のトラフィックが問題になる。あるいは、カメラ 2 台で撮影できない場所が存在する場合などには、カメラをアクティブに動作させる必要があると考える。このようなシステムのスケラビリティについて検討する。謝辞 本研究の一部は、文部科学省科学研究費補助金基盤研究 (A) (2) 1620014、及び、独立行政法人情報通信研究機構受託研究「ユビキタスホームのための人物行動監視システムの実装」の補助を受けている。

## 参考文献

- 1) N. Ukita and T. Matsuyama: "Incremental observable-area modeling for cooperative tracking", 15th International Conference on Pattern Recognition (ICPR2000), pp. 192-196 (2000).
- 2) 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, J84-D-II, 1, pp. 102-110 (2001).
- 3) 鈴木達也, 岩崎慎介, 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", 第 7 回画像の認識・理解シンポジウム (MIRU2004), 第 1 巻, pp. 243-248 (2004).
- 4) K. Oka, Y. Sato, Y. Nakanishi and H. Koike: "Head pose estimation system based on particle filtering with adaptive diffusion control", IAPR Conf. Machine Vision Applications, pp. 586-589 (2005).
- 5) M. Isard and A. Blake: "Condensation - conditional density propagation for visual tracking", International Journal of Computer Vision, 29, 1, pp. 5-28 (1998).