

## Network Augmented Multisensor Assosiation-CONDENSATION: CONDENSATION の自然な拡張による 3 次元空間内での人物頭部の実時間追跡

松元 郁佑<sup>†</sup> 加藤 丈和<sup>†</sup> 和田 俊和<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ていないという食い違いが生じる。この場合、推定される対象の 3 次元位置が大きく変動し、追跡が不安定になりやすい。そこで、我々は空間中で追跡対象の位置に関する仮説を生成し、それらを各画像座標に投影して尤度計算を行うという CONDENSATION の自然な拡張法を提案する。提案手法では、各画像上で求めた尤度の和を各仮説に対する尤度とすることで、オクルージョンや背景などの対象のコントラストが低い場合にも、全体として安定した追跡を実現する。この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、1 人の人物頭部の追跡を 5ms 間隔で実行できること、および上述の安定性を実験的に確認した。

### Network Augmented Multisensor Assosiation-CONDENSATION: Real time human head 3-D tracking by natural expansion of CONDENSATION

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

This paper presents a novel method for human head tracking using multiple cameras. Most existing method 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 propose 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. The distinguishing property of this method is that the viewpoint dependent likelihood is integrated by averaging instead of multiplication. This is because a viewpoint dependent likelihood can be reduced occlusion, which also reduces the integrated likelihood value by multiplication. We implemented a prototype system consisting of six set of PCs and CAMERAs. We confirmed the stability and the efficiency of our method using the prototype system.

#### 1. はじめに

3 次元空間内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、人物が行動する環境内に埋め込まれた複数のカメラ郡による人物追跡技術は追跡対象を最も良く観測できる視点を選択し、遮蔽などの影響を回避して安定な追跡が実現できるという可能性があることから、数多くの研究例があり、これらは 2 通りの手法に大きく分けることができる。

1 つめの手法は、各カメラの画像情報 (2 次元情報)

から追跡を行い、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手法<sup>1)</sup>である。これらの手法では、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、容易に拡張可能である。一方、各カメラの追跡失敗が生じた場合や、異なる視点で推定された対象位置が単一の 3 次元物体に対応しない場合、各カメラの追跡結果に矛盾が生じ、全体として追跡が安定しにくい。

2 つめの手法は、全てのカメラで得られた画像から 3 次元情報を復元し、それを用いて追跡する手法である。この場合、画像そのものを一台の PC に集約する必要がある。PC クラスタにより画像自体を PC 間で交換して追跡する手法<sup>2)3)</sup>では、高速な並列処理により、数台のカメラでは実時間の追跡や運動推定を実現している。本研究ではこのような 3 次元復元を目的と

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

せず、効率的な処理による3次元追跡を目的とする。

これらの手法は、2次元の画像情報から3次元空間情報を復元するアプローチを取っており、前述のような問題点が顕著に現れてしまうと考えられる。これらに対し、森らの手法<sup>4)</sup>では床平面に立った人物3次元モデルを表現し、画像のシルエット検出を行い、モデルとシルエットとの対応関係を用いることで人物の追跡を実現している。この手法では、カルマンフィルタによって追跡予測を行っているため、各カメラで用いることのできる画像特徴は単一になり、画像を有効に扱えないという問題点が残る。

これらに対して、我々は人物頭部の3次元空間中の位置を追跡対象の状態として考え、この事後確率を求めるアプローチを導入する。この考え方に従い、CONDENSATION<sup>5)</sup>の自然な拡張法である、NAMA-CON(Network Augmented Multisensor Association CONDENSATION)を提案する。これは、図1のように、3次元空間中での仮説生成と、それら各仮説を各画像平面へ投影して得られる2次元画像座標における尤度推定を行い、各カメラの尤度を統合することによって、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。また、尤度の統合によって、複数カメラの情報を有効に用いることができる。つまり、提案手法は従来手法のように各カメラで得られる画像情報を元に3次元位置を復元するのではなく、直接3次元空間中の確率分布を求めることで追跡を行う手法である。

このような3次元空間中の位置に対する事後分布を推定する方法は鈴木らによる手法<sup>6)</sup>でも用いられている。しかし、この手法ではステレオカメラを用いて3次元位置を計算し、3次元中に仮定したモデルと計測位置の距離と、その位置での画素値を用いて評価しているが、この方法でも画像情報からの3次元復元を行っており、また画角も狭くなるため観測範囲が狭くなる。これに対し、本研究では3次元空間と画像平面の対応を投影によって表現しているため、ステレオカメラで行う3次元位置の測定は必要なく、また普通の単眼カメラを用いるため、観測範囲は狭小にならない。

このように、本研究では先行研究とは異なるアプローチによって複数カメラを用いた3次元位置の推定法を実現する。本論文では、提案手法によって安定な追跡を実現し、追跡手法としては十分なフレームレートを得ることに成功したことを理論的、実験的に示す。

## 2. CONDENSATION の概要

本研究は人物の頭部を追跡することを目的としている。求める頭部の位置を3次元実空間中の確率分布

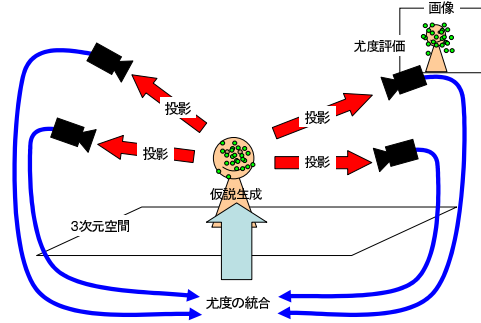


図1 NAMA-CON の概要  
Fig. 1 outline of NAMA-CON

として推定し、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。本章では、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 $t$ のパラメータを $\mathbf{X}_t$ と定義し、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 $\mathcal{Z}_t = \{\mathbf{Z}_1, \dots, \mathbf{Z}_t\}$ 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、 $p(\mathbf{Z}_t | \mathbf{X}_t)$ はパラメータ $\mathbf{X}_t$ 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。

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

$p(\mathbf{X}_t | \mathbf{X}_{t-1})$ は時間が経過するときの分布の推移確率を表している。

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。これにより、事後分布を効率良く推定する。つまり、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。

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

ただし、 $\delta$ はディラックデルタ関数である。

本研究では、サンプリング点 $\mathbf{s}_t^{(n)}$ を3次元空間中の頭部位置の仮説とし、それらを画像上に投影して尤

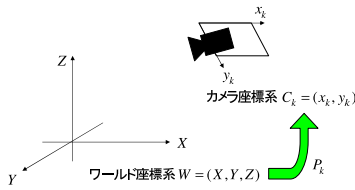


図2 ワールド座標系とカメラ座標系

Fig. 2 world coordinate system and camera coordinate system

度評価を行うことで人物頭部位置の推定を行う。

### 3. NAMA-CON

人物頭部の3次元空間追跡を行うためには、各画像の情報が必要である。複数カメラを用いた従来手法は第1章で示した通り、各画像の情報を元に3次元追跡を行う方法であった。各画像は2次元であるため、その情報を用いて3次元空間の情報を知るためには、2次元から3次元への逆変換が必要となり、これには複雑な処理が必要となる。

そこで、本研究では、3次元空間から2次元の画像平面への投影のみでこの対応関係を記述できるNAMA-CONを提案する。この方法は、前述のCONDENSATIONの自然な拡張法であり、以下、前章の式を元に提案手法について説明する。

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として、ランダムサンプリングを行い、3次元空間中の仮説を生成し( $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ の推定)、それらを各カメラ座標に投影して尤度推定(カメラ $k$ に投影した仮説を $\mathbf{x}_{t(k)}$ 、画像特徴を $\mathbf{z}_{t(k)}$ とすると、これは $p(\mathbf{z}_{t(k)} | \mathbf{x}_{t(k)})$ の推定になる)を行う。さらに、この尤度を統合することによって、各カメラの情報が統合される( $p(\mathbf{Z}_t | \mathbf{X}_t)$ の推定)。これら仮説と尤度によって求めたい確率分布 $p(\mathbf{X}_t | \mathcal{Z}_t)$ の推定が行われる。提案手法では、3次元空間中に仮説を発生させるため、3次元空間と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、人物頭部の3次元空間中の位置を正しく追跡できる。

本章では、このNAMA-CONについてシステム構築例を元に説明していく。

#### 3.1 座標系

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2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

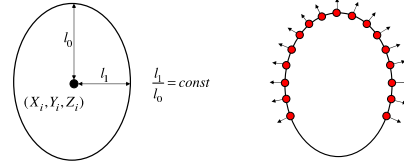


図3 楕円モデル

図4 用いる楕円上の法線

Fig. 3 ellipse model Fig. 4 normals of ellipse

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 $(X, Y, Z, 1)$ に対応するカメラ $C_k$ ( $k$ はカメラ番号)座標系の点 $(x_k, y_k, w_k)$ 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。

$$\begin{pmatrix} x_k \\ y_k \\ w_k \end{pmatrix} = P_k \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (4)$$

#### 3.2 モデル

本研究のシステムではPC間で追跡に用いる情報を通信する。この通信路で効率良く通信を行うためには、追跡対象を表現するモデルを簡単にする必要がある。複雑なものになれば、通信量が増大し、システムのスケーラビリティはもちろん、実時間性も失う可能性がある。

そこで、本研究では人物頭部を表現するモデルとして、図3のような回転楕円体モデルを用いる。このモデルは実空間上の地面と垂直な平面上に存在し、中心が $(X, Y, Z)^T$ 、長軸が $l_0$ 、短軸が $l_1$ の $Z$ 方向に細長い楕円である。このモデルを作成する際に、 $l_0$ と $l_1$ については人物の頭部を良く表現できるようにパラメータを経験的に設定した。この中心位置 $(X, Y, Z)^T$ の事後確率を求めることが追跡の目的となる。つまり、3次元実空間上でこの楕円モデルを一つの仮説 $s_t^{(n)} = (X, Y, Z)^T$ としたランダムサンプリングと、その評価を行うことで、離散的な事後確率を推定することが追跡の目的である。

#### 3.3 仮説の画像空間への投影

ある仮説 $i$ 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 $(x_{k(i)}, y_{k(i)}, w_{k(i)})$ について考える。このパラメータは第3.2節に示した楕円モデルをカメラ $C_k$ 座標系上に投影したときの中心座標であり、スケールの情報については含まれていない。そこで、ワールド座標系上の仮説の座標から $Z$ 方向にモデ

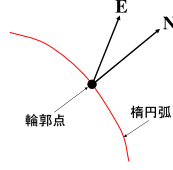


図 5 楕円上の法線  $N$  とエッジ勾配  $E$  の関係  
Fig. 5 relationship of normal at ellipse  $N$  and edge gradient  $E$

ルの長軸分だけ変化させた値を同様に線形変換させ、これによって得られた点  $(x'_{k(i)}, y'_{k(i)}, w'_{k(i)})$  と画像上での仮説のパラメータ  $(x_{k(i)}, y_{k(i)}, w_{k(i)})$  との距離を用いて仮説の画像上でのスケール  $l_{img0}$  を求める。ただし、カメラ座標系に変換した際に、 $w_{k(i)}$  項によって点を正規化しないといけないので、画像上での座標  $(p_{x_i}, p_{y_i})$  は  $(p_{x_i}, p_{y_i}) = (x_i/w_i, y_i/w_i)$  となり、その他も同様に変換した。以上のことから  $l_{img0}$  は次のように求めることができる。

$$l_{img0} = \sqrt{(p_{x_i} - p_{x'_i})^2 + (p_{y_i} - p_{y'_i})^2} \quad (5)$$

この  $l_{img0}$  は画像上での仮説の長軸を表し、楕円の偏平率を一定としていることから、短軸  $l_{img1}$  も求めることができる。

### 3.4 仮説の尤度評価

ワールド座標系の仮説  $i$  のパラメータをカメラ座標系に変換する方法について、節 3.3 に示した。本節ではこの座標を用いて画像特徴について尤度評価を行うことで、仮説  $i$  の尤度 (重み)  $\pi_i$  を求める方法について説明する。

本研究では、図 4 のような仮説の評価に楕円上の点 (これを輪郭点という) における法線とエッジ勾配の内積を用いる。

まず、仮説  $i$  の座標  $(p_{x_i}, p_{y_i})$  と、節 3.3 で得られた長軸、短軸によって以下の楕円方程式が得られる。

$$\begin{cases} x = l_{img1} \cos(\theta) \\ y = l_{img0} \sin(\theta) \end{cases} \quad (6)$$

この方程式により、ある  $\theta$  をとったときの楕円上の輪郭点  $\mu$  における法線方向  $N_\mu = (\partial y / \partial \theta, -\partial x / \partial \theta)$  が求まる。また、エッジ勾配は画像  $I(x, y)$  の  $x, y$  それぞれの方向についての微分によって計算する。輪郭点  $\mu$  上に置けるエッジ勾配を  $E_\mu = (\partial I(x, y) / \partial x, \partial I(x, y) / \partial y)$  とすると、求める尤度  $\pi_i$  は次のようになる。

$$\pi_i = \sum_{\mu} \frac{N_\mu \cdot E_\mu}{\|N_\mu\| \|E_\mu\|} \quad (7)$$

この尤度  $\pi_i$  は図 5 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、サンプル点  $i$  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 4 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。

### 3.5 尤度の統合

提案手法では、3 次元の仮説を各カメラ座標に投影し尤度推定を行う。これらは各カメラで独立した情報であるので統合する必要がある。この統合の様子を示したのが図 6 である。一般的にはこれは同時確率であり、各カメラで得られた尤度の積で推定し、本章のはじめに示したパラメータを用いて、以下のように表すことができる。

$$p(\mathbf{Z}_t | \mathbf{X}_t) = \prod_k p(\mathbf{z}_{t(k)} | \mathbf{x}_{t(k)}) \quad (8)$$

このように積で表す場合、図 6(2.a) のように全てのカメラで対象を観測できる場合は問題が無いが、同図 (2.b) のようにオクルージョンが生じた場合に、あるカメラで得られる尤度が急激に下がってしまい、全体の尤度も下がってしまうという問題がある。そこで、同図 (1.b) のように各画像の尤度の和によって統合することで、ある画像で尤度が下がった場合でも全体の尤度としては急激に下がらないようにする。これは、以下のように表すことができる。

$$p(\mathbf{Z}_t | \mathbf{X}_t) = \sum_k p(\mathbf{z}_{t(k)} | \mathbf{x}_{t(k)}) \quad (9)$$

この式は確率的には正しくなくなるが、求めたい尤度の分布がゆるやかになって安定化する。

## 4. PC 間の通信方法の検討

本章では各 PC 間の通信における問題点とその解決策を述べ、それぞれ具体的な方法を示す。本研究で用いるカメラと PC は図 7 のように接続されており、各 PC は 1Gbps イーサネットで接続されている。

### 4.1 各計算機における通信方法

本研究では、3 章に示したとおり、対象の位置を表す事後確率を求めるために、仮説の生成とその尤度評価を行う必要がある。そこで、これらを行う PC を分けて考える。つまり、ある PC がランダムサンプリングを行い、それを尤度評価するためにカメラと接続された PC に対し仮説を送信し、その仮説を受け取った PC が尤度評価を行う。

本研究では、仮説の生成を行うマスター PC と、その仮説の評価を行うスレーブ PC を用意した。つまり、

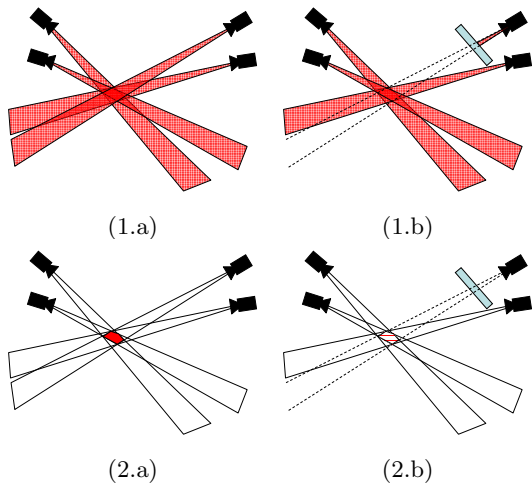


図 6 尤度評価の統合:(1.a) 尤度の和, オクルージョン無し. (1.b) 尤度の和, オクルージョン有り, (2.a) 尤度の積, オクルージョン無し, (2.b) 尤度の積, オクルージョン有り  
 Fig. 6 likelihood unification:(1.a)sum of likelihood, expect occlusion. (1.b)sum of likelihood, include occlusion. (2.a)product of likelihood, expect occlusion. (2.b)product of likelihood, include occlusion.

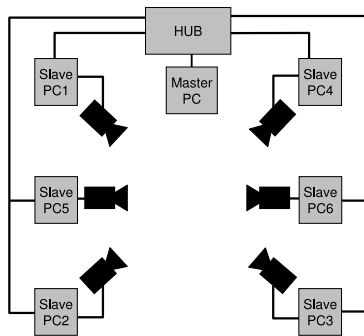


図 7 カメラと計算機の接続  
 Fig.7 connection between camera and PC

マスター PC とスレーブ PC との通信を考えることになり, スレーブ PC 同士での通信は考慮しなくても良い. この PC 構成によって行う通信のタイミングは, 理論的には図 8 のようになる. まず, マスター PC だけで前の時刻の事後確率 (仮説とその尤度評価によって得られる重みの集合) を用いた現在の時刻の仮説を発生させる. これを全スレーブ PC へと送信する. これを受け取ったタイミングでスレーブ PC で得られた画像特徴を評価し, これをそれぞれの仮説の重み (尤度) としてマスター PC へと返信する. このようにして得られた仮説と重みが求める事後確率になる.

このようにマスター PC とスレーブ PC の役割をはっきりさせることで, システムを単純な方法によ

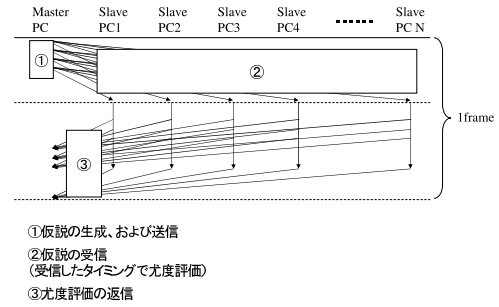


図 8 PC 間の通信タイミング  
 Fig. 8 timing of random sampling and unification of likelihood

て構築することができ, これが処理の負担分散とシステムをシンプルにすることにつながる.

#### 4.2 通信の安定化に関する検討

提案手法では, 仮説とその重みを通信する必要があり, これらは多数の packets として通信される. これを 1Gbps のイーサネットを用いて通信するため, スレーブ PC における仮説の受信時や, マスター PC における尤度 (重み) の受信時には, PC の受信バッファが溢れてしまい, その結果, 情報の取りこぼしが発生してしまう. あるいは, packets の送信とその受信を同時に行えないため, 通信の効率が悪くなる可能性がある. 本研究ではこれら 2 つの問題に対して, 以下の方法を用いる.

- 大量 packets の問題

仮説と尤度の packets をそれぞれある程度まとめて通信する方法を用いる. これによって, packets の数を減少させることができ, 取りこぼしを少なくすることができる.

- 送受信における効率の問題

packets の受信と尤度の計算, 送信を並列化して, 無駄な待ち時間が発生しないようにする. また, 全部のサンプルの尤度を一度に送信すると, マスターに通信が集中するので, 尤度計算を行ったサンプルから順に尤度を送信して通信集中を避けている.

### 5. 実験結果

本研究では, 前章の NAMA-CON 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 9 に示すようにカメラを配置し, それぞれ CPU が Pentium 4 3.4GHz, メモリが 1Gbytes のスレーブ PC に接続されている. また, 図 10 に用いたカメラを示す. さらに, マスター





図 9 実際のカメラの配置  
Fig. 9 camera location



図 10 用いたカメラ  
Fig. 10 camera which used

PC も同様の PC を用いた．これらを 3 章で示した方法を用いてシステムを構築し，その追跡の様子と軌跡を示す．また，4 章で示した通信方法による通信タイミングを，このシステム上で追跡実験で実際に計測した結果を示し，図 8 と比較し，考察する．

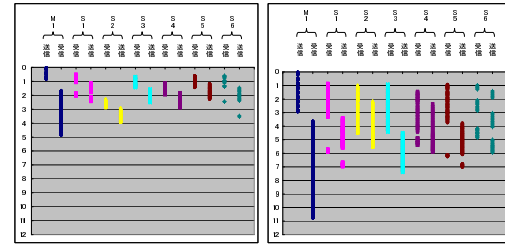
### 5.1 追跡結果

図 11 に追跡の様子を示す．これらはそれぞれが各カメラで撮影した画像であり，そこに写っている人物の頭部にある多数の点が仮説を画像平面上に投影したの位置を表す．また，大きな楕円はサンプル点のそれぞれの画像に対する期待値である．これにより，提案手法による人物の頭部を追跡できたことが分かる．

また，図 12 に，図 11 とは別のシーケンスを用いたときの 3 次元位置推定の結果を示す．これは 3 次元空間中のワールド座標系における事後確率の期待値を示しており， $X - Y$  を床面としている．図 12(a) はワールド座標系の  $X - Y$  平面を上から見た様子であり，右方向が  $X$  軸の正方向であり，下方向が  $Y$  軸の正方向である．さらに，図 12(b) は  $X - Z$  平面を，(c) は  $Y - Z$  平面を見た様子である．この人物の移動シーケンスでは，しゃがんだりする動作も含まれており，この図によれば，地面に対して水平方向だけでなく，垂直方向に対しても正しい位置を推定できていることが分かる．

### 5.2 通信タイミング

図 13(a) に実際の追跡時の通信タイミングを示す．この表では，縦方向に時間軸を取り，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 がマスター PC を表し，送信は仮説の送信，受信はスレーブ PC から返信される尤度の受信，S1 から S6 までは各スレーブ PC を表し，受信はマスター PC からの仮説の受信，送信はマスター PC への尤度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，マスター 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，仮説数は多ければより具体的な事後確率を求めることができるが，今回は経験的に安定に追



(a) 150 samples

(f) 450 samples

図 13 1 フレームの処理に要する通信のタイムチャート

Fig. 13 Time chart of communication in 1 frame

跡が行えると判断できる 150 個と，それを増加させた場合の 450 個を用いた．

これらによれば，マスター PC が仮説を送信し，それぞれの尤度評価が全スレーブ PC から返信し終えるまでに 150 個の場合で 5ms 弱，450 個の場合で 11ms の時間を要することになる．つまり，1 フレームの処理を 5ms 未満で終わっていることになり，1 人の人物を追跡する場合において，十分な実時間性を確認できる．

また，各スレーブ PC では仮説の受信と尤度の送信のタイミングを見ると，それぞれに対して処理の並列化が行えており，今回のようなシステムに対してマルチスレッド化を行うことは有効な手段であるといえる．このことから，並列化による処理時間の余裕を，今回の実験では用いなかった，顔向き推定などの他の処理へ割り振ることも可能である．

さらに，図 8 の通信タイミングの理論値と比較しても，タイムロスが考慮されていない以外は，マスター PC，スレーブ PC それぞれの通信が理論通り効率良く行えているといえ，提案手法の有効性を確認することができる．

### 5.3 尤度統合における実験および検討

本研究では各カメラで得られた尤度を統合する必要がある．3.5 節では，その方法に対する説明を行ったが，本節では実験的な検証を行う．この検証のために人物が撮影された各カメラの画像に対して，尤度を 3 次元空間上の全ての点に関して求めることを行った．これを行うことで，ある 3 次元位置を探索したときの尤度を調べることができ，また，人物頭部の周辺での尤度分布を知ることができる．

図 14 にその結果を示す．ここで，人物頭部の中心は  $(X, Y, Z) = (320, 150, 160)$  に存在するものとする．また，同図の左側が尤度を和によって表現した場合，右側が積によって表現した場合を示す．同図 (a)(b) では， $Z = 150$  の  $X - Y$  平面に平行な面上の尤度であり，上方向が  $Y$  軸方向，右方向が  $X$  軸方向を表す．

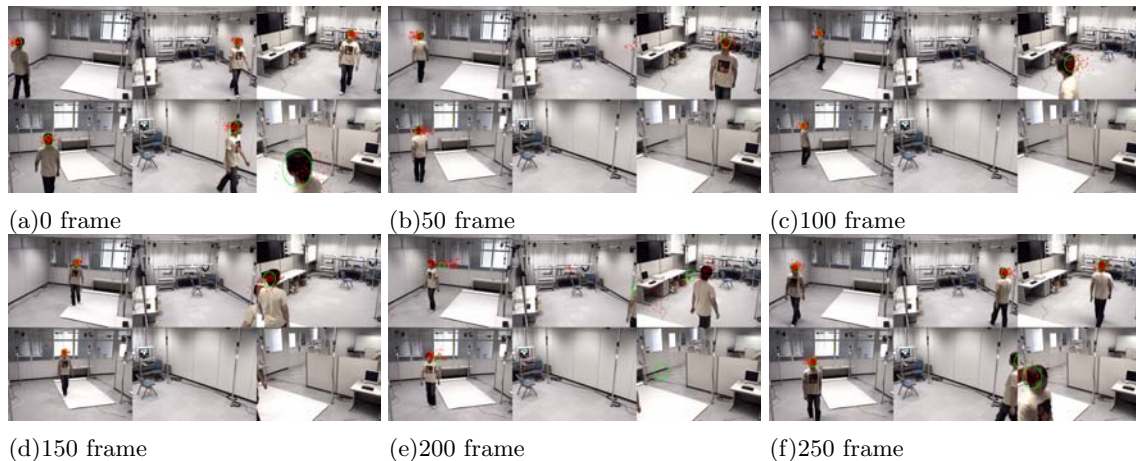


図 11 追跡の様子  
Fig. 11 a look of tracking

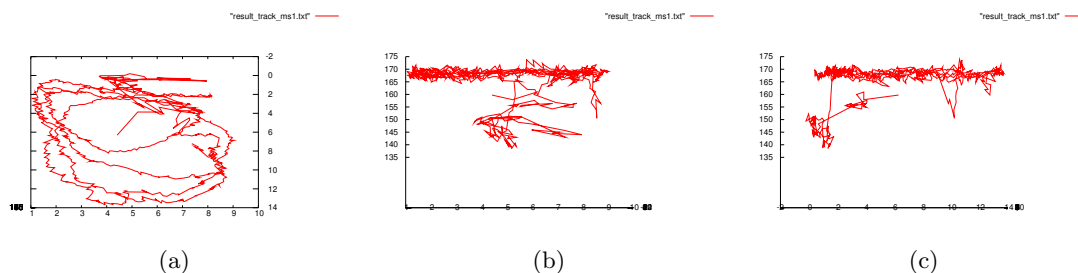


図 12 追跡結果:(a) $X - Y$  平面, (b) $X - Z$  平面,  $Y - Z$  平面  
Fig. 12 Tracking result:(a) $X - Y$  plane, (b) $X - Z$  plane, (c) $Y - Z$  plane

また、同図 (c)(d) は、 $X = 320$  の平面を表しており、上方方向が  $Z$  軸方向、右方向が  $Y$  軸方向を表す。これらによれば丁度人物頭部の位置でピークを示している。

尤度を和によって統合する場合、頭部以外の場所でも大きな値が現れており、これらが影響する可能性が分かった。逆に、積によって統合する場合では、頭部以外の場所には大きなピークが現れていないが、ピーク位置からの減衰が激しい。提案手法では、仮説の位置とその尤度によって事後確率を表現しているため、ブロードな尤度分布が得られなければ、頭部が素早く動いたときに対応できない可能性がある。

また、これら进行评估するために、実際にオクルージョンが発生する場合の追跡実験を行った。図 15 にその結果を示す。同図では、左側の系列が尤度の和で統合を行った場合、右側の系列が積で統合を行った場合で、どちらも (b) でオクルージョンが発生した後を示す。ただし、オクルージョンが発生しているのはそれぞれの左下のカメラ画像だけであり、5 つのカメラからは観測可能で、1 つだけでオクルージョンが発生

している。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、和の場合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. 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、これを 3 次元空間におけるランダムサンプリングによる仮説の生成と、それらを画像空間へ投影した尤度評価によって、人物頭部の 3 次元空間中の位置を追跡する、NAMA-CON を提案した。従来手法と比べ、アルゴリズムとしても、実験装置も単純であり、計算機間の通信も効率的に行うことができた。これにより、単一の人物であれば、1 フレームあたり 5ms の安定な追跡処理を実現し、本論文ではこれを実験的に示すことに成功した。また、尤度統合に関する検討を行い、その特性を実験的に示し、2 つの統合方法を用いた追跡実験を行い、和による統合方

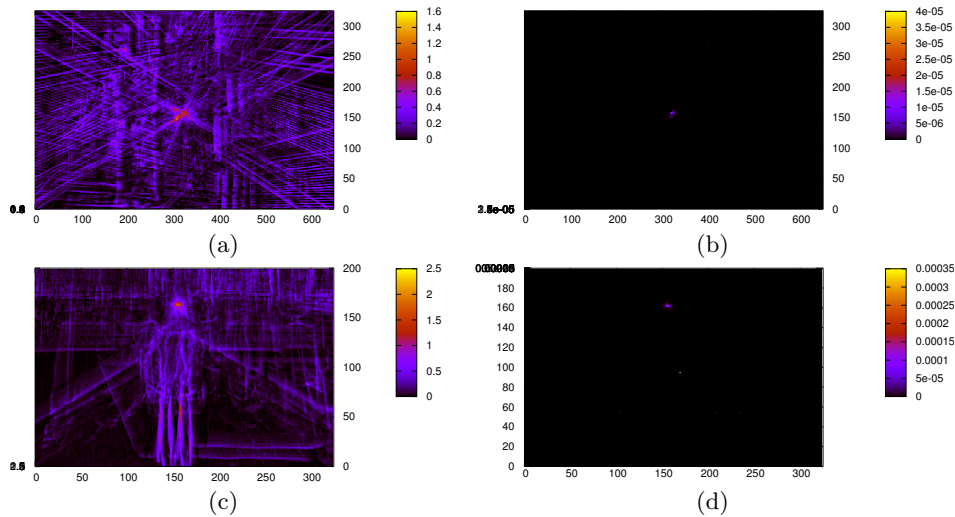


図 14 観測範囲内の尤度マップ: (a) 和による表現,  $Z = 160$  での平面, (b) 積による表現,  $Z = 160$  の平面, (c) 和による表現,  $X = 320$  の平面, (d) 積による表現,  $X = 320$  の平面

Fig. 14 Likelihood map of observal area: (a)sum of likelihood, plane  $Z = 160$ , (b)product of likelihood, plane  $Z = 160$ , (c)sum of likelihood, plane  $X = 320$ , (d)product of likelihood, plane  $X = 320$

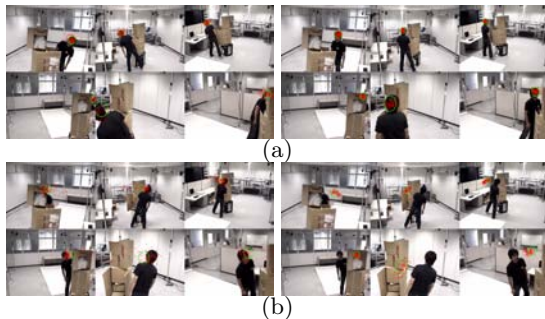


図 15 オクルージョンを含む場面での追跡結果: 左側, 和で統合した場合, 右側, 積で統合した場合

Fig. 15 tracking result at scene that including occlusion:left side/summation unification, right side/product unification

法が隠れに強いことを示すことができた。

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。本論文では仮説数を変化させることで擬似的にシステムの拡張性について考察したが, 実際にスレーブ PC を増加させたときや, その他のセンサを用いた大規模システムに導入したときの動作を確認する予定である。また, この場合のスケラビリティをもたせるために, ネットワークの構成や通信方法について検討する。さらに, 尤度統合はその他の方法も含め, これから最適なものを実験的に求めていく予定である。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) 谷口倫一郎, 和田俊和: “PC クラスタを用いた実時間多視点映像処理システム - 実時間モーションキャプチャ・三次元復元システム -”, 日本ロボット学会誌, pp. 427–432 (2001).
- 3) 松山隆司: “分散強調視覚プロジェクト”, 日本ロボット学会誌, pp. 416–419 (2001).
- 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, J84-D-II, 1, pp. 102–110 (2001).
- 5) M. Isard and A. Blake: “Condensation – conditional density propagation for visual tracking”, International Journal of Computer Vision, **29**, 1, pp. 5–28 (1998).
- 6) 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).