

視野を共有しないカメラ群を用いた人物行動軌跡の推定

小林 貴訓[†] 佐藤 洋一[†] 杉本 晃宏[§]

[†] 東京大学生産技術研究所
〒 153-8505 東京都目黒区駒場 4-6-1
E-mail : {yosinori, ysato}@iis.u-tokyo.ac.jp

[§] 国立情報学研究所
〒 101-8430 東京都千代田区一ツ橋 2-1-2
E-mail : sugimoto@nii.ac.jp

あらまし

安全・安心な社会環境の実現のため、監視カメラ映像の解析技術への期待は大きい。しかし、監視カメラの普及が進む一方で、広大な人々の活動空間や、プライバシーの問題などにより、人物の行動すべてをカメラで観察することは困難であることが多い。そのため、環境に疎に分散配置された視野を共有しないカメラ群からの情報を統合し、人物の行動を計測する技術が重要となる。環境に疎に分散配置された視野を共有しないカメラ群による人物の行動計測では、複数のカメラで観察される人物の同一性推定や、カメラで観察されない箇所における人物の行動推定などの課題がある。このうち本研究では後者の課題に焦点をあて、環境の知識をシステムに与えることで、屋内における人物の行動軌跡をカメラで観察されない箇所においても良好に推定する手法を提案する。

キーワード：分散疎配置カメラ群，行動軌跡推定，環境モデル，MAP 推定

Tracking people by using distributed cameras with non-overlapping views

Yoshinori Kobayashi[†] Yoichi Sato[†] Akihiro Sugimoto[§]

[†]Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku,
Tokyo 153-8505, JAPAN
E-mail : {yosinori, ysato}@iis.u-tokyo.ac.jp

[§]National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo 101-8430, JAPAN
E-mail : sugimoto@nii.ac.jp

Abstract

A sparse network of multiple cameras can cover large environment for monitoring object's activity. To track objects successfully by using distributed cameras, we need to estimate objects' trajectory even in unobserved areas, and also need to establish correspondence between objects captured in different cameras. This paper provides a method for estimating trajectories of people by using distributed cameras with non-overlapping views. Trajectories of people are estimated by considering an evaluation function derived from a motion model, an observation model and an environment model.

Keywords : non-overlapping cameras, trajectory estimation, environment model, MAP estimation

1 はじめに

安全・安心な社会環境の実現のために、監視カメラの普及が進んでいる。犯罪捜査に監視カメラ映像が利用されるなど、監視カメラの活用への期待は大きく、また、得られた大量の映像を自動的に処理するための解析技術は多くの場面で必要とされている。しかし一方で、広大な人々の活動空間や、プライバシーの問題などにより、人物の行動すべてをカメラで観察することは困難であ

ることが多い。このような状況で、人物の行動を計測するためには、環境に疎に分散配置された視野を共有しないカメラ群からの情報を統合し、人物の行動を計測する技術が重要となる。

環境に密に配置された視野を共有するカメラ群による観察から、人物の行動を計測する研究は従来から行われている (例えば [6, 9, 10] など)。しかし、環境に疎に分散配置された視野を共有しないカメラ群からの情報

を統合し、人物の行動を計測している例は少ない。

環境に疎に分散配置された視野を共有しないカメラ群による人物の行動計測には、次の二つの課題があると考えられる。一つは、複数のカメラで観察される人物の同一性推定の課題である。これは、複数のカメラで観察された人物が同一人物であるかを、人物が一方のカメラで観察された後に、他方のカメラでどのように観察されるかを手掛りとして推定するものである。もう一つは、カメラで観察されない箇所における人物の行動推定の課題である。これは、カメラで観察されない箇所における人物の行動を、観察された少ない情報から補間し、推定するものである。

視野を共有しないカメラ群による人物の行動計測の実現には、これらの課題を解決する必要があるが、本研究では、カメラで観察されない箇所における人物の行動推定の問題に着目し、環境の知識をシステムに与えることで、屋内の人物の行動軌跡をカメラで観察されない箇所においても良好に推定する手法を提案する。このような人物の行動推定の手法は、公共施設における不審人物の検出など、セキュリティシステムへの応用が期待され、安全・安心な社会環境の実現のため、意義が大きいと考えられる。

本稿では、2節において視野を共有しないカメラシステムについて関連研究を概観する。次に、3節にて本研究での提案手法について述べ、4節にて提案手法を用いた実験とその結果について言及する。最後に、5節にてまとめと今後の課題について述べる。

2 関連研究

視野を共有しないカメラシステムには、それぞれのカメラから得られた情報を独立に扱い、情報の収集や提示に着目したシステムと、カメラ群から得られた情報を統合的に扱うことで、新たな知見を得ようとするものがある。情報の収集や提示に着目したシステムでは、VSAM Project[1]が良く知られているが、ここでは本研究でのアプローチを鑑み、後者について着目する。

カメラ群から得られた情報を統合し、カメラで観察されない箇所の人物の行動を推定するために有用な手法は、大きく次の三種類が挙げられる。複数のカメラで観察される対象の振舞いから、カメラの配置やトポロジを推定する手法 [5, 7]、複数のカメラで観察される人物の同一性を推定する手法 [2, 3, 4]、カメラで観察されない箇所における人物の行動軌跡を推定する手法 [8] である。

2.1 カメラの配置やトポロジを推定する手法

カメラの配置やトポロジの推定は、大量かつ広範囲にカメラを配置した場合など、全てのカメラの位置や姿勢を計測できない場合に有用な手法である。

Makrisら [5] は、それぞれのカメラの位置や姿勢が分からない場合に、カメラ画像中で観察される対象の振舞いから、カメラで観察している空間同士の間隔を推定する手法を提案している。視野を共有しない複数のカメラで撮影された映像において、自動車などの対象が出現する点、消失する点をノードとして扱い、各ノードにおいて、対象の出現、消失のイベントが、どの程度の頻度で発生しているかを考慮してノード間の間隔を推定し、最終的にカメラのトポロジを推定している。

また、Rahimiら [7] はカメラで観察する対象の移動速度などが詳しく分かっている場合に、カメラ画像中で観察される対象の振舞いからカメラの状態を推定できるとし、天井から下方を観察しているカメラ群の配置を、対象の既知の動きから推定する手法を提案している。

2.2 人物の同一性を推定する手法

複数のカメラで観察される人物の同一性推定は、視野を共有しないカメラ群を利用して、複数の人物を観察する場合に有用な手法である。

池亀ら [2] や Javedら [4] は視野を共有しない複数のカメラで観察される人物を対応づけ、同一性を推定する手法を提案している。推定には、人物の出現・消失時の位置、移動方向・速度、色、及びカメラ間の平均移動時間などの情報を確率的に用いており、これらは同一人物であることがあらかじめ分かっている状況で事前に学習し、獲得している。

また、Javedら [3] は一方のカメラの視野境界が他方のカメラの視野とどのような関係にあるかを利用して、人物を追跡する手法を提案している。一方のカメラの視野境界線を他方のカメラの視野に関係づけ、人物がどの視野境界線を越えてフレームアウトしたかを考慮することで、次にどのカメラに出現するかを推定している。

2.3 人物の行動軌跡を推定する手法

カメラで観察されない箇所における人物の行動軌跡推定は、広い範囲を少ないカメラで観察する場合など、全体を観察できない状況で人物の行動を観察する場合に有用な手法である。しかし、同一性推定に関する研究に比べ、カメラで観察されない箇所における人物の行動推定はほとんど試みられていない。

Rahimi ら [8] は屋内に分散配置した視野を共有しないカメラ群を利用して人物の行動軌跡を推定する手法を提案している．対象の移動速度が既知である場合、カメラ画像中で観察される対象の振舞いから、対象とカメラの距離が推定でき、2台のカメラ間の距離が既知である場合、2台のカメラによる観察によって、対象の移動速度が推定できる．この考えを一般化し、人物の移動がガウスマルコフ過程でモデル化できる場合に、複数のカメラによる観察から、カメラで観察されない箇所における人物の行動軌跡を推定する手法を提案している．

しかし、Rahimi らの手法では、壁などの障害物の配置状況を考慮していないため、推定した人物の行動軌跡が壁を通過してしまうなどの問題がある．また、センサとしてカメラを壁面に水平に設置する必要があり、一般的な監視カメラのように天井付近に配置したカメラを用いた議論はなされていない．

3 提案手法

本研究では、推定する人物の行動軌跡として、屋内に設置された監視カメラで観察することを想定して、人物頭部の移動軌跡に着目する．また、観測には、位置や姿勢が既知である環境に疎に分散配置された視野を共有しないカメラ群を用い、人物一人の軌跡を推定対象とする．さらに、環境の知識を考慮することで、屋内における人物の行動軌跡をカメラで観測されない箇所においても良好に推定する手法を提案する．

3.1節、及び3.2節ではRahimi ら [8] と同様の考えに基づいて、人物頭部の3次元移動軌跡推定を行うための定式化を行う．

3.1 人物頭部の移動軌跡

ある時刻 t における人物頭部の状態量 \mathbf{x}_t を式 (1) のように定義する．

$$\mathbf{x}_t = \begin{bmatrix} u_t & \dot{u}_t & v_t & \dot{v}_t & w_t & \dot{w}_t \end{bmatrix}^\top \quad (1)$$

ここで、 u_t, v_t, w_t は世界座標系における人物頭部の位置、 $\dot{u}_t, \dot{v}_t, \dot{w}_t$ はそれぞれの速度成分を表す．

また、観測開始から時刻 T まで ($t=1 \dots T$) の軌跡 \mathbf{X} を \mathbf{x}_t の状態集合として表現すると、軌跡 \mathbf{X} は式 (2) のような $1 \times 6T$ の行列として表現される．

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top & \mathbf{x}_2^\top & \dots & \mathbf{x}_T^\top \end{bmatrix}^\top \quad (2)$$

この軌跡 \mathbf{X} は時間経過に伴う位置の履歴を含んでいるため、人物頭部の移動軌跡の推定は、この軌跡 \mathbf{X} を推定する問題に帰着される．

3.2 運動モデルと観測モデルによる評価

3.2.1 運動モデル

人物頭部の動きは急激に変化しないと仮定し、運動モデルを式 (3) のように準備する．

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{v}_t \quad (3)$$

ここで \mathbf{A} は、時刻 t から時刻 $t+1$ において、人物頭部の状態量の変化をモデル化した状態遷移行列である．また、 \mathbf{v}_t は平均 0 共分散行列 Σ_v を持つノイズであり、 \mathbf{A} 、 Σ_v は例えば以下のような行列である．

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Sigma_v = \text{diag} \left[0.1 \quad 1 \quad 0.1 \quad 1 \quad 0.1 \quad 1 \right]$$

これより、 \mathbf{x}_{t-1} が与えられたときの \mathbf{x}_t の確率を式 (4) のように表現することができる．

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = N(\mathbf{x}_t | \mathbf{A}\mathbf{x}_{t-1}, \Sigma_v) \quad (4)$$

ここで $N(\mathbf{x}_t | \mathbf{A}\mathbf{x}_{t-1}, \Sigma_v)$ は、 \mathbf{x}_t の分布が平均ベクトル $\mathbf{A}\mathbf{x}_{t-1}$ 共分散行列 Σ_v の多変量正規分布に従うことを表す．

さらに、 \mathbf{x}_t の状態集合である軌跡 \mathbf{X} についても確率で表現でき、事前確率として式 (5) のように定義する．

$$p(\mathbf{X}) = \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{x}_{t-1}) = N(\mathbf{X} | \mathbf{0}, \Lambda_{\mathbf{X}}) \quad (5)$$

ここで $\Lambda_{\mathbf{X}}$ は以下で表される行列である．

$$\Lambda_{\mathbf{X}} = \begin{bmatrix} \mathbf{A}^\top \Sigma_v \mathbf{A} & -\mathbf{A}^\top \Sigma_v & 0 & \dots \\ -\Sigma_v \mathbf{A} & \Sigma_v + \mathbf{A}^\top \Sigma_v \mathbf{A} & -\mathbf{A}^\top \Sigma_v & \dots \\ 0 & -\Sigma_v \mathbf{A} & \Sigma_v + \mathbf{A}^\top \Sigma_v \mathbf{A} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

3.2.2 観測モデル

時刻 t における人物頭部の3次元位置の観測を、各カメラ画像への射影 (図1) とノイズの和で記述し、観測モデルを式 (6) のように準備する．

$$\mathbf{y}_t^i = \pi^i(\mathbf{x}_t) + \omega_t \quad (6)$$

ここで $\pi^i(\mathbf{x}_t)$ は、時刻 t における状態量 \mathbf{x}_t から i 番目のカメラの画像座標 $[m_t^i, n_t^i]^\top$ への射影を行う関数であり、要素 m_t^i, n_t^i は、次式で表される。

$$m_t^i = f^i \frac{\mathbf{R}_y^i(\mathbf{C}\mathbf{x}_t - \mathbf{p}^i)}{\mathbf{R}_x^i(\mathbf{C}\mathbf{x}_t - \mathbf{p}^i)}, n_t^i = f^i \frac{\mathbf{R}_z^i(\mathbf{C}\mathbf{x}_t - \mathbf{p}^i)}{\mathbf{R}_x^i(\mathbf{C}\mathbf{x}_t - \mathbf{p}^i)}$$

\mathbf{p}^i は i 番目のカメラの位置であり、 \mathbf{C} は時刻 t における人物頭部の状態量 \mathbf{x}_t から位置成分を取り出すための以下の行列である。

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

\mathbf{R}^i は i 番目のカメラの姿勢を表す回転行列であり、 \mathbf{R}_x^i は \mathbf{R}^i による計算結果から X 軸に対応する要素を取り出すことを表している。 f^i は i 番目のカメラの焦点距離である。また、 ω_t は平均ベクトル $\mathbf{0}$ 、共分散行列 Σ_ω を持つノイズであり、 Σ_ω は以下で表される。

$$\Sigma_\omega = \text{diag} \left[\sigma_\omega^2 \quad \sigma_\omega^2 \right]$$

ここで σ_ω^2 はカメラ画像における観測の水平、垂直方向の分散であり、簡単のために同じ値を採用している。

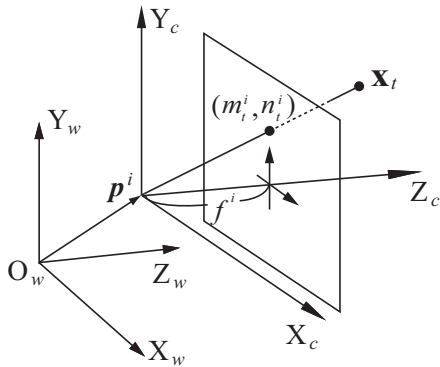


図 1: カメラ画像への射影

これより、時刻 t における状態量 \mathbf{x}_t を持つ人物頭部が i 番目のカメラに画像座標 y_t^i で観測される確率 $p(y_t^i|\mathbf{x}_t)$ を式 (7) のように表すことができる。

$$p(y_t^i|\mathbf{x}_t) = \begin{cases} N(y_t^i|\pi^i(\mathbf{x}_t), \Sigma_\omega) & \text{if } \ell^i(\mathbf{x}_t) \\ \delta(y_t^i - \phi) & \text{otherwise} \end{cases} \quad (7)$$

ここで、 $\ell^i(\mathbf{x}_t)$ は、 i 番目のカメラ視野内に状態量 \mathbf{x}_t を持つ人物頭部が観測される場合に真となる関数であり、 i 番目のカメラ視野内で人物頭部が観測されないとき、 y_t^i は ϕ をとる。

全てのカメラ ($i = 1 \dots N$) の観測開始から時刻 T まで ($t = 1 \dots T$) の観測の集合を \mathbf{Y} とすると、尤度を以下のように定義できる。

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{t=1}^T \prod_{i=1}^N p(y_t^i|\mathbf{x}_t)$$

これを時刻 t で i 番目のカメラ視野内で観測される場合 ($(t, i) \in o$) と観測されない場合 ($(t, i) \notin o$) で分け、ある軌跡 \mathbf{X} に対して観測 \mathbf{Y} が得られる確率を式 (8) のように表す。

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{(t,i) \in o} N(y_t^i|\pi^i(\mathbf{x}_t), \Sigma_\omega) \prod_{(t,i) \notin o} (1 - \ell^i(\mathbf{x}_t)) \quad (8)$$

ここで $(t, i) \in o$ は時刻 t において i 番目のカメラで観測されることを表す。

3.2.3 運動モデルと観測モデルによる評価関数

ベイズの法則に基づき、運動モデル $p(\mathbf{X})$ 、観測モデル $p(\mathbf{Y}|\mathbf{X})$ を用いて、観測 \mathbf{Y} が得られたときの軌跡 \mathbf{X} の条件付き確率 $p(\mathbf{X}|\mathbf{Y})$ を求める。

$$p(\mathbf{X}|\mathbf{Y}) \approx p(\mathbf{X}) p(\mathbf{Y}|\mathbf{X}) =$$

$$N(\mathbf{X}|\mathbf{0}, \Lambda_{\mathbf{X}}) \prod_{(t,i) \in o} N(y_t^i|\pi^i(\mathbf{x}_t), \Sigma_\omega) \prod_{(t,i) \notin o} (1 - \ell^i(\mathbf{x}_t))$$

これを整理して、確率 $p(\mathbf{X}|\mathbf{Y})$ が高いほど、尤もらしい軌跡として高い評価を与える評価関数 $O_o(\mathbf{X})$ を導出する (式 (9))。

$$O_o(\mathbf{X}) = \mathbf{X}^\top \Lambda_{\mathbf{X}}^{-1} \mathbf{X} + \frac{1}{\sigma_\omega^2} \sum_{(t,i) \in o} (|y_t^i - \pi^i(\mathbf{x}_t)|)^2 \quad (9)$$

ここでは簡単のため、観測モデルにおいて、観測されない場合の評価を無視している。また、整理の途中で、正負の変換を行っているため、評価関数 $O_o(\mathbf{X})$ の値は小さいほど尤もらしい評価を与える。

ここまでの式変形の詳細は文献 [8] を参照されたい。

3.3 環境モデルの導入

これまでのモデルに加え、さらに環境モデルとして、人物頭部の存在確率に基づく重み付けを行う。

通常、人物は壁を通過することはなく、机や棚などの障害物を越えて移動することもほとんどないと考えられる。また経験的に、人物が通路やドアを通過する際は、壁面や障害物の直近ではなく、広く、通りやすい箇所を通過すると考えられる (図 2)。

このことから、人物が壁面や障害物に近づくほど大きくなり、壁面や障害物の位置で非常に大きなペナルティ

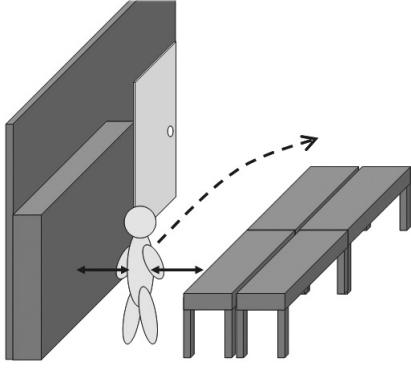


図 2: 環境モデル

を与える関数を環境モデルとして準備する．これを，状態量 \mathbf{x}_t を持つ人物頭部と障害物との距離を $d_h(\mathbf{x}_t)$ とし，ペナルティを与える関数を $e_h(d_h(\mathbf{x}_t))$ で表す．

また，人物頭部の高さについても，歩いている間は大きな変化は起こらないと予測されることから，人物頭部の高さは，平均的な身長から予測される人物頭部の高さから大きな乖離を持たないと仮定し，この乖離に対してペナルティを与える関数を準備する．これを，状態量 \mathbf{x}_t を持つ人物頭部と，平均的な身長から予測される人物頭部の高さとの距離を $d_v(\mathbf{x}_t)$ とし，ペナルティを与える関数を $e_v(d_v(\mathbf{x}_t))$ で表す．

時刻 t における人物頭部の状態量 \mathbf{x}_t に対する環境モデルによる評価は， $e_h(d_h(\mathbf{x}_t))$ と $e_v(d_v(\mathbf{x}_t))$ の和で表し，さらに，観測開始から時刻 T まで ($t = 1 \dots T$) の軌跡 \mathbf{X} に対する環境モデルによる評価関数を $O_e(\mathbf{X})$ とし，式 (10) のように与える．

$$O_e(\mathbf{X}) = \sum_t (e_h(d_h(\mathbf{x}_t)) + e_v(d_v(\mathbf{x}_t))) \quad (10)$$

関数 $e_h(d_h(\mathbf{x}_t))$ と関数 $e_v(d_v(\mathbf{x}_t))$ における水平方向及び鉛直方向の距離と，与えるペナルティの関係を図 3 に示す．

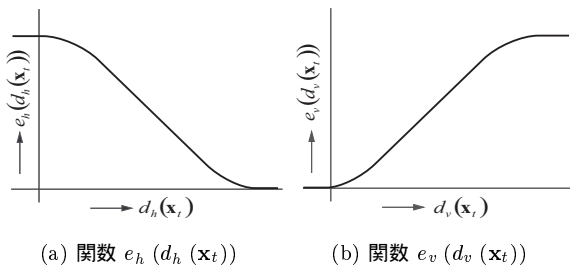


図 3: 距離とペナルティの関係

このようにして準備した環境モデルをあわせて評価することで，人物頭部の軌跡が障害物を通過したり，平

均的な身長から大きくずれてしまう誤推定が回避できると考えられる．

3.4 各モデルによる評価の統合と最適化

これまでの評価関数を統合し，軌跡 \mathbf{X} が $\mathbf{x}_1 \dots \mathbf{x}_T$ の状態集合として与えられたとき，運動モデル，観測モデル，環境モデルによる評価を行い，尤もらしい軌跡 \mathbf{X}^* を推定する．評価関数の統合は式 (11) のようにそれぞれの評価関数の和で表現し，この評価関数の値が最も小さくなる軌跡 \mathbf{X}^* を推定する．

$$\begin{aligned} \mathbf{X}^* = \arg \min_{\mathbf{X}} (O_o(\mathbf{X}) + O_e(\mathbf{X})) = \\ \arg \min_{\mathbf{X}} \left(\mathbf{X}^\top \Lambda_{\mathbf{X}}^{-1} \mathbf{X} + \frac{1}{\sigma_\omega^2} \sum_{(t,i) \in o} (|\mathbf{y}_t^i - \pi^i(\mathbf{x}_t)|)^2 \right. \\ \left. + \lambda \sum_t (e_h(d_h(\mathbf{x}_t)) + e_v(d_v(\mathbf{x}_t))) \right) \quad (11) \end{aligned}$$

ここで λ は環境モデルにより与えられるペナルティの調整に用いる変数である．

この評価関数の最小化には，非線形計画法を用いる．本手法で用いる評価関数は凸関数ではないため，非線形計画法を適用した場合，局所的最適解に収束してしまう可能性が考えられる．この問題を回避するために，評価関数の最適化を始めるための初期軌跡 \mathbf{X}_0 を経験的に複数与え，それぞれに対して最適化を行い，得られた評価関数の値が最小となる軌跡を採用する．人物頭部がカメラ 1 とカメラ 2 で観測された場合，初期軌跡 \mathbf{X}_0 は次のように与える．

a) カメラ 1 とカメラ 2 の間に障害物がない場合

カメラ 1 及びカメラ 2 の光軸が平均的な身長から予測される人物頭部の高さで交わる点を A, B とし，A-B をカメラ 1 で人物頭部が観測されてからカメラ 2 で観測されるまでの観測フレーム数で等間隔に分割した点をそれぞれの時刻における初期人物頭部位置とする．分割した 2 点間の差分をそれぞれの時刻における速度として初期軌跡 \mathbf{X}_0 を生成する (図 4a) ．

b) カメラ 1 とカメラ 2 の間に障害物がある場合

カメラ 1 及びカメラ 2 の光軸が平均的な身長から予測される人物頭部の高さで交わる点を A, B とし，経験的に経路として人物頭部が通過する位置にホールドポイント H1, H2 を置く．A-H1, H1-B の距離に応じて，カメラ 1 で人物頭部が観測されて

からカメラ 2 で観測されるまでの観測フレーム数を配分し、配分されたフレーム数によって A-H1, H1-B を等間隔で分割した点をそれぞれの時刻における初期人物頭部位置とする。また、分割した 2 点間の差分をそれぞれの時刻における速度として初期軌跡 X_0 を生成する。H2 を通過する場合も同様に生成する (図 4b)。

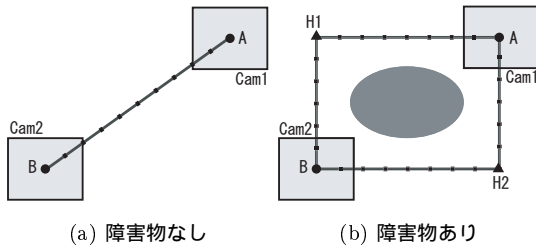


図 4: 初期軌跡

4 実験

4.1 シミュレーション

これまでの提案手法に基づき、人物の行動軌跡推定のシミュレーションを行った。

4.1.1 観測データの取得

シミュレーションは WindowsXP を搭載した PC 上で動作するシミュレータを利用して行った。シミュレーションでは入力として、マウスのドラッグによる軌跡を人物の模擬行動軌跡として使用した。ただし、マウスによるドラッグでは鉛直方向の高さを与えることができないため、状態量の生成では、一定間隔でサンプリングされた水平方向位置に対して、一定の人物頭部位置の高さを付加した。また、模擬行動軌跡の入力では、実際の人物の行動を模擬する観点から、室内に配置された壁や机などの障害物を避けて入力を行った。

用いた室内モデルは研究室内部を模擬したもので、広さは 920cm×1290cm である。

4.1.2 シミュレーション結果

シミュレータへの入力により記録された観測データを用いて、軌跡の推定を行った。評価関数の最小化には MATLAB の非線形計画法による最適化関数を利用した。推定結果の例を図 5(a) に示す。

図 5 において実線が、マウスのドラッグによって生成された人物の模擬行動軌跡であり、人物の右上方から左下方への移動を模擬している。点線が推定された軌跡である。灰色の矩形は壁や棚、机の配置を示しており、右

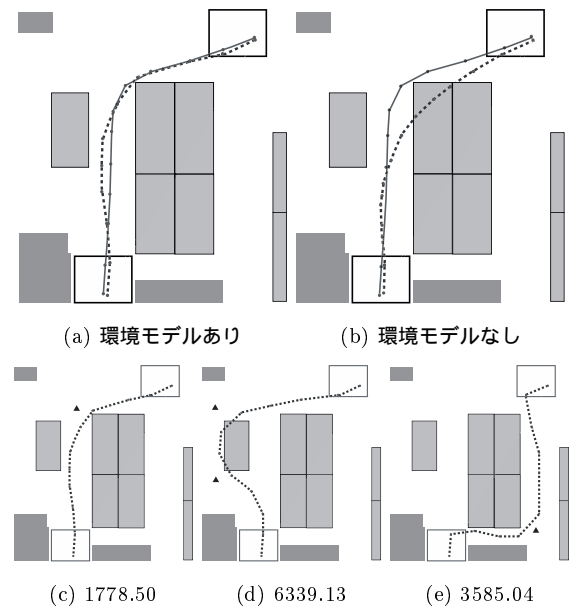


図 5: 軌跡推定のシミュレーション

上方と左下方の矩形領域はカメラの視野を表している。

図 5(b) は環境モデルによる評価を行わずに推定を行った結果である。図より明らかに軌跡が机の上を通過していることが分かる。日常的に人物が机の上を通過することがない状況においてこのような推定結果は好ましくない。

また、推定時に入力した初期軌跡は三種類であり、図 5(c)(d)(e) にそれぞれの初期軌跡を入力した場合に推定された軌跡と評価関数の値を示した。初期軌跡は図 5(c)(d)(e) に黒三角形で示したホールドポイントを通る軌跡である。評価関数の値から、最終的な推定軌跡として図 5(c) の軌跡が採用されている。

4.2 実画像を用いた実験

実画像を用いた実験は研究室内天井に IEEE1394 カラーカメラ (Point Grey Research 社製 Flea) を設置して行った。カメラの設置箇所は図 6 に示す廊下側室入口 (A)、窓側室への通路 (B)、窓側室の窓際 (C)、窓側室給湯設備の前 (D) の 4 箇所であり、4 台のカメラは視野を共有していない。また、カメラ A, B と C, D は焦点距離の異なるレンズを使用しており、各カメラはそれぞれ異なる姿勢で下方を撮影している。各カメラの映像は 320×240 画素、毎秒 15 フレームの映像で得ることとし、汎用 PC (CPU Intel Pentium4 3.2GHz, Memory 2GByte, OS WindowsXP) に IEEE1394 インタフェースカードを組み込み、合計 4 台のカメラの映像を 1 台の PC で記録した。

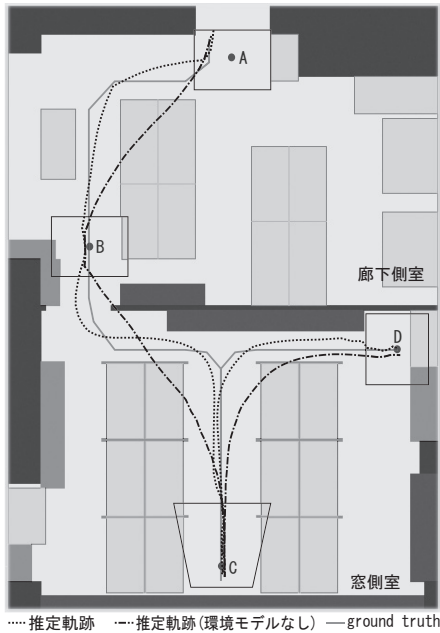


図 6: 室内レイアウト図と推定された軌跡

実験ではあらかじめ決められた軌跡を人物が移動した際の映像を記録した。人物は、室外から廊下側室に入り (A)、通路を通過して窓側室へ入り (B)、窓側室の窓際で向きを変え (C)、窓側室の給湯設備の前 (D) まで移動する。また、得られた映像から人物頭部の位置の計測を容易にするため、白い帽子を着用し実験を行った。

4.2.1 観測データの取得

人物頭部の位置の検出は、背景差分及び帽子色である白色の抽出を行うことで検出した。背景差分により、背景画素より明るい画素を前景として抽出し、さらに抽出された前景の白色の領域の重心をとることで画像中での人物頭部の位置とした。実験で得られた各カメラの画像の例を図 7 に示す。図 7 の黒円は検出された人物頭部の位置を表している。

各カメラで得られた人物頭部の位置データは、最適化計算時間を考慮してすべては使用せず、5 フレーム毎の位置データを観測データとして使用した。実際には、カメラ 4 台の各 101 フレームの観測データを使用して、人物が室内を約 40 秒間移動した際の軌跡の推定を行った。

また、推定に用いる環境モデルは室内レイアウト図からあらかじめ作成しておき、室内の座標系は図 6 の左下床面を原点として、横方向を X 軸、縦方向を Y 軸、床面からの高さを Z 軸とした。

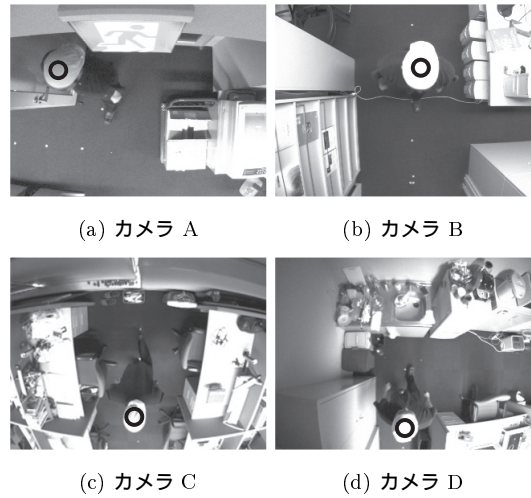


図 7: カメラ画像の例

4.2.2 実験結果

シミュレーション時と同様に評価関数の最小化には MATLAB の非線形計画法による最適化関数を利用した。

実験で記録した映像を利用して推定した人物の行動軌跡の推定結果を図 8 に示す。また、図 6 にも推定結果を重ねて表示した。図 8 及び図 6 において、黒色の点線が本手法を用いて推定を行った結果である。灰色の実線は、人物が実験で移動したあらかじめ決められた軌跡あり、ground truth として表示している。また、環境モデルの導入による効果を明示するため、同じ観測データを用いて環境モデルなしで推定を行った結果を黒色の一点鎖線で表示している。

図 6 においてカメラの設置箇所付近で推定軌跡が ground truth に近づいている。これは評価関数の観測モデルにより、推定軌跡が拘束されているためと考えられる。一方、カメラで観測されない部分においては ground truth と差が大きくなっているが、軌跡は滑らかに接続されており、運動モデルによる拘束が有効に働いていると考えられる。また、環境モデルなしで推定した結果と比較して、机や壁などの障害物を避けた人物の行動軌跡を推定できていることから、環境モデルによる拘束が有効に働いていると考えられる。また、鉛直方向について乖離が少ないことから環境モデルによる拘束が有効に働いていることが分かる。

カメラで人物頭部が観測されたフレームは全体の 3 割程度であるが、運動モデルや環境モデルにより軌跡を拘束することで、人物の行動について得られる情報が少ない場合でも、良好な軌跡が推定できた。

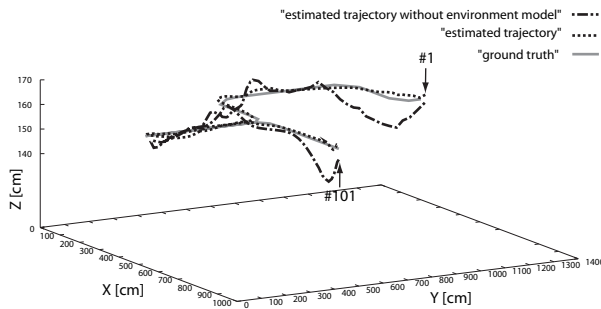
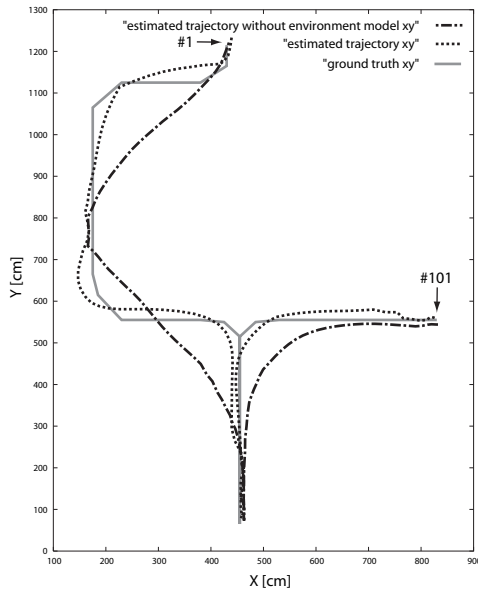


図 8: 軌跡の推定結果

5 おわりに

本稿では、視野を共有しないカメラ群を用いて、人物の行動軌跡を、運動モデルと観測モデルに環境モデルを加えた評価関数の最適化を行うことで推定する手法を提案した。

運動モデルは、人物の動きは急激に変化しないと仮定してモデル化した。環境モデルは、建築図面やレイアウト図などから壁や机の位置を取得し、それらの位置には人物は存在しないなどの知識をペナルティとして経験的に付与することでモデル化した。カメラによる観測から、いくつかの経験に基づいた行動軌跡を初期軌跡として入力し、運動モデル及び観測モデル、環境モデルに基づく評価関数を、非線形計画法を用いて最適化することで尤もらしいユーザの行動軌跡を推定した。このとき、机や壁などの環境を考慮することにより、ユーザの行動軌跡が机や壁を通過してしまう誤推定を回避することが可能となった。

今後は、得られたカメラ映像から柔軟に運動モデルを

構築することで、より適したモデルの適用による行動軌跡推定を行うことを計画している。また、環境モデルにおいても、カメラなどによる観測からの学習により、最適なペナルティを獲得するなどの手法の検討を行う予定である。

なお、本研究の一部は、科学研究費補助金 13224051 の補助を受けて行った。

参考文献

- [1] R. Collins, A. Lipton, and T. Kanade, "A System for Video Surveillance and Monitoring," Proceedings of American Nuclear Society(ANS) Eighth International Topical Meeting on Robotic and Remote Systems, pp.25-29, 1999.
- [2] 池亀幸久, 平野真, 玉木徹, 山本正信, "視野重複のないマルチカメラを用いた映像中の人物抽出と人物の確率的歩行経路推定," 電子情報通信学会言語理解とコミュニケーション (NLC) 研究会技報, pp.43-48, February 2005.
- [3] O. Javed, Z. Rasheed, O. Alatas and M. Shah, "Knight: A Real Time Surveillance System for Multiple Overlapping and Non-Overlapping Cameras," IEEE International Conference on Multimedia and Expo(ICME) 2003, pp.I-649-652, 2003.
- [4] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, "Tracking Across Multiple Cameras With Disjoint Views," Proceedings of International Conference on Computer Vision(ICCV) 2001, pp.II-952-958, 2001.
- [5] D. Makris, T. Ellis, and J. Black, "Bridging the Gaps between Cameras," Proceedings of Computer Vision and Pattern Recognition(CVPR) 2004, pp.II-205-210, 1999.
- [6] 中澤篤志, 加藤博一, 日浦慎作, 井口征士, "分散視覚エージェントを用いた複数人物追跡システム," 情報処理学会論文誌, Vol.42, No.11, pp.2699-2710, November 2001.
- [7] A. Rahimi, B. Dunagan, and T. Darrell, "Simultaneous Calibration and Tracking with a Network of Non-Overlapping Sensors," Proceedings of Computer Vision and Pattern Recognition(CVPR) 2004, pp.187-194, 2004.
- [8] A. Rahimi, B. Dunagan, and T. Darrell, "Tracking People with a Sparse Network of Bearing Sensors," Proceedings of European Conference on Computer Vision(ECCV) 2004, pp.507-518, 2004.
- [9] 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 DII, Vol.J88-DII No.8, pp.1592-1600, August 2005.
- [10] 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," 情報処理学会 CVIM 研究会論文誌, vol.43, no.SIG11, pp.64-79, December 2002.