

パーティクルフィルタを用いた歩き方による個人識別

江本 光晴[†] 林 朗 末松 伸朗 岩田 一貴

広島市立大学大学院情報科学研究科

〒731-3194 広島県広島市安佐南区大塚東3-4-1

概要 歩き方から個人識別を行う。従来手法は、画像（または画像特徴量）マッチングを基礎とするため、カメラ視点に依存する。テストデータと訓練データのカメラ視点が大きく異なる場合、その識別精度は低くなる。我々はカメラ視点に依存しない識別手法を提案する。歩行動作の位相およびカメラ視点を状態変数とする状態空間モデル Cyclic motion model を用いる。識別候補者それぞれのモデルパラメータを学習し、歩行を表現する。Particle Filter を用いて観測画像列に対応する状態変数を推定する。そして、モデル尤度を用いて識別を行う。

Human Gait Identification using a Particle Filter

Mitsuharu Emoto, Akira Hayashi, Nobuo Suematsu and Kazunori Iwata
Graduate School of Information Sciences, Hiroshima City University
Uzuka-Higashi 3-4-1, Asa-Minami-Ku, Hiroshima, 731-3194, JAPAN

Abstract We challenge the human identification problem from gait and body shape. Conventional methods depend on the camera viewing direction since they are based on matching images, their accuracy is low when there is a big difference between the camera viewing direction of the test and training. We propose a novel method that does not depend on the camera viewing direction. We develop a state space model called "cyclic motion model" whose state variables are the phase and the camera viewing direction. We learn model parameters for each candidate person, and represent their walking. To identify a person from the observed image sequence, we first compute the model likelihoods using a particle filter. We then identify the person from the model likelihoods.

1 はじめに

頑健なセキュリティを目的として、指紋や網膜など生体情報を用いた個人識別が行われている。これらは高い精度での識別が可能である一方、被識別者自らが専用機器に触れなければならない。人間の歩行は誰もが一見同じように見える。しかし、個々の歩行スタイルは、それぞれの歩き方や身体の特徴に影響される。歩行を撮影した画像列は個人識別に有効な情報といえる [9]。また、1台のビデオカメラから撮影できるため、被識別者への負担はない。

従来手法の多くは、固定単眼カメラで被識別者を撮影し、その画像列から識別を行う [1] [7] [9] [11]。

これらは、画像（または画像特徴量）マッチングを基礎とするため、カメラ視点（被識別者から見たカメラの方位角）に依存する。テストデータと訓練データのカメラ視点と大きく異なる場合、その識別精度は低くなる。任意な方向で歩く人物を正確に識別することは難しい。実際に、カメラ視点とその変化が既知な環境は多くない。カメラ視点に依存しない個人識別は重要である。

Cuzzolin らは Bilinear model を用いて未知なカメラ視点を推定し、識別を行う [2]。しかし、「曲がる」などカメラ視点に変化する歩行は対応してない。また、Shakhnarovich らは正面および側面の画像を合成することでさまざまなカメラ視点から識別を行うが、複数台のカメラが必要である [10]。

[†] e-mail: emoto@robotics.im.hiroshima-cu.ac.jp

本研究では、状態空間アプローチを用いて、カメラ視点に依存しない識別手法を提案する。歩行運動の周期性を仮定し、周期運動の位相およびカメラ視点を状態変数（隠れ変数）とする状態空間モデル Cyclic motion model (CMM) を定義する。固定単眼カメラから被識別者を撮影した画像（特徴量）列を観測変数とする。識別候補者それぞれのモデルパラメータを学習し、歩行パターンを CMM で表現する。Particle Filter (PF) を用いて、観測画像列に対応する状態変数を推定する。そして、モデル尤度を計算し、そのモデル尤度から個人を識別する。PF は、状態推定問題を効率的に解くための近似計算手法である。Kalman Filter とは異なり、ガウス分布を仮定する必要がない。状態変数の初期値が未知である個人識別では有効な手法となる。

本論文は以下のように構成される。最初に問題設定を 2 章に示す。そして、提案手法の概要を 3 章で述べ、その詳細を 4 章および 5 章に示す。6 章および 7 章にて実験結果を報告し、手法の有効性について考察する。最後に結論を 8 章で述べる。

2 問題設定

K 人の識別候補者を考える。その候補者の一人を被識別者 $\{k | 1 \leq k \leq K\}$ とし、その人物が任意に歩く様子を撮影する。そして、その観測画像列から、被識別者がどの候補者であるか推定する。ここで、時刻 t における画像特徴量ベクトルを \mathbf{y}_t とし、観測画像列を $\mathbf{y}_{1:t} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t)$ とする。

訓練データとして、 V 通りのカメラ視点から撮影した候補者それぞれの画像列 $\{D_{(k,v_i)} | 1 \leq k \leq K, 1 \leq i \leq V\}$ が事前に得られると仮定する。

$$D_{(k,v_i)} = (\tilde{\mathbf{y}}_{1|v_i}, \tilde{\mathbf{y}}_{2|v_i}, \dots, \tilde{\mathbf{y}}_{T|v_i}) \quad (1)$$

ここで、カメラ視点を $\{v_i | 0 \leq v_i < 2\pi, 1 \leq i \leq V\}$ と表記し、カメラ視点 v_i から得られる時刻 t の画像特徴量ベクトルを $\tilde{\mathbf{y}}_{t|v_i}$ とする。

3 提案手法

3.1 状態空間モデル

CMM は周期性を仮定するため、状態変数 \mathbf{x}_t に周期運動の位相 $\phi_t \in (0, 2\pi)$ を持つ [5]。さらに、カ

メラ視点を表す変数 $\psi_t \in (0, 2\pi)$ を加えることで、カメラ視点の変化に対応する。時刻 t における状態変数を $\mathbf{x}_t = (\phi_t, \psi_t)^T$ 、観測変数を \mathbf{y}_t とする。このとき、状態方程式は式 (2) で与えられる。

$$\begin{cases} \mathbf{x}_t \sim \Pr(\mathbf{x} | \mathbf{x}_{t-1}, \lambda_k) \\ \mathbf{y}_t \sim \Pr(\mathbf{y} | \mathbf{x}_t, \lambda_k) \end{cases} \quad (2)$$

ここで、 λ_k は候補者 k のモデルを示す。

人間動作を表現する場合、3 次元姿勢が状態変数によく用いられるが、状態変数が高次元になるため、膨大な計算量が問題となる。CMM は高次元な姿勢を位相で表現することで、その計算量を軽減する。

3.2 概要

提案手法は学習と識別の 2 つのフェーズからなる。学習フェーズは以下のように構成される。

1. 訓練データ $D_{(k,v_i)}$ から位相を定義する
2. モデル λ_k それぞれについて、状態遷移確率 $\Pr(\mathbf{x} | \mathbf{x}_{t-1}, \lambda_k)$ と観測確率 $\Pr(\mathbf{y} | \mathbf{x}_t, \lambda_k)$ を訓練データから学習する

また、識別フェーズは次のようになる。

1. 初期分布 $\Pr(\mathbf{x}_0 | \lambda_k)$ をそれぞれ仮定する
2. PF を用いて、観測画像列のモデル尤度 $L(\lambda_k)$ を候補者それぞれについて計算する
3. モデル尤度 $\{L(\lambda_k) | 1 \leq k \leq K\}$ を用いて、観測画像列から被識別者を推定する

4 学習フェーズ

4.1 位相の定義

周期性を仮定するため、訓練データ $D_{(k,v_i)}$ から位相を定義する。歩行速度が常に一定とは限らないため、以下のように位相を定義する。なお、表記の簡単化のため、 v_i の添字 i を省略する。

1. $D_{(k,v)}$ を周期ごとに区切り、 M 個の周期からなる部分列 $\{D_{(k,v)}^{(i)} | 1 \leq i \leq M\}$ に分割する。なお、1 周期に満たない部分列は削除する。

$$\begin{aligned} D_{(k,v)} &= \left((\tilde{\mathbf{y}}_{1|v}, \dots, \tilde{\mathbf{y}}_{T_1|v}), (\tilde{\mathbf{y}}_{T_1+1|v}, \dots, \tilde{\mathbf{y}}_{T_1+T_2|v}), \right. \\ &\quad \left. \dots, (\tilde{\mathbf{y}}_{\sum_{m=1}^{M-1} T_m+1|v}, \dots, \tilde{\mathbf{y}}_{\sum_{m=1}^{M-1} T_m+T_M|v}) \right) \\ &= \left(D_{(k,v)}^{(1)}, D_{(k,v)}^{(2)}, \dots, D_{(k,v)}^{(M)} \right) \quad (3) \end{aligned}$$

ここで, T は $T \geq \sum_{i=1}^M T_i$ を満たす.

- 動的時間伸縮 (Dynamic Time Warping: DTW) [8] を用いて, $D_{(k,v)}^{(i)}$ と $D_{(k,v)}^{(j)}$ (なお, $i \neq j$ かつ $1 \leq i, j \leq M$) のフレーム間の対応と DTW 距離 $d^2(D_{(k,v)}^{(i)}, D_{(k,v)}^{(j)})$ を求める.
- プロトタイプ周期 $D_{(k,v)}^{(P_k)}$ を部分列から選ぶ.

$$P_k = \operatorname{argmin}_{1 \leq i \leq M} \sum_{j=1}^M d^2(D_{(k,v)}^{(i)}, D_{(k,v)}^{(j)}) \quad (4)$$

- プロトタイプ $D_{(k,v)}^{(P_k)}$ の各フレームに対応する位相 $\{\phi_t \mid 1 \leq t \leq T_{P_k}\}$ を次式で定義する.

$$\phi_t = \frac{2\pi}{T_{P_k}}(t-1) \quad (5)$$

- 非プロトタイプ $\{D_{(k,v)}^{(i)} \mid 1 \leq i \leq M, i \neq P_k\}$ の各フレームの位相を, それぞれ対応するプロトタイプ周期のフレームの位相とする.

4.2 モデルパラメータの学習

状態遷移確率 $\Pr(\mathbf{x}|\mathbf{x}_{t-1}, \lambda_k)$ は, ガウス分布 $N(\boldsymbol{\mu}_x^{(k)}, \Sigma_x^{(k)})$ で近似する. 位相 ϕ_t はプロトタイプ周期 $D_{(k,v)}^{(P_k)}$ とほぼ同様に遷移し, カメラ視点 ψ_t は緩やかに変化すると仮定する. なお, 位相とカメラ視点に相関は無いとする.

$$\boldsymbol{\mu}_x^{(k)} = (\phi_{t-1} + 2\pi/T_{P_k}, \psi_{t-1})^T \quad (6)$$

$$\Sigma_x^{(k)} = \operatorname{diag}\left(\left((2\pi/T_{P_k})/3\right)^2, (c_v^{(k)})^2\right) \quad (7)$$

ここで, $c_v^{(k)}$ はカメラ視点の分散であり, 経験的に求める.

観測確率 $\Pr(\mathbf{y}|\mathbf{x}_t, \lambda_k)$ は, ガウス分布 $N(\boldsymbol{\mu}_y^{(k)}, \Sigma_y^{(k)})$ で近似する. 訓練データから状態変数 \mathbf{x}_t と観測変数 \mathbf{y}_t の関係を学習した RBF ネットワーク $f_k(\mathbf{x}_t)$ から, $\boldsymbol{\mu}_y^{(k)}$ を補間近似する.

$$\boldsymbol{\mu}_y^{(k)} = f_k(\mathbf{x}_t) \quad (8)$$

なお, 共分散行列 $\Sigma_y^{(k)}$ は経験的に求める.

CMM はモデル λ_k について, これら確率分布のパラメータ θ_k を持つ.

$$\theta_k = \left(\boldsymbol{\mu}_x^{(k)}, \Sigma_x^{(k)}, \boldsymbol{\mu}_y^{(k)}, \Sigma_y^{(k)}\right) \quad (9)$$

5 識別フェーズ

観測画像列 $\mathbf{y}_{1:T}$ から識別するため, $\mathbf{y}_{1:T}$ におけるモデル尤度 $\{L(\lambda_k) \mid 1 \leq k \leq K\}$ を計算する. そして, そのモデル尤度から被識別者 k^* を推定する.

$$k^* = \operatorname{argmax}_{1 \leq k \leq K} L(\lambda_k) \quad (10)$$

モデル尤度 $L(\lambda_k)$ は, 次式のように定義される (簡単化のため, モデル λ_k の添字 k を省略する).

$$\begin{aligned} L(\lambda) &= \log \Pr(\mathbf{y}_{1:T}|\lambda) \\ &= \sum_{t=1}^T \log \Pr(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \lambda) \end{aligned} \quad (11)$$

ここで,

$$\begin{aligned} \Pr(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \lambda) &= \int \Pr(\mathbf{y}_t|\mathbf{x}_t, \lambda) \Pr(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \lambda) d\mathbf{x}_t \\ &= \iint \Pr(\mathbf{y}_t|\mathbf{x}_t, \lambda) \Pr(\mathbf{x}_t|\mathbf{x}_{t-1}, \lambda) \\ &\quad \Pr(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \lambda) d\mathbf{x}_t d\mathbf{x}_{t-1} \end{aligned} \quad (12)$$

5.1 PF を用いたモデル尤度計算

PF は, 状態推定問題を効率的に解くための近似計算手法である [3]. 状態空間に散布されたパーティクル (重み付きサンプル) により, 事後確率分布の近似を行う. 時刻 t の i 番目のサンプルを $\mathbf{x}_t^{(i)} = (\phi_t^{(i)}, \psi_t^{(i)})^T$, その重みを $w_t^{(i)}$ とする. N 個の重み付きサンプル集合 $\mathcal{X}_t|\lambda = \{\mathbf{x}_t^{(i)}, w_t^{(i)} \mid 1 \leq i \leq N\}$ を用いて, 事後確率分布 $\Pr(\mathbf{x}_t|\mathbf{y}_{1:t}, \lambda)$ は次式で近似計算される. なお, $w_t^{(i)}$ は $\sum_{i=1}^N w_t^{(i)} = 1$ を満たす.

$$\Pr(\mathbf{x}_t|\mathbf{y}_{1:t}, \lambda) \approx \sum_{i=1}^N w_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) \quad (13)$$

ここで,

$$\delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) = \begin{cases} 1 & \text{if } \mathbf{x}_t = \mathbf{x}_t^{(i)} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

式 (13) から, 式 (12) の $\Pr(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \lambda)$ を以下のように近似する.

$$\Pr(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \lambda) \approx \sum_{i=1}^N w_{t-1}^{(i)} \Pr(\mathbf{y}_t|\tilde{\mathbf{x}}_t^{(i)}, \lambda) \quad (15)$$

$\{\tilde{\mathbf{x}}_t^{(i)} \mid 1 \leq i \leq N\}$ は, 遷移確率 $\Pr(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}, \lambda)$ に従って, $\{\mathbf{x}_{t-1}^{(i)} \mid 1 \leq i \leq N\}$ を遷移させたサンプルを示す.

モデル尤度 $L(\lambda)$ は、次式のように近似される。

$$L(\lambda) = \sum_{t=1}^T \log \Pr(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \lambda) \\ \approx \sum_{t=1}^T \log \left(\sum_{i=1}^N w_{t-1}^{(i)} \Pr(\mathbf{y}_t | \tilde{\mathbf{x}}_t^{(i)}, \lambda) \right) \quad (16)$$

5.2 PF のアルゴリズム

時刻 $t-1$ の $\mathcal{X}_{t-1|\lambda}$ が与えられると、以下の手順を N 回繰り返すことで、 $\mathcal{X}_{t|\lambda}$ が計算される。

1. $\mathcal{X}_{t-1|\lambda}$ で近似される分布に従って、時刻 $t-1$ のサンプル $\mathbf{x}_{t-1}^{(i)}$ をリサンプルする。

$$\mathbf{x}_{t-1}^{(i)} \sim \mathcal{X}_{t-1|\lambda} \quad (17)$$

2. 状態遷移確率 $\Pr(\mathbf{x}|\mathbf{x}_{t-1}, \lambda)$ に従って、時刻 t のサンプル $\mathbf{x}_t^{(i)}$ を生成する。

$$\mathbf{x}_t^{(i)} \sim \Pr(\mathbf{x}|\mathbf{x}_{t-1}^{(i)}, \lambda) \quad (18)$$

3. 観測確率 $\Pr(\mathbf{y}|\mathbf{x}_t, \lambda)$ を用いて、サンプル $\mathbf{x}_t^{(i)}$ の重み $w_t^{(i)}$ を計算する。

$$w_t^{(i)} \approx \Pr(\mathbf{y}_t | \mathbf{x}_t^{(i)}, \lambda) \quad (19)$$

初期集合 $\mathcal{X}_{0|\lambda}$ は、一様分布からサンプリングする。なお、重みは均等に $1/N$ とする。

6 シミュレーション

提案手法の有効性を示すため、シミュレーション実験を行う。また、従来手法の1つである Baseline アルゴリズムと比較し、その結果を報告する。

Baseline アルゴリズムは、観測されるプロープ (テストデータ) と事前に用意されたギャラリー (訓練データ) から計算されるマッチングスコアを用いて、個人の識別を行う [9]。マッチングスコアは単純にシルエット画像間の類似度から計算される。それぞれのカメラ視点が大きく異なる場合、その識別精度は期待できない。そこで、複数のカメラ視点から撮影された画像列を訓練データとして扱えるように手法を改善する。この詳細は付録 A に示す。

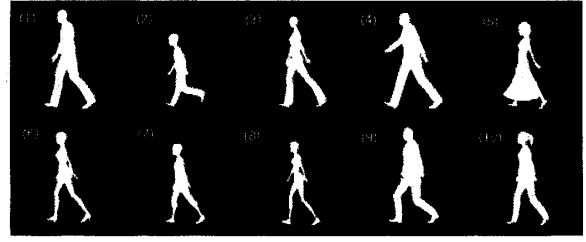


図 1: CG ソフトで作成した候補者のシルエット画像。左上の数字はそれぞれの固有番号 k を示す。

6.1 データ

市販 CG ソフト [4] を用いて、図 1 が示す 10 体の候補者を生成する。また、6 種類の歩行特徴を生成し、特徴 1 を候補者 1 および 7、特徴 2 を候補者 3, 5 および 10、特徴 3 を候補者 6 および 8、特徴 4 を候補者 2、そして、特徴 5 を候補者 4 に割り当てる。

6.1.1 画像特徴量ベクトル

画像特徴量として Hu's seven moments を用いる。この特徴量はシルエット画像から計算され、スケール、並進および回転に不変という性質を持つ [6]。

6.1.2 訓練データ

以下のようにして、訓練データ $\{D_{(k,v_i)} | 1 \leq k \leq 10, 1 \leq i \leq 8\}$ を取得する。各候補者をトレッドミルで歩かせ、固定単眼カメラで撮影する。カメラは図 2(a) のように、正面 ($v = 0$)、左斜め前 ($v = \pi/4$)、左 ($v = \pi/2$)、そして、右斜め前 ($v = 7\pi/4$) に配置する。シルエット画像を考えると、画像それぞれを左右反転させることで、対照方向のカメラからの画像を合成できる。これにより、8通りのカメラ視点 $v_i \in \{(i-1)\pi/4 | 1 \leq i \leq 8\}$ からの画像列を取得する。候補者 $k = 1$ の訓練データの一部を図 2(b) に示す。訓練データから、モデルパラメータ $\{\mu_{\mathbf{x}}^{(k)}, \Sigma_{\mathbf{x}}^{(k)}, \mu_{\mathbf{y}}^{(k)} | 1 \leq k \leq 10\}$ を学習する。また、経験的に $c_v^{(k)} = 2\pi/32$, $\Sigma_{\mathbf{y}}^{(k)} = \text{diag}(0.3, 0.5, 1.8, 0.5, 1.3, 0.8, 0.4)$ とする。

6.1.3 テストデータ

テストデータとして、50 個の画像列を生成する。図 3(a) で示される 5 種類の歩行パターンに沿って、それぞれの候補者を歩かせる。テストデータの一部を図 3(b) に示す。これらのカメラ視点は、パター

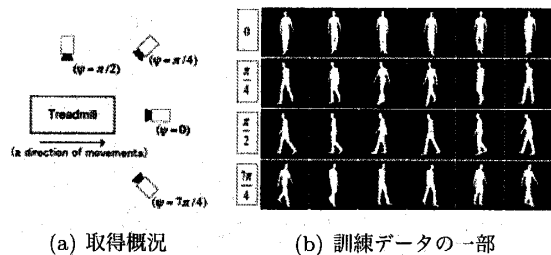


図 2: 訓練データ

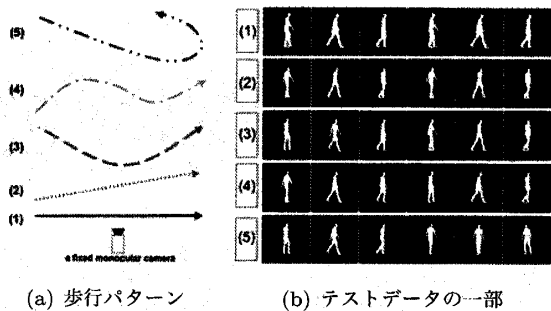


図 3: テストデータ

表 1: 識別精度

| テストデータ | 提案手法 | 従来手法 |
|------------|--------|--------|
| 1 | 100 % | 100 % |
| 2 | 100 % | 70.0 % |
| 3 | 100 % | 60.0 % |
| 4 | 90.0 % | 50.0 % |
| 5 | 90.0 % | 50.0 % |
| 1, 2 | 100 % | 80.0 % |
| 3, 4, 5 | 93.3 % | 53.3 % |
| 2, 3, 4, 5 | 95.0 % | 57.5 % |
| 合計 | 96.0 % | 66.0 % |

ン 1 とパターン 2 では一定だが、その他のパターンでは変化する。また、パターン 1 のみ、そのカメラ視点が訓練データに含まれる。簡単化のため、歩行パターン i のテストデータをテスト i と表記する。

6.2 結果

モデル λ_k それぞれの状態空間モデルに 100 個のサンプルを与える。つまり、合計 1000 個のサンプルから識別を行う。提案手法および Baseline アルゴリズムにおける識別精度を表 1 に示す。また、被識別者 $k = 5$ におけるテスト 1 およびテスト 5 でのモデル尤度の遷移を図 4 に示す。なお、この図は正確な識別結果を示す。

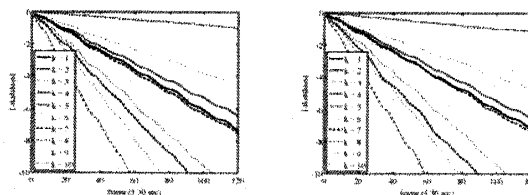


図 4: 被識別者 $k = 5$ におけるモデル尤度の遷移

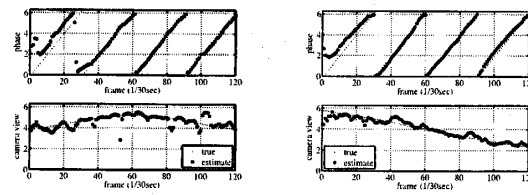


図 5: 被識別者 $k = 1$ における状態変数の推定

6.2.1 テストデータ 1 および 2

Baseline アルゴリズムのテスト 2 における識別率は 70% である。テスト 2 のカメラ視点は訓練データ (ギャラリー) に含まれていないため、Baseline アルゴリズムでは誤識別が起こる。提案手法では、状態変数と画像特徴量の関係を RBF ネットワークにより関数近似するため、識別精度はその近似精度に影響される。提案手法はテスト 1 および 2 共に 100% である。これは近似が正確に行われたことを示す。6.3 節にて、RBF ネットワークの近似精度について、さらに議論する。

6.2.2 テストデータ 3, 4 および 5

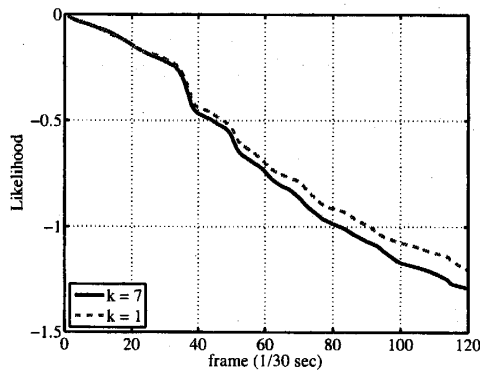
これらテストデータのカメラ視点は時間変化する。変化に対応していない Baseline アルゴリズムは、提案手法と比べて、低い識別精度となる。正確に識別するためには、カメラ視点の変化を正確に推定しなければならない。推定値 k^* が与えられると、事後確率分布 $\Pr(\mathbf{x}_t | \mathbf{y}_{1:t}, \lambda_{k^*})$ 、つまり、サンプル集合 $\mathcal{X}_t | \lambda_{k^*}$ の最頻値から、状態変数 $\mathbf{x}_t = (\phi_t, \psi_t)^T$ の推定値 $\hat{\mathbf{x}}_t$ が決まる。

$$\hat{\mathbf{x}}_t \approx \text{mode}(\mathcal{X}_t | \lambda_{k^*}) \quad (20)$$

被識別者 $k = 1$ におけるテスト 4 およびテスト 5 での状態変数の推定値と真値を図 5 に示す。状態変数は正確に推定されている。序盤に推定誤差が見られるが、時間と共に小さくなっている。

表 2: RBF ネットワークの近似精度

| | | | | | |
|-----|-------|-------|-------|-------|-------|
| 候補者 | 1 | 2 | 3 | 4 | 5 |
| 誤差 | 0.114 | 0.103 | 0.130 | 0.100 | 0.097 |
| 候補者 | 6 | 7 | 8 | 9 | 10 |
| 誤差 | 0.121 | 0.085 | 0.078 | 0.126 | 0.101 |

図 6: 被識別者 $k = 7$ におけるテストデータ 5 でのモデル尤度の遷移

6.3 考察

RBF ネットワーク $f_k(x_t)$ を用いた観測確率パラメータ $\mu_y^{(k)}(x_t)$ の補間精度を検証する。なお、隠れユニットは 40 個である。CG ソフトを用いて、さまざまなカメラ視点 $v \in \{(i-1)\pi/4 \mid 1 \leq i \leq 8\} \cup \{(j-1)\pi/6 \mid 1 \leq j \leq 12\}$ からの画像列を生成する。これら画像列を $f_k(x_t)$ の真値とし、相関誤差を計算する。すべての候補者において、その相関誤差が小さいことが表 2 から分かる。

提案手法は、テスト 2 およびテスト 5 における被識別者 $k = 7$ の識別に失敗する。被識別者 $k = 7$ を候補者 $k = 1$ と間違える。テスト 5 でのモデル尤度の遷移を図 6 に示す。それぞれのモデル尤度は同等に変化する。両対象の歩行特徴は一致し、体型は比例する。スケール不変な画像特徴量を用いるため、これら対象を上手く区別することは難しい。新たな画像特徴量を加えることで、克服できるだろう。

7 実験

実際の識別問題を想定した予備実験を示す。前章にて提案手法の有効性が示されるが、シミュレーションである。実画像では、撮影環境やシルエット画像生成による画像ノイズが精度に影響する。



図 7: 候補者 (実画像)

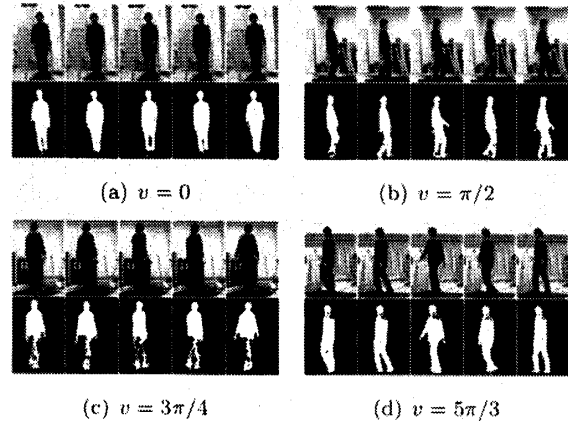


図 8: 訓練データ (実画像)

7.1 データ

図 7 が示す候補者 3 人からデータを取得する。なお、画像特徴量ベクトルはシミュレーションと同様に、Hu's seven moments を用いる。

7.1.1 画像ノイズの対応

シルエット画像の生成における画像ノイズは以下のようにして、軽減する。単純な背景差分で求めたシルエット画像に対して、膨張と収縮の処理を行い、軽微なノイズを削除する。そして、その画像にラベリング処理を行い、画像内の領域（連結した画素集合）を調べる。検出された領域から面積最大となる領域を人体として、その領域のみを抽出する。しかしながら、人体のみを完璧に抽出することは困難である。

7.1.2 訓練データ

6.1.2 節と同様にして、訓練データ $\{D_{(k,v_i)} \mid 1 \leq k \leq 3, 1 \leq i \leq 8\}$ を取得する。撮影概況および訓練データ（シルエット画像）の一部を図 8 に示す。

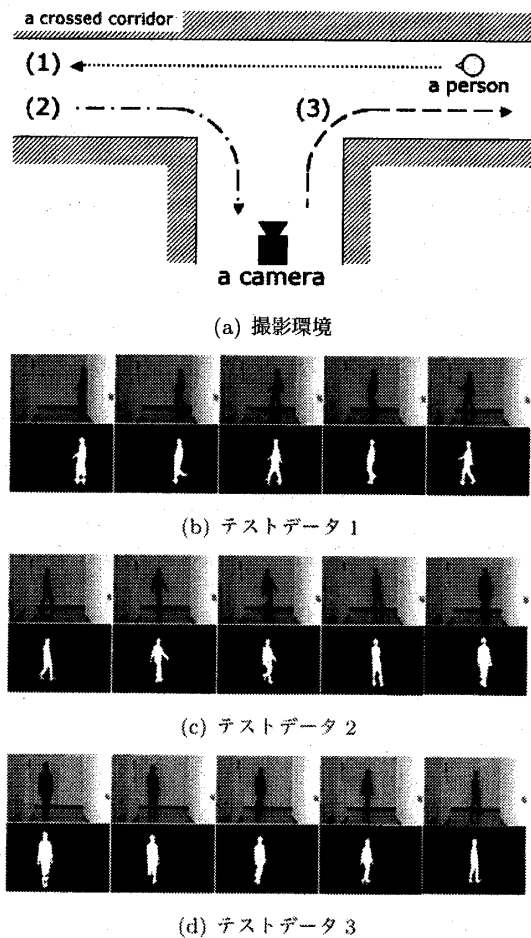


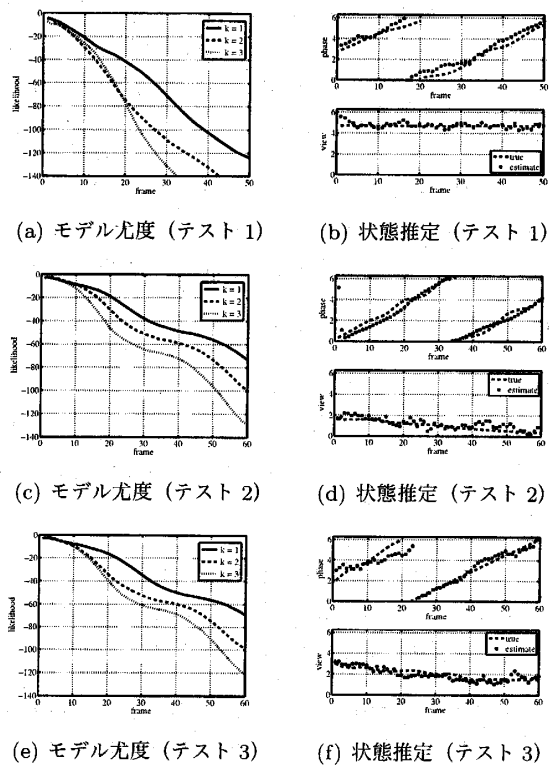
図 9: テストデータ (実画像)

7.1.3 テストデータ

丁字交差の廊下を考える。図 9(a) が示す 3 種類の歩行パターンに沿って、候補者を歩かせる。候補者 $k=1$ におけるテストデータの一部を図 9 に示す。

7.2 結果

被識別者 $k=1$ におけるモデル尤度の計算結果と状態変数の推定結果を図 10 に示す。なお、サンプルは候補者それぞれに 100 個である。テストデータそれぞれにおいて、モデル尤度 $L(\lambda_1)$ が最大となり、被識別者 $k=1$ は正確に識別される。また、状態変数の推定結果はシミュレーションと比べると誤差が見られるが、よく推定されている。提案手法は、画像ノイズが起る実環境でも適用できることがわかる。

図 10: 被識別者 $k=1$ における実験結果

8 まとめ

歩き方による個人認識について取り組んだ。従来手法の多くは画像マッチングを基礎とするため、カメラ視点に依存した。テストデータと訓練データのカメラ視点が大きく異なると、その識別精度は低くなる。

我々はカメラ視点に依存しない識別手法を提案し、その有効性を実験により示した。位相およびカメラ視点を状態変数に持つ状態空間モデル CMM を提案し、候補者それぞれのモデルパラメータを学習し、歩行パターンを表現した。PF を用いて、観測画像列に対応する状態変数を推定した。そして、それぞれのモデル尤度を計算し、そのモデル尤度から識別した。

参考文献

- [1] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *IEEE Conf. on Automatic Face and Gesture Recognition*, pages 366–371, 2002.

- [2] F. Cuzzolin and S. Soatto. Using bilinear models for view-invariant identity recognition from gait. Technical Report CSD-TR050007, University of California at Los Angeles, 2005.
- [3] A. Doucet, N. Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001. ISBN 0-387-95146-6.
- [4] e frontier Inc. POSER 6.
- [5] M. Emoto, A. Hayashi, and N. Suematsu. Efficient posture estimation using the particle filter. In *Brain-Inspired Information Technologies*, 2005.
- [6] M. K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. information Theory, IT-8*, pages 179–187, 1962.
- [7] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntorr, A. K. Roy-Chowdhury, V. Krhüger, and R. Chellappa. Identification of humans using gait. *IEEE Trans. on Image Processing*, 13(9):1163–1173, 2004.
- [8] L. Rabiner and B. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, 1995. ISBN 0-130-15157-2.
- [9] S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. Bowyer. The human id gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.
- [10] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 439–446, 2001.
- [11] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

A Baseline アルゴリズム

Baseline アルゴリズムは、シルエット画像列間のマッチングスコアを用いる識別手法である [9]。プローブ (テスト) シルエット画像列 $S_P = \{S_P(1), S_P(2), \dots, S_P(M)\}$ とギャラリー (訓練) シルエット画像列 $S_G = \{S_G(1), S_G(2), \dots, S_G(N)\}$ のマッチングスコア $\text{Sim}(S_P, S_G)$ は、次のように計算される。ここで、プローブ S_P における周期の長さを N_{gait} 、プローブ部分列を $S_{P_t} = \{S_P(t), S_P(t+1), \dots, S_P(t+N_{\text{gait}}-1)\}$ ($1 \leq t \leq M - N_{\text{gait}} + 1$) とする。

1. i 番目のプローブ画像 $S_P(i)$ と j 番目のギャラリー画像 $S_G(j)$ とのスコアを $\text{FrameSim}(S_P(i), S_G(j))$ とする。

$$\text{FrameSim}(S_P(i), S_G(j)) = \frac{\text{Num}(S_P(i) \cap S_G(j))}{\text{Num}(S_P(i) \cup S_G(j))}$$

ここで、 $\text{Num}(S)$ はシルエット画像 S における人体領域のピクセル数である。

2. プローブ部分列 S_{P_t} とギャラリー列 S_G とのスコア $\{\text{Corr}(S_{P_t}, S_G)(l) \mid 1 \leq l \leq N - N_{\text{gait}} + 1\}$ を計算する。

$$\begin{aligned} \text{Corr}(S_{P_t}, S_G)(l) \\ = \sum_{j=0}^{N_{\text{gait}}-1} \text{FrameSim}(S_P(t+j), S_G(l+j)) \end{aligned}$$

3. マッチングスコア $\text{Sim}(S_P, S_G)$ を計算する。

$$\text{Sim}(S_P, S_G) = \text{median}_t \left(\max_l \text{Corr}(S_{P_t}, S_G)(l) \right)$$

Baseline アルゴリズムは、訓練データとカメラ視点が異なるテストデータを扱えない。そこで、Baseline アルゴリズムを改良する。さまざまなカメラ視点から撮影された画像列を訓練データとして、最も一致するカメラ視点の画像列を選択する。

V 個のカメラ視点から撮影された候補者 $\{k \mid 1 \leq k \leq K\}$ の訓練データを $S_G = \{D_{(k,v_i)} \mid 1 \leq i \leq V\}$ として、候補者 k におけるマッチングスコアを $\text{Sim}_k(S_P, S_G)$ と定義する。

$$\text{Sim}_k(S_P, S_G) = \max_{1 \leq i \leq V} \text{Sim}(S_P, D_{(k,v_i)})$$

そして、次式から被識別者を決定する。

$$k^* = \operatorname{argmax}_{1 \leq k \leq K} \text{Sim}_k(S_P, S_G)$$