

## 逐次学習型顔画像認識における SMC サンプル数の動的制御

Pruned Resampling: a new SMC resampling algorithm for Sequential Face Recognition

松井 淳†, ‡  
Atsushi Matsuiサイモン クリッピングデル†  
Simon Clippingdale松本 隆‡  
Takashi Matsumoto

## 1. まえがき

近年、動物体追跡などの分野で、非線形・非ガウス型状態空間モデルである逐次モンテカルロ法が注目されている。逐次モンテカルロを含むモンテカルロ法の近似誤差は、注目する確率分布から抽出するサンプルの総数が無限大となる極限において0に漸近するが、計算コストはサンプル数に比例して増大する。すなわち、近似精度と計算コストはトレードオフの関係にあり、両者を直接的に左右するサンプル数の設定は、逐次モンテカルロ法の実装において、きわめて重要な問題である。本稿では、同手法の具体的応用として筆者らが提案した逐次学習型顔画像認識をとりあげ、サンプル数を統計的指標に基づいて効率的に削減する手法を提案する。

## 2. 逐次学習型 Bayes 的顔画像認識

筆者らが提案した逐次学習型顔画像認識[1]は、Elastic Graph Matching[2][3]をベースとした可変テンプレートマッチングにおいて、顔テンプレートに付随するいくつかのパラメータを確率変数とみなし、それらの確率分布から、階層的 Bayes 推定[4][5]にもとづいて顔テンプレートの出現確率(事後確率)を計算する。顔テンプレートは、登録顔画像上に定義した  $N$  個の特徴点の配置  $\mathbf{x}_A \in \mathbb{R}^{2 \times N}$  と、各特徴点における  $K$  方位・ $R$  解像度のガボール・ウェーブレット係数  $\mathbf{c}^A \in \mathbb{C}^{N \times K \times R}$ 、および人物 ID で構成される。特徴点配置の具体例を図1に示す。

連続した入力画像列  $y_{1:n} = \{y_1, y_2, \dots, y_n\}$  が逐次的に与えられる場合について考える。  $n$  番目の入力画像上の特徴点配置を  $\mathbf{x}_n \in \mathbb{R}^{2 \times N}$  とおき、各特徴点の位置  $\mathbf{x}_n$  で計算したガボール・ウェーブレット係数を  $\mathbf{c}^B \in \mathbb{C}^{N \times K \times R}$  とおく。Wiskott ら[2]の手法は、 $\mathbf{c}^A$ ,  $\mathbf{c}^B$  の間の類似度を各々の解像度  $r$  に関する成分ベクトル  $\mathbf{c}_r^A$ ,  $\mathbf{c}_r^B$  ごとに定義し、各々のノルムで各々正規化したもとの内積で評価しているが、これはベクトル  $\mathbf{c}_r^B$  を指向性データとみなすことを意味する。そこで本稿では、指向性データの自然な生成モデルの一つである von Mises-Fisher 分布[6]を用いて尤度を定義する。

$$P(y_n | \mathbf{x}_n, \beta_{n:R}, \mathcal{H}_j) = \frac{1}{R} \sum_{r=1}^R \frac{1}{Z_b(\beta_{n,r})} \exp \left( \beta_r \frac{\langle \mathbf{c}_r^A, \mathbf{c}_r^B \rangle}{\|\mathbf{c}_r^A\| \|\mathbf{c}_r^B\|} \right) \quad (1)$$

ここで、 $\langle \mathbf{c}_r^A, \mathbf{c}_r^B \rangle$  は二つの複素ベクトル  $\mathbf{c}_r^A$ ,  $\mathbf{c}_r^B$  の内積をあらわす。また、 $\mathcal{H}_j$  は仮定した  $j$  番目の顔テンプレート、 $\beta_{n:R} = \{\beta_{n,1}, \beta_{n,2}, \dots, \beta_{n,R}\}$  は時刻  $n$  における尤度関数の鋭さを決定するハイパーパラメータをあらわす。  $Z_b$  は von Mises-Fisher 分布の規格化定数である。

$\mathbf{x}_n$  の予測確率分布は、筆者らが行った予備実験の結果[7]をふまえて、多次元正規分布で定義する。

$$P(\mathbf{x}_n | \alpha_n, T_n, \mathcal{H}_j) = \frac{1}{Z_\alpha(\alpha_n)} \exp \left( -\frac{\alpha_n}{2} (T_n^{-1}(\mathbf{x}_n) - \mathbf{x}_j^A)' \Lambda_j^{-1} (T_n^{-1}(\mathbf{x}_n) - \mathbf{x}_j^A) \right) \quad (2)$$

ここで、 $\Lambda_j$  は正規特徴点配置空間上での特徴点配置の共分散行列をあらわす。  $\alpha$  は事前分布の鋭さを決定するハイパーパラメータであり、 $Z_\alpha$  は規格化定数である。

$T_n$  は、顔テンプレートの特徴点配置  $\mathbf{x}_j^A$  が張る空間から、 $\mathbf{x}_n$  が張る空間への写像をあらわす。以下、表記の簡略化のため、 $T_n$  を決定する4つのパラメータの組  $(r_n, \theta_n, u_n, v_n)$  を記号  $T_n$  であらわす。  $T_n$  の各パラメータは、入力画像上の顔の中心位置と大きさ、画像平面内の回転を決定する。本稿では、顔全体の動きが極端に激しくない場合を想定し、 $T_n$  の逐次更新則  $P(T_n | T_{n-1})$  として次のような確率的ダイナミカルシステムを考える。

$$\begin{aligned} r_n &= r_{n-1} + v_r, & v_r &\sim N(0, \delta_r^2), \\ \theta_n &= \theta_{n-1} + v_\theta, & v_\theta &\sim N(0, \delta_\theta^2), \\ u_n &= u_{n-1} + v_u, & v_u &\sim N(0, \delta_u^2), \\ v_n &= v_{n-1} + v_v, & v_v &\sim N(0, \delta_v^2). \end{aligned} \quad (3)$$

尤度関数および事前確率の分布の鋭さを決定するハイパーパラメータについては、正值性の必要条件をふまえて、次のような予測モデル  $P(\alpha_n, \beta_{n:R} | \alpha_{n-1}, \beta_{n-1:R})$  を仮定する：

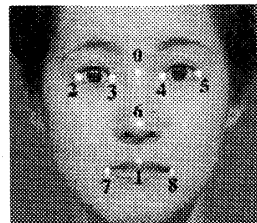
$$\begin{aligned} \log \alpha_n &= \log \alpha_{n-1} + v_\alpha, & v_\alpha &\sim N(0, \delta_\alpha^2), \\ \log \beta_{n,r} &= \log \beta_{n-1,r} + v_\beta, & v_\beta &\sim N(0, \delta_\beta^2). \end{aligned} \quad (4)$$

データ  $y_{1:n}$  が与えられたもとの  $\mathcal{H}_j$  の事後確率は、Bayes の定理より、次式によって再帰的に与えられる。

$$P(\mathcal{H}_j | y_{1:n}) = \frac{P(y_n | y_{1:n-1}, \mathcal{H}_j) P(\mathcal{H}_j | y_{1:n-1})}{P(y_n | y_{1:n-1})} \quad (5)$$

システムは、逐次的に与えられた入力画像列  $y_{1:n}$  についての事後確率が最大となるモデル  $\mathcal{H}_{MP}$  を、時刻  $n$  における認識結果として出力する。

$$\mathcal{H}_{MP}^{(n)} = \arg \max_{\mathcal{H}} P(\mathcal{H} | y_{1:n}) \quad (6)$$

図1: 特徴点の配置 ( $N=9$ )

† NHK 放送技術研究所 (知能情報処理)

‡ 早稲田大学理工学術院

モデルの初期分布  $P(\mathcal{H}_j|y_0)$  は、登録した  $N_{persons}$  個の顔モデルについて等確率:  $P(\mathcal{H}_j|y_0) = 1/N_{persons}$  とする. 式(5)にあらわれる  $\mathcal{H}_j$  の尤度は次式で与えられる.

$$\begin{aligned} P(y_n | y_{1:n-1}, \mathcal{H}_j) \\ = \int P(y_n | x_n, \beta_{n,1,R}, \mathcal{H}_j) P(x_n | \alpha_n, T_n, \mathcal{H}_j) \\ P(T_n | T_{n-1}) P(\alpha_n, \beta_{n,1,R} | \alpha_{n-1}, \beta_{n-1,1,R}) \\ d(x_n, \alpha_n, \beta_{n,1,R}, T_n) \end{aligned} \quad (7)$$

この計算は多次元連続空間に関する積分であり、解析的に計算を実行することは不可能であるが、本稿では逐次モンテカルロ法[8][9] (Sequential Monte Carlo, SMC法) により近似する. 次のような *proposal distribution*,  $\pi$  を考える.

$$\begin{aligned} \pi(x_n, \alpha_n, \beta_n, T_n) \\ = P(x_n | \alpha_n, T_n, \mathcal{H}_j) P(T_n | T_{n-1}, \Sigma_T) P(\alpha_n | \alpha_{n-1}, \sigma_\alpha) P(\beta_n | \beta_{n-1}, \sigma_\beta) \end{aligned} \quad (8)$$

$\pi$  に従う  $N_j$  個のサンプルが抽出されたとき、式(7)の積分は次式で近似的に計算可能である.

$$\begin{aligned} P(y_n | y_{1:n-1}, \mathcal{H}_j) \\ = \int P(y_n | x_n, \beta_{n,1,R}, T_n, \mathcal{H}_j) P(x_n, \alpha_n, \beta_{n,1,R}, T_n | y_{1:n-1}, \mathcal{H}_j) \\ d(x_n, \alpha_n, \beta_{n,1,R}, T_n) \\ \approx \int P(y_n | x_n, \beta_{n,1,R}, T_n, \mathcal{H}_j) \\ \times \sum_{i=1}^{N_j} \tilde{w}_n^{(i)} |_{\mathcal{H}_j} \delta(\|(x_n, \alpha_n, \beta_{n,1,R}, T_n) - (x_n^{(i)}, \alpha_n^{(i)}, \beta_{n,1,R}^{(i)}, T_n^{(i)})\|) \\ d(x_n, \alpha_n, \beta_{n,1,R}, T_n) \\ = \sum_{i=1}^{N_j} P(y_n | x_n^{(i)}, \beta_{n,1,R}^{(i)}, T_n^{(i)}, \mathcal{H}_j) \tilde{w}_n^{(i)} |_{\mathcal{H}_j} \end{aligned} \quad (9)$$

ここで、 $\tilde{w}_n^{(i)} |_{\mathcal{H}_j}$  は提案分布と事後確率分布との誤差を補正する *normalized importance weight* であり、式(8)の *proposal distribution* を用いる場合、次式で計算される.

$$w_n^{(i)} |_{\mathcal{H}_j} = P(y_n | x_n^{(i)}, \alpha_n^{(i)}, \beta_{n,1,R}^{(i)}, T_n^{(i)}, y_{1:n-1}, \mathcal{H}_j) \times w_{n-1}^{(i)} |_{\mathcal{H}_j} \quad (10)$$

$$\tilde{w}_n^{(i)} |_{\mathcal{H}_j} = \frac{w_n^{(i)} |_{\mathcal{H}_j}}{\sum_{k=1}^{N_{persons}} \sum_{q=1}^{N_k} w_n^{(q)} |_{\mathcal{H}_k}} \quad (11)$$

### 3. 提案手法

逐次モンテカルロ法では、時間の経過とともに増大する *importance weight* の分散を抑制するために、サンプルの再抽出 (リサンプル) が行われる. サンプルの総数  $N_{total}$  は、モンテカルロ近似の精度を決定する一つの重要な要素であり、多くの場合、リサンプルのアルゴリズムは  $N_{total}$  が不変となるように設計される. 本稿の目的は、このリサンプルの処理において、サンプルの総数を統計的指標に基づいて動的に制御することで、最終的な認識精度への影響を最小限にとどめつつ、計算コストを効果的に削減することである. 以下、表記が複雑になることを避けるため、個々のサンプルを  $S^{(i)}$  であらわし、 $j$  番目のモデルに属する  $i$  番目の *normalized importance weight* を  $\tilde{w}_j^{(i)}$  であらわす.

リサンプルの具体的な実現方法はいくつか提案されているが、本稿では  $N_{total}$  が不変の、一般的な手法の一つである Residual Resampling[10][11]について考える. このアルゴリズムは、次の手順に従ってリサンプルを実現する.

1. 各々のサンプル  $S^{(i)}$  から、 $k^{(i)} = \lfloor N_{total} \times \tilde{w}_j^{(i)} \rfloor$  個のサンプルを複製する. ( $\lfloor x \rfloor$  は  $x$  の整数部)
2. 残りの  $R = \sum_{i=1}^{N_{total}} (N_{total} \times \tilde{w}_j^{(i)}) - k^{(i)}$  個のサンプルを、 $(N_{total} \times \tilde{w}_j^{(i)}) - k^{(i)}$  に応じた確率で抽出する.
3. 操作1.と2.で抽出した全てのサンプルの *importance weight* を  $\sum_{i=1}^{N_{total}} \tilde{w}_j^{(i)} / N_{total}$  に設定する.

このアルゴリズムは、再抽出するサンプルの総数  $N_{total}$  が不変で、かつ、個々のサンプルの抽出が *importance weight* の大小に応じておこなわれる. このとき、リサンプル後の個々のモデル間に属するサンプルの数の比率は、リサンプル前のそれらの *importance weight* の和の比率と同じである.

本稿で取り上げる問題では、各時刻におけるモデル事後確率分布のピークを求めることが最終的な目的であり、事後確率分布の裾野の部分の推定誤差は最終的な認識性能にほとんど影響しないと考えられる.

そこで、本稿では、リサンプル後に最も多くのサンプルが割り当てられるモデル、すなわち、リサンプルを行う時点でモデル周辺尤度最大となるモデルに属するサンプル数  $N_j$  が一定となる処理を考える. リサンプルによって各モデルに割り当てられるサンプルの数は、リサンプル前の *normalized importance weight* の和に比例することから、リサンプル後のサンプルの総数  $N_{total}$  は次式で計算される.

$$\begin{aligned} N_{total} &= N_j \times \frac{\sum_{k=1}^{N_{persons}} \sum_{i=1}^{N_k} \tilde{w}_k^{(i)}}{\sum_{i=1}^N \tilde{w}_j^{(i)}} \\ &= N_j \times \frac{\sum_{k=1}^{N_{persons}} P(y_n | y_{1:n-1}, \mathcal{H}_k)}{P(y_n | y_{1:n-1}, \mathcal{H}_j)} \end{aligned} \quad (12)$$

このとき、モデル周辺尤度最大の  $j$  番目以外のモデルについては、リサンプルの前後で割り当てられるサンプルの数は同一または減少する. したがって、このリサンプルのアルゴリズムは、モデル事後確率分布のピークからはなれたモデル、すなわち、最終的な認識結果として選ばれる可能性の低いモデルを統計的な指標にもとづいて枝狩り (Pruning) する一つの手法であると捉えることができる.

### 4. 比較実験

#### 4.1 実験条件

提案手法 (以下、Pruned Resampling と呼ぶ) の有効性を検証するため、Residual Resampling との比較実験を行った. 評価データは、10人の被験者の発話時の顔、1秒間を用いた. 各々の画像は、スタジオ照明・単色背景 (ブルーバック) の条件で撮影した正面顔の動画画像から、タテ240画素×ヨコ360画素のサイズ、30フレーム/秒のフレームレートで取得した. 登録データは、評価データと同じ条件で撮影した無表情・正面の顔画像1枚/人を用意した. 顔の位置情報の初期値  $T_0 = (r_0, \theta_0, u_0, v_0)$  は、あらかじめ手動で抽出した結果を用いた. 予測モデルの分散とハイパーパラメータの初期値は、 $\alpha_1^2 = \alpha_2^2 = 1.0$ ,  $\alpha_3^2 = 0.1$ ,  $\alpha_4^2 = 0.0001$ ,  $\alpha_5^2 = \alpha_6^2 = 0.01$ ,  $\alpha_0 = 10.0$ ,  $\beta_0 = 1.0$ , とした. サンプルの総数の初期値は一つの顔テンプレートあたり3600個に統一した.

4.2 実験結果

実験結果を表1に示す。表1より、提案手法 ( Pruned Resampling ) が従来手法 ( Residual Resampling ) と同じ認識性能を、約半分の処理時間で実現していることがわかる。次に、モデルごとのサンプルの分布の推移の様子をグラフにまとめ、図2 (従来手法) および図3 (提案手法) に示す。図2と図3を比較すると、モデル間のサンプル数の比率はほぼ同じであるが、ヒストグラムの高さの合計が示すサンプルの総数は、提案手法では時間の経過とともに指数関数的に減少する様子がわかる。その一方、周辺尤度最大となるモデル (図3での  $j=10$  のモデル) については、サンプル数がほぼ一定に保たれていることから、モンテカルロ近似の精度が最尤モデルについて一定の水準に保たれているとみなせる。

表1: 顔認識実験の結果 (顔認識誤り率)

Resample algorithm	Residual Resampling	Pruned Resampling
ID Error Rate	2.3 %	2.3 %
Proc. time	2329 sec.	1135 sec.

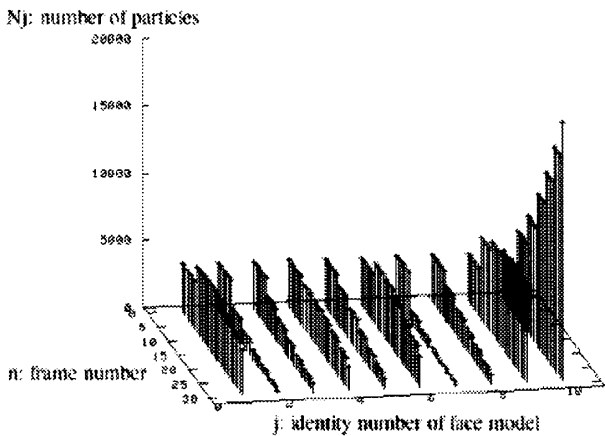


図2: 従来手法 ( Residual Resampling ) によるモデルごとのサンプル数の時間発展 (正解 ID=10)

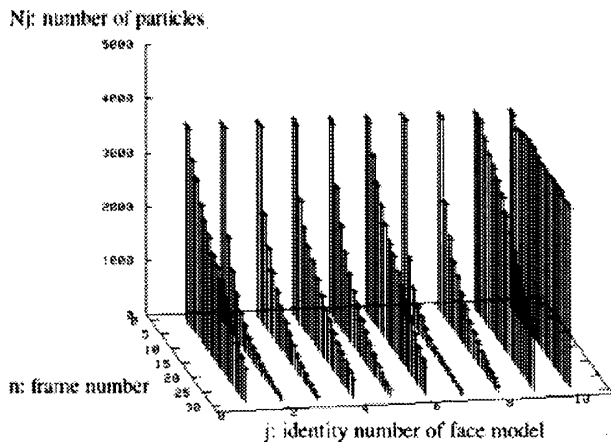


図3: 提案手法 ( Pruned Resampling ) によるモデルごとのサンプル数の時間発展 (正解 ID=10)

5. まとめ

逐次モンテカルロ法によるモデル比較の問題において、モデル周辺尤度に基づくサンプル数の動的制御法を提案した。各時刻における最尤モデルに属するサンプルの数が一定となるよう、リサンプルの段階で総サンプル数を調整することで、モデル事後確率分布の推定への影響を最小限にとどめたまま、計算コストの効率的な削減を可能にした。発話時の連続顔画像を用いた顔画像認識における比較実験の結果、サンプルの総数を固定した従来の手法に比べ、提案手法は、新たな顔認識誤りを生じることなく、処理時間を平均で約半分にまで削減した。

今後は、頭部や人物全体の動きを想定したパラメータの運動モデルの導入や、本実験で既知とした入力画像の開始・終了点 (カット点) の自動検出、ならびに、顔の位置情報の自動検出法について検討していきたい。

文献

- [1] 松井淳, S. Clippingdale, 藤井真人, 八木伸行 「逐次モンテカルロ法を用いた Bayes 的顔画像認識」, 信学技報, PRMU2005-250, pp.119-124 (2006).
- [2] L. Wiskott, J. M. Fellous, N. Krüger and von der C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," TR96-08, Institut für Neuroinformatik, Ruhr-Universität Bochum (1996).
- [3] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven and von der C. Malsburg, "The Bochum/USC Face Recognition System And How it Fared in the FERET Phase III Test," Face Recognition: From Theory to Applications, Springer (1998).
- [4] D.J.C. Mackay, "Information Theory, Inference, and Learning Algorithms," Cambridge University Press (2003).
- [5] R.M. Neal, "Bayesian Learning for Neural Networks," Lecture Notes in Statistics, Springer, pp.55-98 (1996).
- [6] I.S. Dhillon and S. Sra, "Modeling Data using Directional Distributions," TR-03-06, University of Texas at Austin (2003).
- [7] 松井淳, S. Clippingdale 「表情変化の事前情報を利用した可変テンプレートマッチングによる顔画像認識」, 信学技報, PRMU2003-161, pp.61-65 (2003).
- [8] A. Doucet, N.D. Freitas, and N. Gordon, "Sequential Monte Carlo Methods in Practice," Springer, (2001).
- [9] J.S. Liu, "Monte Carlo Strategies in Scientific Computing," Springer, pp.53-77 (2001).
- [10] D. Whitley, "A genetic algorithm tutorial," Stat. Comput. Vol. 4, pp.65-85 (1994).
- [11] J. Liu and R. Chen, "Sequential Monte-Carlo methods for dynamic systems," Vol. 93, pp.1032-1044 (1998).