ハミルトニアンモンテカルロ法による Bayes 的顔画像認識

Bayesian Face Recognition using a Hamiltonian Monte Carlo Method

正会員 松井 淳^{†,††}, 正会員 クリッピングデル サイモン[†],
 鶴澤 史貴^{†††}, 正会員 松本 隆^{††}

Atsushi Matsui^{†,††}, Simon Clippingdale[†],
 Fumiki Uzawa^{†††} and Takashi Matsumoto^{††}

Abstract We developed a new algorithm for face recognition using a Bayesian framework. In our algorithm, posterior distributions are computed using a Hamiltonian Monte Carlo method. The face features we used were those used in our previous work based on the elastic graph matching method. While our previous method attempted to optimize facial feature point positions to maximize a similarity function between each model and face region in the input sequence, our new approach evaluates posterior distributions of models conditioned on the input sequence. Experimental results show that our new algorithm out-performs the Eigenface method in terms of both identification errors and processing costs. Our new algorithm eliminates almost all identification errors on sequences showing individuals smiling, although such data was not used in training.

キーワード：顔画像認識，可変テンプレートマッチング，表情変化，Bayes 的手法，HMC 法，ロバスト性

1. ま え が き

顔画像認識の応用範囲は幅広く，セキュリティシステムのバイオメトリクス認証，画像データベースの自動インデキシング，対話型ロボットの人物識別などへの適用が期待される．顔画像認識は，その目的に応じて，入力画像の様々な変動に対応できなければならない．

筆者らは，放送素材映像の自動インデキシングを目的とした顔画像認識システム¹⁾²⁾を試作し，顔の向きの変化に対応可能な顔画像認識を実現した．これは可変テンプレートマッチングを利用しており，入力画像の変形にある程度対応可能な手法であるが，様々なパラメータを事前に最適化する必要があった．それゆえ登録人物の追加・削除，照明や背景などの撮影条件の変化が生じた場合，その都度パラメータの調整を行う必要がある．また，表情などで顔が

大きく変形する場合には認識誤りが著しく増加する傾向があった．

本稿では，このような問題を解決するため，システムの内部パラメータを確率変数とみなし，顔テンプレートに付随するパラメータの事後確率分布を，階層的 Bayes 推定の枠組みから捉え，その確率分布のサンプルをハミルトニアンモンテカルロ法で採取する．その上で尤度を事後確率分布に関して周辺化（積分）することで，顔画像認識を遂行する（Bayesian HMC）．筆者らの従来手法¹⁾²⁾における一部のパラメータに対して提案手法を適用し，パラメータの最適化問題の解決を図る．従来手法，および統計的顔画像認識法として代表的な Eigenface³⁾ との比較実験を行い，提案手法の有効性を検証する．

2. 可変テンプレートマッチング

筆者らが開発した顔画像認識システム¹⁾²⁾は，Elastic Graph Matching⁴⁾⁵⁾のデータ表現をベースとした可変テンプレートマッチングを行う．顔テンプレートは，登録顔画像上に定義した N_{nodes} 個の特徴点 $x_A \in R^{2 \times N_{nodes}}$ と，各特徴点における N_{orns} 個の方位と N_{resns} 段階の解像度のガボール・ウェーブレット係数 $c^A \in C^{N_{nodes} \times N_{orns} \times N_{resns}}$ ，および人物 ID から構成される．特徴点配置の具体例を図 1 に示す．

システムに登録した各テンプレートと入力画像上の顔画像との各解像度 r における類似度を式 (1) で定義する．

2004 年 8 月 25 日，ICPR2004 (Cambridge, U.K.) で発表
 2005 年 1 月 7 日受付，2005 年 4 月 22 日再受付，2005 年 6 月 13 日採録

[†]NHK 放送技術研究所

(〒157-8510 世田谷区船 1-10-11, TEL 03-5494-3127)

^{††}NHK 岡山放送局

(〒700-8621 岡山市駅元町 15-1, TEL 086-214-4700)

^{†††}早稲田大学 理工学術院

(〒169-8555 新宿区大久保 3-4-1, TEL 03-5286-3377)

[†]NHK Science and Technical Research Laboratories

(1-10-11, Kinuta, Setagaya-ku, Tokyo 157-8510, Japan)

^{††}NHK Okayama Broadcasting Station

(15-1, Ekimotomachi, Okayama 700-8621, Japan)

^{††}Faculty of Science and Engineering, Waseda University

(3-4-1, Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan)

図 1 特徴点の配置 ($N_{nodes} = 9$)
Example of node location x_A .

図 2 ガボル・ウェーブレット (左: 実数部 右: 虚数部)
Gabor wavelets (left: real part, right: imaginary part).

$$S_{A,B}^r = 1 - \alpha_f \left(1 - \frac{\langle c_r^A, c_r^B \rangle}{|c_r^A| |c_r^B|} \right) - \alpha_s \frac{\sqrt{E_{A,B}}}{\lambda_r} \quad (1)$$

式 (1) の右辺第 2 項は特徴量の類似度を、第 3 項は特徴点の歪みをあらわす。また $\langle c_r^A, c_r^B \rangle$ は二つの複素ベクトル c_r^A, c_r^B それぞれの解像度 r に関する成分の内積をあらわし、定数 λ_r は解像度 r におけるガボル・ウェーブレットの波長である。各々のガボル・ウェーブレット係数は、2次元のガボル・ウェーブレットと各特徴点の周囲の輝度との畳み込み積分である。解像度 r 、方位 n のガボル・ウェーブレットは式 (2) で定義される。

$$g_n^r(x) = \frac{k_r^2}{\sigma^2} \exp\left(-\frac{k_r^2 \|x\|^2}{2\sigma^2}\right) \times \left[\exp\left(i(k_n^r)^T x\right) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (2)$$

ここで k_n^r はガボル・ウェーブレットの中心周波数にあたる波数ベクトルであり、式 (3) で定義される。図 2 にガボル・ウェーブレットの例を示す。

$$k_n^r = k_r \begin{pmatrix} \cos(\frac{n\pi}{N_{orns}}) \\ \sin(\frac{n\pi}{N_{orns}}) \end{pmatrix}, \quad \begin{matrix} 0 \leq n < N_{orns} = 8 \\ 0 \leq r < N_{resns} = 5 \end{matrix} \quad (3)$$

$E_{A,B}$ はテンプレートの特徴点配置 x_A に対する入力画像上の特徴点配置 x_B の変形の度合いを評価する歪みエネルギーである。筆者らが提案したシステム¹⁾²⁾では、個々の特徴点についての特徴量の内積値から推定した信頼度を対角成分に配置した実対角行列 $W \in R^{2N_{nodes} \times 2N_{nodes}}$ を用いて、式 (4) に示す重み付きユークリッド距離で歪みエネルギーを定義している。

$$E_{A,B} = (x_B - x_A)^T W (x_B - x_A) \quad (4)$$

式 (1) における α_f 、 α_s は、それぞれ特徴量の類似度、歪みエネルギーの重みであり、経験的に固定された定数である。システムの顔検出部がテスト画像上の顔領域を検出すると、顔認識部はその大きさを正規化した入力顔画像上に

特徴点 $x_B \in R^{2N_{nodes}}$ を各々のテンプレートを初期状態として配置し、特徴点ごとに特徴量の類似度が最大になるまで変形する。特徴点の変形後、式 (1) で定義される類似度 $S_{A,B}^r$ を最低解像度から順に計算し、閾値 θ を超えた場合のみ各人物のスコアを式 (5) に従って累積し、次の解像度に進む。 $S_{A,B}^r$ が閾値以下の場合、その解像度でスコアの累積をやめる。ここで式 (5) の関数 f は経験的に設計した区分線形関数であり、式 (6) で示される。システムは最終的にスコア最大を与えるテンプレートの人物 ID を出力し認識結果とする。

$$Score = \prod_{r=1}^{N_{resns}} f(S_{A,B}^r) \quad (5)$$

$$f(S_{A,B}^r) = \begin{cases} 1 & (0 \leq S_{A,B}^r < \theta) \\ \frac{S_{A,B}^r - \theta}{1 - \theta} (M - 1) + 1 & (\theta \leq S_{A,B}^r < 1) \\ 0 & (0 < \theta < 1, 1 < M) \end{cases} \quad (6)$$

3. 提案手法

3.1 階層的 Bayes 推定

拘束条件の項を持つ評価関数に基づいたパラメータの最適化問題は、Bayes 推定⁶⁾における事後確率最大化と対応づけることができる。

$$P(w|D_{train}, \mathcal{H}) = \frac{P(D_{train}|w, \mathcal{H})P(w|\mathcal{H})}{P(D_{train}|\mathcal{H})} \quad (7)$$

ここで w は、システムのパラメータをあらわし、 D_{train} は学習データをあらわす。また $\mathcal{H}$ はモデルをあらわす。Bayes 的枠組みでは尤度関数 $P(D_{train}|w, \mathcal{H})$ を、式 (1) 右辺第 2 項に示した特徴量の類似度を用いて定義することができる。同様に、事前確率分布 $P(w|\mathcal{H})$ は式 (1) 右辺第 3 項の特徴点の配置に関する拘束項を用いて定義できる。パラメータ w の最適化は Bayes 推定における事後確率最大化と等価である。筆者らが提案した手法¹⁾²⁾では、式 (1) の最大化によってパラメータを最適化した後、式 (5)、(6) で定義した経験的な評価基準を用いてモデル (顔テンプレート) を比較する。本稿で問題としている顔画像認識では、具体的なパラメータ w の設定値よりも、与えられたデータからいかにして尤もらしいモデル $\mathcal{H}$ を選択するかが重要である。

本提案では、この問題に対してパラメータの確率分布を、モデルの選択基準に反映させた、階層的 Bayes 推定⁶⁾を適用する。筆者らの従来手法¹⁾²⁾では、式 (7) に示した事後確率を最大化するパラメータ 1 点のみを用いて認識を行うのに対し、提案手法ではパラメータ w の事後確率分布を推定し、パラメータ空間全体にわたる期待値計算 (周辺化) を実行する。認識対象である評価データ D_{test} が与えられたとき、モデル $\mathcal{H}$ の事後確率分布は、Bayes 公式を階層的に適用することにより、次式のように自然な形で導かれる。

$$P(\mathcal{H}|D_{test}) = \frac{P(D_{test}|\mathcal{H})P(\mathcal{H})}{P(D_{test})} \quad (8)$$

ここで式 (8) における右辺の分子第 1 項は式 (7) の右辺の分母であり、次式で定義される。

$$P(D_{test}|\mathcal{H}) = \int P(D_{test}|w, \mathcal{H})P(w|D_{train}, \mathcal{H})dw \quad (9)$$

学習データ D_{train} が与えられたときのパラメータ w の事後確率分布は、評価データ D_{test} に依存しないため、事前に計算可能である。本提案では、モデルの事後確率の計算をパラメータの確率分布の推定 (学習) と、式 (9) の周辺化 (認識) の二つの処理に分離し、前者の計算を認識対象の画像が与えられる前に実行する。システムは、学習データを用いて事前に推定したパラメータの確率分布にもとづき、評価データについての事後確率最大を与えるモデル $\mathcal{H}_{MP}$ を次式のように推定結果として出力する。

$$\mathcal{H}_{MP} = \arg \max_{\mathcal{H}} P(\mathcal{H}|D_{test}) \quad (10)$$

3.2 ハミルトニアンモンテカルロ

階層的 Bayes 推定を行うには式 (9) の積分を行う必要がある。この計算は、多次元連続空間の積分であり、ここで考えている問題を含め多くの場合、解析的に計算を実行することは不可能であることが多い。本提案では、この積分をハミルトニアンモンテカルロ法⁷⁾⁸⁾ (Hamiltonian Monte Carlo Method, 以下 HMC 法と略す) による近似計算を用いて実現する。HMC 法は、本稿が対象とする画像認識とは全く異なる分野においても、比較的困難な問題に対して成功裏に適用されている統計的計算手法である⁹⁾¹⁰⁾。

HMC 法が、既存のメトロポリス法に比べて優れている点の一つは、後者のいわゆる random walk behavior を大幅に改善している点である。一般に、メトロポリス法で用いられるステップサイズは非常に小さい値が選ばれる。高次元で大きなステップサイズを用いると、確率の極めて低い領域にサンプル点が落ち込むことが多く、その結果、全体の確率分布を捉えるのが困難になる。HMC 法では、メトロポリス法におけるステップ以外に対象としている確率密度関数から自然に導かれるハミルトニアン力学系による確定的な遷移を行い、極力、分布全体を捉えようとするものである。

HMC 法では、確率変数 x とその速度 u で決定されるハミルトニアン

$$H(x, u) = E(x) + K(u) \quad (11)$$

を考え、 $H(x, u)$ を用いて次式で定義される x と u の同時確率に従うサンプルを抽出する。

$$\begin{aligned} P_H(x, u) &= \frac{1}{Z_H} \exp(-H(x, u)) \\ &= \frac{1}{Z_H} \exp(-E(x)) \exp(-K(u)) \end{aligned} \quad (12)$$

ここで $E(x) = -\log(P(x))$ は x の位置エネルギーをあらわし、 $K(u) = u^T u / 2$ はその運動エネルギーをあらわす。HMC 法のアルゴリズムの概要を以下に示す。

1. 確率変数 x の初期状態を $x^{(t=0)}$ とする。
2. 一様乱数 $v \sim \mathcal{U}_{[0,1]}$ と、ガウス分布に従う乱数 $u^* \sim \mathcal{N}(0, I_{n_s})$ を得る。
3. $x_0 = x^{(t)}$, $u_0 = u^* + \rho \Delta(x_0) / 2$ とおく。
4. $l = 1, \dots, L$ について順次 x_l, u_l を計算し、サンプルの候補 (x_L, u_L) を得る (leapfrog 法)。

$$x_l = x_{l-1} + \rho_l u_{l-1},$$

$$u_l = u_{l-1} + \rho_l \Delta(x_l).$$

ただし、

$$\rho_l = \begin{cases} \rho & (l < L), \\ \rho/2 & (l = L), \end{cases}$$

$$\Delta(x) = \delta E(x) / \delta x.$$

5. (x_L, u_L) と $(x^{(t)}, u^{(t)})$ の各々の同時確率の比

$$a = \frac{P_H(x_L, u_L)}{P_H(x^{(t)}, u^{(t)})}$$

を計算する。 $a > v$ の場合、 (x_L, u_L) を受理し、 $(x^{(t+1)}, u^{(t+1)}) = (x_L, u_L)$, $t = t + 1$ とする。それ以外の場合は (x_L, u_L) を棄却する。

6. R 個のサンプルが得られるまで操作 2 から操作 5 を繰り返す。

ここで $\rho(>0)$ は、leapfrog のステップサイズをあらわすパラメータ、 L は leapfrog の回数をあらわすパラメータである。確率分布 $P(x)$ に従うサンプル $\{x^{(r)}\}_{r=1}^R$ が得られた場合、関数 $\phi(x)$ の期待値 $\langle \phi(x) \rangle$ は次の計算式で近似可能である。

$$\langle \phi(x) \rangle = \int P(x) \phi(x) dx \cong \frac{1}{R} \sum_{r=1}^R \phi(x^{(r)}) \quad (13)$$

このとき、式 (13) で得られる期待値 $\langle \phi(x) \rangle$ の分散は σ^2/R であり、サンプル数 R が無限大のとき 0 に収束する。収束する速さは被積分変数 x の次元に依存しない。

3.3 パラメータ分布の推定 (学習)

学習データ $D_{train} = \{t_1, t_2, \dots, t_Q\}$ 上の N_{nodes} 個の特徴点の座標を x 、学習対象の人物の顔テンプレートを $\mathcal{H}$ とする。 D_{train} の尤度関数を、式 (1) における特徴点の類似度に関する項を用いて次式の混合分布で定義する。

$$\begin{aligned} P(D_{train}|x, \beta, \mathcal{H}) &= \frac{1}{Q} \sum_{q=1}^Q P(t_q|x, \beta, \mathcal{H}) \\ &= \frac{1}{Q} \sum_{q=1}^Q \frac{\exp\left(-\beta \left(1 - \frac{\langle c^A, c_q^B \rangle}{\|c^A\| \|c_q^B\|}\right)\right)}{Z_q(x, \beta)} \end{aligned} \quad (14)$$

ここで β は、尤度関数の鋭さを決定するハイパーパラメータであり、 $Z_q(x, \beta)$ は q 番目の学習データについての尤度関数の正規化定数である。 Q はモデル $\mathcal{H}$ ごとに与えられる学習データ t_q の総数をあらわす。また $\langle c^A, c^B \rangle$ は二つ

図 3 表情変化のサンプル画像から抽出した特徴点の位置変化
Samples of feature points x from a set of images of various facial expressions.

の複素ベクトル c^A, c^B の内積をあらわす。 x の事前確率分布は、筆者らが行った従来手法における歪みエネルギー項に関する実験¹¹⁾をふまえ、次式で定義する。

$$P(x|\alpha, \mathcal{H}) = \frac{1}{Z(\alpha)} \exp\left(-\frac{\alpha}{2\lambda_1^2} (x - \bar{x})^T C^{-1} (x - \bar{x})\right) \quad (15)$$

ここで C は、様々な表情変化のサンプル顔画像から手動で抽出した特徴点の座標をもとに推定した x の共分散行列である。 C の推定に用いた特徴点の座標データの例を図 3 に示す。

$\bar{x}$ は各特徴点の座標 x の平均をあらわし、 λ_1 は最低解像度のガボール・ウェーブレットの波長をあらわす定数である。 α は事前分布の鋭さを決定するハイパーパラメータである。 $Z(\alpha)$ は正規化定数であり、以下の式で求められる。

$$Z(\alpha) = (2\pi)^{N_{nodes}} \lambda_1 \sqrt{\frac{\det C}{\alpha}} \quad (16)$$

このとき x, α, β の同時事後確率分布 $P(x, \alpha, \beta | D_{train}, \mathcal{H})$ は次の式であらわされる。

$$\begin{aligned} P(x, \alpha, \beta | D_{train}, \mathcal{H}) &= \frac{P(D_{train} | x, \beta, \mathcal{H}) P(x | \alpha, \mathcal{H}) P(\alpha, \beta | \mathcal{H})}{P(D_{train} | \mathcal{H})} \\ &= \frac{1}{Q} \sum_{q=1}^Q \frac{P(t_q | x, \beta, \mathcal{H}) P(x | \alpha, \mathcal{H}) P(\alpha, \beta | \mathcal{H})}{P(D_{train} | \mathcal{H})} \\ &= \frac{1}{Q} \sum_{q=1}^Q \frac{P(x, \alpha, \beta | t_q, \mathcal{H}) P(t_q | \mathcal{H})}{P(D_{train} | \mathcal{H})} \end{aligned} \quad (17)$$

ただし、

$$P(x, \alpha, \beta | t_q, \mathcal{H}) = \frac{P(t_q | x, \beta, \mathcal{H}) P(x | \alpha, \mathcal{H}) P(\alpha, \beta | \mathcal{H})}{P(t_q | \mathcal{H})} \quad (18)$$

とおいた。

3.4 モデル比較 (認識)

モデル $\mathcal{H}$ に対応する人物についての N 枚の連続画像である評価データ $D_{test} = \{d_1, d_2, \dots, d_N\}$ が逐次的に与えら

れるとき、時刻 n (第 n フレーム) の画像データ d_n の尤度を式 (19) で計算する。

$$P(d_n | \mathcal{H}) = \iiint P(d_n | x, \beta, \mathcal{H}) P(x, \alpha, \beta | D_{train}, \mathcal{H}) dx d\alpha d\beta \quad (19)$$

同じ $\mathcal{H}$ についての学習データ t_q を用いて事前に HMC 法で抽出した R 個のサンプル

$$\{x_q^{(r)}, \alpha_q^{(r)}, \beta_q^{(r)}\}_{r=1}^R \sim P(x, \alpha, \beta | t_q, \mathcal{H}) \quad (20)$$

を抽出したとき、式 (19) の積分は次式で近似できる。

$$\begin{aligned} P(d_n | \mathcal{H}) &= \frac{1}{Q} \iiint P(d_n | x, \beta, \mathcal{H}) \\ &\quad \sum_{q=1}^Q \frac{P(x, \alpha, \beta | t_q, \mathcal{H}) P(t_q | \mathcal{H})}{P(D_{train} | \mathcal{H})} dx d\alpha d\beta \\ &\cong \frac{1}{QR} \sum_{q=1}^Q \sum_{r=1}^R \frac{P(d_n | x_q^{(r)}, \beta_q^{(r)}, \mathcal{H}) P(t_q | \mathcal{H})}{P(D_{train} | \mathcal{H})} \end{aligned} \quad (21)$$

個々の学習データ t_q の出現確率が q について一様であると仮定すると、 $P(t_q | \mathcal{H}) / P(D_{train} | \mathcal{H}) = \text{const.}$ とおけるので、式 (21) は次式であらわされる。

$$\begin{aligned} P(d_n | \mathcal{H}) &\propto \frac{1}{QR} \sum_{q=1}^Q \sum_{r=1}^R P(d_n | x_q^{(r)}, \beta_q^{(r)}, \mathcal{H}) \end{aligned} \quad (22)$$

時刻 n における i 番目の顔テンプレート $\mathcal{H}_i$ の事後確率を、直前の時刻 $n-1$ における $\mathcal{H}_i$ の事後確率と d_n の尤度を用いて次式で定義する。

$$\begin{aligned} P(\mathcal{H}_i | D_n) &= \frac{P(d_n | \mathcal{H}_i) P(\mathcal{H}_i | D_{n-1})}{P(d_n | D_{n-1})} \\ &= \frac{P(d_n | \mathcal{H}_i) P(\mathcal{H}_i | D_{n-1})}{\sum_{j=1}^{N_{persons}} P(d_n | \mathcal{H}_j) P(\mathcal{H}_j | D_{n-1})} \\ &\cong \frac{P(\mathcal{H}_i | D_{n-1}) \sum_{q=1}^Q \sum_{r=1}^R P(d_n | x_q^{(r)}, \beta_q^{(r)}, \mathcal{H}_i)}{\sum_{j=1}^{N_{persons}} P(\mathcal{H}_j | D_{n-1}) \sum_{q=1}^Q \sum_{r=1}^R P(d_n | x_q^{(r)}, \beta_q^{(r)}, \mathcal{H}_j)} \end{aligned} \quad (23)$$

評価データが与えられる前の事前確率 $P(\mathcal{H} | d_0)$ は特別な理由がない限り等確率とする。

$$P(\mathcal{H} | d_0) = \frac{1}{N_{persons}} \quad (24)$$

ここで $N_{persons}$ はシステムに登録された人数である。

4. 認識実験

4.1 実験条件

提案手法（以下、Bayesian HMC と呼ぶ）の有効性を検証するため、可変テンプレートマッチングにもとづく従来手法¹⁾²⁾（以下、EGM と呼ぶ）、および、統計に基づく顔画像認識手法としてよく知られている Eigenface³⁾を比較対象とした認識実験を行った。実験に用いるデータは、認識対象の評価データ、モデル $\mathcal{H}$ を作成するための登録データ、パラメータ w の分布を推定するための学習データの3種類を用意した。評価データは、10人（男性8人、女性2人）の被験者について撮影した3種類の連続静止画像“Neutral”（無表情），“Talking”（発話），“Smiling”（笑顔）を用意した。各々の画像は、スタジオ照明・単色背景（グレー）の条件で撮影した動画（毎秒30フレーム）から、タテ240画素×ヨコ360画素のサイズで取得した。登録データには、評価データ“Neutral”と同じ条件で撮影した別の画像を用意し、特徴点を手動でプロットした。学習データは、評価データ“Talking”から4フレーム間隔で抜き出した画像20枚/人を用意した。なお、認識実験を行う上で必要となる顔の位置情報は、評価データについて10フレームおきに手動で顔の位置と大きさを抽出したリストをあらかじめ用意した。

表1に評価データの詳細を、表2に登録データと学習データの詳細をそれぞれ示す。なお、個々の人物でフレーム数が異なっているが、これは極力自然な状態での顔画像を取得するために、被験者に時間的な拘束条件を与えずに撮影を行ったためである。

Bayesian HMC で HMC 法の計算に用いる leapfrog のパラメータは、いくつかの設定を試行した結果、学習データから得られる x, α, β のサンプルについての同時事後確率のピークが最大となった $L=5, \rho=0.1$ とした。ハイパーパラメータ α の事前確率分布 $P(\alpha|\mathcal{H})$ は、式(25)で示すガンマ分布を用いる。ガンマ分布は、ハイパーパラメータ α についての正值条件 $\alpha > 0$ を満たし、かつ、なだらかな形状をもつことから、HMC 法を含むモンテカルロ法では

表1 評価データ
The test image data.

type	number of frames		
	min.	max.	total (10 persons)
Neutral	30	170	796
Talking	80	299	1508
Smiling	40	259	1287
TOTAL	150	589	3591

表2 登録データと学習データ
Registration data and training data.

Data set	number of images
Registration data	10 (1 per person)
Training data*	200 (20 per person)

* Set of face images sub-sampled from “Talking” test data.

図4 従来手法におけるパラメータ α_f と顔認識誤り率
ID error rate vs EGM parameter α_f .

しばしば用いられる分布族である¹³⁾。なお、本実験でのガンマ分布のパラメータは $p=2, \sigma=0.1$ とする。

$$P(\alpha|\mathcal{H}) = \begin{cases} \frac{1}{\Gamma(p)\sigma^p} \alpha^{p-1} e^{-\frac{\alpha}{\sigma}} & (\alpha \geq 0) \\ 0 & (\alpha < 0) \end{cases} \quad (25)$$

ハイパーパラメータ β は、式(14)の尤度関数の規格化定数 $Z_q(x, \beta)$ に含まれる。考えている尤度関数により、規格化定数の解析表示が可能な場合とそうでない場合があり、本稿では後者になっている。本実験では、従来手法において、 β に対応するパラメータ α_f を評価データについて最適化した値 ($\beta=5.0$) に固定する。従来手法において α_f を変化させたときの認識結果を図4に示す。

登録人物ごとに得た特徴点の位置のサンプルは、あらかじめ用意した顔位置のリストを用いて、各評価データの画像平面上に拡大・縮小および平行移動した後、各々の人物（モデル $\mathcal{H}$ ）の事後確率を計算する。ある一人の人物について、20枚の学習データを与えたときのHMC法で得た特徴点のサンプルの分布の様子を図5に示す。

一方、Eigenfaceの固有ベクトル（固有顔）の計算には、統計的信頼性の向上と少数の学習データへの過学習の抑制を図るため、表2の登録データに加えて、ともに正確な照度の管理が可能なスタジオにおいて、均等な照明条件のもとで撮影された、HOIP顔画像データベース¹²⁾（男女各150人）と、我々が独自に整備した顔画像データベース（男性72人、女性44人）を学習データに追加した。ただし、そ

図5 学習データから得た特徴点配置のサンプル
Example of the distribution of $\{x_q^{(r)}\}$.

図 6 顔画像データベースの平均顔
Mean image of the face database.

図 7 顔画像データベースの固有顔 (固有値の降順, 上位 10 個)
Example of Eigenfaces (top 10 in eigenvalue order).

これらのうち眼鏡をかけていない正面顔のみを用いた (合計 368 枚, うち男性 184 人, 女性 184 人). 顔登録には, 表 2 に示した登録データと学習データの両方 (合計 210 枚) を用いた. Eigenface に用いる登録データ, 学習データ, および評価データの顔画像は, 両目の瞳の位置を基準としてタテ 100 画素×ヨコ 100 画素の大きさにあらかじめ正規化した. 図 6 に学習データから得られた平均顔を示す. 図 7 に固有顔 (固有値の降順, 上位 10 個) を示す.

4.2 実験結果

従来手法 (EGM) において, α_s を変化させた場合の顔認識誤り率を図 8 に示す. 図 8 のグラフにおいて, 各々の評価データごとに描かれた認識誤り率の曲線が異なる α_s の値において最小値をとっている. この実験結果より, 従来手法では表情や発話動作による顔の変形の度合いや内容に応じて異なる最適値が存在するといえる.

続いて, Bayesian HMC において, 1 枚の学習データあたりに抽出するサンプルの数を変化させた場合の顔認識誤り率を図 9 に示す. 図 9 に示した認識結果のグラフより, 提案手法では HMC 法で得るサンプルの数が増えるに従って, 顔認識誤りが単調に減少している様子がわかる.

図 8 従来手法 (EGM) の評価データごとの顔認識誤り率
($\alpha_f = 5.0$)
ID error rate vs EGM parameter; α_s ($\alpha_f = 5.0$).

図 9 提案手法 (Bayesian HMC) における HMC 法の学習
画像 1 枚当たりのサンプル数と顔認識誤り率
ID error rate vs Number of HMC sample sets (per training
image).

図 10 Eigenface における固有ベクトル (固有顔) の数と顔
認識誤り率
ID error rate vs Number of Eigenfaces.

Eigenface において, 固有ベクトルの数 M を変化させた場合の顔認識誤り率を図 10 に示す. ただし, Eigenface での認識結果は, 一人あたり 21 通り登録した顔テンプレートのうち, ひとつでも固有ベクトル空間上での評価データとのユークリッド距離が最小となった場合は, 正解とした. 図 10 に示した Eigenface の認識結果のグラフにおいて, 固有顔の個数=5 までは認識に用いる固有顔の個数が増えるに従って顔認識誤り率が減少しているが, 固有顔の個数を 5 から 6 に増やした際に, 顔認識誤り率が 1% 程度増加しており, それ以上固有顔の個数を増やした場合には, 顔認識誤り率の変化はほとんどみられず, ほぼ一定値に収束している. Eigenface で用いる固有顔の数が多いほど計算量も増えるため, 何らかの方法で事前に最適値に設定する必要があるといえる.

従来手法 (EGM), 提案手法 (Bayesian HMC), および Eigenface の顔認識誤り率 (3 種類の評価データの平均値) を, 認識に要した全処理時間を横軸にしたグラフとして図 11 に示す. ただし, 処理時間は評価データ全体の撮影時の時間 (全フレーム数 × 1/30 秒) と, それぞれの手法で認識の処理に要した実時間との比率 (Real time factor, RTF) で評価した. なお, それぞれの認識実験は CPU: Xeon 2.8 GHz, memory: 2GB の計算機で行った.

図 11 Eigenface, Bayesian HMC, EGM それぞれの顔認識誤り率と処理時間 (TOTAL) ID-error-rate and real-time factor(TOTAL).

図 11 のグラフより, Bayesian HMC は EGM の 10 倍～15 倍の処理時間を要したが, 各々の曲線の下限が示す認識誤りの最小値に関しては, 著しく改善したことがわかる. 一方, Eigenface に対しては, Bayesian HMC の実験結果がすべて原点に近い領域にプロットされていることから, Bayesian HMC が Eigenface よりも高い認識精度を速い処理時間で実現可能であることがわかる.

表 3 に, EGM, Bayesian HMC, Eigenface それぞれについて顔認識誤り率が最小となった設定での認識結果と, 処理時間 (RTF) を示す. 登録データと同じ条件で撮影した評価データ “Neutral” については, 三つの手法のいずれもすべての顔を正しく認識した. 一部を学習に用いた評価データ “Talking” については, EGM で 2.97%, Eigenface で 1.66% の顔認識誤りが発生したのに対して, Bayesian HMC では全ての顔を正しく認識した. 学習に用いていない評価データ “Smiling” については, EGM で 22.6% Eigenface で 2.02% の顔認識誤りが発生したのに対して, Bayesian HMC の顔認識誤りは 0.08% であり, 未知の変動 (表情変化) を含んだ入力に対して, 大幅に頑健性が改善した.

処理時間については, EGM で RTF が 2.1 だったのに対して, Bayesian HMC では 22.9 であり, 約 10 倍の処理時間がかかった. 一方, Eigenface と Bayesian HMC の比較では, 前者の RTF が 22.9 倍だったのに対して, 後者ではそれよりも約 2 割多い 27.4 倍であり, 両者の計算コストの面での大きな違いはなかった.

ある 1 枚の学習画像に対して, ハイパーパラメータ α の初期値を 3.0 から 0.15 まで, 6 通り変化させた場合のハイパーパラメータ α のサンプルの軌跡を図 12 に示す. 異なる初期値から出発させたそれぞれの α の軌跡は, サンプルの数が 10 以上の領域で $\log \alpha = -2.0$ 付近の値にサンプルが分布している. この結果は, 可変テンプレートマッチングにおける歪みエネルギー項の重みに相当するハイパーパラメータ α がその初期値の設定によらず, 事後確率に従って自動的に推定されることを示す.

表 3 顔認識実験の結果 (顔認識誤り率)

Face recognition results (ID error rate)

INPUT	EGM ($\alpha_f=5.0, \alpha_g=0.4$)	Bayesian HMC ($\beta=5.0, R=100$)	Eigenface ($M=5$)
Neutral	0.00 %	0.00 %	0.00 %
Talking	2.97 %	0.00 %	1.66 %
Smiling	22.6 %	0.08 %	2.02 %
TOTAL	8.6 %	0.03 %	1.42 %
RTF	2.1	22.9	27.4

図 12 異なる初期値を与えた場合の α のサンプルの軌跡 Trajectories of sample set $\{\alpha_q^{(r)}\}$ starting from different initial states.

5. む す び

登録した人物ごとのモデルの事後確率を, 階層的 Bayes 推定と HMC 法を用いて計算する顔画像認識法を提案した. 従来, 試行錯誤で決定していたパラメータの一部を確率変数と捉え, その確率分布を学習データから推定する具体的なアルゴリズムを示した. 実験の結果, 提案手法が未知の変化を含んだ評価データに対して, Eigenface よりも高い認識精度を速い処理時間で実現できることを示した.

今後は, 動画追跡の分野で近年注目されている Sequential Monte Carlo (SMC) 法⁸⁾⁹⁾¹⁴⁾を用いて, 逐次的に与えられる評価データからパラメータの確率分布を動的に更新を検討する. SMC 法の適用により, 本実験で手動で与えた顔領域を, 自動的に追跡する顔認識システムの実現を目指す. また, 複数の顔の向きに対応するためのモデル $\mathcal{H}$ の改良, および, 本稿で固定としたハイパーパラメータ β の確率分布の推定方法の検討に取組み, 提案手法の適用範囲の拡大と更なる顔認識精度の改善を目指す.

本研究を進めるにあたり, 伊庭幸人氏 (統計数理研究所), Arnaud Doucet 氏 (Cambridge University) から多くの助言を頂いた. 深く感謝したい.

〔文 献〕

- 1) S. Clippingdale and T. Ito: "A Unified Approach to Video Face Detection, Tracking and Recognition", Proc. ICIP'99, pp.662-666 (1999).
- 2) S. Clippingdale and T. Ito: "Partial automation of database acquisition in the FAVRET face tracking and recognition system using a bootstrap approach", Proc. MVA2000, pp.5-8 (2000).
- 3) M. Turk and A. Pentland: "Eigenfaces for recognition", Journal of Cognitive Neuroscience, **3**, 1, pp. 71-86 (1991).
- 4) 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).
- 5) 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, New York (1998).
- 6) D.J.C. MacKay: "Information Theory, Inference, and Learning Algorithms", Cambridge University Press, Cambridge (2003).
- 7) R.M. Neal: "Bayesian Learning for Neural Networks", Lecture Notes in Statistics, Springer, New York, pp.55-98 (1996).
- 8) C. Andrieu, N.D. Freitas, A. Doucet and M.I. Jordan: "An Introduction to MCMC for Machine Learning", Machine Learning, **50**, pp.5-43 (2003).
- 9) J.S. Liu: "Monte Carlo Strategies in Scientific Computing", Springer, New York, pp.53-77 (2001).
- 10) Y. Nakada, T. Matsumoto, T. Kurihara and K. Yosui: "Bayesian Reconstructions and Predictions of Nonlinear Dynamical Systems via the Hybrid Monte Carlo Scheme", Signal Processing, Elsevier, **85**, pp.129-145 (2005).
- 11) 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).
- 12) 財団法人ソフトラビジャパン: 「顔画像データベース」
http://www.hoip.jp/web_catalog/top.html
- 13) A. Matsui, S. Clippingdale, F. Uzawa and T. Matsumoto: "Bayesian Face Recognition using a Markov Chain Monte Carlo Method", Proc. ICPR2004, **3**, pp.918-921 (2004).
- 14) M. Isard and A. Blake: "Contour tracking by stochastic propagation of conditional density", Proc. ECCV'96, pp.343-356 (1996).

まつい あつし
松井 淳 1994 年, 早稲田大学理工学部電気工学科卒業。1996 年, 同大学大学院修士課程了。同年, NHK 入局。札幌放送局を経て, 1998 年より同放送技術研究所勤務。音声認識, 画像認識の研究に従事。2004 年, 早稲田大学大学院博士後期課程入学 (社会人特別選考)。正会員。

サイモンクリッピングデル
Simon Clippingdale 1982 年, 英国バーミンガム大学卒業。1988 年, 英国ウォーリック大学博士課程了。1996 年, NHK 入局。同年より同放送技術研究所勤務。現在, NHK 放送技術研究所知能情報処理主任研究員。画像認識の研究に従事。博士 (工学)。正会員。

うざわ ふみき
鵜澤 史貴 2004 年, 早稲田大学理工学部電気電子情報工学卒業。同年, NHK 入局。現在, NHK 岡山放送局勤務。在学中, NHK 放送技術研究所にて顔画像認識の研究に従事。

まつもと たかし
松本 隆 1966 年, 早稲田大学理工学部電気工学科卒業。1970 年, ハーバード大学大学院応用数学修了。1977 年～1979 年, カリフォルニア大バークレー・電気工学・計算機科学科研究員。1980 年, 早稲田大学理工学部電気工学科教授。2003 年～2004 年, ケンブリッジ大学研究員。現在, 早稲田大学理工学術院教授。博士 (工学)。IEEE 会員 (Fellow)。正会員。