

摂動特徴量による顔画像に対する形状モデルフィッティング

木下 航一^{†a)} 小西 嘉典[†] 勞 世竝[†] 川出 雅人[†]
 村瀬 洋^{††}

Perturbation Feature and It's Application to Shape Model Fitting
 for Facial Images

Koichi KINOSHITA^{†a)}, Yoshinori KONISHI[†], Shihong LAO[†], Masato KAWADE[†],
 and Hiroshi MUSASE^{††}

あらまし 顔画像に対して、高速で高精度に形状モデルフィッティングを行う手法を提案する。形状モデルのフィッティング手法は多数提案されているが、顔画像に対するフィッティングは影や表情変化などのノイズの影響を受けやすく、高精度な結果が得られにくいという課題があった。また多数の繰返し演算を必要とするため、リアルタイム処理の実現は難しかった。我々はこれらの課題に対して (1) 形状モデル上でのサンプリング点の構造的配置 (2) 特徴量による形状パラメータの摂動量推定、を行うフィッティング手法を提案する (1) によってノイズの影響を受けにくい局所特徴量を有効に活用でき、影や表情変化などに対する頑健性が向上する。また (2) によって特徴量から一度の行列演算で形状パラメータの修正量を求めることが可能となり、処理時間が削減できる。公開データベースによる実験の結果、従来手法と比較して約 20 倍の高速化を実現しつつ、影や表情変化などのノイズのある顔画像に対して従来手法を上回る検出性能を示した。

キーワード 特徴量, 顔画像, 特徴点検出, 形状モデル

1. ま え が き

顔画像上から、目や鼻、口の端点等の位置を検出する技術は一般に顔特徴点検出 (Facial Feature Detection) あるいは顔器官検出 (Facial Parts Detection) などと呼ばれ、顔認証、表情推定などの顔画像処理を実現する上で重要な役割を担っている。しかしその重要性にもかかわらず、リアルタイム処理が可能で、様々な撮影条件 (例えば影や表情変化など) に対して頑健な性能をもつ、実用的な手法は報告されていない。

顔特徴点位置の正確な検出には、個々の特徴点周りの局所的な情報と、それら特徴点配置に関する大域的な情報を適切に組み合わせる利用することが重要であると考えられている [3]。この観点から成功を収め

た手法は、Cootes ら [1], [2] の Active Shape Model (ASM) と Active Appearance Model (AAM) であろう。これらの手法は、顔特徴点配置を形状モデル (Shape Model) と呼ばれるモデルで表現する。形状モデルとは、特徴点配置の座標集合に対して主成分分析を適用し、これによって得られた基底ベクトルのうち固有値の大きなものだけで、座標表現を再構成したものである。形状モデルを用いることにより特徴点分布が低次元で表現可能になる。このことは、データ量の圧縮だけでなく、検出結果が顔らしい形状を保つような優れた拘束条件として有効に活用される。

両者ともに顔特徴点検出技術として活用可能であり、これらを基礎とした改良手法が近年多数提案されている。しかしながら、形状モデルを顔画像上に適合させる処理 (フィッティング) には多数の繰返し計算が必要になり、リアルタイム処理は難しい。また照明による影や表情変化を含む顔画像に対してのフィッティング性能は大幅に低下することが多く、これらの課題は、これまで提案された改良手法でも解決されていない。

例えば ASM を基礎にした手法として STASM [8]

[†] オムロン株式会社技術本部, 木津川市
 Research & Development HQ, Omron Corporation, 9-1
 Kizugawadai, Kizugawa-shi, 619-0283 Japan

^{††} 名古屋大学大学院情報科学研究科, 名古屋市
 Graduate School of Information Science, Nagoya University,
 Furo-cho, Chikusa-ku, Nagoya-shi, 460-3807 Japan

a) E-mail: kino@ari.ncl.omron.co.jp

を例にとると、この手法では特徴量として多重解像度画像における法線方向輝度パターンを用いることによって、フィッティングの精度向上が図られている。しかしながら影や表情変化時の検出精度には依然大きな低下が見られ、また特徴点探索や形状モデルの修正に多数の繰返し計算を必要とするため、処理時間の面でも課題が大きい。

本論文では、これら二つの問題に対して、以下二つの特徴をもつ新たな形状モデルフィッティング手法を提案する。

(1) 形状モデル上でのサンプリング点の構造的配置

照明変動による影、表情変化によるしわなどは特徴量に対してノイズとなり、フィッティングの性能を低下させる要因となる。特徴点から離れた(もともとはっきりした特徴のない)領域においては、こういった変動は支配的な影響をもたらすが、もともとはっきりとした特徴のある領域ではノイズの影響は相対的に弱くなる。したがって特徴点周辺の局所的情報は、影や表情変化などに対して比較的頑健であり、フィッティングの際重要な情報となる。提案するサンプリング手法により、形状モデル全域を疎にカバーしながら、各特徴点周辺の局所的な特徴に関しては密な情報をもつ特徴量集合が得られる。これによって影や表情変化などのノイズに対して頑健なフィッティングが実現できる。

(2) 特徴量による形状パラメータの摂動量推定

正解位置からずれた状態の形状モデルの特徴量と、そのときのずれ量の関係を正準相関分析によって学習する。この関係を利用することにより形状モデルのずれ状態(摂動状態)が特徴量から推定できる。これによって特徴量から一度の行列演算で形状修正を行うことが可能となり、処理時間を削減できる。

提案する形状モデルフィッティング手法は以上二つの処理によって、各ノード周りで一定の構造をもってサンプリングされた特徴量によって形状モデルを制御する。そのため我々は、この形状モデル制御方法を Active Structured Appearance Model (ASAM) と呼ぶ。

本論文では以下、2. において形状モデルについて概説し、3. で(1)の特徴量サンプリング手法について説明する。4. において(2)の原理となる特徴量による摂動状態推定について述べ、5. で(2)の形状パラメータ推定手法について説明する。6. では実験結果を示し、7. において考察を行う。

2. 顔形状モデル

顔特徴点配置は、適切な方法で移動、回転、縮小の操作を行って正規化した後、主成分分析 (Principal Component Analysis: PCA) を用いることにより低次元で表現可能となる。今、ある特徴点の座標を $[x_m, y_m]$ とし、全特徴点の座標をまとめて $\hat{\mathbf{x}} = [x_1, y_1, x_2, y_2, \dots, x_M, y_M]^T (\in \mathcal{R}^{2M})$ と書く。ここで M は特徴点の数である。また、ある顔画像 n についての特徴点座標を $\hat{\mathbf{x}}_n$ とし、正規化を行った後の特徴点座標を \mathbf{x}_n と表す。分散共分散行列 Σ は、

$$\Sigma = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T \quad (1)$$

となる。 N は顔画像の数である。任意の正規化特徴点ベクトルは標準固有値問題、

$$\Sigma \Phi = \Phi \Lambda \quad (\Phi^T \Phi = \mathbf{I}) \quad (2)$$

の解として得られる正規直交基底 Φ を用いて、以下のように表せる。

$$\mathbf{x} \approx \bar{\mathbf{x}} + \tilde{\Phi} \tilde{\mathbf{b}} \quad (3)$$

ここで、 $\tilde{\Phi}$ は上位 k 個の固有値に対応する固有ベクトルから構成した基底ベクトル、 $\tilde{\mathbf{b}}$ はそれに対応するパラメータベクトルである。したがって、元画像上の特徴点座標は、

$$\begin{aligned} \hat{\mathbf{x}} &\approx F(t_x, t_y, t_\theta, t_s, \tilde{\mathbf{b}}) \\ &= F(\mathbf{p}) \end{aligned} \quad (4)$$

として相似変換パラメータと基底ベクトルに関するパラメータの関数として近似的に表現することが可能である。ただし、 $\mathbf{p}^T = [t_x, t_y, t_\theta, t_s \mid \tilde{\mathbf{b}}^T]$ である。ここで、相似変換に関するパラメータを“Pose パラメータ”、基底ベクトルにかかるパラメータを“Shape パラメータ”と呼ぶ。

3. 構造的レティナサンプリング

ASM に基礎を置く手法においては、特徴量は輪郭の法線方向の 1 次元上でサンプリングされることが多い。このような特徴量は比較的次元であるため情報量が少なく、信頼性のある局所情報とはいえない。そのため影や表情変化による顔のしわなどのノイズに影響されて良い性能を得にくいという課題があった。

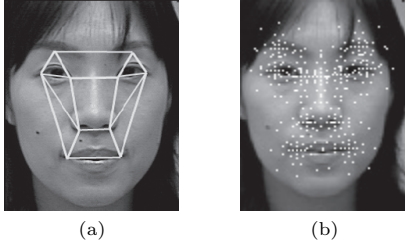


図 1 (a) 形状モデルの例, (b) 対応するサンプリング点配置

Fig. 1 (a) Shape model. (b) Corresponding sampling points.

一方 AAM に基礎を置くアピアランススペースのモデルフィッティング手法においては, 特徴量サンプリングは平均形状に変換したモデル上で均一に行われることが一般的である (例えば [12]). そのため特徴量は高次元となり, また形状変換のための演算が必要となり高速な処理は難しかった. 更に得られた特徴量は, モデルの摂動に対して変化が少なく, 必ずしもフィッティングに有効でない (例えば頬や額の部分など) 領域を多く含んでいるため, ノイズによる悪影響を受けやすいという課題があった.

このような課題に対応するため, 我々は形状モデルの各ノード周りに放射状にサンプリング点を配置するサンプリング手法を導入する.

特徴点の周りに放射状にサンプリング点を配置する手法は, Smeraldi ら [4] によって考案され, 特徴点検出に対して有効であることが示されている. ただし彼らはこの点配置を各特徴点に対して独立に適用し, 形状モデルと関連づけて利用されることはなかった.

顔特徴点周辺の局所的情報は, 影や表情変化などに対して比較的頑健であり, フィッティングの際重要な情報となる. 提案する手法は顔全域を疎にカバーしながら, 各特徴点周りでは密な特徴量サンプリングを行う. これにより比較的低次元で, ノイズに頑健な, フィッティングに有効な局所情報を顔全域から抽出することができる.

ここでこの手法を, 構造的レティナサンプリング (Structural Retinotopic Sampling) と呼ぶ. 図 1 に形状モデルの例と, この上でのサンプリング点配置の例を示す.

本論文では, 形状パラメータ p のとき, 上記方法により画像から特徴量 f を取得するサンプリング操作を

$$f = S(p) \quad (5)$$

と表記する.

4. 特徴量による摂動状態の推定

特徴点検出において特徴量サンプリングは, そのサンプリング位置 (あるいは角度, 大きさ) が正解か不正解かを判別する目的で行われるのが一般的である (例えば [4]). 例えば目の位置検出であれば, 事前に正解とされる目の位置においてサンプリングした特徴量と, それ以外の不正解とされる場所でサンプリングした特徴量によって識別器を学習しておき, 位置を変えながら逐次サンプリングを行って正解・不正解を判定する. したがって正解位置に到達するまでに多数の繰返し計算が必要となる.

これに対して我々は, 特徴量から, その特徴量がサンプリングされた位置の正解点に対する相対位置を推定する手法を提案する. これによって正解位置までの移動量が推定できるため, 探索処理が大幅に削減可能となる.

なお本章では説明の簡略化のため, 特徴点一点の位置に着目して摂動状態推定の原理説明を行う. その後, 5. において本原理を用いた形状モデルの摂動状態推定について説明を行う.

4.1 摂動特徴量

ここでは単純な場合を例にとり, サンプリングされた特徴量から摂動状態を推定可能であることを示す. 図 2(a) にサンプリング点配置の例を示す. サンプリング点は十字に配置された 5 点とし, 1 画素ずつ間隔を空けて配置されているものとする. 数字はサンプリング点の番号を示す. 図 2(a) 下図は, このサンプリング点を方形上に配置した状態を示す. 画像は 2 値画像とする. 今, この状態を基準位置として, サンプリング点配置が上下左右に摂動した状態でサンプリングを行う. 図 2(b) に摂動ベクトルと, 各状態でサンプリングした特徴量を示す. 例えば, L を例にとると基準位置から 1 画素左にサンプリング位置が移動することにより, 1, 2, 3, 5 のサンプリング点は方形から外れ, 4 のみが方形の画素と重なる. 図 2(c) はこれらの特徴量について, 3, 4, 5 の 3 点の状態について特徴量空間にプロットしたものである (黒=1, 白=0 とする).

このように, 特徴量は摂動状態によって特徴量空間内で異なった位置を示す. そしてその位置は特徴量がサンプリングされた摂動状態を反映したものになる.

ここで, 基準点から摂動させた状態で取得した特徴

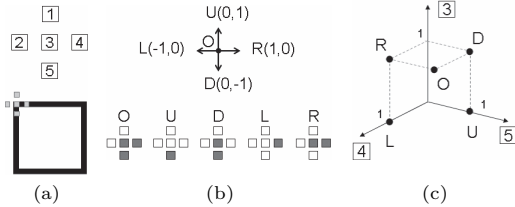


図 2 (a) サンプリング点配置 (上) と基準位置に配置した状態 (下), (b) 摂動ベクトル (上) と各摂動状態での特徴量 (下), (c) 特徴量空間. 各軸は数字で示したサンプリング点の状態を示す (黒=1, 白=0)

Fig. 2 (a) Sampling points and base position. (b) Perturbation vector and features at each perturbation. (c) Feature space. Each axis shows state of sampling point indicated by number.

量を摂動特徴量と呼ぶ。

4.2 摂動状態の推定

摂動特徴量に関してサンプリング点を十分多くとれば、位置・角度等の微妙な摂動を反映する低次元の特徴量空間を得ることができる。このような空間は PCA によって求めることができる。

今、あるサンプリング点配置で取得した特徴量を f_p と表す。ある特徴点を基準としてサンプリング位置を摂動させて特徴量 f_1, \dots, f_P を得たとき、この平均ベクトルを \bar{f} とすると、この特徴量分布の分散共分散行列 Σ は、

$$\Sigma = \frac{1}{P} \sum_{p=1}^P (f_p - \bar{f})(f_p - \bar{f})^T \quad (6)$$

となる。 P は摂動サンプリングの数である。これに対する標準固有値問題、

$$\Sigma \Phi = \Phi \Lambda \quad (\Phi^T \Phi = I) \quad (7)$$

を解くことによって、この空間の正規直交基底 Φ が得られる。このとき上位 k 個の固有値に対応する少数の固有ベクトルのみから構成した基底ベクトル $\tilde{\Phi}$ は、サンプリング時の摂動による特徴量変化を反映した部分空間となる。いったんこの空間が求めれば、サンプリングされた特徴量をこの空間に射影することで、摂動状態の推定が可能となる。

図 3 にこの手法によって求めた部分空間の例を示す。図 3 (a) は対象画像と、基準位置におけるサンプリング点配置を示している。特徴量次元は 37 である。このとき、基準点を中心として $x, y = \pm 0.1R$, $\theta = \pm 30 \text{ deg}$ の範囲で一様乱数によってサンプリング点を摂動させ、

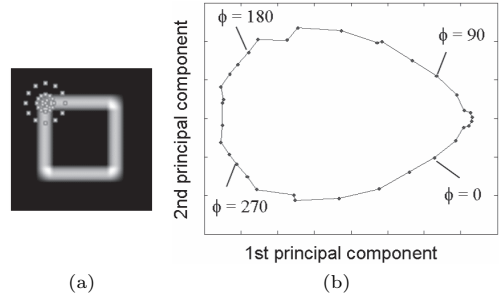


図 3 (a) テスト画像とサンプリング点配置, (b) 特徴量空間

Fig. 3 (a) Test image and sampling points. (b) Feature space.

5000 個の特徴量を得た。ここで R は図 3 (a) に示した方形の 1 辺の長さであり、 θ はサンプリング点配置の回転を表す。このようにして得られた特徴量集合に対して、上記の手続きにより部分空間を求めた。次に、各摂動状態での特徴量が部分空間内でどのように分布しているのかを知るため、基準点を中心とする半径 $0.1R$ の円周上の各位置 ($\phi = 0, 10, 20, \dots, 350 \text{ deg}$) で同一の点配置による特徴量サンプリングを行い ($\theta = 0$ に固定)、これを部分空間に射影した。図 3 (b) にこの結果を示す。横軸は第 1 主成分、縦軸は第 2 主成分に相当し、図中の ϕ は、サンプリング位置を示している。サンプリング時の摂動状態が、空間内での位置に反映されていることが確認できる。

4.3 顔画像における摂動特徴量

摂動特徴量の部分空間を用いることにより、特徴量から摂動状態を推定可能であることを、顔画像を用いた実験により説明する。図 4 (a) に実験に用いた顔画像の例と、サンプリング点配置を示す。基準位置は右目の顔中央側端点とした。

このサンプリング点配置にランダムな移動、回転、拡大縮小を加え、意図的に摂動状態での特徴量を得る。200 名の顔画像について、1 人当たり 5 回、計 1000 個の特徴量を得た。なお摂動の範囲は移動については両目間隔の 20% 以内、回転については $\pm 30 \text{ deg}$ 以内、拡大縮小については 0.5 ~ 1.5 倍の範囲とした。またこの実験では特徴量は画素の輝度値とした。この特徴量分布に対して上述の手続きを行い部分空間を得た。

図 4 (b) は、上記 200 名とは別の 100 名の顔画像について、基準位置から顔の中央寄り、外側寄りにそれぞれ水平に両目間隔の 20% 移動した点でサンプリングした特徴量を、第 1 主軸と第 2 主軸で張られる空間に

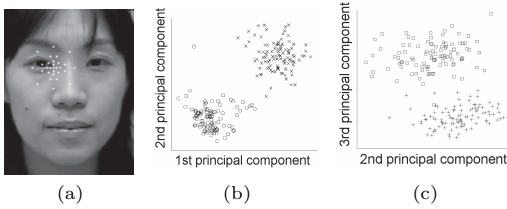


図 4 (a) サンプリング点配置 . (b), (c) 摂動状態での特徴量のプロット (\circ : 顔中央寄り, \times : 顔外側寄り, \ominus : -30 deg, \oplus : $+30$ deg)
 Fig. 4 (a) Sampling points. (b), (c) Feature vector plots on the feature space.

射影しプロットしたものである . また図 4 (c) は同様に $+30$ deg, -30 deg 回転させた配置での特徴量を第 2, 第 3 主軸で張られる空間に射影したものである . 特徴量は基準位置からの摂動方向によって空間内で明確に分離していることが分かる . したがって本実験について, 特徴量の部分空間内の位置より, 摂動方向とその距離を推定できることが分かる . このことは, 摂動特徴量の部分空間を用いることにより, 摂動状態を推定可能であることを示唆する結果であるといえる .

5. 形状モデルの摂動状態推定

前章においては各特徴点を独立に取り扱ったが, 形状モデル全体について特徴量と形状モデルパラメータの摂動状態を関連づけて学習することで, 形状モデルの摂動量を直接推定することが可能である . これによって形状モデルがどれだけ正解形状からずれているかの推定ができ, 大幅な形状修正による高速なフィッティングが実現できる .

特徴量と形状モデルパラメータはどちらも多次元であるため, 両者の相関関係の推定には多次元変量の回帰学習手法が必要となる . ここではその一種である正準相関分析 (Canonical Correlation Analysis : CCA) によってこれを行う .

以下, CCA について概説した後, 提案する学習手順とフィッティング手順について述べる .

5.1 正準相関分析

p 次元の変量ベクトル $\mathbf{x} = [x_1, \dots, x_p]^T$ と q 次元の変量ベクトル $\mathbf{y} = [y_1, \dots, y_q]^T$ があるとき, この同時分布を考え, その分散共分散行列を

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \quad (8)$$

とする . これらの変量ベクトルの任意の線形結合に

よって生成される新変量

$$\mathbf{u} = \mathbf{a}^T \mathbf{x}, \quad \mathbf{v} = \mathbf{b}^T \mathbf{y} \quad (9)$$

を考えたとき, 両者の相関が最大になるような係数ベクトル \mathbf{a}, \mathbf{b} を求める . そのためには共分散

$$\text{Cov}(\mathbf{u}, \mathbf{v}) = \mathbf{a}^T \Sigma_{XY} \mathbf{b} \quad (10)$$

を最大にするような \mathbf{a}, \mathbf{b} を求めればよい . この問題は, 両者の分散を 1 に標準化し, ラグランジュの未定乗数法を用いることで一般固有値問題を解く問題に帰着される .

今, この固有値問題を解いて得られた第 i 固有値に対応する固有ベクトルを $\mathbf{a}_i, \mathbf{b}_i$ と書く . $p > q$ とし, 第 q 正準変量まで求めるとすれば, 元の変量空間 \mathbf{x}, \mathbf{y} から新変量 \mathbf{u}, \mathbf{v} への変換はそれぞれ

$$\begin{aligned} \mathbf{u} &= [\mathbf{a}_1, \dots, \mathbf{a}_q]^T \mathbf{x} \\ &= \mathbf{A}^T \mathbf{x} \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{v} &= [\mathbf{b}_1, \dots, \mathbf{b}_q]^T \mathbf{y} \\ &= \mathbf{B}^T \mathbf{y} \end{aligned} \quad (12)$$

となる . \mathbf{u} から \mathbf{v} への線形回帰式は,

$$\begin{aligned} \mathbf{v} &= \text{diag}[\lambda_1, \dots, \lambda_q] \mathbf{u} \\ &= \Lambda \mathbf{u} \end{aligned} \quad (13)$$

で与えられる . 以上より, $\mathbf{x} \Rightarrow \mathbf{y}$ の写像変換は

$$\mathbf{y} = \mathbf{G} \mathbf{x} \quad (\text{ただし}, \mathbf{G} = (\mathbf{B}^T)^{-1} \Lambda \mathbf{A}^T) \quad (14)$$

のように表される .

5.2 学習手順

学習は以下の手順で行う . なお, 学習画像にはあらかじめ特徴点の正解座標が入力されているものとする .

- (1) $i = 1, n = 1$ とする .
- (2) n 枚目の顔画像に対して, 式 (3) の関係を用い正解座標群をパラメータ空間に射影し, 正解位置でのモデルパラメータ \mathbf{p}_{fit} を求める .
- (3) 乱数により摂動 $\Delta \mathbf{p}_i$ を加えたパラメータ $\mathbf{p}_{\text{err}} = \mathbf{p}_{\text{fit}} + \Delta \mathbf{p}_i$ をもつ摂動モデルを生成する .
- (4) $\mathbf{f}_i = \mathcal{S}(\mathbf{p}_{\text{err}})$ により特徴量を取得する .
- (5) $i \leftarrow i + 1$
- (6) (3) 以降を R 回繰り返す . R は乱数の発生回数 .
- (7) (2) 以降を $n = N$ まで繰り返す (N は顔画

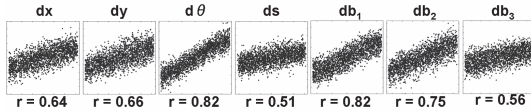


図 5 形状モデルパラメータ摂動量の正解値（横軸）と推定値（縦軸）の散布図．Pose パラメータ (t_x, t_y, t_θ, t_s), 及び Shape パラメータ (b_1, b_2, b_3) の分布を示す．各図とも縦軸は横軸の 1/2 倍のスケール．

Fig. 5 Scatter plot of model parameter perturbation. (horizontal) ground truth, (vertical) estimated value.

像の枚数)．

(8) $\Delta \mathbf{p}$ と f の同時分布を考え, CCA により変換行列 \mathbf{G} を求める．

図 5 に, 変換行列 \mathbf{G} によるパラメータ摂動量の推定値と正解値との散布図の例を示す．各グラフとも横軸が正解摂動量, 縦軸が推定摂動量であり, 各図の下にそれぞれの相関値を示した．図より相関値はおおむね 0.5 以上の正の値を示しており, 有効な変換行列が学習できていることが分かる．

なお図より, 推定誤差のばらつきは摂動量によらず一定となる傾向が見られる．これは学習を行った範囲では, 特徴量のもつばらつきが, 摂動量の大きさにかわらず一定であるためであると考えられる．

5.3 フィッティング手法

以上の学習によって, 特徴量と摂動によるパラメータ誤差を関連づける変換行列が得られた．本節では, これを用いてフィッティングを行う手法について述べる．なおここでは何らかの顔検出手段により既に画像上での顔位置は検出できているものとする．

(1) $i = 1$ とし, $\mathbf{p}_i = \mathbf{p}_{\text{init}}$ とする (例えば $\mathbf{p}_{\text{init}} = [\hat{t}_x, \hat{t}_y, \hat{t}_\theta, \hat{t}_s \mid \mathbf{0}^T]^T$, $\hat{t}_x, \hat{t}_y, \hat{t}_\theta$ は顔検出によって求めた顔中心座標と顔の傾き, \hat{t}_s は顔の大きさに応じた平均的な値)．

(2) $\mathbf{f}_i = S(\mathbf{p}_i)$ により特徴量を取得する．

(3) 変換行列 \mathbf{G} と式 (14) より誤差の推定値 $\Delta \mathbf{p}_i$ を得る．

(4) $\mathbf{p}_{i+1} = \mathbf{p}_i - \delta \Delta \mathbf{p}_i$ によりパラメータの修正を行う．

(5) $i \leftarrow i + 1$

(6) (2) 以降を終了条件を満足するまで繰り返す．なお, 終了条件としては $i > R$ (R は繰返し上限回数) とする方法, $\|\Delta \mathbf{p}_i\| < \varepsilon$ とする方法, またあらかじめ正解パラメータでの特徴量識別器を学習しておき,

表 1 使用した顔画像データベース

Table 1 Test databases.

	撮影環境	表情	画像枚数
BioID-DB	屋内, 均一照明	目や口の開閉	1,521 枚
Snap-DB	屋外 自然光を含む様々な環境	笑顔を含む様々な表情	2,325 枚
YaleB-DB	屋内, 照明変化	無表情	601 枚 (顔向き正面のみ)
INC-DB	屋内, 均一照明	目を閉じる, 口を開く等の表情変化 (詳細は図 6 に記載)	各表情 300 枚

この識別器の出力により終了判定する手法等が考えられる．

6. 評価実験

提案手法の性能を評価するための実験を行った．評価には, 公開データベース (DB) である BioID Face Database [9], The Extended Yale Face Database B [10] に加え, 我々が独自に収集した表情変化 DB (INC-DB) 及びスナップ写真 DB (Snap-DB) を用い, 表情変化や照明の変化に対する性能の比較を行った．各 DB の詳細については表 1 に示した．学習には, 照明 (屋外含む), 表情, 顔向きなどを変化させて我々が独自に撮影した画像約 10,000 枚を使用した．

なお我々は, ノイズに対する頑健性や計算の高速性から, 特徴量として Haar-like 特徴量 [5] を用いた．具体的には, 各サンプリング位置で方向, パターンの異なる数種類の Haar-like 特徴量をサンプリングする．特徴量が j 種類, サンプリング点数が k のとき, 特徴量次元は jk となる (本実験では $j = 6, k = 235$)．

従来手法の実装プログラムとしては STASM [8] を使用した．提案手法は Appearance を推定する手法ではなく, 特徴点位置の検出を目的とした手法であるため, 前者を目的とした AAM ではなく, 後者を目的とした ASM ベースの手法との性能比較を行った．STASM では, 多重解像度画像における法線方向輝度パターンのマッチングを利用し形状モデルのフィッティングを行う．ASAM 及び STASM は C/C++ 言語を用いて実装し, PC (PentiumD 3.2 GHz CPU) 上で実行した．まず入力画像に対して顔検出 [11] を行い, 顔方形の中心に平均形状モデルを初期位置として配置する．その初期位置から ASAM あるいは STASM を用いた形状モデルのフィッティングを行い, 左右の目尻, 目頭及び口の左右端点の計 6 点の位置を検出した．

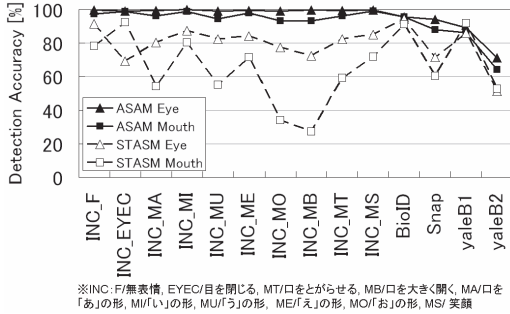


図 6 提案手法 (ASAM) 及び従来手法 (STASM) による検出性能

Fig. 6 Experimental results.

顔器官の端点位置は定義が難しく、入力する人によっても多少のずれが生じる。そこで、正解点を中心として半径が両目間隔の 10% の円内に入っているものを正解とした。

図 6 に検出率を示した。検出率については、各器官における左右の端点の平均値を示した。The Extended Yale Face Database B については、照明に対する影響を比較するため、顔向き正面の画像のうち、照明角度が左右 20 度、上下 30 度以内のもの (yaleB1) とそれ以外のもの (yaleB2) に分けて評価を行った。

図より両手法とも、表情変化及び照明変化の少ない DB (INC_F, BioID, yaleB1) においては同程度の検出率を示している。しかし目を閉じる (INC_EYEC)、口を開ける (INC_MA, MO, MB) といった表情変化時に従来手法では性能が大幅に低下しているのに対して、提案手法は安定して高い検出率を示していることが分かる。また激しい照明変動を含む DB (yaleB2) に対しては両手法とも大きな性能の低下を示しているが、提案手法の検出率は従来手法よりも高い値を示している。また表情変化及び照明変化を複合的に含む Snap-DB に対しても、提案手法が大幅に高い性能を示している。以上の結果から、提案手法は照明変化による影や、表情の変化に対して従来手法と比較して頑健であるといえる。

また平均処理時間は提案手法が 0.017 (s)、従来手法が 0.264 (s) であった。提案手法の処理時間は従来手法の約 20 分の 1 であり、大幅な高速化が実現できている。また形状モデル修正の回数は提案手法が 21 回、従来手法では 32 回であり、1 回の形状修正に要する平均処理時間はそれぞれ 0.81 (ms)、8.25 (ms) であった。



図 7 形状モデルフィッティングの成功例
Fig. 7 Model fitting results.

7. 考 察

本章では提案手法がもつ、(1) 構造的レティナサンプリング、(2) 特徴量からの摂動状態推定、それぞれの処理に関して、検出精度、処理速度に対する効果を実験結果より考察する。

7.1 構造的レティナサンプリング

影や表情変化によるしわなどのノイズは、フィッティングの性能を低下させる要因となる。頬や額など、もともとはっきりした特徴のない領域の特徴量は、モデルの摂動に対して変化が少なく、フィッティングに有効な情報をもたないことが多い。そのため相対的にノイズが主要な変動成分となりやすい。モデル全域を均等にサンプリングする手法では、こういった領域の特徴量を多く含むため、ノイズに対して脆弱となりやすい。

それに対して各特徴点周辺の局所的な画像特徴は、ノイズに対してそれほど大きく変化しないことが多い (例えば、鼻の穴は照明変化でも表情変化でも似た特徴を保つ)。しかしながら局所的な特徴だけに注目すると、正解位置からの少しのずれに対してもフィッティングが困難になりやすく、また、周辺情報を用いないため局所解に陥りやすいという課題がある。形状モデル全体で特徴点周辺を密にサンプリングする本手法は、特徴点周辺に比重を置いた特徴量が得られるため影や表情変化に頑健な性質をもちながら、周辺のサンプリングによりずれに対する頑健性を高め、更に形状モデル全体の情報を用いることで局所解に陥る危険性を低減させており、高精度のフィッティングを実現する上で有効であるといえる。

図 7 に、表情変化、照明環境の変化を含めた様々な状態に対するフィッティング結果例を挙げる。提案手法は笑顔、口を大きく開けるといった表情変化に対しても頑健にフィッティング可能である。その一方、目

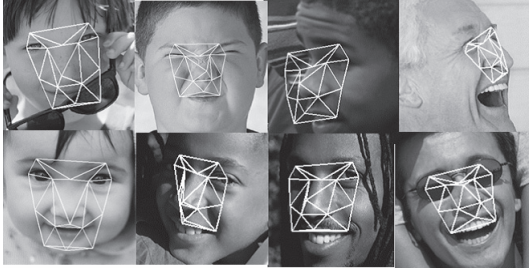


図 8 形状モデルフィッティングの失敗例
Fig. 8 Model fitting failure results.

を強く閉じて顔をゆがめるような表情変化に対しては失敗することがある。これは、重要な局所情報である目の特徴が大きく変化してしまうことによって、フィッティングが困難になったためと考えられる。同様の理由により、特徴点周囲の局所特徴を大きく変化させるような影がある場合や、サングラス・マスク等によって目や口が完全に隠れた場合も、フィッティングが困難となる。失敗したフィッティングの例を図 8 に示す。

7.2 特徴量からの摂動状態推定

従来手法においては、各ノードでの特徴点探索、形状モデルの微調整を繰り返しながら徐々に正解に近づいていくという処理を行うため、比較的長い処理時間を必要とした。これに対して提案手法によるフィッティングでは、各ノード個別での探索が不要であり、特徴量から一度の行列演算で形状修正を行うことが可能なため、処理時間を大幅に削減できている。

なお回帰推定により形状修正することは、上述の利点をもつ反面、摂動状態の推定に失敗した場合、正解とは別方向に大きく動くことでモデルの摂動が学習範囲を超え、フィッティングが発散する危険性をもつ。これを防止するため、5.3 のステップ (4) において十分小さな重み δ の値を選択し、一度に大きな形状修正が起らないようにすることが有効である。しかしながら δ を小さい値に設定するとフィッティングに多数の形状修正が必要となる。現在は徐々に重みの修正を行いながら性能評価を繰り返し最適値を求めているが、より効率的な重みの設定方法については今後検討が必要である。

また相関の学習に関して、本論文では CCA の例を示したが、Support vector 回帰、Relevance vector machine 等他の多変量回帰の手法によっても学習は可能であると考えられる。その際の検出性能の検証は今後の課題である。

8. む す び

構造的レティナサンプリングと特徴量からの摂動状態推定による、新たな形状モデルフィッティングの枠組みを提案した。提案手法は照明変化や表情変化など様々なノイズに対して頑健な性能を示しながら、従来手法と比較して大幅な高速化を実現している。今後はより幅広い顔向き、表情変化への対応とともに、サングラスやマスク等の顔の隠れに対する性能向上が課題である。これによって、デジタル画像機器、セキュリティ機器等でのよりいっそうの活用が期待される。また一方、人体や工業部品等に対して提案手法を適用することで、これらの対象に対しても高速、高精度な姿勢推定を実現していくことも今後の課題である。

文 献

- [1] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models — Their training and application," *Comput. Vis. Image Understand.*, vol.6, no.1, pp.38–59, 1995.
- [2] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *Computer Vision - ECCV' 98*, vol.II, Freiburg, Germany, 1998.
- [3] D. Cristinacce and T. Cootes, "Facial feature detection using adaboost with shape constraints," *Proc. 14th British Machine Vision Conference*, pp.231–240, Norwich, England, 2003.
- [4] F. Smeraldi and J. Bigun, "Retinal vision applied to facial features detection and face authentication," *Pattern Recognit. Lett.*, vol.23, pp.463–475, Feb. 2002.
- [5] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.
- [6] R. Donner, M. Reiter, G. Langs, P. Peloschek, and H. Bischof, "Fast active appearance model search using canonical correlation analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.28, no.10, pp.1690–1694, Oct. 2006.
- [7] G. Langs, P. Peloschek, R. Donner, M. Reiter, and H. Bischof, "Active feature models," *Proc. 18th International Conference on Pattern Recognition*, vol.1, pp.417–420, 2006.
- [8] S. Milborrow, *Locating Facial Features with Active Shape Models*, Master's thesis, University of Cape Town (Department of Image Processing), 2007.
- [9] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz, "Robust face detection using the Hausdorff distance," *Proc. Third International Conference on Audio- and Video-based Biometric Person Authentication*, Springer, Lect. Notes Comput. Sci., LNCS-

- 2091, pp.90–95, Halmstad, Sweden, June 2001.
- [10] K.-C. Lee, J. Ho, and D. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.27, no.5, pp.684–698, May 2005.
- [11] T. Yamashita, S. Lao, N. Tabata, H. Saito, and M. Kawade, “A fast omni-directional face detection system,” *IEEE 10th International Conference on Computer Vision (ICCV 2005), Demo*, 2005.
- [12] I. Matthews and S. Baker, “Active appearance models revisited,” *Int’l J. Computer Vision*, vol.60, no.2, pp.135–164, 2004.
- [13] 木下航一, 小西嘉典, 勞世広, 川出雅人, “3D モデル高速フィッティングによる顔特徴点検出・頭部姿勢推定,” 第 11 回画像の認識・理解シンポジウム, July 2008.
(平成 22 年 8 月 4 日受付, 11 月 8 日再受付)



木下 航一

平 8 神戸大・工・システム工卒。平 10 同大学院博士前記課程了。同年オムロン(株)入社。主として画像認識の研究開発に従事。平 21 画像センシングシンポジウム高木賞受賞。



小西 嘉典

平 12 京大・総合人間学卒。平 14 同大学院人間・環境学研究科博士前期課程了。平 17 同研究科博士後期課程指導認定退学。平 19 オムロン(株)入社。以来、顔画像認識技術、映像監視技術の研究開発に従事。



勞 世広

1984 中国浙江大・電気卒。1992 京都大学工学研究科博士課程中退。同年オムロン(株)入社。コンピュータビジョンの研究に従事。画像センシングシンポジウム高木賞受賞。



川出 雅人 (正員)

昭 56 名工大・計測工学卒。同年立石電機(現オムロン)入社。CAD/CAM/CAE の知能化研究開発に従事。平 2 米国カーネギーメロン大訪問研究員。平 6 日本帰国後、顔・人の視覚による理解技術の研究に従事。平 20 技術戦略立案、技術協創促進、技術事業化支援の業務に従事。画像センシングシンポジウム高木賞(平 21)受賞。IEEE 会員。



村瀬 洋 (正員:フェロー)

昭 53 名大・工・電気卒。昭 55 同大学院修士課程了。同年日本電信電話公社(現 NTT)入社。平 4 から 1 年間米国コンピュータ大客員研究員。平 15 から名古屋大学大学院情報科学研究科教授。現在に至る。文字・図形認識、コンピュータビジョン、マルチメディア認識の研究に従事。工博。昭 60 本会学術奨励賞、平 6 IEEE-CVPR 最優秀論文賞、平 7 情報処理学会山下記念研究賞、平 13 高柳記念奨励賞、平 13 本会ソサエティ論文賞、平 14 本会業績賞、平 15 文部科学大臣賞、平 16 IEEE Trans. MM 論文賞、平 22 前島賞、他受賞。IEEE フェロー、情報処理学会会員。