

顔変形をともなう 3 次元頭部姿勢の単眼推定

菅 野 裕 介^{†1} 佐 藤 洋 一^{†1}

本稿では、顔変形を含む 3 次元頭部姿勢の単眼カメラによる実時間推定手法を提案する。本手法は、顔形状の個人内変動（変形）と個人間変動（個人差）のモードを分離した多重線形顔形状モデルの下で、2 つの推定手法の統合により実現される。1 つは時間的に変化する姿勢・変形パラメータに対するパーティクルフィルタを用いた時系列推定であり、もう 1 つは人物に依存する個人差パラメータに対するバンドル調整の枠組みを用いた逐次的な推定である。このような統合により顔形状の変形と個人差を実時間で同時に推定することを可能にし、不特定多数の人物に対する顔変形と 3 次元頭部姿勢の実時間推定を実現している。本稿では実際のビデオ入力を用いた評価実験を行い、その結果についても報告する。

Person-independent Monocular Tracking of Face and Facial Actions

YUSUKE SUGANO^{†1} and YOICHI SATO^{†1}

This paper presents a monocular method of tracking faces and facial actions using a multilinear face model that treats interpersonal and intrapersonal shape variations separately. We created this method by integrating two different frameworks: particle filter-based tracking for time-dependent facial action and pose estimation and incremental bundle adjustment for person-dependent shape estimation. This combination together with multilinear face models is the key to tracking faces and facial actions of arbitrary people in real time with no pre-learned individual face models. Experiments using real video sequences demonstrate the effectiveness of our method.

1. はじめに

人間の 3 次元頭部姿勢および顔形状の実時間推定はコンピュータビジョンの分野における重要な研究課題であり、HCI や ITS、ビデオ監視技術など様々な分野で注目を集める技術になっている。実環境での応用を想定した場合、大きな表情変動が起きた場合にも安定して頭部姿勢と顔形状の変化を追跡することのできる技術が重要になる。さらに運用コストの観点から見ると、単独のカメラで推定が可能であり、かつ顔変形を推定するうえで個人ごとの事前準備を必要としないような技術が望ましい。本研究の目的は、これらの条件を満たすような、不特定多数の人物を対象とした 3 次元頭部姿勢および顔変形の単眼推定手法を提案することにある。

顔変形を含む頭部姿勢推定技術に関しては、顔変形モデルを利用したモデルベースの手法が数多く提案されている^{5),8),9),17),18)}。主成分分析などの手法を用いて顔変形を少数の基底の重み付き線形和として表現することにより、基底に対する重みとしての変形パラメータとモデル姿勢を推定するアプローチをとる。しかし、過去に提案されたモデルベースの推定手法のほとんどが、追跡前に取得したデータに基づく各ユーザ固有の変形モデルを利用していた^{5),8),9),17)}。実際の応用を考えるうえで、ユーザごとに長い時間をかけて事前に変形モデルを準備するのは現実的ではない。

この問題に対する 1 つの解法としては、ユーザごとの 3 次元変形基底を入力画像から直接推定するアプローチが考えられる。Bregler ら¹⁾ は、Tomasi-Kanade の因子分解法を非剛体の形状推定に拡張することで、変形に関する事前の知識なしで 2 次元の特徴点軌跡からの形状復元を実現している。しかし、このアプローチは事前に追跡された特徴点軌跡を前提とするため、こうした手法を実時間の 3 次元頭部姿勢推定に直接応用することは難しいといえる。また、岡ら¹⁸⁾ の手法では複数台のカメラを用いて 3 次元顔形状データを取得することで、実時間推定と並行して変形モデルの構築を行うことを可能にしている。しかし、事前に正確なキャリブレーションがなされた複数台のカメラを必要とするため、設置コストの観点から見ると不十分な点があった。

一方、こうした線形形状モデルを複数人の顔形状データをもとに構築することで、汎用的な顔形状モデルを用いた推定アプローチをとることが可能になる^{6),16)}。しかし、2 次元の AAM (Active Appearance Model) に関して汎用モデルと個人専用モデルの比較を行った Gross らの論文⁶⁾でも指摘されているように、汎用モデルはパラメータ推定における収束性能の低下という点で本質的な欠点を持つ。さらに汎用モデルは必然的に、本来個人には起

^{†1} 東京大学生産技術研究所

Institute of Industrial Science, The University of Tokyo

こりえないスケール変動や回転などの変形要素を含むことになる。これらの要素は深さ方向の運動を含む頭部姿勢との区別が難しく、特に 3 次元形状モデルを用いた単眼推定の場合には推定精度を劣化させる大きな原因になる。近年では、Zhu ら¹⁶⁾ がこのような汎用 3 次元形状モデルを利用した実時間頭部姿勢推定手法を提案している。ここでは実画像およびシュミレーションデータを用いてモデルの再投影誤差に関する精度評価が行われているが、上記のように汎用 3 次元モデルを扱ううえで大きな問題となる 3 次元頭部姿勢の推定精度に関しては、定量的な誤差評価がなされていない。

他方、もう 1 つの汎用モデル構築アプローチとして、複数人の変形に含まれる 2 つの要素をそれぞれ異なるパラメータ（個人内変動に対する変形パラメータと個人間変動に対する個人差パラメータ）により表現したパラメータ分離顔形状モデルをあげることができる。前述のような線形モデルと比較すると、パラメータ分離モデルには振舞いの異なる 2 つのパラメータ、すなわち、時間的に変化する変形パラメータとユーザごとに一定の値をとる個人差パラメータを切り分けて扱うことが可能になるという大きなメリットがある。Dornaika らの手法⁴⁾ ではパラメータを分けたアニメーション用 3D メッシュモデルを用いて、変形パラメータの時系列推定を行っている。しかし、推定の中で個人差パラメータを調整する枠組みは備えていないため、追跡開始前に何らかの手法で決定する必要がある、任意人物対応への取り組みはなされていない。また、Vlasic らの手法¹³⁾ では多重線形のパラメータ分離モデルに対して複数フレームの解析により個人差を決定するアプローチをとっている。ただし彼らの手法は動画の後処理を前提としており、実時間姿勢推定の枠組みには適用できない。さらに、DeCarlo ら²⁾ はモデルベースオブティカルフローの推定残差を利用して、個人差も含めたパラメータ全体をフレームごとに調整する手法を提案している。しかし、彼らの手法は計算コストが高く、実時間での推定は実現されていない。

このように、これまで提案されてきたモデルベースの実時間推定手法にとって最大の問題は、不特定多数の人物に対する高精度な推定を実現できない点にある。現在単眼カメラでの実時間推定を実現している手法のほとんどは線形モデルを前提としているが、これを汎用モデルに拡張した場合推定精度の劣化が大きな問題となる。一方パラメータ分離モデルを用いることで精度劣化の主な原因であるパラメータの曖昧性を解消することができるが、任意ユーザへの対応を可能にするためにはユーザごとに異なる個人差パラメータを事前情報なしで推定する枠組みが必要になる。しかし、従来の手法では変形および個人差パラメータを実時間で同時に推定する枠組みは提案されていなかった。これに対して本手法では、多重線形顔形状モデルに対して図 1 に示すような 2 つのパラメータ推定の枠組みを統合して

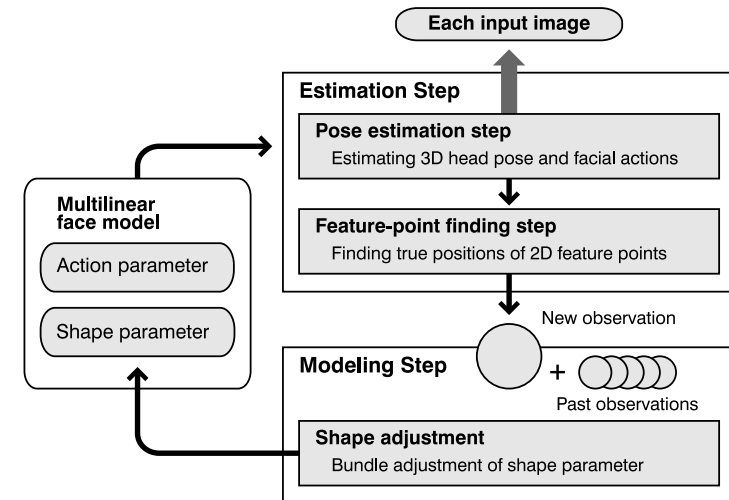


図 1 システムの概要
Fig. 1 System overview.

用いることにより、顔変形を含む 3 次元頭部姿勢推定と同時に個人差パラメータの調整を実現する。

1 つ目は時系列推定部（*Estimation step*）であり、パーティクルフィルタを利用した変形パラメータと 3 次元頭部姿勢の推定、および画像中の 2 次元特徴点座標の再計算を行う。これにより、部分遮蔽や深さ方向の運動が起こった場合でも、3 次元状態推定と特徴点抽出を安定して実行することが可能になる。

2 つ目は形状調整部（*Modeling step*）であり、時系列推定部で得られた 2 次元特徴点座標をもとにした逐次的なバンドル調整により個人差パラメータの最適化を行う。過去に観測されたフレームを合わせて調整に用いることで、頭部姿勢との区別が困難な要素を含む個人差に対しても安定した形状推定を可能にしている。形状調整部で更新された個人差パラメータは次フレームの時系列推定部で用いられ、これを繰り返すことで漸進的な精度改善および変形推定の個人対応が可能になる。

以上のようなパーティクルフィルタによる時系列推定と逐次的なバンドル調整の統合により、本手法ではユーザごとの事前準備をともなうことなく、顔変形を含む 3 次元頭部姿勢の実時間推定を実現する。本研究は、このような 2 つの手法の統合による推定アプローチを提

案し、顔形状の個人差成分と変形成分の実時間推定を不特定多数の人物に対して実現した最初の研究例であると考えられる。

以下、本稿の構成は次のようになる。まず、2 章では本手法で用いるパラメータ分離モデルの構築手法について述べる。次に、3 章で形状調整部、4 章で時系列推定部の処理の詳細について説明する。さらに 5 章で提案手法を用いた評価実験の結果について述べ、最後に 6 章で本稿のまとめについて述べる。

2. パラメータ分離顔形状モデルの事前構築

本手法では追跡システムの構築に先立ち、変形と個人差のモードが分離された多重線形顔形状モデルを構築する。まず本章では、この顔形状モデルの詳細とその作成方法について述べる。

本手法において顔形状は、ユーザ頭部に固定されたモデル座標系で定義される K 個の特徴点の 3 次元座標により構成された形状ベクトル $M \in \mathbb{R}^{3K}$ と、追跡開始時にユーザごとに登録される各特徴点のテンプレート画像 T により表現される。なお、本稿で特徴点の数は $K = 10$ で固定されており、各点は両目の両端と口の両端、両鼻孔、そして両眉の内側の端点に対応する。図 2 中、写真内の + マークが各特徴点を示している。

本手法では、変形と個人差の 2 つのパラメータにより任意の顔形状ベクトル M を記述するために、N-mode SVD (Singular Value Decomposition) を利用した多重線形モデルを構築する。多重線形モデルは Vasilescu ら^{11),12)} による画像解析への応用を受けてコンピュータビジョンの領域でも様々な分野への応用が試みられており、以下本研究における N-mode SVD の定式化も Vasilescu らの研究をもとにしたものである。

多重線形モデルは、図 3 のようなデータテンソル $\mathcal{T} \in \mathbb{R}^{3K \times S \times A}$ をもとに構築する。同図中、Feature points のモードには顔形状ベクトル M が、Action および Shape の各モードにはそれぞれ変形（個人内変動）と個人差（個人間変動）が対応する。すなわち、Shape 軸に垂直な断面には同じ人物の A 個の異なる変形データが、Action 軸に垂直な断面には S 人の異なる人物の同じ変形が、それぞれ対応付けられた形で並んでいる。

N-mode SVD を用いることで、データテンソル $\mathcal{T}$ は次のような積の形に分解できる。

$$\mathcal{T} = \mathcal{C} \times_{\text{feature}} U_{\text{feature}} \times_{\text{shape}} U_{\text{shape}} \times_{\text{action}} U_{\text{action}} \quad (1)$$

$$= \mathcal{M} \times_{\text{shape}} U_{\text{shape}} \times_{\text{action}} U_{\text{action}} \quad (2)$$

$\times_i$ はテンソルと行列のモード i における積を表し、具体的にはテンソルのモード i 空間内の各ベクトルに関して行列との積をとる演算を示す。式 (2) は通常の SVD を 3 階以上のテ

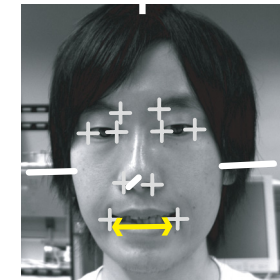


図 2 モデル化対象となる顔変形の例
Fig. 2 Example of facial deformation.

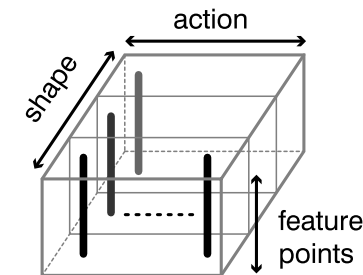


図 3 データテンソル
Fig. 3 Data tensor.

ンソルに対して一般化したものであり、 U_i の列ベクトルはモード i 空間の正規直交基底に対応し、その相互関係を記述したコアテンソル $\mathcal{C} \in \mathbb{R}^{3K \times S \times A}$ は SVD における特異値行列に対応する。 U_{feature} とコアテンソル $\mathcal{C}$ との積 $\mathcal{M}$ は、特徴点ベクトルのモードに関する基底を含む顔形状モデルテンソルである。

さらに、変形と個人差の次元数を削減した基底を用いることで、次のようにデータの近似表現が得られる。

$$\mathcal{T} \approx \mathcal{M} \times_{\text{shape}} \tilde{U}_{\text{shape}} \times_{\text{action}} \tilde{U}_{\text{action}} \quad (3)$$

このように近似されたモデルテンソル $\mathcal{M}$ の各基底に対して任意の重みを与えることで、任意の顔形状ベクトルを記述することができるようになる。

データテンソルを構築するためには、まず複数の人物から顔変形のデータを取得する必要

がある．本手法では，形状データの獲得に複数カメラに基づく顔変形推定手法¹⁸⁾を利用した．モデル構築のサンプルとなる S 人の人物は， K 個の顔特徴点が自動的に追跡された状態で 2 種類の運動を行う．1 つは口を横に広げる水平方向の動きであり，もう 1 つは眉と口の垂直方向の動きである．その後，各運動の最小位置と最大位置の間で等間隔に 5 個，計 $A = 10$ 個の形状を記録されたデータの中から自動的に抽出する．

以上の処理により獲得した S 人 $\times$ A 個の形状データに関して，最初にサンプル全体の平均 $\bar{M}$ を差し引いたうえで，データテンソル $\mathcal{T}$ を構築する．さらに，式 (3) のように変形 ($A \rightarrow A'$) と個人差 ($S \rightarrow S'$) の次元数を削減したモデルテンソルを算出する．すると，任意の個人差パラメータ $s \in \mathbb{R}^{S'}$ と変形パラメータ $a \in \mathbb{R}^{A'}$ ，そして平均形状 $\bar{M}$ を用いることで，次のように単独の顔形状ベクトル M を記述することができる．

$$M = \bar{M} + \mathcal{M} \times_{\text{shape}} s^T \times_{\text{action}} a^T \quad (4)$$

ここで，式 (3) における $\check{U}_{\text{shape}} = (\check{s}_1, \dots, \check{s}_S)^T$ ， $\check{U}_{\text{action}} = (\check{a}_1, \dots, \check{a}_A)^T$ の各行は，データテンソルに含まれる $S \times A$ 個の形状ベクトルの各々に対応するパラメータベクトルになっている．本手法では，モデル構築の段階で $\check{s}_1$ から $\check{s}_S$ の平均 $\bar{s}$ と標準偏差 σ_s ，および $\check{a}_1$ から $\check{a}_A$ の平均 $\bar{a}$ と標準偏差 σ_a を算出する．こうして得られるパラメータの分布に関する情報を，バンドル調整における調整範囲の制約やパーティクルフィルタにおける仮説の拡散・評価に利用している．

以上の手順により作成した顔形状モデルを用いることで，ユーザ依存の個人差 s と時間依存の変形 a ，そして世界座標系におけるモデル姿勢 p により任意のユーザに対して顔変形と顔姿勢を表現することができる．以降の章では，このモデルを用いた個人差調整および顔姿勢推定の枠組みについて説明する．

3. バンドル調整の枠組みによる個人差調整

本章では，顔形状モデルに対して個人差の調整を行う，形状調整部 (図 1) の処理について述べる．バンドル調整は最尤推定手法の 1 つであり，複数フレームにまたがる投影誤差関数を最小化することで 2 次元の特徴点軌跡に対する 3 次元空間内パラメータの最適化を行う．剛体形状推定の枠組みとしては顔形状構築¹⁴⁾ や実時間顔姿勢推定¹⁰⁾ への応用が見られ，また，基底を含めた最適化を非剛体形状推定に利用する例も提案されている³⁾．

一方，本研究では過去の手法とは異なり，多重線形モデルのもとでバンドル調整の枠組みを時間的に変化しない個人差パラメータの推定に利用する．このとき，実時間処理の中で安

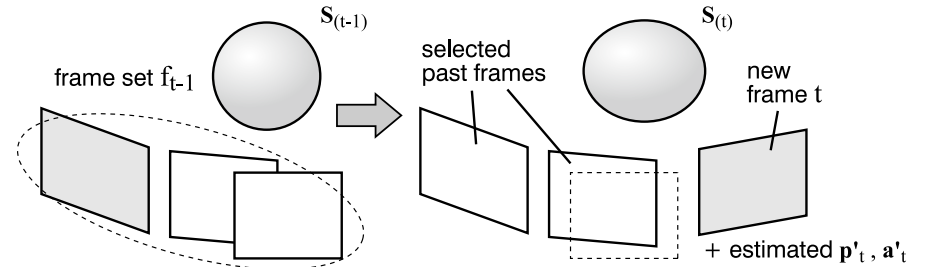


図 4 逐次的なバンドル調整の実行
Fig. 4 Flow of incremental bundle adjustment.

定した推定を実現するために，本手法では 2 つの工夫を導入している．1 つは時系列推定部の姿勢推定結果を利用した調整用フレーム集合の選択的な構築であり，もう 1 つはパラメータの分布に基づく調整範囲の拘束条件設定である．以下，その具体的な処理について説明する．

3.1 調整用フレーム集合の逐次的構築

本手法の顔形状モデルを用いた場合，バンドル調整の問題設定は次のようになる．まず，フレーム i におけるユーザ顔部の状態は，世界座標系からモデル座標系への並進と回転を表す 6 次元顔姿勢ベクトル p_i と変形パラメータベクトル a_i ，そして時間に依存しない個人差パラメータベクトル s によって記述できる． a_i, s から式 (4) によって決まる顔形状ベクトル M_i について，その投影点は次のように算出することができる．

$$m_i = \mathcal{P}(p_i, M_i(a_i, s)) \quad (5)$$

$\mathcal{P}$ はカメラパラメータにより事前に決定される投影関数であり， M_i に顔姿勢 p_i に応じた並進や回転を適用したうえで， M_i の各特徴点を画像平面に投影する．すなわち， m_i は K 個の 2 次元座標で構成される $2K$ 次元ベクトルとなる．一方，フレーム i における真の特徴点 2 次元座標を $\hat{m}_i$ とする．これは時系列推定部において姿勢推定の結果をもとに探索した座標であり，計算方法の詳細に関しては 4 章で述べる．すると，次のような誤差関数が定義できる．

$$F_t = \sum_{i \in f_t} \|\hat{m}_i - m_i(p_i, a_i, s)\|^2 \quad (6)$$

f_t は図 4 のように，第 t フレーム時点の調整に利用する n 枚のフレーム集合を示し，式 (6) は f_t 全体における特徴点投影誤差の和を示す．

本手法では、前回の調整に利用したフレーム集合の中から 1 つを新たなフレーム t で置き換えることでフレームごとに新たな調整用フレーム集合 f_t を生成し、これを利用した逐次的なバンドル調整を行う (図 4)。このとき、新たに追加されるフレーム t に対応する姿勢および変形に関しては時系列推定部の推定結果 p'_t, a'_t を初期値として与える。以降はこれをもとに、投影誤差が小さくなるようなパラメータ差分を繰り返し算出することでパラメータの調整を実行する。

同様のアプローチとして、Zhang ら¹⁵⁾ は連続したフレームを用いた逐次的バンドル調整を提案している。しかし、実時間追跡の中で得られる連続フレームは多くの場合各画像の見えがほぼ変化しないため、深さ方向の曖昧さを十分解消できないことが考えられる。そこで、できる限り多様な角度から顔形状を評価するために、本手法では考えられる n 通りの組合せに関して顔向きの分散が最も大きくなる組合せを調整用の集合として採用する。このフレーム選択を繰り返すことにより、追跡を続けるに従って調整フレーム集合における姿勢のばらつきが増加するような枠組みを実現する。

3.2 パラメータ制約をともなう誤差関数最小化

次に、式 (6) の具体的な最小化手法について説明する。本手法では、各パラメータの調整範囲に対する制約条件 C_{p_i}, C_{a_i}, C_s をともなう LM (Levenberg-Marquardt) アルゴリズム⁷⁾ を利用した次のような最小化を行う。

$$\min_{\{p_i\}, \{a_i\}, s} F_t, \quad p_i \in C_{p_i}, \quad a_i \in C_{a_i}, \quad s \in C_s \quad (7)$$

このとき、前述のとおり時系列推定部の推定結果が初期値として与えられる変形および姿勢パラメータに関してはすでにほぼ正確な推定が行われていると考え、次のように初期値 $\hat{p}$ から一定の範囲内に調整が制限されるような制約を設定する。

$$C_{p_i} = \{p_i \mid \hat{p}_i - \lambda_p \leq p_i \leq \hat{p}_i + \lambda_p\} \quad (8)$$

調整範囲を示す定数ベクトル λ_p は、現在は経験的に値を設定している。変形に関する制約 C_{a_i} も同様に設定する。これに対して、個人差パラメータに関しては 2 章で算出したパラメータの分布をもとに、次のように比較的緩やかな制約を与える。

$$C_s = \{s \mid \bar{s} - 2\sigma_s \leq s \leq \bar{s} + 2\sigma_s\} \quad (9)$$

すなわち、正規分布を仮定した場合に分布全体の 95% に収まるような制限を与えることで、必要以上の変形が起こるのを防いでいる。

最後に、以上の処理により調整された個人差パラメータを s' とすると、過去の調整結果の平均として次フレームの推定で用いる個人差パラメータ $s_{(t)}$ を算出する。

$$s_{(t)} = \frac{t-1}{t} s_{(t-1)} + \frac{1}{t} s' \quad (10)$$

これにより、個人差パラメータに関してはフレーム間の細かい推定変動の影響を抑え、長期的な観測に基づく値を推定に利用している。

4. 姿勢と変形の時系列推定

本章では、時系列推定部 (図 1) の処理について説明する。前章で述べたバンドル調整の枠組みだけでは、いくつかの理由から姿勢と変形を正しく推定することができない。第 1 に、姿勢の変動に対して不安定で、特に深さ方向の運動に対する追従性が低いこと、第 2 に、バンドル調整に用いる特徴点座標の算出を単独に安定して実現するのが困難であること、そして第 3 に、特徴点の遮蔽に対応できないことがその理由としてあげられる。本手法ではこれらの問題に対処するために、変形と姿勢に関してパーティクルフィルタを用いた 3 次元モデルベースの状態推定を行う。

図 1 に示したように、時系列推定部はパーティクルフィルタによる姿勢、変形の推定を行うパラメータ推定部 (*Pose estimation step*) と、この推定結果をもとにバンドル調整に用いるための正確な特徴点位置を再計算する特徴点再計算部 (*Feature-point recalculation step*) により構成される。以下、各ステップの詳細について説明する。

4.1 パーティクルフィルタによる姿勢・変形推定

本節ではパーティクルフィルタを利用した頭部姿勢と顔変形の推定を行うパラメータ推定部の処理について説明する。

逐次推定の対象となるのは時間的に変化するパラメータ p_t, a_t のみであり、個人差パラメータ s はその対象としない。式 (10) により前フレームで算出された個人差パラメータ $s_{(t-1)}$ を用いて、式 (4) は変形パラメータ a_t に関する線形モデルとして表現することができる。

$$M_t = \bar{M} + \mathcal{M}_t a_t \quad (\mathcal{M}_t = \mathcal{M} \times_{\text{shape}} s_{(t-1)}^T) \quad (11)$$

このようにして得られる変形モデルを用いて、各入力画像に対して $(6 + A')$ 次元状態ベクトル $x_t = (p_t^T, a_t^T)^T$ を推定する。

本手法におけるパーティクルフィルタの仮説群 $\{(u_t^{(i)}; \pi_t^{(i)})\} (i = 1 \dots N)$ は、 $(6 + A')$ 次元状態空間内の N 個の仮説 $u_t^{(i)}$ と、各仮説に対応する重み $\pi_t^{(i)}$ で構成される。まず、直前のフレーム $t-1$ の仮説群 $\{(u_{t-1}^{(i)}; \pi_{t-1}^{(i)})\}$ と等速直線運動に基づく動作モデルを用いて新た

な仮説を N 個生成する．

$$\mathbf{u}_t^{(i)} = \mathbf{u}_{t-1}' + \tau \mathbf{v}_{t-1} + \omega \quad (12)$$

このとき、 $\mathbf{u}_{t-1}'$ は $\{(\mathbf{u}_{t-1}^{(i)}; \pi_{t-1}^{(i)})\}$ の中から確率に基づいて選択された仮説であり、 τ はフレーム間の時間間隔、 $\mathbf{v}_{t-1}$ は直前のフレーム $t-1$ で計算されている状態ベクトル $\mathbf{x}$ の速度である．ただし、変形パラメータ $\mathbf{a}_t$ に関しては必ずしも等速直線運動の仮定に一致しないため、 $\mathbf{v}_{t-1}$ の中で $\mathbf{a}_t$ に対応する要素は 0 に設定している．

ω は仮説の拡散の性質を決めるシステム雑音であり、各要素はそれぞれ固有の分散を持つ平均が 0 のガウス雑音としている．頭部姿勢パラメータに対応する部分は岡らの手法¹⁸⁾と同様に、状態ベクトルの速度に応じて適応的に制御する．一方、変形パラメータに対応する部分に関してはモデル作成時に計算した標準偏差に基づき、ガウス雑音の分散を $\kappa\sigma_a$ とする． κ は経験的に 0.2 に設定している．

次に、新たな仮説 $\mathbf{u}_t^{(i)}$ に対応する重み $\pi_t^{(i)}$ を決定する．これは仮説と入力画像との一致度を示す量であり、次のような関数により計算される．

$$\pi_t^{(i)} \propto \exp\left(-\frac{(K - \mathcal{N}(\mathbf{u}_t^{(i)}))^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{1}{2} \sum_{b=1}^{A'} \left(\frac{a_{t,b}^{(i)} - \bar{a}_b}{s_b}\right)^2\right) \quad (13)$$

ここで $\mathcal{N}(\mathbf{u}_t^{(i)})$ はテンプレート画像 T に基づく正規化相関スコアを全 K 個の特徴点において加算したものであり、 $-K$ と K の間の値をとる．第 1 項はこれをガウス関数により評価したもので、標準偏差 σ は経験的に 1.0 に設定している．一方、第 2 項は顔変形パラメータ $\mathbf{a}_t^{(i)}$ についての関数であり、モデルが過度に変形しないような拘束を与えている．このとき、 $a_{t,b}^{(i)}$ 、 $\bar{a}_b$ 、 s_b はそれぞれ $\mathbf{a}_t^{(i)}$ 、 $\bar{\mathbf{a}}$ 、 σ_a の b 番目の要素である．以上の計算を行ったうえで、合計が 1 になるよう $\pi_t^{(i)}$ を正規化する．

最後に、仮説群 $\{(\mathbf{u}_t^{(i)}; \pi_t^{(i)})\}$ を用いて、重みが最大となる仮説の近傍に属する仮説集合の加重平均により現在の状態ベクトル $\mathbf{x}_t$ を求める．

なお、初期状態ベクトル $\mathbf{x}_0$ は前述のバンドル調整の枠組みを利用して決定する．初期化部ではまず、オムロン社の OKAO ビジョンライブラリを利用して入力画像から顔および K 個の顔特徴点座標を自動的に検出する．バンドル調整に必要な n フレーム分の検出に成功すると、式 (6) の最小化により全パラメータの初期化を行う．この場合、バンドル調整におけるパラメータ初期値はあらかじめ設定した値を利用する．5 章で述べる実験では、頭部姿

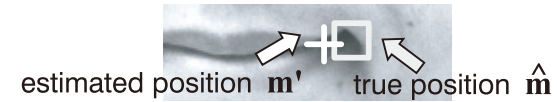


図 5 特徴点位置の再計算

Fig. 5 Finding true feature points.

勢に関しては画面中央に正対する値を、形状パラメータに関しては平均 $\bar{\mathbf{a}}$ 、 $\bar{\mathbf{s}}$ を初期値として与えている．

4.2 2 次元特徴点座標の再計算

次に、本節では特徴点再計算部の処理について説明する．個人差の調整が正しく行われていない場合、4.1 節で推定した特徴点座標は必ずしも正確な位置にならない (図 5)．そこで特徴点座標の推定値 $\mathbf{m}_t'$ をもとに真の 2 次元特徴点座標 $\hat{\mathbf{m}}_t$ を再計算し、これを式 (6) のバンドル調整における観測として利用することにより個人差パラメータの修正を行う．以下、計算方法の詳細について述べる．

本手法では Gokturk らの手法⁵⁾ と同様に次のような誤差関数 E_t を定義し、 E_t の最小化により前フレームとの差分形状 $d\hat{\mathbf{m}}$ を繰り返し計算することで $\hat{\mathbf{m}}_t$ を算出する．

$$E_t = \sum_{\text{ROI}} \left\{ \rho \|\hat{\mathbf{I}}_t - \hat{\mathbf{I}}_{t-1}\|^2 + \|\hat{\mathbf{I}}_t - \hat{\mathbf{I}}_1\|^2 \right\} + \epsilon \|\hat{\mathbf{m}}_t - \mathbf{m}_t'\|^2 \quad (14)$$

第 1 項は、特徴点付近の見えに関する誤差を表す． $\hat{\mathbf{I}}_t \in \mathbb{R}^K$ は $\hat{\mathbf{m}}_t$ に対応した輝度ベクトルであり、 $\hat{\mathbf{I}}_t$ の k 番目の要素は $\hat{\mathbf{m}}_t$ の k 番目の 2 次元座標における入力画像の輝度を表している．直前のフレームとの誤差だけではなく追跡開始時のフレームとの誤差を加えることで特徴点のドリフトを回避しており、これらの比率を表す ρ は経験的に 4 に設定している．注目領域の大きさは 16×16 画素に設定している．

また、第 2 項の $\mathbf{m}_t'$ は 2 次元特徴点座標の推定値であり、4.1 節において推定された状態ベクトル $\mathbf{x}_t$ に基づき、式 (5) により計算される．すなわち、第 2 項は推定座標 $\mathbf{m}_t'$ と真座標 $\hat{\mathbf{m}}_t$ の誤差を表す．この項を加えることで、推定座標の周辺で真の特徴点座標を計算することができる． ϵ は定数であり、本手法では経験的に 4000 に固定している．

5. 評価実験

提案手法の性能を評価するために、いくつかの比較実験を行った．まず、個人用の変形モデルを用いた複眼推定手法¹⁸⁾ との比較により提案手法の推定精度を評価した．さらに、

47 顔変形をとまう 3 次元頭部姿勢の単眼推定

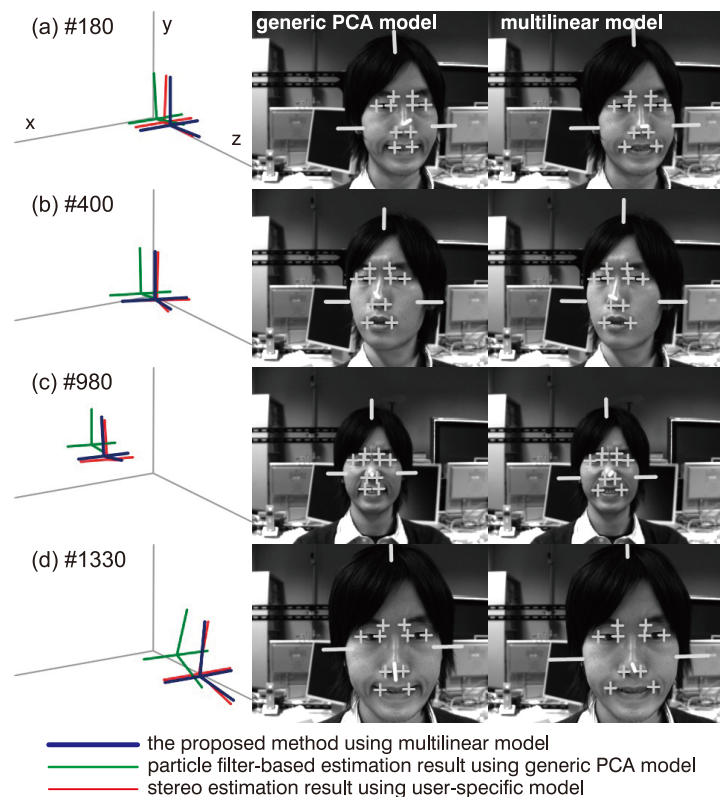


図 6 頭部姿勢推定の比較（右列は提案手法による推定結果，中列は汎用モデルを用いた手法による推定結果であり，左列はこれら 2 つの結果を異なる視点から比較した画像である）

Fig.6 Result images: the right column shows actual estimation results of our method using the multilinear model, and the center column shows results of the generic model-based method. The left column shows these results rendered from a different viewpoint.

PCA により構築した線形顔形状モデルを用いてパーティクルフィルタによる推定を行った結果との比較を行うことにより，文献 16) に代表されるような，変形と個人差を分離しない汎用モデルを用いた手法との性能比較を行った．

実験では，26 人の人物から獲得した各 10 通りの変形サンプルをもとに，個人差パラメータ 15 次元，変形パラメータ 5 次元の顔形状モデルを作成した．さらに，同じデータ集合に

表 1 複眼推定に対する推定誤差の比較
Table 1 Comparison of estimation errors.

[mm]	x	y	z	[deg.]	roll	pitch	yaw
Particle filter-based estimation using the generic PCA model							
Mean	6.14	4.71	51.32	Mean	0.34	6.54	3.34
Std. Dev.	4.88	4.09	38.29	Std. Dev.	0.29	4.71	2.73
Our method using the multilinear model							
Mean	3.26	4.37	20.18	Mean	0.41	3.12	2.33
Std. Dev.	2.62	2.83	11.18	Std. Dev.	0.27	2.49	1.98

に対する PCA により線形汎用モデルを構築し，パラメータの次元数は 20 次元とした．なお，追跡対象となるユーザは，モデルの学習対象である 26 人の中には含まれていない．

本実験では，Intel Core 2 Duo E6700 を搭載した汎用 PC を 1 台使用した．搭載メモリは 3.0 GB であり，OS は Windows XP を利用している．2 台の校正済み IEEE1394 カメラにより撮影した 60 秒（1800 フレーム）分の画像列を入力として利用し，画像サイズは 640×480 画素，画像テンプレート T の大きさは 16×16 画素であった．このとき，追跡対象ユーザはランダムな顔変形と同時に，奥行き方向の大きな姿勢変動を含む頭部運動を行っている．パーティクルフィルタの仮説の総数は 1000，バンドル調整に用いるフレーム数は $n = 7$ とした．LM アルゴリズムにおける繰返しの回数は，初期化部で計 10 回，追跡中は計 5 回に制限した．このとき，初期化部の処理時間は約 90 [ms]，追跡中は約 32 [ms/frame] となった．

図 6 は右列が提案手法による姿勢推定の結果，中列が汎用モデルに基づくパーティクルフィルタによる推定結果であり，左列は 2 つの推定結果について異なる視点から複眼推定との比較を行った図である．さらに，頭部姿勢の推定結果，およびモデル座標系における顔変形の推定誤差に関して，入力画像列全体の詳細なグラフを図 7 に示す．このとき，複眼推定に対する各手法の推定誤差は表 1 に示すとおりである．このように，汎用モデルを用いた推定では形状変動が深さ方向の推定に悪影響を与え，顔変形も正しく推定できないのに対し，提案手法を用いた推定では複眼推定に比べてもほぼ遜色のない推定精度が実現できていることが分かる．

6. おわりに

本稿では，多重線形顔形状モデルを用いた単眼カメラによる実時間頭部姿勢推定手法を提案した．パーティクルフィルタを用いた顔変形・頭部姿勢の実時間推定とバンドル調整の枠

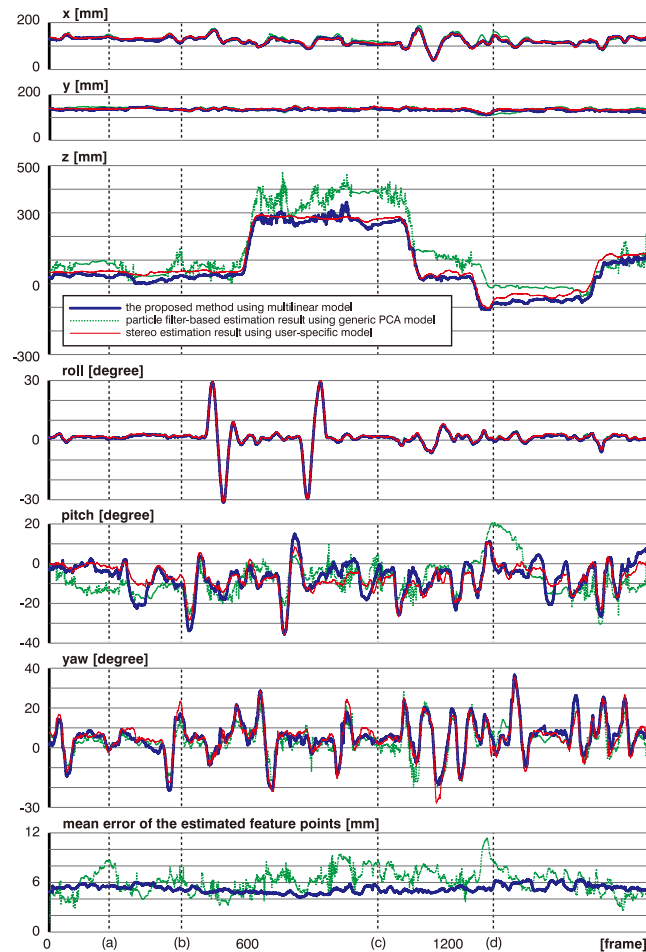


図 7 頭部姿勢推定の結果と顔変形推定の誤差 (x は水平方向, y は鉛直方向, z は深さ方向の並進運動であり, $roll$ は z 軸回り, yaw は y 軸回り, $pitch$ は x 軸回りの回転運動である)
 Fig. 7 Estimation results: x , y , and z are the horizontal, vertical, and depth-directional translation, and $roll$, $pitch$, and yaw are the rotation around the z , y , and x axes, respectively. The bottom graph shows the facial shape estimation error in the model coordinate system.

組みを用いた個人差調整を組み合わせることで、ユーザごとの事前準備をとまなうことなく、顔変形をともなう頭部姿勢推定を実現している。本研究は、このような統合手法により実時間推定を実現した最初の研究例であると考えられる。本手法を用いることで複眼カメラを用いた従来手法と比較しても遜色のない推定精度が実現できることが、評価実験により示された。

本手法の 1 つの課題としては、何らかの頭部姿勢変動や顔変形が得られない限り個人差パラメータの調整が不可能であるという点があげられる。本手法では 2 次元特徴点座標の再計算を行うことで、3 次元変形が正しく推定できなかった場合でも個人差調整のための情報を獲得する枠組みを備えているが、応用する場面によっては十分な情報が得られず、追跡の不安定性が増すことも考える。これは本手法に限らず、ターゲットの統計的情報を逐次的に学習するアプローチが持つ本質的なトレードオフであると考えられ、この問題に取り組むことは今後の重要な課題といえる。そのほかにも、深さ方向の追従性や顔向きを追跡範囲など単眼推定自体の精度改善を図るとともに、視線推定や表情解析など詳細な顔情報解析との統合を見据えて、より実用に適した推定システムの提案を行う予定である。

謝辞 本研究の一部には、オムロン株式会社の顔検出・顔器官検出技術を利用した。

参 考 文 献

- 1) Bregler, C., Hertzmann, A. and Biermann, H.: Recovering non-rigid 3d shape from image streams, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Vol.2, pp.690–696 (2000).
- 2) DeCarlo, D. and Metaxas, D.: Adjusting Shape Parameters using Model-based Optical Flow Residuals, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.6, pp.814–823 (2002).
- 3) DelBue, A., Smeraldi, F., Agapito, L. and Mary, Q.: Non-rigid structure from motion using non-parametric tracking and non-linear optimization, *Proc. IEEE Workshop on Articulated and Non-Rigid Motion*, Vol.1 (2004).
- 4) Dornaika, F. and Davoine, F.: On appearance based face and facial action tracking, *IEEE Trans. Circuits and Systems for Video Technology*, Vol.16, No.9, pp.1107–1124 (2006).
- 5) Gokturk, S.B., Bouguet, J.Y. and Grzeszczuk, R.: A data-driven model for monocular face tracking, *Proc. IEEE Int. Conf. Computer Vision*, Vol.2, pp.701–708 (2001).
- 6) Gross, R., Matthews, I. and Baker, S.: Generic vs. person specific active appearance models, *Image and Vision Computing*, Vol.23, No.11, pp.1080–1093 (2005).

- 7) Lourakis, M.I.A.: levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++ (2004). <http://www.ics.forth.gr/~lourakis/levmar/>
- 8) Matthews, I. and Baker, S.: Active appearance models revisited, *Int. J. Computer Vision*, Vol.60, No.2, pp.135–164 (2004).
- 9) Munoz, E., Buenaposada, J.M. and Baumela, L.: Efficient model-based 3D tracking of deformable objects, *Proc. IEEE Int. Conf. Computer Vision*, pp.877–882 (2005).
- 10) Vacchetti, L., Lepetit, V. and Fua, P.: Stable real-time 3D tracking using online and offline information, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.26, No.10, pp.1380–1384 (2004).
- 11) Vasilescu, M.A.O. and Terzopoulos, D.: Multilinear analysis of image ensembles: Tensorfaces, *Proc. European Conf. on Computer Vision*, pp.447–460 (2002).
- 12) Vasilescu, M.A.O. and Terzopoulos, D.: Multilinear image analysis for facial recognition, *Proc. Int. Conf. Pattern Recognition (ICPR '02)*, Vol.2, pp.511–514 (2002).
- 13) Vlasic, D., Brand, M., Pfister, H. and Popovic, J.: Face transfer with multilinear models, *ACM Trans. Graphics (Proc. ACM SIGGRAPH 2005)*, Vol.24, No.3, pp.426–433 (2005).
- 14) Xin, L., Wang, Q., Tao, J., Tang, X., Tan, T. and Shum, H.: Automatic 3D face modeling from video, *Proc. IEEE Int. Conf. Computer Vision*, Vol.2, pp.1193–1199 (2005).
- 15) Zhang, Z. and Shan, Y.: Incremental motion estimation through modified bundle adjustment, *Proc. IEEE Int. Conf. Image Processing*, Vol.2, pp.343–346 (2003).
- 16) Zhu, J., Hoi, S.C.H. and Lyu, M.R.: Real-time non-rigid shape recovery via active appearance models for augmented reality, *Proc. 9th European Conf. Computer Vision*, pp.186–197 (2006).
- 17) Zhu, Z. and Ji, Q.: Robust Real-Time Face Pose and Facial Expression Recovery, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.681–688 (2006).
- 18) 岡 兼司, 菅野裕介, 佐藤洋一: 頭部変形モデルの自動構築をともなう実時間頭部

姿勢推定, 情報処理学会論文誌: コンピュータビジョンとイメージメディア, Vol.47, No.SIG10 (CVIM15), pp.185–194 (2006).

(平成 19 年 9 月 21 日受付)

(平成 20 年 3 月 10 日採録)

(担当編集委員 牧 淳人)



菅野 裕介

2005 年東京大学工学部電子情報工学科卒業。2007 年東京大学大学院情報理工学系研究科修士課程修了。同年同研究科博士課程入学。現在に至る。コンピュータビジョン, ヒューマン・コンピュータ・インタラクションに関する研究に従事。



佐藤 洋一 (正会員)

1990 年東大学工学部機械工学科卒業。1997 年カーネギーメロン大学大学院計算機科学部ロボティクス学科博士課程修了。Ph.D. in Robotics. 同年より東京大学生産技術研究所研究機関研究員, 講師, 助教授を経て, 現在, 同大学大学院情報学環准教授。コンピュータビジョン, ヒューマン・コンピュータ・インタラクション, コンピュータグラフィックスに関する研究に従事。2008 年電子情報通信学会論文賞, 2006 年電子情報通信学会論文賞, 1999 年情報処理学会山下記念研究賞, 1999 年日本バーチャルリアリティ学会論文賞等を受賞。電子情報通信学会, 日本バーチャルリアリティ学会, ACM, IEEE 各会員。