

フリーハンド撮影された多視点スナップショットからの顔形状復元

Reconstruction of Facial Shape from Freehand Multi-viewpoint Snapshots

鈴木 誠 司[†], 正会員 斎 藤 英 雄[†], 持 丸 正 明^{††}

Seiji Suzuki[†], Hideo Saito[†] and Masaaki Mochimaru^{††}

Abstract We propose a method that can reconstruct both facial poses and the facial shape from freehand multi-viewpoint snapshots. This method is based on Active Shape Models (ASM), which is a technique that uses a facial shape database. Most ASM methods require image in which the facial pose is known, but our method does not require this information. First, we chose an initial shape by selecting the model from the database which is most similar to the input images. Then, we improve the model by morphing it to better fit the input images. Next, we estimate the pose of the face using the morphed model. Finally we repeat the process, improving both the facial shape and the facial pose until the error between the input and the computed result is minimized. Through experimentation, we show that our method reconstructs the facial shape within 3.5mm of the ground truth.

キーワード: 顔, 3D 復元, 多視点, Active Shape Model, Principal Component Analysis, 最適化

1. ま え が き

以前より物体の3次元形状を復元する研究が盛んに行われてきた。レンジスキャナなどのハードウェアデバイスを用いると、高精度に3次元形状を測定することが可能であるということが知られている¹⁾。また、プロジェクタを用いて特殊なパターンを物体に投影することで、物体の形状を復元する手法もある²⁾³⁾。しかしこういった特殊なデバイスを用いることは利便性に欠け、復元対象も自ずと制限される。これらの問題を解決するため、カメラで得られた画像のみから形状を復元する手法が、コンピュータビジョンの分野において数多く提案されている。

従来より、Shape from Shading, Shape from Textureといった単視点画像から物体の3次元形状を復元する手法が提案されてきた。しかし、物体の反射特性や照明条件などに対する拘束や前提条件が多く実用的なものではなかった。それに対し、三角測量や両眼視差の考え方を導入し、複

数視点画像から物体の3次元形状を復元する Stereo Vision という手法がある。こちらはカメラの厳密な位置合わせや画像間の対応点の取得が困難であるという問題があった。

一方、顔に特化した3次元形状の復元法についても数多くの研究がなされている。特に顔は人間にとって非常に重要な情報を持っており、顔の形状を求めることができれば様々な応用が期待できる。

画像からの顔の3次元形状復元の一例として、ビデオ等の連続画像から形状を復元する Structure from Motion という手法がある⁴⁾。これは、1台のカメラで撮影された同一人物の画像列から形状を復元しようとするものである。しかし、この手法は復元に際し、各フレーム間の対応点を求めるために Optical Flow を用いている。顔は頬などテクスチャの均一な部分が大部分を占めており、Optical Flow による対応点算出が正確に行われず、復元精度が悪いという問題があった。

そこで、単に同一人物に対して撮影された画像を利用するだけではなく、あらかじめ収集しておいた人間の頭部3次元形状情報を持つデータベースを主成分分析し、低次元の固有空間から形状を推定する Active Shape Model (ASM) という手法が提案されている。ASM の例として、顔の3次元形状・テクスチャ情報を主成分分析し形状を推定する手法^{5)~7)}がある。これらの手法は、テクスチャ情報を持つ顔の3次元モデルを高々1枚の入力画像(単視点画像)にフィッティングさせることによって復元を行っている。す

2008 年 8 月 30 日受付, 2008 年 12 月 15 日最終受付, 2008 年 12 月 22 日採録

[†] 慶應義塾大学大学院理工学研究科

(〒223-8522 横浜市港北区日吉 3-14-1, TEL 045-563-1141)

^{††} 産業総合研究所デジタルヒューマン研究センター

(〒135-0064 江東区青海 2-41-6, TEL 03-3599-8201)

[†] Graduate School of Science and Technology, Keio University

(3-14-1, Hiyoshi, Kohoku-ku, Yokohama-shi 223-8522, Japan)

^{††} Digital Human Research Center, AIST

(2-41-6, Oume, Koto-ku, Tokyo 135-0064, Japan)

なわち、別視点から捉えた画像を用いておらず、幾何学的情報を無視してしまっている。よって復元結果のアピラランスは尤もらしく見えるが、形状としては幾何学的整合性に欠けるものとなってしまっている。

一方、多視点画像に対してASMを適用した手法^{8)~10)}が提案されている。しかし、これらの手法は顔のカメラに対する姿勢が既知であるという条件を用いている。すなわち、キャリブレーションされたカメラを用いた特殊な環境下での撮影が要求される。そのため、顔の姿勢が未知であるフリーハンドで簡易に撮影されたスナップショットなどからは形状を復元することができなかった。

本稿ではこれらの問題を踏まえ、こういった未校正の多視点画像から顔の3次元形状を復元するASMの一手法¹¹⁾を基とした、より実用的な手法を提案する。すなわち、提案手法は以下のような特徴を持つ。

- レンジスキャナやプロジェクタといった特殊なデバイスを使用することなく、カメラで撮影された画像のみを入力とし顔の形状を復元する。
- Stereo Vision などのようなキャリブレーションされたカメラシステムを必要とせず、フリーハンドで撮影されたスナップショットから形状を復元することが可能である。

上記のように、提案手法ではフリーハンドで撮影されたスナップショットを入力として復元を行うことが可能であるが、その際でもカメラの内部パラメータは既知であるとする。

なお、提案手法ではいくらかの情報をユーザが手動で入力する必要があるが、安価で一般的なデバイスであるデジタルスチルカメラ1台のみの利用で顔の形状を復元することが可能であるため、自宅で手軽に用いることができるツールとなる。近年、仮想現実世界において自身のアバタを登場させる機会が増えている。提案手法は、このような仮想現実のためのアバタ作成などに応用されることが想定される。また、監視カメラ映像などから不審者の顔の形状を復元し、これにより顔の認識を行うシステムや眼鏡のオーダーメイドシステム¹²⁾などへの応用が考えられる。

さらに、近年、顔の置換システム¹³⁾が提案されているが、顔の姿勢が異なる場合においての利用が困難であるという問題点があった。提案手法は顔の形状を復元することで、自由な姿勢で置換できるようになるため、この問題を解決する一手法となり得る。

なお、著者らの所属する慶應義塾の創立150年を記念した展示会「2008年 慶應連合三田会大会 大慶應博」において、著者らは実際にこのシステムを作成し展示を行っている。この展示システムは以下のようなものである。

- 事前に提案手法を用いて慶應義塾創立者である福澤諭吉の顔形状を復元しておく。
- 会場において来場者の写真を展示会スタッフと一緒に撮影する。

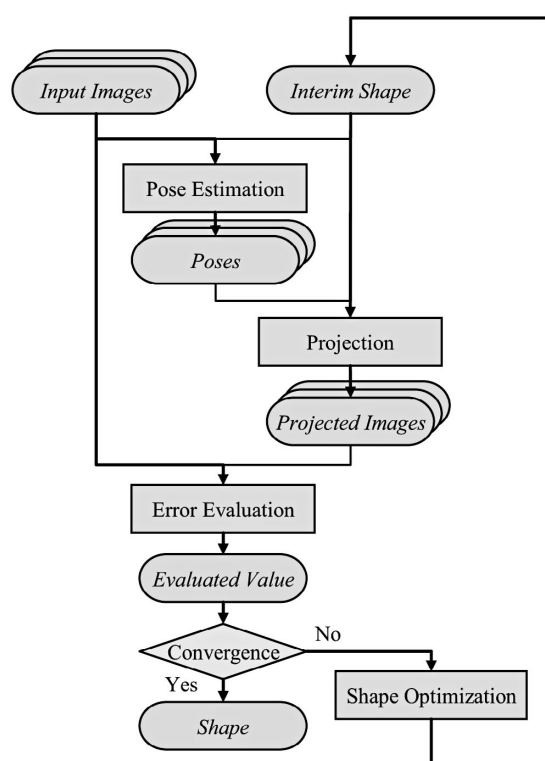


図1 本手法の流れ
Flowchart of this method.

iii. 展示会スタッフの顔を福澤諭吉に置き換える。

これにより、来場者はあたかも福澤諭吉と撮ったかのような写真を得ることができる。このように、ある人物のスナップショットを著名人の顔に置き換えるといったエンターテインメント分野での利用も可能となる。

2. 手 法

図1に本手法の大まかな流れを示す。基本的には、あらかじめデータベースから初期の形状を生成しておき、この形状を徐々に入力画像に近づくように変化させ最適な形状を推定していくということになる。

まず、暫定的な形状（初回は初期形状；後述）を用いることにより、カメラの顔に対する姿勢を各入力画像に対して大まかに推定する。次に大まかに推定された姿勢を用いて、各入力画像と同様の位置に暫定的な形状を投影し、CGで描画する。このようにして得られた投影画像と入力画像の組それぞれにおいて、その画像同士の相違度を2.6に示す評価関数で計算する。その評価値の合計が小さくなるように形状を変形し最適化していく。

最適化の初期段階においては、暫定的な形状は復元対象人物の真の形状とは大きく異なるため、姿勢推定において正しい姿勢を推定できない場合がある。よって、初期段階では姿勢はあくまで大まかに推定されることになるが、形状が最適化されるにつれ、より真の形状に近づき、姿勢も次第に安定して正確に求められるようになってくる。さらに、姿勢が正確に求められるようになってくると、形状の

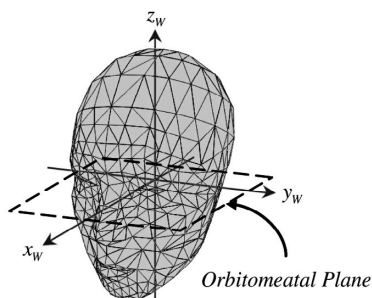


図 2 耳眼水平座標系
Orbitomeatal plane coordinate system.

復元精度も上がる．このように姿勢推定，形状最適化の処理を繰り返すことで，復元誤差を減少させる．

最終的に形状が最適化されてくると，上述の通り姿勢および形状は安定してくる．すると形状最適化の処理で計算される評価値が収束する．評価値が収束したときの形状を復元結果とする．

以上のように，本手法は入力画像と投影画像とのエネルギー評価値が最も低くなるような姿勢と形状を推定するエネルギー最小化問題に帰着される．すなわち，形状を $\mathbf{x}$ ，入力画像と投影画像との相違度（評価値）を計算する評価関数を f ，相違度を y としたとき， $y = f(\mathbf{x})$ なる y を最小にする引数 $\mathbf{x}$ を求める問題となる．

このとき，引数である形状 $\mathbf{x}$ はなるべく低い次元であることが望ましい．そのため本手法では ASM の手法に基づき，データベースを主成分分析することにより，形状 $\mathbf{x}$ のパラメータ数を減少させている．2.1 ではこのデータベースの構成について，2.2 では主成分分析の手法について簡単に説明する．さらに，初期形状生成，姿勢推定，形状最適化，評価関数についてそれぞれ，2.3，2.4，2.5，2.6 で述べる．

2.1 データベース

ここでは，本手法で用いるデータベースの構成について述べる．本データベースは，解剖学的知見に基づいた人物の頭部頂点データからなる．各人物の頭部データは，レンジスキャナで撮影された約 20 万の頂点から，解剖学的に重要な情報を持つ 100 点を指定し，残りの 330 点を内挿することで得られた形状モデルである．解剖学の専門家が対象人物の顔を実際に触診し，重要な点にランドマークを貼り付けて計測を行うことで，この解剖学的に重要な頂点を求めている．このようにして測定されたすべての形状モデルは図 2 に示す，耳眼水平座標系によって正確に位置合わせされている．

本手法では，この 430 点の頭部データのうち，顔部分のみの 260 点を用いる．本データベースの頂点データ，および復元に用いる顔の特徴点を図示したものを図 3 に示す．この形状モデルを男女各 52 人分保有しており，それぞれが各性別におけるデータベースを構成している．

なお，本データベースは著者の一人，持丸正明の所属す

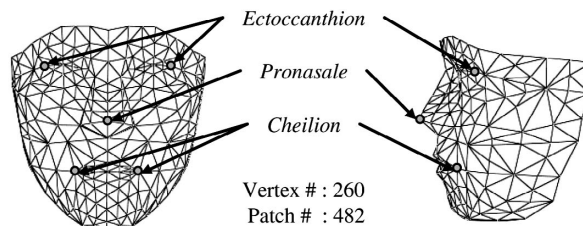


図 3 データベースにおける形状モデル
Shape model in this database.

る産業総合研究所デジタルヒューマン研究センターにて作成されたものである．

2.2 主成分分析

主成分分析とは，できるだけ情報の損失なしに，多変量データを少数個の総合的なパラメータで代表させる方法である．前述の通り，形状最適化の際に形状のパラメータ数を減少させる必要があるが，ここではその手法について述べる．

今，データベース内に， n 個の頂点で構成される形状データ $\mathbf{x} = [x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_n, y_n, z_n]^T$ が m 人分ある．これをそれぞれ， $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ とする．ただし今回用いたデータベースにおいては， $n = 260$ ， $m = 52$ である．これらのデータを主成分分析した結果，得られた主成分ベクトルを第 1 主成分から順に $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k$ とする．このとき， $k = \min(3n, m)$ である．また， $\mathbf{p}_i \in \mathbb{R}^{3n}$ ($1 \leq i \leq k$) である．

得られた主成分ベクトルを用いると，任意の形状データ $\mathbf{x}$ は，データベースの平均形状 $\bar{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i$ を用いて，

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{i=1}^k s_i \mathbf{p}_i \quad (1)$$

と表現することができる．このとき，この各主成分ベクトルに対応する主成分得点 s_i を各要素とするベクトル $\mathbf{s} \in \mathbb{R}^k$ は $\mathbf{x}$ と 1 対 1 対応の関係にあり，互いに一意に形状を表現するパラメータとなる．

この $\mathbf{s}$ の上位 l ($1 \leq l < k$) パラメータのみを取り出したベクトル $\hat{\mathbf{s}} = [\hat{s}_1, \hat{s}_2, \dots, \hat{s}_l]^T$ で形状を表現すると，

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \sum_{i=1}^l \hat{s}_i \mathbf{p}_i \quad (2)$$

として，形状を予測値 $\hat{\mathbf{x}}$ として復元できる．これ以後， $\hat{\mathbf{x}}$ と $\hat{\mathbf{s}}$ はいつでも相互に変換可能であることに注意されたい．

このようにして，本来 $3n$ すなわち 780 次元で表現される形状を， l 次元にまで圧縮することが可能となる．以後の操作においては，入力画像に最もフィットする最適な l 個の形状パラメータ $\hat{\mathbf{s}}$ を推定することが目的となる．

2.3 初期形状生成

ここでは，最適化における初期形状の生成方法について述べる．考えられる最も単純な方法としては，データベースの平均顔 $\bar{\mathbf{x}}$ を初期形状として用いる方法である．しか

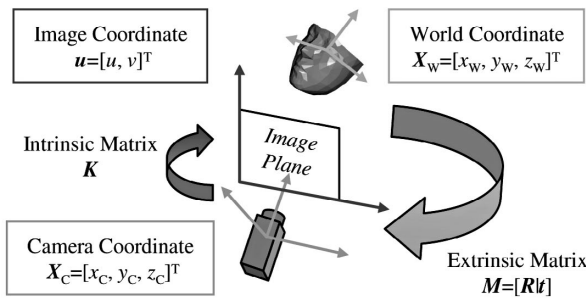


図 4 座標系間の関係
Relationships among coordinate systems.

し、復元対象人物の真の形状が平均から遠い場合には、評価値が局所解に陥り、得られる形状が本人のものに最適ではない可能性が高くなってしまいう問題がある。

そこで本手法では、入力画像に最もフィットするデータベース上の形状を初期形状とする。これは平均顔形状 $\bar{x}$ と比べて、より真の形状に近い初期形状であると考えられる。そのため、最適化の過程で遭遇する局所解の数が比較的少ないと予想される。よって、各局所解に陥る確率こそ同等ではあるものの、たとえ局所解に陥ってしまった場合でも、復元形状がより真の形状に近いことが期待される。

さて、具体的には、以下の操作を行うことにより入力画像に最もフィットするデータベース上の形状を得る。まずデータベースに登録された m 人分の形状 $x_1, x_2, \dots, x_m$ を固有空間に投影し、 $\hat{s}_1, \hat{s}_2, \dots, \hat{s}_m$ を求める。その後、この各形状 $\hat{s}_i$ ($1 \leq i \leq m$) に対して姿勢推定を行い、投影画像を作成することで、評価値 $y_i = f(\hat{s}_i)$ を計算する。そして、この評価値 y_i が最小となる形状 $\hat{s}_i$ を初期形状 $\hat{s}_{init}$ とする。

2.4 姿勢推定

ここでは各入力画像におけるカメラの顔に対する姿勢を推定する方法について説明する。図 4 のように、画像座標系、カメラ座標系、世界座標系をそれぞれ $u = [u, v]^T$, $X_C = [x_C, y_C, z_C]^T$, $X_W = [x_W, y_W, z_W]^T$ とする。このとき、一般に各座標系は同時座標表現で、

$$\tilde{u} \simeq K \tilde{X}_C \quad (3)$$

$$\tilde{X}_C \simeq M \tilde{X}_W \quad (4)$$

という関係を満たす。ただし、 $K, M = [R|t]$ はそれぞれカメラの内部パラメータと外部パラメータであり、各要素を

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

$$M = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \quad (6)$$

と表すこととする。このとき、世界座標系は形状 $\hat{x}$ が記述

されている耳眼平面座標系にあたる。

さて、入力画像における特徴点座標、内部パラメータ、外部パラメータをそれぞれ u, K, M とする。本手法においては、内部パラメータ K は既知であるとする。また、特徴点座標 u はユーザのクリックにより得られているとする。さらに、特徴点の世界座標系における 3 次元位置 X_W は暫定的な形状 $\hat{s}$ より得られる。このとき、外部パラメータ M を、 u と X_W の特徴点 5 組（鼻尖点・外眼角点・口角点；図 3 参照）から、Zhang の手法¹⁴⁾により算出する。このようにして算出された外部パラメータ M は、カメラと顔の 3 次元的な位置関係を表すものであり、これを算出したことにより顔の位置を特定できたことになる。

2.5 形状最適化

得られた姿勢から、式 (3)(4) より、

$$\begin{aligned} \tilde{u} &\simeq K M \tilde{X}_W \\ &\simeq P \tilde{X}_W \end{aligned} \quad (7)$$

となり、射影行列 P が求まるので、これを用いて暫定的な形状 $\hat{s}$ を入力画像と同じ姿勢で投影し、CG で描画する。このようにして得られた投影画像は、 P で表される画像上の位置に $\hat{s}$ を描画したものとなるが、このとき P は $\hat{s}$ により計算されることに注意されたい。そのため、結局、投影画像は $\hat{s}$ のみに依存することになる。

さて、次に入力画像と投影画像との相違度を 2.6 で示す評価関数で計算し、これを最適化する。最適化には Levenberg-Marquardt 法 (LM 法) を用いる。この方法で、後述する四つの評価関数 $f_1, f_2, \dots, f_4$ の値の 2 乗和、

$$f(\hat{s}) = \sum_{i=1}^4 \{\alpha_i f_i(\hat{s})\}^2 \quad (8)$$

を最小とする形状 $\hat{s}_{opt} = \arg \min_{\hat{s}} f(\hat{s})$ を求める。なお、 α_i ($1 \leq i \leq 4$) は経験的に得られたそれぞれの評価関数に対する重み付け定数である。このようにして得られた形状 $\hat{s}_{opt}$ が最終的な出力となる。

2.6 評価関数

形状最適化において用いる評価関数について述べる。評価に際し、各入力画像に対して、五つの特徴点座標、顔部分のシルエット画像、外輪郭画像を用意する必要がある。なお、外輪郭とは撮影された顔と背景との境界のことを指す。一方、顔と髪の毛・耳などとの境界のことを内輪郭と呼ぶことにする。

また、以下では簡単のため、1 枚の入力画像に対する評価値の計算方法を示すが、実際の入力画像は複数枚あるので、各入力画像の評価値を統合する必要がある。しかし、入力画像はスナップショットであり、画像に占める顔の割合が、カメラと顔の 3 次元的な位置関係によってそれぞれ異なっている。形状最適化を目的とする評価関数においては、形状の相違度のみを評価するべきであり、画像上における顔の面積に依存するべきではない。そのため、本手法では以

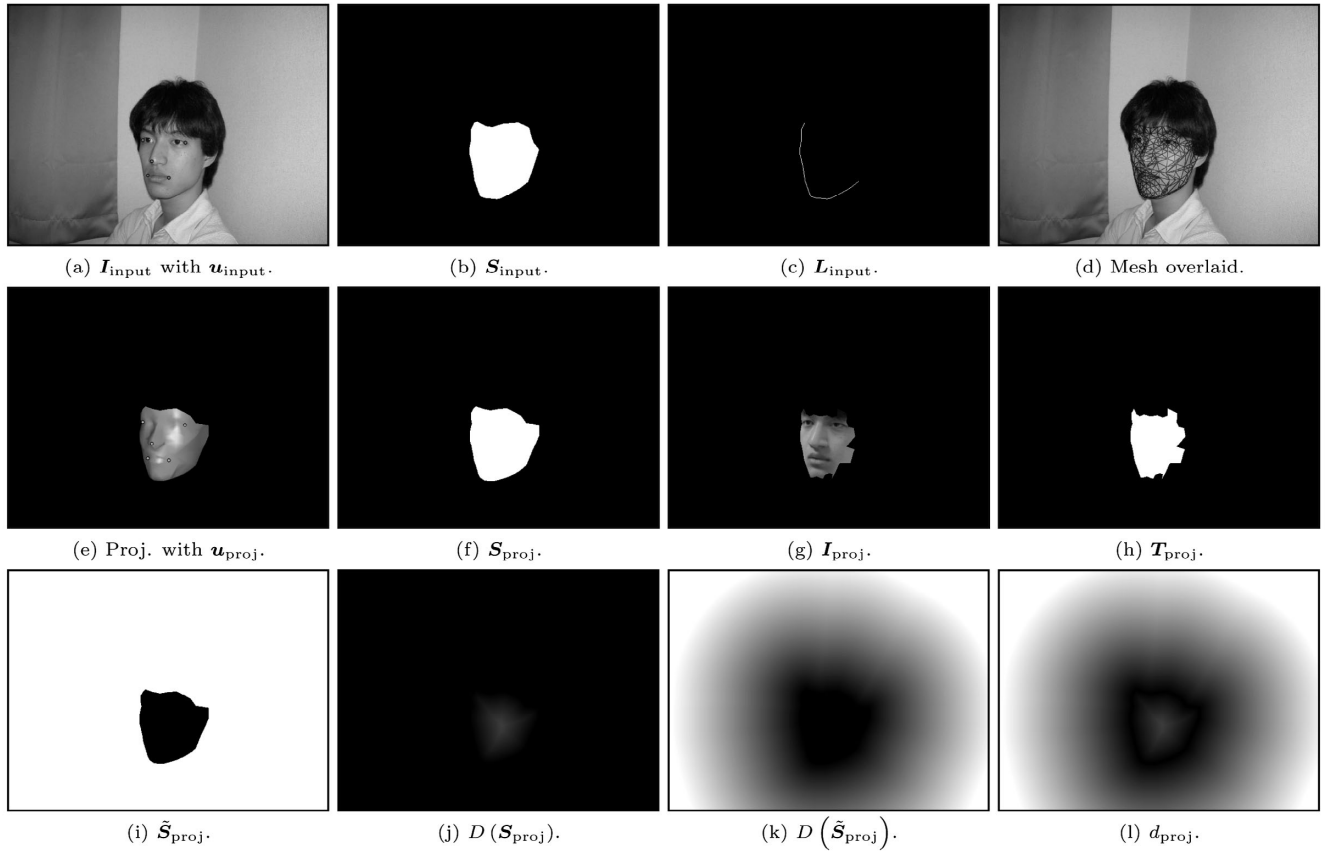


図 5 入力画像と復元結果
Input image and reconstructed result.

下に示す評価関数のうち、入力画像を用いる三つの評価関数 f_2, f_3, f_4 において、これに依存しないよう正規化を行っている。最終的には、この正規化された各入力画像に対する評価値を平均することで統合し、全入力画像の評価値とする。なお、具体的な正規化の手法については、以下に示す各評価関数の定義と共に述べる。

(1) 顔尤度評価関数

顔尤度評価関数は引数の形状の、顔らしさを評価する。具体的には、データベースに登録された形状集合との相関値を評価値とする。

この関数は引数の形状ベクトルを $\hat{s} = [\hat{s}_1, \hat{s}_2, \dots, \hat{s}_l]^T$ として、

$$f_1(\hat{s}) = 1 - \exp\left(-\frac{1}{2} \sum_{i=1}^l \frac{\hat{s}_i^2}{\lambda_i}\right) \quad (9)$$

と定義される。このとき、 λ_i ($1 \leq i \leq l$) は各主成分を表す固有ベクトル $\mathbf{p}_i$ に対応する固有値で、その主成分における学習データの分散を表す。

つまり、形状 $\hat{s}$ がデータベースに登録されている形状集合から離れると評価値が上がっていく関数である。ただしこのとき、人の顔の形状が平均形状を中心とした正規分布の確率密度で存在すると仮定している。

この関数によって、最適化中の形状が他の評価関数によって極端な変形を施され、顔らしくなくなることを防ぐことができる。

(2) 輪郭評価関数

輪郭評価関数は、引数の形状がどれだけ入力画像の輪郭に一致しているか、その相違度を評価する。

この評価関数を、

$$f_2(\hat{s}) = \frac{\iint_A d_{\text{proj}}^{(u,v)} dudv}{w \iint_A dudv} \quad (10)$$

と定義する。ただし、入力画像における外輪郭画像を $\mathbf{L}_{\text{input}}$ 、投影画像におけるシルエット画像を $\mathbf{S}_{\text{proj}}$ として、

$$A = \{(u, v) | \mathbf{L}_{\text{input}}^{(u,v)} = \text{white}\} \quad (11)$$

$$d_{\text{proj}}^{(u,v)} = D(\mathbf{S}_{\text{proj}})^{(u,v)} + D(\tilde{\mathbf{S}}_{\text{proj}})^{(u,v)} \quad (12)$$

$$w = \frac{\sqrt{f_x^2 + f_y^2}}{t_3} \quad (13)$$

である。ここで、 $\tilde{\mathbf{S}}$ は 2 値画像 $\mathbf{S}$ の明度反転画像を、 $D(\mathbf{S})$ は $\mathbf{S}$ のユークリッド距離変換画像をそれぞれ意味する。明度反転とは明度値をネガボジ反転させる操作を指し、白画素を黒画素に、黒画素を白画素にそれぞれ変換する。また、ユークリッド距離変換とは、各画素から最も近い黒画素までの距離を計算し、その距離をそれぞれ新たな明度値とする操作のことである。なお、ここで投影画像のシルエット $\mathbf{S}_{\text{proj}}$ は形状 $\hat{s}$ によって定まるため、 $\hat{s}$ の関数 $\mathbf{S}_{\text{proj}}(\hat{s})$ といえることに注意されたい。

それでは、この式 (10) の意味について順を追って説

明する。まず、基本的に評価を行っている部分は分子の $\iint_A d_{\text{proj}}^{(u,v)} dudv$ である。これは、入力画像の外輪郭上の各点から投影画像のシルエットまでの距離の総和を意味している。

これに対し、分母は正規化項である。 w はカメラパラメータから計算される値で、顔が画像としてどの程度の大きさで投影されているかを表現するものである。 $\iint_A dudv$ は入力画像の外輪郭 $\mathbf{L}_{\text{input}}$ の長さを表す。この二つの正規化項により、顔の画像上の大きさや外輪郭の長さに依存せずに輪郭の相違度を評価することが可能となる。

以上のようにこの関数を定義することにより、形状を入力画像の輪郭に沿うように変化させる。なお、単なる XOR などによる輪郭評価¹⁵⁾とは異なり、内輪郭が最適化されることによって、真に最適化されるべき外輪郭の最適化が行われないことを防ぐことができる。

(3) 特徴点評価関数

特徴点評価関数は入力画像と投影画像の特徴点の2次元座標の相違度を評価する。

入力画像、投影画像における特徴点座標をそれぞれ $\mathbf{u}_{\text{input}}, \mathbf{u}_{\text{proj}}$ とする。このとき、投影画像における特徴点座標 $\mathbf{u}_{\text{proj}}$ は形状 $\hat{s}$ によって定まるため、先程と同様に、 $\hat{s}$ の関数 $\mathbf{u}_{\text{proj}}(\hat{s})$ といえる。

さて、五つの特徴点の画像座標ベクトルをそれぞれ、 $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(5)}$ として、特徴点評価関数は

$$f_3(\hat{s}) = \frac{1}{5w} \sum_{i=1}^5 \left\| \mathbf{u}_{\text{input}}^{(i)} - \mathbf{u}_{\text{proj}}^{(i)} \right\| \quad (14)$$

と定義することができる。

つまり、各特徴点間のユークリッド距離の総和を評価値とする。分母は先程と同様な正規化項である。この関数により鼻の高さなどを的確に評価し、形状を変形する。

(4) テクスチャ評価関数

テクスチャ評価関数は、輪郭や特徴点だけでは評価しきれない細かい部分を評価するための関数である。

まず、射影行列 $\mathbf{P}$ を用いて、最も正面向きと判断された入力画像を形状 $\hat{s}$ にテクスチャとして貼り付ける。次に、このようにしてテクスチャマッピングされた形状を各入力画像の姿勢でそれぞれ投影する。

このとき、テクスチャ評価関数は入力画像およびテクスチャマッピングされた投影画像をそれぞれ $\mathbf{I}_{\text{input}}, \mathbf{I}_{\text{proj}}$ として、

$$f_4(\hat{s}) = \frac{\iint_B \left\| \mathbf{I}_{\text{input}}^{(u,v)} - \mathbf{I}_{\text{proj}}^{(u,v)} \right\| dudv}{\iint_B dudv} \quad (15)$$

と定義する。ただし、

$$B = \left\{ (u, v) \mid \mathbf{S}_{\text{input}}^{(u,v)} = \mathbf{T}_{\text{proj}}^{(u,v)} = \text{white} \right\} \quad (16)$$

である。また、 $\mathbf{T}_{\text{proj}}$ はテクスチャマッピングを行ったパッチかどうかを表すマスク画像である。

このとき、分子 $\iint_B \left\| \mathbf{I}_{\text{input}}^{(u,v)} - \mathbf{I}_{\text{proj}}^{(u,v)} \right\| dudv$ はテクスチャが貼れている比較可能な領域に対する、明度値の差の絶対値の総和 (SAD) である。つまり、入力画像と投影画像が画像としてどれだけ似ているかを評価する。分母 $\iint_B dudv$ は正規化項で、比較可能な領域の面積を表し、これで割ることで、顔の画像上の大きさに依らずにテクスチャの評価をすることができる。この評価関数により、輪郭や特徴点以外の細かい部分をも最適化し、形状を変形する。

3. 実験および結果

本節では、上記の提案手法を用いて行った実験、およびその結果を示す。3.1 においては、実際にフリーハンドで撮影された多視点のスナップショットから顔の3次元形状を復元する実験を行う。また3.2では、この復元結果を用いて、人物の顔を入力画像にはない新しい視点から見る事が可能となることを示す。さらに3.3においては、復元結果の応用例として顔の置換システムについて言及する。

3.1 フリーハンド多視点スナップショットからの復元

提案手法の有効性を示すため、実際にフリーハンドで撮影された多視点スナップショットから人物の顔の3次元形状を復元する実験を行った。入力画像として解像度 640×480 のものを4枚用いた。これらの画像はデジタルスチルカメラで簡易に撮影された外部未校正の画像である。

本実験においては、累積寄与率が90%を超えるパラメータ数として $l = 20$ 主成分を用いた。また、式(8)における復元に用いるパラメータを、 $\alpha_1 = 1.0$, $\alpha_2 = 6.0$, $\alpha_3 = 3.0$, $\alpha_4 = 0.1$ とした。これらの復元に用いるパラメータは2.5でも述べた通り、経験的に決定したものである。ここでいう経験的とは、以下のような手順を指す。

- i. 様々な入力画像に対して、評価値 $\alpha_i f_i(\hat{s})$ ($1 \leq i \leq 4$) がそれぞれほぼ同じ値を示すような α_i を求める。
- ii. 復元結果が入力画像に近づくように各パラメータ α_i をさらに変更して、最適なパラメータを決定する。

このときの入力画像および復元結果を図5に示す。ただし、ここでは4枚の入力画像のうち、1枚を代表として挙げる。図5(a)は特徴点画像座標を重畳表示した入力画像である。図5(b)および図5(c)はそれぞれ入力画像のシルエット画像および外輪郭画像である。

これらを入力とし形状を復元した結果が図5(e)である。この画像は、復元された形状を推定された入力画像と同じ姿勢で投影したものである。また、図5(f)は投影画像のシルエットを示し、図5(g)および図5(h)は、それぞれ、最も正面向きとされた入力画像をテクスチャとして貼り付けた投影画像、およびテクスチャが貼り付けられた領域を表すマスク画像を表す。図5(i)および図5(j)はそれぞれ投影画像のシルエットの明度反転画像とユークリッド距離変換画像である。図5(k)はこの明度反転画像をさらにユークリッド距離変換したものであり、これらは形状最適化の評価関

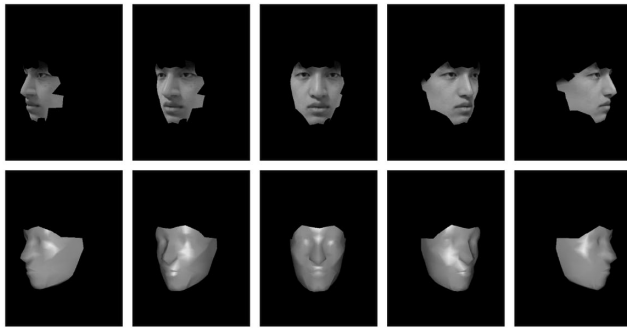


図 6 生成された自由視点画像
Generated free-view-point images.

数で使われる図 5(l) d_{proj} を計算するために使われる。

さらに、復元結果を入力画像にメッシュとして重畳したものが図 5(d) である。この結果を見るとわかるように、人物の輪郭に沿った姿勢および形状が復元されていることがわかる。また、眼鼻口などの特徴的な部分についてもメッシュがフィットしていることが確認できる。このとき、図 5(a) と図 5(g) を比較すると、そのアピアランスは PSNR (Peak Signal-to-Noise Ratio) で 24.4 dB であった。

なお、提案手法ではユーザがクリックした特徴点に基づいて姿勢推定が行われているが、ユーザのクリック誤差 (1–2 ピクセル程度) による影響はほとんどないことが経験的に確認されている。

また、本実験における計算時間は Windows XP SP3, Intel Core 2 Duo 6700 (2.66GHz), 3.5GB RAM 環境下において 10 分程度であった。ASM を用いた Blanz らの手法⁷⁾は、実験条件こそ異なるものの、50 分の計算時間がかかるとしている。

3.2 自由視点画像の生成

上記のように、入力画像の姿勢と形状を復元できると、復元形状に入力画像をテクスチャとして貼り付け、自由な視点から顔を見ることが可能となる。最も正面向きの入力画像をテクスチャとして用い、自由な視点で投影した画像の例を図 6 に示す。上段は生成した自由視点画像、下段はそのときの形状を示している。このように入力画像にはない向きから人物の顔を見ることができる。

特に中央の列のように真正面からの画像を作成することはとても有用である。多くの顔画像を用いた人物認証システムは、正面顔を入力として要求する。すなわち、入力画像が斜めのものしかない場合、このシステムには適用できないこととなる。しかし、このような場合でも、本手法を用いて顔の姿勢および形状を復元し、図 6 のように正面向きの画像を生成することで、このようなシステムに適用することが可能となる。

3.3 顔の置換システム

自由視点画像の生成が可能となると、さらに顔の置換を行うことが可能となる。まず最初に、入力画像および置換結果を図 7 に示す。図 7(a) は入力である被置換画像にその

特徴点座標を重畳したものである。この特徴点座標はユーザがクリックすることにより得られる。図 7(d) は各入力に対し、顔を置換した結果を示す。このとき、置換元画像は 3.1 の実験で復元を行った形状とテクスチャを用いている。

置換の流れとしては、まず被置換画像の特徴点座標 u_{input} および、置換元形状 $\hat{s}_{opt}$ より、2.4 で示した手法で被置換人物の顔の姿勢を推定する。このようにして推定された姿勢を用いてメッシュを重畳した結果を図 7(b) に示す。ただし、このメッシュは置換元人物の形状を投影したものであり、被置換人物の形状を復元しているわけではない。

次に、推定された姿勢を用いてテクスチャマッピングされた投影画像 I_{proj} を作成する。しかし、このようにして作成されたテクスチャマッピングを施された投影画像を、そのまま被置換画像に重ね合わせると、肌の色の違いや境界の不連続性により図 7(c) のように不自然な画像が得られてしまう。そこで、置換する領域における I_{proj} の明度値の平均および分散が被置換画像のものと等しくなるよう、RGB それぞれのチャンネルにおいて補正する。その後、アルファ値を徐々に変化させることで境界を滑らかに繋ぎ合わせ、違和感を取り除いている。

置換結果を見てみると、違和感なく顔が置き換えられていることが確認できる。ここで、提案手法においては被置換画像の特徴点座標のみを入力としており、色および境界の補正は全自動で行われている。ただし、光源の推定などは行っていないため、被置換画像が特殊な照明条件下にある場合には不自然さが生じる場合がある。

さて、このような顔の置換システムには様々な応用例が考えられている。例えば、従来は被写体のプライバシーを保護する方法としてモザイク処理などが一般的であったが、顔を置き換えることで、より自然にこれを行うことが可能となる。また、集合写真において、全員が目を開けている、あるいは笑顔であるといった画像を作成するために有用である。つまり、目をつぶってしまった人の顔を、その本人が目を開けている別の写真を用いて置き換えるといった利用方法である。

しかし、Bitouk らの置換手法¹³⁾においては、顔の形状を復元せずにこれが行われていた。そのため、置換元画像と被置換画像との顔の姿勢が類似していることが前提として要求されていた。多少の姿勢の違いがある場合には、2 次元画像処理的に補正を行っているが、形状を復元していないため、幾何学的に正確でないという問題があった。そこで顔の姿勢に依存しない顔の置換手法が必要とされていたが、上記のように提案手法においては、置換に際して置換元の人物の顔の形状を復元してから被置換画像の姿勢を推定して投影し重畳する方法を取っている。よって、図 7 のように顔の姿勢が異なる場合でも置換を行うことが可能であり、この問題を解決するための一手法となり得る。



図 7 顔の置換
Face swapping.

4. 検 討

本節では、復元結果の精度を定量的に評価し、検討する。4.1 では、形状が既知の人物の画像を入力とし、提案手法によりその形状を復元する。そして、この正解形状と復元形状との誤差を評価する。4.2 では復元形状における誤差が顔面上でどのように分布しているかについて示し、提案手法の利点および欠点について検討する。

4.1 レンジデータとの比較

復元結果の精度を定量的に評価するため、あらかじめレンジスキャナで形状が測定された、形状既知な人物の画像を元に復元を行う実験を行った。対象となる人物は男性三人、女性二人の計五人であり、それぞれ各性別で構成されるデータベースを用いた。なお、実験条件は3.1の場合と同一である。

レンジスキャナで測定された形状を正解形状とし、正解形状とこのときの復元形状との誤差を計算した。その結果を表1に示す。なお、ここでいう誤差とは、復元形状上の各頂点から正解形状の表面までのユークリッド距離を全頂点において平均した値であり、単位はmmである。

また、一般的には復元形状と正解形状は異なる座標系で表現されるため、ICP (Iterative Closest Point) アルゴリズムなどの手法を用いて、各形状の位置合わせを行う必要があるが、本手法を用いた復元形状は正解形状と同じ耳眼水平座標系で表現されるため、このような位置合わせは行っていない。

さて、表1で各列は実験を行った五人の人物を表す。また、上段は正解形状と復元形状との誤差を意味している。中段は正解形状とデータベース上の各形状との誤差のうち、最小の値であり、比較のために示してある。つまり、正解形状と復元される人物に最も近い人物の形状との誤差である。一方、下段は正解形状とデータベース上の各形状との誤差を、登録された全形状において平均した値を示している。すなわち、これは正解形状が一般的な形状に対してどの程度の誤差を示すかを表している。

表1からわかるように、本手法の復元形状は正解形状に対して平均3.5mmの復元誤差を持つ。最も正解形状に近いデータベース上の形状と比較すると誤差が大きくなってしまっている人物が多いものの、すべての人物において、復元形状は一般的な形状と比べて、正解形状に近づいていることがわかる。

さらに、このときの復元誤差には主成分分析によるパラメータ数圧縮の際の再投影誤差も含まれている。この再投影誤差がどの程度の値を占めるかを検討した。まず、各性別におけるデータベースにおいて、各形状 $\mathbf{x}_i$ ($0 \leq i \leq m$) を固有空間に投影し、 $l = 20$ 主成分を用いて再投影した予測値 $\hat{\mathbf{x}}$ を作成する。次に $\mathbf{x}_i$ を正解形状として、予測値 $\hat{\mathbf{x}}$ との誤差を計算した。すると、男性では平均1.3mm、女性では平均1.1mmであった。よって、本手法の復元誤差3.5mmに対して、再投影誤差がこれだけ含まれるということが確認された。

表 1 形状の評価
Shape evaluation.

Person ID	Male 1	Male 2	Male 3	Female 1	Female 2
Reconstructed	3.1	3.3	3.2	3.9	3.9
Min Database	3.5	3.2	2.9	3.0	2.7
Mean Database	4.9	5.0	4.3	4.6	4.7

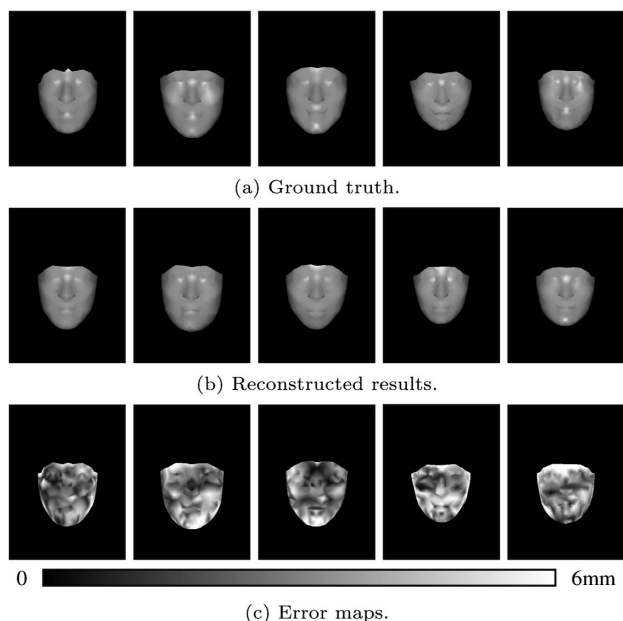


図 8 復元結果と誤差マップ
Reconstructed results and error maps.

4.2 誤差の分布

次に、4.1 で復元した形状の誤差が顔面上でどのように分布しているかを検討する。この五人の形状と誤差の分布を図 8 に示す。この図において、各列は実験を行った各人物に対応する。また、図 8(a) は各人物の正解形状、図 8(b) は復元形状、図 8(c) は誤差マップを表す。誤差マップは明るい部分ほど誤差が大きいということを意味している。

この図を見ると、顔の中心部分は誤差が小さく、端にいくほど誤差が大きくなっていることがわかる。これは、復元に用いた輪郭評価関数が外輪郭のみに依存しているためである。顔の端の部分は内輪郭として入力されることが多いため、外輪郭による輪郭評価では、この部分における相違度を評価することができない。そのため、端にいくほど誤差が大きくなってしまったものと考えられる。

5. む す び

本稿では、フリーハンドで多視点から撮影された人物のスナップショットから顔の 3 次元形状を復元する手法を提案した。本手法は特殊なデバイスが必要とせず、市販のデジタルスチルカメラで簡易に撮影された画像を入力とすることができる。一般に、従来は形状復元にはキャリブレーションされたカメラシステムが必要とされ、カメラと顔の位置関係が既知でなければならないという問題があった。しかし、本手法では顔の姿勢推定と形状最適化の処理を繰り返すことにより、この問題を解決し、実際に復元を行えるこ

とを示した。

また、本手法による復元結果を用いると、正面向き画像や顔の置換画像といった新しい画像を作成できることを示した。さらに、復元形状の精度を定量的に評価した結果、レンジスキャナで測定された真の形状に対して、平均 3.5mm の誤差で復元が可能であることが確認された。

しかし現段階では、ユーザに顔の特徴点座標、シルエット画像、外輪郭画像といった入力を要求する。ユーザの負担を減らすためには、これらを入力画像から自動的に抽出することが望ましい。また、外輪郭による評価では顔の端の部分に比較的大きな復元誤差が生まれてしまうことが確認された。評価関数を改善し、このような誤差を削減することが今後の課題となろう。

〔文 献〕

- 1) M. A. Brunsmann, H. A. M. Daanen and K. M. Robinette: "Optimal postures and positioning for human body scanning", Proceedings of International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM '97), pp. 266-273 (1997)
- 2) L. Zhang, N. Snavely, B. Curless and S. M. Seitz: "Spacetime faces: High Resolution Capture for Modeling and Animation", ACM Transaction on Graphics, 23, 3, pp. 548-558 (2004)
- 3) J. P. Siebert, S. J. Marshall: "Human body 3D imaging by speckle texture projection photogrammetry", Sensor Review, 20, 3, pp. 218-226 (2000)
- 4) A. K. R. Chowdhury and R. Chellappa: "Face reconstruction from monocular video using uncertainty analysis and a generic model", Computer Vision and Image Understanding, 91, Issues. 1-2, pp. 188-213 (2003)
- 5) V. Blanz and T. Vetter: "A Morphable Model for the Synthesis of 3D Faces", Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH '99), pp.187-194 (1999)
- 6) S. Romdhani, V. Blanz and T. Vetter: "Face Identification by Fitting a 3D Morphable Model Using Linear Shape and Texture Error Functions", Proceedings of the Seventh European Conference on Computer Vision (ECCV 2002), 4, pp. 3-19 (2002)
- 7) V. Blanz and T. Vetter: "Face recognition based on fitting a 3D morphable model", IEEE Transactions on Pattern Analysis and Machine Intelligence, 25, 9 (2003)
- 8) Y. Hu, D. Jiang, S. Yan, L. Zhang and H. Zhang: "Automatic 3D Reconstruction for Face Recognition", Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04), pp. 843-848 (2004)
- 9) D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang and W. Gao: "Efficient 3D reconstruction for face recognition", Pattern Recognition, 38, Issue. 6, pp. 787-798 (2005)
- 10) B. Amberg, A. Blake, A. Fitzgibbon, S. Romdhani and T. Vetter: "Reconstructing High Quality Face-Surfaces using Model Based Stereo", Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV 2007) (2007)
- 11) 竹内敏夫, 斎藤英雄, 持丸正明: "非同期多視点カメラからの顔の 3 次元復元", 第 10 回 画像の認識・理解シンポジウム (MIRU2007) 予稿集, pp. 1289-1294 (2007)
- 12) M. Mochimaru, M. Kouchi: "Proper Sizing of Spectacle Frames based on 3D Digital Faces", XVth Triennial Congress International Ergonomics Association (IEA2003) (2003)
- 13) D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur and S. K. Nayar: "Face Swapping: Automatically Replacing Faces in Photographs", ACM Transactions on Graphics, 27, 3, Article 39 (2008)
- 14) Z. Zhang: "A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 11, pp. 1330-1334 (2000)
- 15) B. Moghaddam, J. Lee, H. Pfister and R. Machiraju: "Model-Based 3D Face Capture with Shape-from-Silhouettes", Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03) (2003)



すずき せいじ
鈴木 誠司 2008 年，慶應義塾大学理工学部情報
工学科卒業。現在，慶應義塾大学大学院理工学研究科前
期博士課程に在学中。コンピュータビジョン，特に 3 次
元形状復元に関する研究に従事。



さいとう ひでお
斎藤 英雄 1992 年，慶應義塾大学大学院博士課
程修了。同年，同大学助手，1995 年，専任講師，2001
年助教授，2006 年，情報工学科教授。コンピュータビ
ジョン，画像処理に関する研究に従事。1997～1999 年，
日本学術振興会海外特別研究員として米国カーネギーメ
ロン大学ロボティクス研究所へ留学。2000～2003 年科
学技術振興事業団さがけ 21『情報と知』研究員。博士
(工学)。正会員。



もちまる まさあき
持丸 正明 1993 年，慶應義塾大学大学院博士課
程修了。博士（工学）。同年，工技院生命工学工業技術
研究所入所。2001 年，改組により，産業技術総合研究所
デジタルヒューマン研究ラボ。2003 年より，現職。専門
はバイオメカニクス，人間工学。人体の形状，運動の計
測とモデル化，産業応用の研究に従事。
