

三次元顔形状データベースを用いた無表情顔形状からの表情変化推定

倉立尚明

ATR 人間情報科学研究所

E-mail: kuratate@atr.jp

無表情の三次元顔モデルに対してその人物の顔らしい表情変化を与えた新たな顔形状を生成するためには、熟練したアニメータによる形状変形や、筋肉モデルなどのような高度なモデルを用いる必要がある。本研究では容易に同様の表情変化形状を得るため、複数表情を含む三次元顔形状データベースを利用し、データベースから得た主成分分析を用いることにより、未知の顔形状に対して容易にデータベースに含まれる表情変形を推定する手法を紹介する。推定顔表情を利用すれば、更に顔アニメーションの生成が可能である。

Estimation of Face Expressions from Neutral 3D Faces Using 3D Face Database

TAKAKI KURATATE

ATR International, Human Information Science Laboratories

E-mail: kuratate@atr.jp

Creating various realistic facial expressions from neutral 3D face models requires careful deforming operations by skilled animators or sophisticated computational models such as muscle or performance driven models. In this paper, we demonstrate a statistical approach for estimating facial expressions using a 3D face database. The database has 9 different face postures from over one hundred people and is analyzed by principal component analysis. Any new neutral 3D face can be easily represented with these principal components and its deformation characteristics estimated from faces in the database that are close to the input face in the parameterized space. The estimated facial expressions are quite natural, and the same method can also be applied to non-human faces like statues or dolls. Additional principal component analysis of the estimated faces can then be used to create 3D face animation.

1 はじめに

近年、個人レベルでのコンピューターアニメーションの生成が盛んとなってきている。中でも人物のリアルな CG アニメーションの利用は顕著であり、特に映画や TV 画面中ではフル CG によるキャラクターが実際の俳優とともに現れることはごく普通のこととなりつつある。このように、実際の人物と違和感なく存在したり、またより自然なインターフェースを目的とした人物アニメーションを生成するためには、特に発話においてはより実物の人物に近い発話顔面運動を生成することが重要となる。このためには、任意の CG 顔モデルに対してそのモデルらしい顔の動きを生成する技術が必要となってくる。逆に言えば、いくら精巧な顔モデルが得られたとしても、その人物として不自然な表情しか生成できないようであれば、アニメーションとしては不完全であると言える。

この問題を解決するための手段として、熟練した

アニメータにより表情を生成することが考えられるが、あくまでアニメータ個人の主観的な結果であるために製作者によっては結果が異なることが考えられ、本当にその人物らしい表情であるかは不確定であり、また人手による作業では大量の人物データへの対応も困難である。

また筋肉モデルのようなより高度な計算機モデルを定義して表情生成を行うことが考えられる [1, 2, 3, 4, 5]。しかし、対象が異なるたびに筋肉構造を定義する必要があり、また筋肉構造の定義方法の違いによる生成結果の違いや、モデルそのものが全ての人物に適用できるとは限らない。

対象人物の自然な顔形状を再現する点から、特徴的な表情をあらかじめレーザーレンジファインダなどの三次元計測装置を用いて計測することも考えられるが、実際に表情を生成できる人物に限定されるため、彫刻や人形、任意の CG キャラクターなどの全ての顔形状についてはさまざまな表情を計測することは困難である。

本論文ではより統計学的なアプローチとして、予め複数の人物から特定の表情群を計測した三次顔形状データベースを利用し、未知の人物顔あるいは顔形状物体の三次元形状に対し、データベースに含まれる似通った人物の表情変形を適用することにより、これら未知顔の表情を推定する手法について解説する。さらに本手法により得られた推定顔形状に対して主成分分析を行い、Facial Motion Mappingを適用することにより、アニメーション生成も可能である [6, 7]。

2 三次顔形状データベース

我々は人間の発話の生成知覚機構について、視覚情報が与える影響について調査を行っており、これまでに正確な顔面運動や頭部運動が音声情報と高い相関があり、発話の知覚に対しても強い影響があることを示してきた [8]。しかしあくまで少数被験者から得た計測結果のみを対象としているため、より多くの被験者から音声、顔面および頭部運動、顔形状など様々なデータを取得する必要がある。

この研究調査の一環として、より多くの被験者から発話に関する運動データと形状データの取得を試みている。そのなかで、三次元顔形状の特徴を取得し、被験者間での発話に関する顔形状変形の共通性や個人性、性別、民族性、風土性などのパラメータ化やこれらの違いの定量化を行うことを目的として、三次元顔形状データベースを構築している。

これまでに発話顔形状の分析を主目的として、九種類の顔表情を三百名以上の被験者(男女、日本人、非日本人等を含む)から計測している。計測した表情は無表情、口を開けた笑顔、口を閉じた笑顔、母音/i/, /o/, /u/, 口を大きく開ける、歯を噛み締めて唇を開ける、そして唇の突き出しの合計9形状で、これまでに二百名近くが一般化メッシュを適合データベースに登録され、様々な分析が行えるようになっている [9]。

2.1 データベース化と主成分分析

計測された三次元形状データは、顔の特徴などとは無関係なサンプリングデータである。このため三次元形状データを様々な手法で容易に解析できるよう、同じデータ構造へと変換を行う。この変換には、目・眉・鼻・唇・顎などの基本的な特徴線を抽出し、それらを元にあらかじめ作成した一般化メッシュを個々の計測形状へと適合させている [10, 11]。唇領域については得られるデータの解像度や計測ノイズの影響などの問題により、現段階では本データベースには輪郭情報のみを保持し、ポリゴン構造は持っていない。しかし唇領域は発話の際の視覚情報提示の要素として欠くことができないため、別途口腔内部を含む詳細なモデルを構築しており、将来的には本

データベースの一部として導入する予定である。また眼球、耳などの領域とテクスチャ画像については現在は分析の対象とはしていないが、将来的にはこれも分析を行う予定である。個々の被験者についてこれらの適合化作業が終了すればこれをデータベースに取り入れ、様々な分析・合成へと応用が可能である。ここでは以下に主成分分析を適用する手法について解説する。

ノード数 N の一般化メッシュを適合させることにより、全ての顔形状はその形状を構成する N 個の三次元座標値を用いて以下のように列ベクトルで定義することができる。

$$\mathbf{f} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N \ \mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_N \ \mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_N]^T \quad (1)$$

このベクトル化された顔形状を用い、全ての顔形状や分析を目的とするサブグループ(特定の表情や人種、性別、年齢など)に対して、その集合を定義するインデックス番号 $g(1), g(2), \dots, g(K)$ により次のグループベクトル \mathbf{F} が定義できる。

$$\mathbf{F} = [\mathbf{f}_{g(1)}, \mathbf{f}_{g(2)}, \dots, \mathbf{f}_{g(K)}] \quad (2)$$

ここで、このグループにおける任意の顔形状 \mathbf{f}_i をグループ内の K 個のサンプル形状の平均 $\boldsymbol{\mu}_f$ からの差分 \mathbf{f}_{m_i} により定義する。

$$\mathbf{f}_i = \mathbf{f}_{m_i} + \boldsymbol{\mu}_f, \quad \boldsymbol{\mu}_f = \frac{1}{K} \sum \mathbf{f}_k \quad (3)$$

これより新たにこのグループにおける基本顔形状ベクトル \mathbf{F}_M が定義できる。

$$\mathbf{F}_M = [\mathbf{f}_{m_1}, \mathbf{f}_{m_2}, \dots, \mathbf{f}_{m_K}] \quad (4)$$

この \mathbf{F}_M について、共分散行列 \mathbf{C}_f は以下のように表すことができる。

$$\mathbf{C}_f = \mathbf{F}_M \mathbf{F}_M^T \quad (5)$$

この共分散行列 \mathbf{C}_f に対して次の特異値分解(singular value decomposition)を行うことにより、各固有ベクトルを \mathbf{F}_M の線形独立な主成分として求めることができる。

$$\mathbf{C}_f = \mathbf{U} \mathbf{S} \mathbf{U}^T. \quad (6)$$

ここで \mathbf{U} は正規化されたユニタリ行列で、各列が固有ベクトルすなわち主成分を表す。また \mathbf{S} は対角行列で、対角成分が固有値を表すものである。

この主成分 \mathbf{U} を用いることにより、このグループにおける差分顔形状は \mathbf{f}_{m_i} はそれらの線形結合として表現することができる。この線形結合の係数は、以下の式によりベクトル $\boldsymbol{\alpha}_i$ として求めることができる。

$$\boldsymbol{\alpha}_i = \mathbf{U}^T \mathbf{f}_{m_i} \quad (7)$$

以上より、基本顔形状ベクトル \mathbf{F}_M を主成分に分解した係数ベクトル \mathbf{A} を得ることができる。

$$\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_K] \quad (8)$$

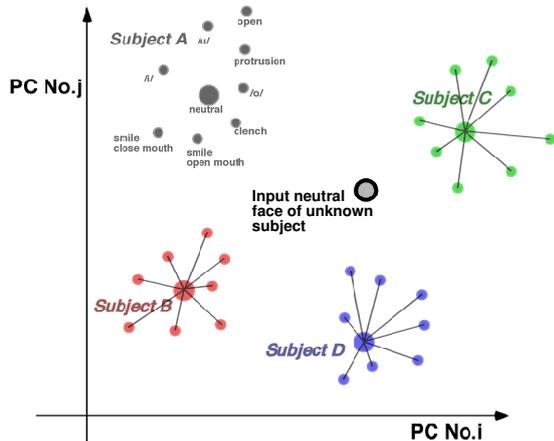


図 1: 主成分空間における被検者ごとの表情分布の概念図

本報告では、アジア系:男性 69 名/女性 36 名、白人系:男性 45 名/女性 34 名、黒人系:男性 2 名/女性 1 名を含む 187 名 x 9 表情 = 1683 顔形状が登録された状態のデータベース全体について主成分分析を行った結果を用いている。またこれまでの予備実験では、上位 30 から 50 主成分程度あれば主観的にもとの顔形状に近い形状を再合成することかできることがわかっている [9]。つまり、30-50 次元の主成分空間を用いることにより、主観的に個々の顔表情を分離した処理を行うことが可能である。

すなわち、上位 k 個の主成分 U_k を用いて主成分分析に用いた m 人の人物から計測した s 種類の顔形状の合計 $M = m \times s$ 個の顔形状について、ここで、 i 番目の顔の線形結合係数 α_i は個々の主成分に対する合成係数 $a_{ij} (j = 1..k)$ から構成される列ベクトルとして表現できる。

$$\alpha_i = [a_{i1} \ a_{i2} \ \dots \ a_{ik}]^T \quad (9)$$

3 未知無表情形状入力と類似顔選定

次に、データベースには存在しない未知人物の無表情形状あるいは顔状物体の形状データがデータベースと同一のメッシュ構造で得られたとする。

このとき、入力顔は式 (1) と同様に列ベクトル f_u で表現でき、先と同様に平均からのずれ f_{m_u} へと分離できる。

$$f_{m_u} = f_u - \mu_f \quad (10)$$

この平均からのずれを、上位 k 個の主成分 U_k の線形結合で表現することにより、入力顔をデータベース内の顔同様に上位 k 個の主成分空間に定義することができる。ここで、この未登録人物の単一の顔表情形状に対応する線形結合係数 β は以下の式で求めることができる。

$$\beta = U_k^T (f_u - \mu_f) = [b_1 \ b_2 \ \dots \ b_k]^T \quad (11)$$

これにより、式 (9) で表されるデータベースに登録されている無表情顔と β について、 k 次元の主成分空間において距離を比較することにより容易に類似する顔を検索することができる。

4 表情推定

主成分空間におけるデータベース中の人物の顔表情は、無表情を基準にその周辺に多くの顔表情が分散して存在しており、人物が異ると無表情の位置および表情の分布がそれぞれ異って現れる。 k 次元中の 2 次元についてこれを模式的に示すと図 1 のようになる。この主成分空間中で未知人物の無表情顔が式 (11) により定義できるので、この近傍の人物の表情の分布を利用して、目標となる人物の表情を推定することが可能である。

まず、未知人物の無表情とデータベース中の人物の無表情との距離を小さい順に並べることにより、入力顔に近い順に顔表情形状を持つ人物のリストを作成することができ、このリストに従い、似通った一人もしくは複数の人物の無表情から他の表情への主成分空間における変化ベクトルを重み付け合成することにより、入力顔における表情変化を推定することができる。本報告では、上位 5 名の人物に対し、距離の小さい順に 0.50, 0.25, 0.10, 0.10, 0.05 の割合で重み付け合成を行うこととした。

この重み付けに加え、目標人物に似た無表情を持つ人物が n 人選定されたことにより、個々の選定人物の顔表情変形ベクトルを算出しておく必要がある。この顔表情変形ベクトルとは、個々の人物の無表情から、その人物の他の表情への変化を主成分空間のベクトルとして表現したものである。

選定人物一名につき、 $k \times (s - 1)$ の表情変形ベクトル ϵ_i が得られる。

$$\epsilon_i = [\alpha_{p_1} - \alpha_q \ \dots \ \alpha_{p_{s-1}} - \alpha_q] \quad (12)$$

ここで、 q は大規模三次元顔形状データベース内における、選定人物の無表情のインデックス番号、 $p_1 \dots p_{s-1}$ は選定人物の無表情以外のインデックス番号を示す。

この ϵ_i を選定人物 n 人全てについて求め、その人物に対する重み付け係数 w_i より、入力人物における顔表情変形ベクトル ϵ_0 は以下の式により求めることができる。

$$\epsilon_0 = \sum_{i=1}^n w_i \epsilon_i \quad (13)$$

この顔表情変形ベクトル ϵ_0 の各列に対し、入力無表情顔の線形結合係数 β を加えることにより、列ご

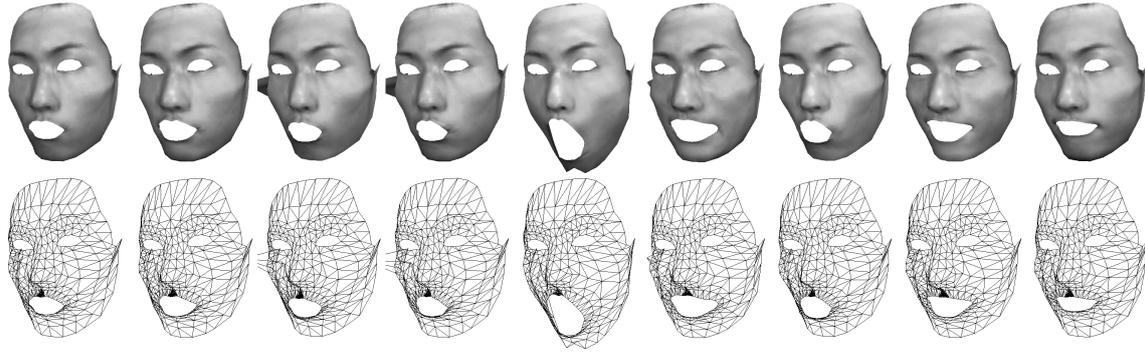


図2: 実人物の計測形状に一般化メッシュを適合した結果 ; ; 左より順に無表情、/i/、/o/、/u/、口を大きく開ける、歯を噛み締めて唇を開ける、唇の突き出し、口を開けた笑顔、口を閉じた笑顔

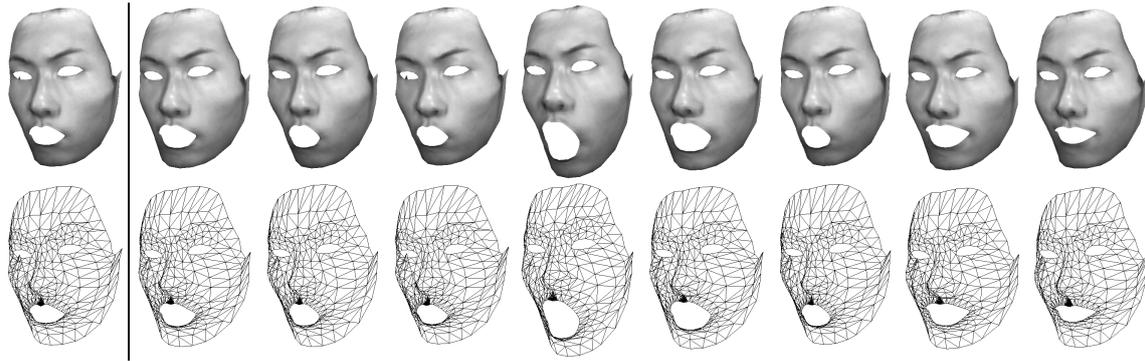


図3: 図1の被検者の無表情(左端)より、データベースを利用して推定した各表情

とに個々の表情に対応する線形結合係数 β_i を求めることができる。

$$\beta_i = [b_1 + e_{1i} \ b_2 + e_{2i} \ \dots \ b_k + e_{ki}]^T \quad (14)$$

この β_i より、データベースから得た平均顔形状 μ_f と主成分 k を用い、入力人物の推定表情の三次元形状を以下の式により求めることができる。

$$f_{iu} = U_k \beta_i + \mu_f \quad (15)$$

推定可能な表情は、もともとのデータベースに含まれる表情に限定はされるが、推定された表情からに対してさらに主成分分析をかけることでその人物の顔の基本的な変形成分を抽出でき、アニメーションへの応用が可能である。

5 表情推定実験

表情推定の検証のため、まず表情データを計測してはいるがデータベースには登録されていない人物について実験を行った。図2が実際に計測された日本人男性被検者の九種の顔形状に対して一般化メッシュを適合した結果で、左から順に無表情、母音/i/、/o/、/u/、

口を大きく開ける、歯を噛み締めて唇を開ける、唇の突き出し、口を開けた笑顔、口を閉じた笑顔である。一部の表情で計測時に右耳に大きなノイズが乗っている。図3が、左端の無表情顔をもとに本手法により他の8表情を推定した例である。テクスチャは無表情のものをそのまま他の表情に利用している。口を大きく開ける形状では開け具合に大きな差があり、また歯を噛み締めて唇を開ける形状でも微妙な差はあるものの、それぞれ対応した表情が推定できていることがわかる。現在利用している一般化メッシュでは唇と目については含めていないため、この例ではこれらの部位が欠けた状態となっている。

次に、無表情顔のみ得られた人物のデータと、表情を作り出すことが出来ない彫刻および人形について三次元計測を行ったデータについて本手法を適用した。これらのデータについては予めデータベース登録時と同様に特徴線を元に一般化メッシュを適合し、その適合結果に対して表情推定を行う。図4が一般化メッシュを適合する前の元の形状で、左から順に実人物、ヴィーナス像、ダミー人形、ヨーダ人形である。図5左端がこれらに対して一般化メッシュを適合した結果の基準となる無表情顔で、右に向って先と同様の表情生成を行った例である。この例では、目のポリゴンおよび別途唇輪郭線などから生成

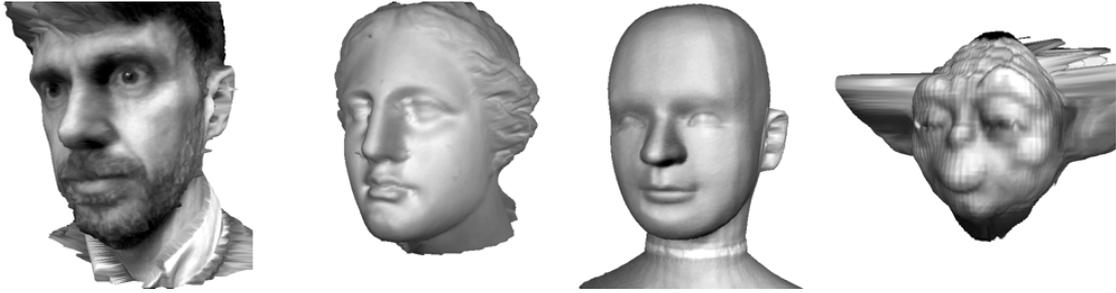


図 4: 表情推定を行った対象データ：左より順に実人物、ヴィーナス像、ダミー人形、ヨーダ人形

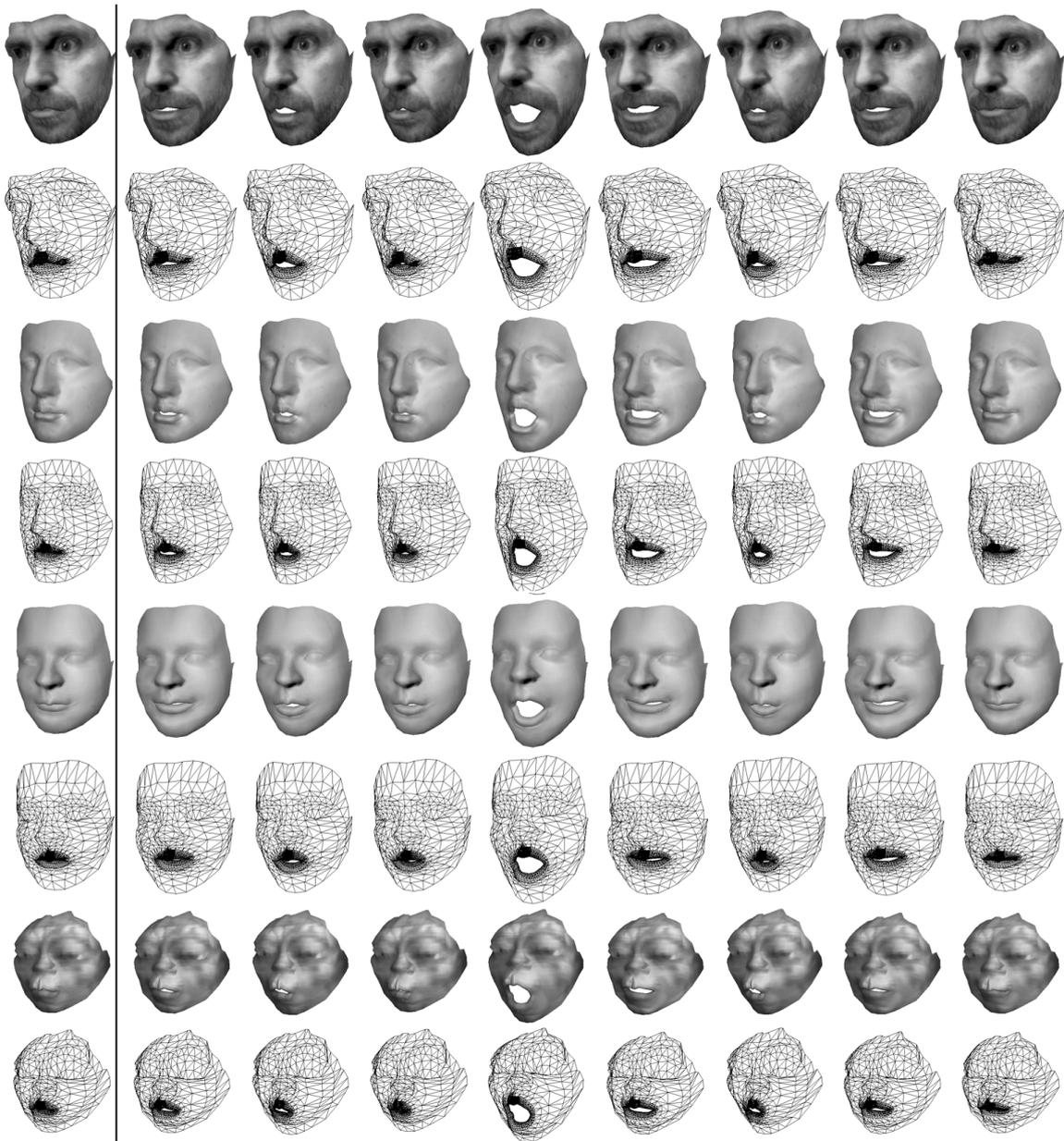


図 5: 無表情 (左端) より推定した表情 ; /i/, /o/, /u/, 口を大きく開ける、歯を噛み締めて唇を開ける、唇の突き出し、口を開けた笑顔、口を閉じた笑顔

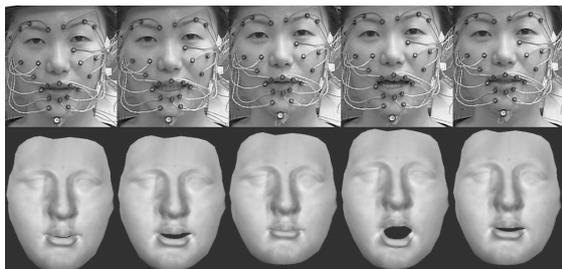


図 6: ヴィーナス像の推定表情結果から得た主成分を利用してアニメーションを生成した例

される唇モデルを付加してあるが、いずれも推定された顔形状のみから自動的に生成されたものである。このように、人形であっても自然な表情が推定可能であることがわかる。

以上のように、本表情推定を用いれば無表情のみから違和感の無い表情を生成することができる。さらに得られた推定表情に対して対象人物や人形ごとに主成分分析を行うと、本手法の特性上、上位二主成分として顎の上下運動と唇の丸め／広がりというデータベース内の人物と同様の傾向が得られる [9]。この主成分の類似性を用いた Facial Motion Mapping により [6, 7]、他の人物の動きを容易にマッピングすることが可能である。図 6 に、ヴィーナス像を用いてアニメーションを生成した例を示す。

6 おわりに

本報告では、三次元顔形状データベースの主成分分析結果を利用し、データベースには含まれていない人物や彫刻や人形などの顔状物体の無表情三次元顔形状からその人物らしい表情変形を推定する手法を示した。推定される表情はデータベースに含まれる表情に限定されるが、得られた表情か主成分を抽出することにより、顔アニメーション生成への応用が可能である。今回の報告では無表情顔の三次元データが存在する人物や人形などを例に推定を行ったが、本データベースの応用の一つでもある写真を用いて三次元顔形状を推定する技術と組合せることにより、写真のみから三次元顔表情を推定することが可能である。今後は生成された顔形状の定量的な評価と主観評価により推定形状の評価、またテクスチャを含めたデータベースの分析と本手法への応用についても検討する予定である。

7 謝辞

本研究は通信・放送機構の研究委託「人間情報コミュニケーションの研究開発」により実施したものである。

参考文献

- [1] Keith Waters. A muscle model for animating three-dimensional facial expression. *Computer Graphics*, 22:17–24, 1987.
- [2] Prem Kalra, Angelo Mangili, Nadia Magnenat-Thalmann, and Daniel Thalmann. Smile: A multilayered facial animation system. *Proc. IFIP WG*, pages 189–198, 1991.
- [3] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Realistic modeling for facial animation. *Proceedings of SIGGRAPH95*, pages 55–62, 1995.
- [4] 森島繁生, 世良元, and Demetri Terzopoulos. 物理法則に基づいた筋肉モデルによる口唇形状の制御. *NICOGRAPH 論文集*, pages 219–229, 1996.
- [5] K. Hahler, J. Haber, and H.P. Seidel. Geometry-based muscle modeling for facial animation. *Graphics Interface 2001*, pages 37–46, 2001.
- [6] 倉立尚明 Eric Vatikiotis-Bateson. Facial motion mapping による発話顔アニメーションの分析生成. *画像電子学会論文誌*, 32(4):355–368, 2003.
- [7] Takaaki Kuratate, Eric Vatikiotis-Bateson, and Hani Yehia. Cross-subject face animation driven by facial motion mapping. *Proceedings of CE2003: Advanced Design, Production and Management Systems*, pages 971–979, 2003.
- [8] Hani C. Yehia, Philip E. Rubin, and Eric Vatikiotis-Bateson. Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26:23–44, 1998.
- [9] 倉立尚明, Guillaume Vignali, and Eric Vatikiotis-Bateson. 大規模三次元顔形状データベースと cg アニメーション応用. *Visual Computing グラフィックスと CAD 合同シンポジウム*, pages 105–110, 2003.
- [10] 倉立尚明, Hani Yehia, and Eric Vatikiotis-Bateson. 主成分の線形結合による音声同期顔アニメーション. *Visual Computing グラフィックスと CAD 合同シンポジウム*, pages 115–120, 1998.
- [11] Takaaki Kuratate, Hani Yehia, and Eric Vatikiotis-Bateson. Kinematics-based synthesis of realistic talking faces. *AVSP'98*, pages 185–190, 1998.