

少数の制御点により直観的操作が可能な 3次元顔形状変化モデルの構築

山本 拓人^{1,a)} 船富 卓哉^{1,b)} 飯山 将晃^{1,c)} 美濃 導彦^{1,d)}

概要: 人の顔は表情の変化に伴い、形状を複雑に変化させる。本研究では、容易な操作によって様々な表情に制御可能な、実在の人物の3次元顔形状変化モデルを構築することを目指す。これまで提案されてきたモデルでは、多数かつ直観的でない制御点によって変形操作を実現していたため、変形操作が容易でなかった。これに対し、モデルを制御する制御点をできるだけ少なくし、また Action Unit に基づいて制御点を定義すれば、直観的で容易な操作が可能になる。提案手法では、ワンショットスキャン法によってさまざまな表情における顔の3次元形状を獲得し、これを用いて個人差を反映した形状変化モデルを構築する。実験により3次元顔形状変化モデルの精度について評価を行い、3.4mm程度の精度で顔形状を再現可能なことを確認した。

キーワード: 3次元形状モデル, 顔形状, Facial Action Coding System

YAMAMOTO TAKUTO^{1,a)} FUNATOMI TAKUYA^{1,b)} IIYAMA MASAOKI^{1,c)} MINOH MICHIIHIKO^{1,d)}

1. はじめに

人物の顔はその表情の変化に伴い、複雑で立体的な形状変化が起こる。こうした複雑な形状変化は、一見そのモデル化が困難なように思われるが、実際には前頭筋や眼輪筋、頬筋などといった、いくつかの表情筋によって引き起こされている。この事実を、顔の3次元形状変化をいくつかのパラメータによってモデル化可能であることを示唆している。

このパラメータを、なるべく少数で、かつパラメータ変化が形状変化と直観的に対応するよう定義できれば、顔の3次元形状をモデルに従って人手で容易に変形操作できる。表情の変化に伴う顔形状の変化を直観的に記述した方法として、Facial Action Coding System (FACS) [1] と呼ばれるものが提案されている。[1] では、観察に基づいて Action Unit (AU) と呼ばれる単位形状変化を定義した。FACS は、こうした AU の組み合わせにより、全ての顔形状変化を表現する方法である。AU は主観的な観察に基づ

いて、直観的に分かりやすい名前によって定義されており、これに対応するようにパラメータを定義すれば、顔形状変化との対応が直観的に理解しやすいものとなる。本研究では、顔表面の点である制御点を、AU に対応するような目、口などの顔器官周辺でいくつか定義し、これらの3次元位置の変位をパラメータとして用いてモデル化を行う。制御点の変位に対する顔の3次元形状変化をモデル化したものを、ここでは顔形状変化モデルと呼ぶ。

実在の人物の顔形状変化モデルが得られれば、映画・ゲームなどの映像中で実在の人物の様々な表情を表現する CG を容易な変形操作で得ることができる。こうして得られた顔形状変化モデルは、実写と CG の違和感無い融合が可能となる。

しかしながら実在の人物の顔の3次元形状、またその変化の様子は個人ごとに大きく異なる。よって、顔形状変化モデルは、各個人ごとに異なるものを構築する必要がある。これに対し本研究では、実在の人物の様々な表情をした顔の3次元形状を多数獲得し、この観測データに基づいて顔形状変化モデルを構築する。

以上をまとめ、本研究では、以下の3点を全て満たす顔形状変化モデルの構築を目指す。

- (1) 少数の制御点により直観的な変形操作が可能である。

¹ 京都大学

Kyoto Univ.

a) yamamoto@mm.media.kyoto-u.ac.jp

b) funatomi@mm.media.kyoto-u.ac.jp

c) iiyama@econ.kyoto-u.ac.jp

d) minoh@mm.media.kyoto-u.ac.jp

- (2) 実際に計測することで得られた顔形状に基づき、個人差を表現可能である。

以降2節では、本研究の取り扱う問題について数学的に定義する。3節では顔形状変化モデル構築の従来手法について紹介し、本研究の位置づけについて述べる。4節で適切な制御点の位置について検証する。5節では顔形状変化モデルの構築手法について詳説し、6節ではその評価を行う。最後に7節でまとめる。

2. 顔形状変化モデルの定式化

顔の3次元形状は、 N 個の3次元点を並べた $3 \times N$ 次元ベクトルである点群 $\mathbf{S} = (\mathbf{s}_1^T, \dots, \mathbf{s}_N^T)^T$ と、その各3次元点で成す面によって構成される。ここでは1つの面を3つの3次元点の接続によって成す3角形で表わす。ある表情に対応する3次元形状の全ての面についての接続関係の集合をここではメッシュ G と呼ぶ。本研究では顔の形状変化は、 \mathbf{S} の各点の座標のみの変化であると定義し、メッシュ G は不変であるとする。すなわち、この顔の3次元形状のトポロジーは表情変化に対して不変であるとする。

人が表情を変化させると、顔の形状も変化する。ここでは表情と顔形状は1対1に対応するものとし、任意の表情 t における3次元形状の点群 \mathbf{S}_t を、基準となるある表情 b における形状の点群 $\mathbf{S}^b = (\mathbf{s}_1^{bT}, \dots, \mathbf{s}_N^{bT})^T$ を用いて以下のように表現する。

$$\mathbf{S}_t = \mathbf{S}^b + \mathbf{D}_{\mathbf{S},t} \quad (1)$$

ここで $\mathbf{D}_{\mathbf{S},t} = (\mathbf{d}_{\mathbf{s}_1,t}^T, \dots, \mathbf{d}_{\mathbf{s}_N,t}^T)^T$ は \mathbf{S}^b から \mathbf{S}_t への各頂点の変位を表す。また \mathbf{S}^b は、あらかじめ観測によって得られたデータから作成しておく。具体的な作成方法については5.1節で述べる。 \mathbf{S}^b は欠損、ノイズについてはスムージングや手動による修正によって取り除かれているものとする。

顔形状変化を制御するパラメータとして、本研究は顔の表面上に配置されたいくつかの制御点を用いる。ここでは、制御点の位置によって表情は一意に定まるものとする。制御点は n 個の3次元点を並べた $3 \times n$ 次元ベクトル $\mathbf{H} = (\mathbf{h}_1^T, \dots, \mathbf{h}_n^T)^T$ として表わされる。表情 t での制御点位置 \mathbf{H}_t は、基準表情 b における制御点位置 $\mathbf{H}^b = (\mathbf{h}_1^{bT}, \dots, \mathbf{h}_n^{bT})^T$ を用いて以下のように表現される。

$$\mathbf{H}_t = \mathbf{H}^b + \mathbf{D}_{\mathbf{H},t} \quad (2)$$

ここで $\mathbf{D}_{\mathbf{H},t} = (\mathbf{d}_{\mathbf{h}_1,t}^T, \dots, \mathbf{d}_{\mathbf{h}_n,t}^T)^T$ は、表情 t での各制御点の変位を表わす。

顔形状変化モデルとは、各制御点の変位 $\mathbf{D}_{\mathbf{H},t}$ が与えられた時に、顔形状の各点の変位 $\mathbf{D}_{\mathbf{S},t}$ を得るモデル $M(\mathbf{D}_{\mathbf{H},t})$ であると定義する。モデル M については、解剖学的知見に基づくモデル[2]や顔形状データベースに基づくモデル[3][4][5]、計測データに基づくモデル[6][7]など様々なモ

デル化が考えられるが、本研究では計測データに基づく手法を取る。具体的には、同一人物の異なる表情での点群と制御点位置を観測し、そこから未観測の制御点位置における点群の変位を生成するモデルを学習する。

ここで、 F 個の異なる表情における3次元形状計測で得られた点群を $\mathbf{S}_1^o, \dots, \mathbf{S}_F^o$ (ただし、 $\mathbf{S}_t^o = (\mathbf{s}_{t,1}^{oT}, \dots, \mathbf{s}_{t,N_t}^{oT})^T$) とする。またそれぞれの表情に対応する制御点を計測して得られた制御点位置を $\mathbf{H}_1^o, \dots, \mathbf{H}_F^o$ (ただし、 $\mathbf{H}_t^o = (\mathbf{h}_{t,1}^{oT}, \dots, \mathbf{h}_{t,n}^{oT})^T$) とする。本研究では、実際に人物の顔を計測して得られた顔形状の点群 $\mathbf{S}_1^o, \dots, \mathbf{S}_F^o$ と制御点位置 $\mathbf{H}_1^o, \dots, \mathbf{H}_F^o$ から M を得ることにより、任意の表情 t について、 $\mathbf{D}_{\mathbf{H},t}$ を入力として $\mathbf{D}_{\mathbf{S},t}$ を得ることを目標とする。

ここで、 $\mathbf{S}_1^o, \dots, \mathbf{S}_F^o$ には計測時の欠損やノイズなどが含まれ、また \mathbf{S}^b とは点の数が異なる。ここでは、 \mathbf{S}_t^o に基づいて \mathbf{S}^b を変形することで \mathbf{S}^b の各点 \mathbf{s}_i^b の表情 t における点 $\mathbf{s}_{t,i}$ を求める。これによって、観測点群 \mathbf{S}_t^o における顔形状の変位 $\mathbf{D}_{\mathbf{S},t}^o$ が得られる。

モデル M は制御点の変位 $\mathbf{D}_{\mathbf{H},t}^o$ を説明変数とし、各頂点の変位 $\mathbf{D}_{\mathbf{S},t}^o$ を目的変数とする回帰問題として定式化でき、観測点群 \mathbf{S}_t^o から得られる $\mathbf{D}_{\mathbf{S},t}^o$ と \mathbf{H}_t^o から得られる $\mathbf{D}_{\mathbf{H},t}^o$ の組 $(\mathbf{D}_{\mathbf{S},t}^o, \mathbf{D}_{\mathbf{H},t}^o)$ を学習データとしてモデルを学習する。モデルが示すべき性質として、式3が定式化できる。すなわち、

$$M = \arg \min_M |\mathbf{D}_{\mathbf{S},t}^o - M(\mathbf{D}_{\mathbf{H},t}^o)| \quad (3)$$

以上において、解決すべきは以下の点である。

- (1) 顔形状がどのように変化するのかが操作者によって直観的に理解できるよう、各制御点 \mathbf{h}_i を少数にとどめると同時にその配置についても考慮する必要がある。
- (2) 各形状計測データ \mathbf{S}_t^o については、形状計測を行うにあたり、どのような手法が本研究の目的に適するかを議論する必要がある。
- (3) 一般には非線形となる M の定式化と、その最適化が必要となる。
- (4) \mathbf{S}^b を変形して \mathbf{S}_t^o から各 $\mathbf{s}_{t,i}$ を得る必要がある。

1.については4で議論する。2.については5.1節で議論する。3.は4の制御点の定義の吟味と同時に議論する。4.は5.2節で詳説する。

3. 顔形状変化モデル構築の従来手法

顔形状変化モデルの構築手法は大きく分けて、解剖学的見地に基づいて顔形状変化モデルを構築する手法[2]、あらかじめ作成した顔形状データベースに基づき顔形状変化モデルを構築する手法[3][4][5]、顔形状変化モデルを構築する本人の計測データに基づいて3次元顔形状を表現する手

法 [6][7] の 3 つがある。

解剖学的見地に基づいて顔形状変化モデルを構築する手法として、例えば Zhang ら [2] は、人の顔を骨、筋肉、肌の層に分け、それぞれの変形を制御点の位置変化から計算している。まず、あらかじめテンプレートとして骨のモデル、筋肉のモデル、肌のモデルを作成しておき、制御点の変位に対する肌の変形モデル、肌の変形に対する筋肉の変形モデル、筋肉の変形に対する骨の変形モデルを解析的に求めておく。この後、個人ごとの顔形状に適応した顔形状モデルを、骨、筋肉、肌の各モデルを基準表情の制御点に基づいて変形することによって作成する。

しかしこうして構築された顔形状変化モデルの制御点の定義について、直観的操作に適するかどうか十分な議論がなされていない。また、顔形状変化モデルの精度について、定量的な評価も十分に行われていない。

顔形状データベースに基づく顔形状変化モデル構築の代表的な手法としては、Cao ら [5] によるものを挙げる。これは、カメラ画像を入力として、画像中の制御点位置に基づき、個人ごとに構築した顔形状変化モデルから 3 次元顔形状の構築を目指したものである。様々な表情 e について 3 次元顔形状モデル S_e をあらかじめ作成し、任意の表情 t の 3 次元形状の点群 S_t をこれらの線形重ね合わせで表わした。このような 3 次元顔形状モデルの表現方法を Blendshape と呼ぶ。各 α_i を制御点の変位 $D_{H,t}$ と対応付けることで、顔形状変化モデルを構築している。[5] は制御点として、顔の鼻や目、鼻などの顔器官や前頭部等に配置された計 75 点を用いた。[5] は各個人に対応した Blendshape を得るため、個人ごとに S_e を変形している。[5] では、様々な個人差、表情の 3 次元顔形状の集合である 3 次元顔形状データベースをあらかじめ用意し、 S_e をこのデータベースにおける 3 次元顔形状の線形和として表現することで、個人ごとに異なる S_e を得る。こうして得られた S_e を用いた Blendshape によって、各個人に対応した顔形状変化モデルの構築を試みている。

[5] では個人差ごとに異なる顔形状が得られているものの、これはデータベース中の個人差の重ね合わせであり、その個人差を十分に表現することができない。また制御点については人が手動で操作することを考慮したものではなく、直観的操作のために最適なものとなっているかどうかについては十分な議論がなされていない。

計測データに基づく代表的な手法として、ここでは Bickel らの手法 [7] を挙げる。[7] は顔全体に 89 点のマーカを付与し、これらを制御点とした顔形状変化モデルを形状計測データを基に構築した。[7] はパターン光投影法により各表情の 3 次元形状を獲得した。続いて、各表情において得られた制御点位置に基づいて、基準形状からの大まかな変形を皮膚の曲げ伸ばしに基づき計算した。こうして得られた大まかな変形に、計測データに基づいた学習によって得

られる、しわなどの詳細な変形を組み合わせることで、詳細な顔形状変化モデルを得た。

しかしながら [7] においては、制御点の位置や数について十分な考察がなされていない。制御点として 89 点を用いているが、こうした多数の点を手動で動かすことによる変形操作は容易ではない。また、その位置も顔の変形に対して直観的とは限らない。

以上のように、多くの研究では制御点の位置、数について、直観的な操作が可能ないように定義がなされていない。

本研究では、制御点の位置、数について吟味し、変形操作が容易な顔形状変化モデルを構築する。

4. 直観的な変形操作を可能とする少数の制御点の設定

1 節や 2 節で述べたように、顔形状変化モデルは制御点によって変形操作される。しかし、その数、位置についてはどのようなものが適切かは自明ではない。ここでは、適切な制御点とはどのようなものか検証する。

制御点は、人がこれを用いて顔の変形操作を行うことを前提としている。この変形操作が人にとって容易であるためには、制御点をどのように変位させればどのような顔の変形が得られるのか、おおよそ直観的に理解しやすいような制御点の定義が必要となる。また、視覚的特徴が豊富な位置に制御点を定義することで、今後こうした場所を画像特徴のみをもちいてマーカレスに追跡することが可能になるとも考えられる。さらにその数については、多ければ多いほど人にとって制御が複雑で、煩雑なものとなるため、なるべく少ない数が求められる。加えて、顔形状変化モデルと制御点の定義より、そうした数、位置の制御点のみによって、顔全体の形状変化を十分に表わせることも必要となる。

以上より本研究では、以下の 3 点の要求を満たすものを適切な制御点とする。

- (a) 視覚的特徴が豊富な位置に定義され、かつ各制御点の変位に対して顔の形状変化が直観的に理解しやすい。
- (b) 数が十分に少ない。
- (c) 各制御点の変位のみで顔全体の形状変化を十分に表わせる。

ここではまず (a) について 4.1 節で議論する。(b) と (c) は一般にトレードオフの関係にあるが、これについては 4.2 節で少ない数の制御点から顔形状変化を表現する方法について議論する。4.3 節では制御点位置の獲得方法について述べる。最後に 4.4 節で (b) について制御点の適切な数、またその組み合わせを検証する。

4.1 直観的な変形操作のための制御点の設定

前述の要件 (a) を満たすため、顔の形状変化の直観的な記述方法について考察する。人の顔の形状変化の記述方法として、Facial Action Coding System (FACS) [1] が提案されている。FACS とは様々な顔の表情変化と、それに伴う顔の動きを観察することで、主観的に顔の形状変化を記述したものである。FACS においては、まず、人の顔形状変化を観察に基づいて最小の形状変化単位 (Action Unit: AU) に分類している。AU は瞼や眉等、顔の上半分の形状変化に関するもの (Upper AU: UAU) と唇や顎等、顔の下半分に関するもの (Lower AU: LAU) がそれぞれ定義されている。各 AU は単一、または複数の筋肉の動きに基づいて記述される。全ての表情は、これら AU の組み合わせによって記述される。例えば、「両眉が引っ張られてあがる」という表情は、AU1「眉の内側が上がる」、AU2「眉の外側が上がる」、AU4「眉が低くなり、沈み込む」を用いて AU1+2+4 のように表現される。

FACS は人の顔の形状変化を主観的な観察によって記述したものである。このため、AU は人間にとって直観的に理解しやすい定義がなされている。よって、FACS における AU に対応するように制御点が定義できれば、直観的な変形操作が可能な顔形状モデルが構築可能となる。

制御点の位置として、本研究では各 AU に対応する目や口などの顔器官の周辺を候補として挙げる。AU は目や口、眉、下顎などの動きに対して定義されており、これらの周辺の点を制御点とすることで、AU に対応したものが得られる。またこうした位置は頬や前頭部などに比較して視覚的特徴が豊富であり、将来的に Active Appearance Model (AAM) [8] や Constrained Local Model (CLM) [9][10] などを用いた特徴点抽出手法により、マーカレスに追跡可能となることが期待される。

しかし、その数や組み合わせについては、どのようなものが最適なものは明らかではない。このため 4.4 節で、こうした制御点の数、組み合わせについて検証を行う。

4.2 少数の制御点による顔形状の表現

事例ベースの手法においては、一般的に各制御点の位置を説明変数、顔形状の点群の各点の位置を目的変数として回帰学習を行うことで、顔形状変化モデルを構築する [7]。前述のように、制御点の数はなるべく少ないほうが望ましい。一方、一般には説明変数の数が十分でないと回帰の精度が低下する。本研究では非線形カーネルを用いた Support Vector Regression (SVR) によって顔形状変化モデルの構築を行う。すなわち、目的変数 y を、説明変数ベクトル \mathbf{x} を用いて以下のように表わす。

$$y(\mathbf{x}) = \sum_{m=1}^N \alpha_m K(\mathbf{x}, \hat{\mathbf{x}}_m) \quad (4)$$

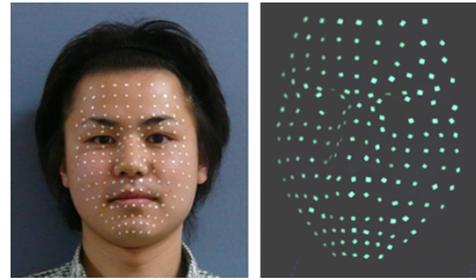


図 1 顔にマーカを付与した様子 (左) と、これを観測した様子 (右)

ここで、 $\hat{\mathbf{x}}_m$ は学習データ、 $K(\mathbf{X}, \mathbf{Y})$ は非線形カーネル関数、 α_m は重みである。本研究では、目的変数と説明変数の取り方として、 $\mathbf{S}_{i,t}$ 中の全点について独立に回帰を行う。すなわち、説明変数は全ての $\mathbf{h}_1, \dots, \mathbf{h}_n$ であり、 $\mathbf{s}_i = (s_i^x, s_i^y, s_i^z)$ とするとき、各 i について s_i^x , s_i^y , s_i^z をそれぞれ目的変数とする。すなわち、 N 個の点群 \mathbf{S} に対して独立な SVR を $3N$ 個学習する。

4.3 制御点位置の獲得と位置合わせによる剛体変換の除去

4.2 節の説明変数として、計測データにおける制御点位置を得る必要がある。従来、顔表面の特徴点を取る方法として、AAM や CLM 等を用いた特徴点抽出手法が提案されている。しかしこうした手法は、その位置を全ての表情に対して高い精度で取得することはできない。このため、本研究の検証においてこうした手法を用いると、制御点としての顔形状の表現能力を正しく評価することができない。

ここでは、顔で制御点に対応する位置に 1mm 角程度の大きさの蛍光マーカを付与し、頑健にこれらの位置を得る。図 1 は、顔に蛍光マーカを付与した様子である。ここで用いる蛍光マーカは、ブラックライトなどによって紫外線を照射すると緑色に発光するものである。形状計測の際には複数台のキャリブレーションされたカメラによって各表情におけるこのマーカを撮影する。得られた画像中において緑色領域を抽出することによって、これらマーカの位置を得る。

なお以降の検証においては、マーカの抽出精度が制御点の性能の評価に含まれてしまうことは好ましくない。これを防ぐため、こうして得られたマーカ位置を手動で修正することにより、こうした制御点位置は最大でもマーカの大きさ程度の、十分小さい誤差のみを含む精度で得られているものとして、以降の検証を行う。

こうして得られた制御点位置には顔全体の回転、並進移動 (合わせて剛体変換と呼ぶ) も含まれる。すなわち、観測された制御点位置 $\mathbf{h}_{t,i}^o$ は、基準点群においてこれに対応する制御点位置 \mathbf{h}_i^b を用いて、

$$\mathbf{h}_{t,i}^o = R_t(\mathbf{h}_i^b + \mathbf{D}_{\mathbf{h},i}) + \mathbf{T}_t \quad (5)$$

となる。ここで、 R_t は 3 次元回転行列、 \mathbf{T}_t は並進ベクトルである。これらは顔形状の変化とは無関係な変位である

ため、こうした制御点の変位を用いて顔形状変化モデルの学習を行うことは好ましくない。これに対し、以下のようにして $\mathbf{D}_{h,i}$ を求めることができる。

$$\mathbf{D}_{h,i} = R_t^{-1}(\mathbf{h}_{t,i}^o - \mathbf{T}_t) - \mathbf{h}_i^b \quad (6)$$

すなわち、 R_t 、 \mathbf{T}_t を求めることができれば、これらの影響を除いた、表情変化による変位のみを求めることができる。ここでは、これを位置合わせと呼ぶ。

本研究では、基準形状から得られるいくつかのマーカ点を位置合わせの基準点として選び、各表情でこれに対応する点を剛体変換して位置合わせすることで、こうした剛体変換による変位を除く。基準形状での基準点の集合を $\mathbf{B}^b = \{\mathbf{b}_1^b, \dots, \mathbf{b}_l^b\}$ とする。また、表情 t でこれに対応する各基準点の集合を $\mathbf{B}_t = \{\mathbf{b}_{t,1}, \dots, \mathbf{b}_{t,l}\}$ とする。このとき、剛体変換のパラメータ R_t 、 \mathbf{T}_t を以下で求める。

$$\arg \min_{R_t, \mathbf{T}_t} \sum_{i=1}^l \|\mathbf{b}_{t,i} - R_t \mathbf{b}_i^b + \mathbf{T}_t\| \quad (7)$$

この位置合わせに用いる点として、本研究では両瞼下、鼻の最も高い点に配置した3点を用いる。この3点は、筆者らの実験により剛体性が高いことが分かっている。

4.4 形状再現誤差が最小となる制御点の探索

ここでは、制御点の適切な数、組み合わせについて実験的に明らかにする。4.1で述べたように、本研究では以下の2点を満たすような点を制御点候補とする。

- (1) FACSにおけるAUに対応する位置である。
- (2) 視覚的情報が豊富な位置である。

ここでは具体的に、このような条件を満たす点として図2に示すように、AUに対応しかつ視覚的特徴が豊富な顔器官周辺の計29点を制御点の候補として設定する。この29点のうちから、以下に述べる実験方法によって顔形状変化モデル構築に最適な数、組み合わせの制御点を策定する。

4.4.1 実験の設定

様々な表情の顔の3次元点群 \mathbf{S}_t を様々な制御点の組み合わせについて各制御点の変位から再現することで、その制御点の組み合わせが顔形状変化モデルを構築する能力を検証可能であると考えられる。しかしながら、 \mathbf{S}_t は5で述べるように、形状計測手法のノイズや欠損などの影響が計測形状に含まれる。ここではこうした影響が顔形状変化モデルの構築に含まれてしまうことを防ぐため、 \mathbf{S}_t についてこうした形状計測手法の精度による影響を取り除いた状態で検証を行う。このため、顔全体に図1に示すように計171点のマーカを付与し、このマーカ171点の3次元位置を顔の3次元形状の点群とみなす。これらの位置についての再現を4.2節で述べた方法により行うことで、制御点の能力を検証する。

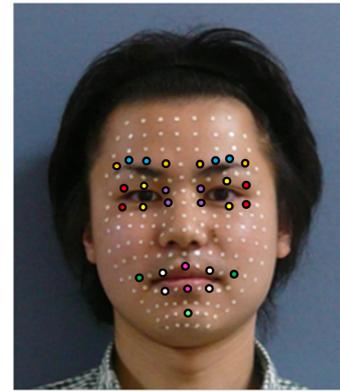


図2 制御点の候補29点を示す。上図において、同じ色で示される制御点候補を同じグループとし、同じグループに属する制御点は必ず同時に用いるとする。

表1 学習データ、テストデータの内訳

	フレーム数	含まれる表情
学習データ	471	両眉を上げる
		片眉を上げる
テストデータ	750	片頬を上げる
		口を大きく開ける
		両眉を寄せる
		片頬を上げる
		口を大きく開ける
		口を尖らせる
		両頬を上げる

ここで用いる学習データとしては、表情変化に伴ってあり得る制御点の動きをなるべく多く含むことで、制御点の動きに対する顔形状の変化の対応を多数得られるようなデータであることが重要である。そこで、ここでは制御点を大きく動かすような表情を学習データとして用いることで、あり得る制御点の動きを多く含むデータを得る。こうした表情を含む学習データとして、今回は連続した471フレームの間表情を変化させながら撮影した形状計測データを用いる。

上記の学習データを用いて顔形状変化モデルを構築し、テストデータと比較することによって提案手法の性能評価を行う。テストデータとしては学習データと同じ位置171点にマーカを付与し、様々な表情を含む500フレーム程度の形状計測データを用いる。学習データ、テストデータに含まれる表情の例を表1に示す。

制御点の顔形状変化モデル構築の能力は、以下の回帰誤差を用いて評価する。

$$E(\mathbf{H}) = \sum_{t=1}^F \sum_{i=1}^N \|\mathbf{s}_{t,i} - (\mathbf{s}_i^b + M(\mathbf{d}_{\mathbf{H},t}))\| \quad (8)$$

ここで、各 $\mathbf{s}_{t,i}$ はマーカを用いることにより、観測データから直接得られていることに注意する。

制御点の候補29点からは $2^{29} - 1 = 536870911$ 通りの制御点の組み合わせが考えられるが、ここでは組み合わせ爆発を抑えるため、これらの候補のうちからいくらかを検

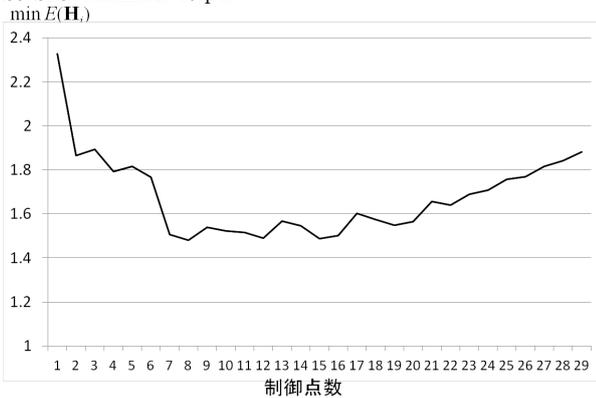


図3 制御点の数と再現誤差の関係。制御点の数を横軸、各制御点数で最も誤差が小さいときの誤差を縦軸で示す

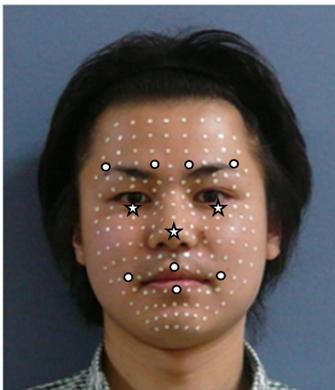


図4 検証の結果得られた最適な制御点8点(丸型マーク)と、位置合わせに用いる基準点3点(星型マーク)

証する候補とする。具体的には、制御点を図2のように、各顔器官の上下・左右の対称性を考慮して9のグループに分ける。これらのグループの組み合わせ $2^9 - 1 = 511$ 通りを、制御点の組み合わせの候補とする。

4.4.2 結果と考察

一般に、制御点の数と顔形状変化モデルの表現能力は密接な関係にあると予想される。そこでまず、制御点の数ごとにその能力を評価する。各制御点の数ごとに、誤差 $E(\mathbf{H})$ が最も小さくなるものを、ここではある制御点数について最も良い組み合わせであるとする。この制御点の数を様々に変化させ、最も良い制御点の数を制御点の数を調べる。制御点の数を横軸、その数での $E(\mathbf{H})$ の最小値を縦軸としたグラフを図3に示す。

本来、制御点の数は多くなるごとに顔形状変化モデルの表現能力は上がり、誤差 $E(\mathbf{H})$ は小さくなると予想される。しかし、図3に示すように、実際には制御点が8個以上の顔形状変化モデルについては、制御点の数に対して $E(\mathbf{H})$ が増加傾向にある。これは、制御点が多くなることによって学習に用いる特徴量の次元数が増加し、過学習が起こったためと考えられる。この結果より、制御点の数は8個が最適であると言える。この8個の組み合わせについて図4に示す。

以降、本研究では制御点として図4に示す8点を用い、

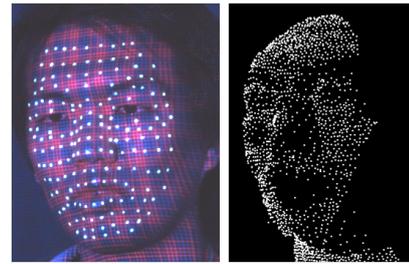


図6 顔にパターン光を投影した様子(左)と、ワンショットスキャン法で得られる点群の一例(右)

顔形状変化モデルを構築する。

5. 観測に基づいた3次元顔形状変化モデルの構築

4節の考察に基づき、本研究では8点の制御点と、形状計測で得られた点群を用いて顔形状変化モデルを構築する。図5に処理の概要を示す。

まず顔形状変化モデルを構築する人物の顔の制御点位置に蛍光マーカを付与し、顔の形状計測と同時に制御点の3次元位置を計測し、 \mathbf{S}^b と \mathbf{H}^b を得る。 \mathbf{S}^b と \mathbf{H}^b については、口や目を閉じた状態の表情を用いて計測すると、上瞼と下瞼、上唇と下唇の間にメッシュが張られた3次元顔形状が得られるが、これは目を開けたり、口を開けたりした表情の3次元形状とメッシュ構造が異なる。このため、本研究では目、口を軽く開けた表情の顔形状と、この時の制御点位置を計測し、これを \mathbf{S}^b 、 \mathbf{H}^b とする。次に、同様に様々な表情下で形状計測と同時に制御点位置を計測し、 \mathbf{S}_i^o と \mathbf{H}_i^o を得る。本研究では、この際の形状計測手法としてワンショットスキャン法 [11][12] を用いる。

5.1 ワンショットスキャン法による詳細な3次元形状の獲得

様々な表情での計測点群 \mathbf{S}_i^o を獲得するため、本研究ではパターン光投影法の一つであるワンショットスキャン法 [11][12] を用いる。グリッド状のパターン光を計測対象の物体に投影し、図6のように物体に投影された様子を複数台のカメラで計測する。投影されたパターンのグリッドを複数カメラ間で対応付けることにより、各グリッド点の3次元位置を獲得することで、物体の3次元形状を点群として得る。図6に得られた点群の一例を示す。また、ワンショットスキャン法は \mathbf{S}_i^o と同時に、面を表現するメッシュ構造 G_i^o も獲得できる。またこの形状計測の際には、計測する顔の制御点にあたる部分に蛍光マーカを貼り付け、これによって各表情における制御点位置 \mathbf{H}_i^o も同時に獲得する。

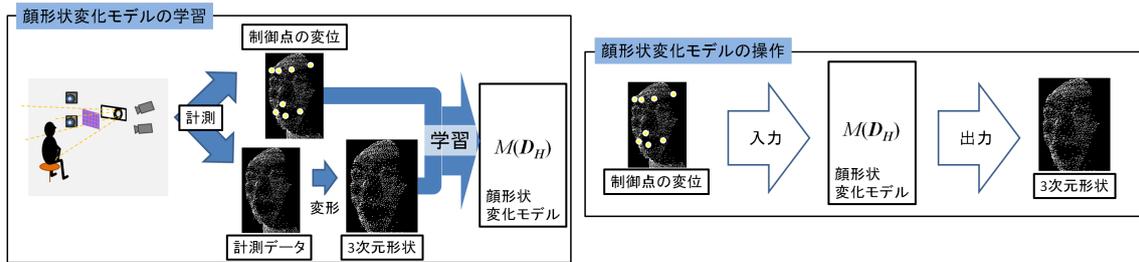


図 5 提案する 3 次元顔形状変化モデル構築手法の概要

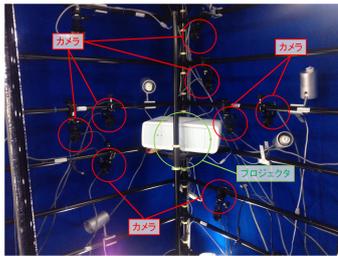


図 7 実験装置を正面から撮影した様子。カメラを前方に 8 台、プロジェクタをその中央に配置した

5.2 Template Mesh Deformation による変形点群の獲得

ある表情 t での点群 S_t について、基準形状からの変位 $D_{S,t}$ を求めるためには、 S_t^b から、 S^b 中の点 s_i^b に対応する点 $s_{t,i}$ を得なければならない。しかしながら S_t^b には計測した際にノイズや欠損などが含まれ、また点の数も計測するたびに異なるため、 $s_{t,i}$ を S_t^b から直接得ることは難しい。

本手法では、Template Mesh Deformation[13][14] によって S^b を変形して S_t^b にフィッティングすることで、 S_t を得る。得られた S_t を用いて、各点の変位 $D_{S,t}$ を以下で求める。

$$D_{S,t} = S_t - S^b \quad (9)$$

5.3 SVR を用いた顔形状変化モデルの構築

こうして得られた $D_{S,t}$ と、マーカの観測によって得られた $D_{H,t}$ の組について、SVR を用いた学習を行うことで顔形状変化モデルを構築する。こうして得られた顔形状変化モデルに、ある表情に対応する制御点の 3 次元位置を入力すると、その表情に対応する顔形状が得られる。

6. 顔形状変化モデル構築の性能評価と考察

提案手法によって構築された顔形状変化モデルが、顔の詳細な 3 次元形状を再現するために充分であるのかについて検証するため、以下のような実験を行った。

被験者 1 人の顔形状を、連続した何フレームかに渡り、同期された 8 台のカメラを用い約 7.5fps で撮影した。被験者の顔に対してはカメラと同方向からプロジェクタによってパターン光が投影されている。図 7 に実験装置を示す。

得られたマーカ点の座標、形状計測結果を用い、位置合

わせと顔形状変化モデルの構築を行う。得られたマーカ点のうち、位置合わせには目の下、鼻の頂点にある 3 点のマーカ位置のみを用い、制御点としては 4.4.1 節で示した 8 点のみを用いた。なお、基準形状の点群 S^b の点の数は約 1700 個である。

今回用いたデータの中には、制御点が大きく動く表情として以下のような表情を含めた。

- 口を大きく開ける
- 口を尖らせる
- 片頬を上げる
- 目を大きく開ける
- 目を強く閉じる

こうした表情を含むデータとして、連続した 100 フレーム程度の形状計測を 4 回行い、これら 4 つのデータを用いた。

これら 4 つのデータのうち、1 つをテストデータ、残り 3 つを学習データとして Cross Validation により顔形状変化モデルの性能評価を行う。ここでは評価指標として、テストデータの形状計測結果である 3 次元点群をどの程度再現できているかを用いる。すなわち、テストデータの各表情の 3 次元点群を $\hat{S}_1, \dots, \hat{S}_F$ (ただし、 $\hat{S}_t = (\hat{s}_{t,1}^T, \dots, \hat{s}_{t,N_t}^T)^T$) とするとき、顔形状変化モデルとテストデータとの各表情 t における、点 $\hat{s}_{t,i}^T$ の誤差 $E(t)_i$ を以下で定義する。

$$E(t)_i = \|\hat{s}_{t,i} - s_{t,min}\| \quad (10)$$

ただし、 $s_{t,min}$ は $M(D_{H,t})$ で再現された点群のうち $\hat{s}_{t,i}$ と最も近い点である。

各表情における誤差 $E(t)_i$ の中央値を図 8 に示す。図 8 において、横軸は各表情、縦軸はその表情における $E(t)_i$ の中央値を示す。この図より、多くの表情において $E(t)_i$ の中央値は 3mm から 4mm 程度の誤差を示すことが分かる。

また、実際に再現された 3 次元形状の一例を図 9 に示す。図 9 左の基準形状 (a) を顔形状変化モデルによって変形させたもの (b) と、実際に観測された形状 (c) とを比較した場合、(b) が (c) とほぼ同じ形状を再現できているのが確認できる。

7. おわりに

本研究では、なるべく少数で、かつ直観的に理解しやすいパラメータにより、容易に制御可能な顔の 3 次元形状全

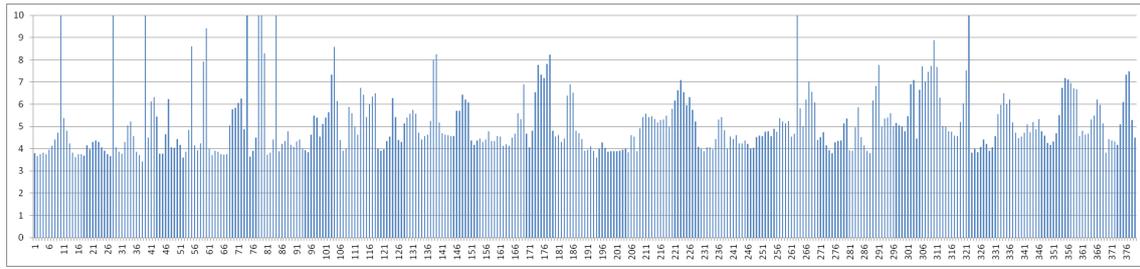


図 8 各表情における $E(t)_i$ の中央値

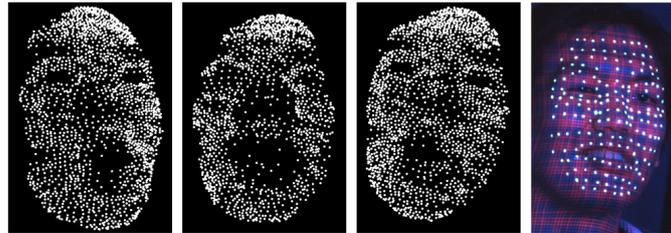


図 9 左から順に、基準形状、形状計測結果、顔形状変化モデルによる再現結果、観測画像

体を制御可能な顔形状変化モデルを構築した。

またこの制御点の定義に基づいて実際に顔形状変化モデルを構築し、その精度について検証を行い、誤差を評価した結果、3,4mm 程度の誤差で顔形状を再現可能であることがわかった。

今後の課題として、精度の向上のため、Template Mesh Deformation の改良が挙げられる。これについては、顔表面のテクスチャ情報などを用いることによって、精度向上が見込めると考えている。

また評価方法として、今後は顔形状変化モデルにより得られた 3 次元顔形状モデルの表面にメッシュを張ることで、リアルな見た目の顔として再現ができていのかどうか確認することも、今後重要である。

謝辞 本研究は科研費 (23700231) の助成を受けたものである。

参考文献

- [1] Ekman, P. and Friesen, W. V.: Measuring facial movement, *Environmental Psychology and Nonverbal Behavior*, Vol. 1, No. 1, pp. 56–75 (1976).
- [2] Zhang, Y., Sim, T., Lim Tan, C. and Sung, E.: Anatomy-based face reconstruction for animation using multi-layer deformation, *Journal of Visual Languages & Computing*, Vol. 17, No. 2, pp. 126–160 (2006).
- [3] Vlastic, D., Brand, M., Pfister, H. and Popović, J.: Face transfer with multilinear models, *ACM Transactions on Graphics (TOG)*, Vol. 24, No. 3, ACM, pp. 426–433 (2005).
- [4] Bouaziz, S., Wang, Y. and Pauly, M.: Online modeling for realtime facial animation, *Acm Transactions On Graphics*, Vol. 32, No. EPFL-ARTICLE-189496 (2013).
- [5] Cao, C., Weng, Y., Lin, S. and Zhou, K.: 3D Shape Regression for Real-time Facial Animation, *ACM Transactions on Graphics - SIGGRAPH 2013 Conference Proceedings*, Vol. 32, No. 41 (2013).
- [6] Joshi, P., Tien, W. C., Desbrun, M. and Pighin, F.: Learning controls for blend shape based realistic facial animation, *ACM SIGGRAPH 2005 Courses*, ACM, p. 8 (2005).
- [7] Bickel, B., Lang, M., Botch, M., Otaduy, M. A. and Gross, M.: "Pose-Space Animation and Transfer of Facial Details", *Eurographics/ ACM SIGGRAPH Symposium on Computer Animation*, pp. 57–66 (2008).
- [8] Cootes, F. T., Wheeler, G. V., Walker, K. N. and Taylor, C. J.: View-based active appearance models, *Image and vision computing*, Vol. 20, No. 9, pp. 657–664 (2002).
- [9] 高野博幸, 山崎彬人, 金出武雄, 出口光一郎: "顔の向きに頑健なフェイスマイニング", "電子情報通信学会論文誌 D", Vol. J96-D, No. 3, pp. 704–712 (2013).
- [10] Saragih, J. M., Lucey, S. and Cohn, J. F.: Deformable Model Fitting by Regularized Landmark Mean-Shift, *International Journal of Computer Vision*, Vol. 91, No. 2, pp. 200–215 (2011).
- [11] Kawasaki, H., Furukawa, R., Sagawa, R. and Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern, *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, IEEE, pp. 1–8 (2008).
- [12] 阿久澤陽菜, 船富卓哉, 飯山将晃, 美濃導彦: "ワンショットスキャン法による獲得形状を利用した様々な姿勢を表現可能な手形状モデルの構築", 第 15 回画像の認識・理解シンポジウム (MIRU2012) (2012).
- [13] Zayer, R., Rössl, C., Karni, Z. and Seidel, H. P.: Harmonic guidance for surface deformation, *Computer Graphics Forum*, Vol. 24, No. 3, Wiley Online Library, pp. 601–609 (2005).
- [14] Stoll, C., Karni, Z., Rössl, C., Yamauchi, H. and Seidel, H. P.: Template deformation for point cloud fitting, *Proceedings of the 3rd Eurographics/IEEE VGTC conference on Point-Based Graphics*, Eurographics Association, pp. 27–35 (2006).