顔の部位識別に基づくマーカレス モーションキャプチャに関する研究*

赤木康宏**, 古川 亮***, 佐川立昌†, 小川原光一††, 川崎 洋**

Marker-less Facial Motion Capture based on the Recognition of the Parts

Yasuhiro AKAGI, Ryo FURUKAWA, Ryusuke SAGAWA, Koichi OGAWARA and Hiroshi KAWASAKI

In this paper, we propose a marker-less motion capture method for a human face. We detect the movement of a face as the deformation from the initial pose to other poses. In such cases, the non-rigid registration algorithms are commonly used to detect the movement of an object. However, since the features of the movement of a face vary in each facial part, it complicates the registration problem. In our method, we recognize the facial parts by using Random Forest algorithm to be simplified to solve the registration problem. Then, we classify the facial parts into five types: nose, mouth, eye, cheek and obstacle. And we define the feature vector of each point of the face by using Fast Point Feature Histograms and a normalized position. We train the Random Forest based on the feature vector to recognize the facial parts from an arbitrary point set data. After the recognition of the facial parts, we detect the movement of the each part by using a non-rigid registration algorithm. Finally, we combine the translations of the each point as the deformation of the face by using the deformation technique based on a Radial Basis Function. In our results, we show that the proposed method enable us to detect the motion of the face more accurately.

Key words: facial motion tracking, non-rigid registration, deformation, facial parts recognition

1. はじめに

人間の顔を対象としたモーションキャプチャ技術は、CGアニメーション制作や表情認識への利用が行われている。モーションキャプチャ技術において最も重要な点は、時間経過により変化する対象形状を連続的に誤りなく追跡することである。顔を対象とする場合では、身体動作のように骨格の動きが主体となる場合とは異なり、表皮の自由変形による影響が相対的に増加するという特徴がある。そして、その変形に関する制約の少なから、変形を追跡することが困難になるという問題がある。そこで、既存の顔モーションキャプチャシステムにおいては、動きを追跡するためのマーカを顔上に配置する解決方法が用いられる。この方法では、頬のように形状のおよび色彩的な特徴の少ない部位であっても、マーカを配置することで形状の変化を頑健に追跡できるという利点がある。しかし、顔そのものの形状変化ではなく、マーカの位置の変化を追跡しているので、顔の3次元形状を別途用意する必要がある、マーカの多数が隠れるような動作は取得できない、および自然なテクスチャ情報を同時に取得できないという問題がある。これらの問題を解決するために、マーカレス顔モーションキャプチャ手法も提案されている。Weiseらの手法¹⁾およびZhangらの手法²⁾では、顔の画像情報から特徴点を抽出し、これをマーカとみなして運動を追跡している。これらの方法では、目や鼻などの画像特徴が取得しやすい部位に対しては有効であるが、頬のように画像上にも特徴のない部位の追跡は困難であり、抽出が容易な特徴点の変化を補間することでその形状を間接的に推定している。よっ

て、特徴抽出の難しい部位においては、撮影対象者の個性が失われるという問題がある。これらの問題を解決するために、本研究ではマーカレス顔モーションキャプチャの新たな手法として、Non-rigid Registration手法に基づき顔の形状変形を追跡するための手法を提案する。Non-rigid Registration手法は、自由変形する物体の対応点問題を解決するための手法であり、一般物体の変形をはじめとし、臓器などの人体形状の変形追跡にも用いられる手法である。この手法は2つの異なる形状同士の対応関係を頑健に求められるという特徴があるが、その頑健さゆえに、形状の変化の大きい部分と少ない部分の境界付近では変化が平均化され、特徴が失われる場合がある。この手法を顔に適用した場合、鼻と唇のような表皮の薄い部位と厚い部位の隣接する領域において同様の問題が生じる。そこで本研究では、顔の3次元形状情報から4種類(鼻、口、目、頬)の部位を識別し、各部位毎に独立してその形状変形を追跡することで、この問題の解決を図る。

Fig. 1 Overview of the marker-less facial motion capture (a)Initial registration and triangulation form the scanned 3D points (b)Facial parts recognition using the random forest (c)Non-rigid registration for tracking the movement of each part (d)Face deformations based on the movement of the parts

* 原稿受付 平成 25 年 4 月 1 日

** 正 会 員 鹿児島大学 (鹿児島県鹿児島市郡元)

*** 広島市立大学 (広島県広島市安佐南区)

† 産業技術総合研究所 (茨城県つくば市梅園)

†† 和歌山大学 (和歌山県和歌山市栄谷)

2. 関連研究

本章では、顔のモーションキャプチャおよび3次元形状の変形追跡に関する研究について述べる。

2.1 顔のモーションキャプチャに関する研究

第1章でも述べたように、顔のモーションキャプチャ手法では時間的に変化する顔上の同一点をどのように追跡するかが重要となる。その解決方法として、顔に反射材やペイント等のマーカを付加する方法と、画像や3次元形状の特徴を分析し、特徴点を定めることでマーカに相当する情報を自動的に生成する手法がある。

前者の方法は、マーカを用いることで撮影時に要する手間が増えるものの、顔の変形を頑健かつ詳細に求めることができるという特徴がある。Huangらの研究では、顔に100個以上のマーカを配置することで、詳細な形状変形を取得することに成功した³⁾。Bickelらの研究では、おでこや目じり等のしわの生じやすい部分にペイントを施すことでしわの発生を検出した⁴⁾。これらの方法では、撮影画像にマーカが写りこんでしまうので、顔そのもののテクスチャ情報を同時には撮影できないという問題がある。また、顔の一部が隠されたり、顔への接触があったりするような動作の取得には適さないという問題もある。

マーカに由来する問題を解決するために、マーカレスで顔の動きを取得する研究が提案されている。Weiseらの研究では、撮影者の顔形状と画像から抽出した特徴点との関連性をあらかじめデータベースに登録しておくことで、画像上の特徴点の位置からリアルタイムに3次元顔形状を生成する手法を提案している¹⁾。Sibbingらの研究では、顔を小領域に分割し、各領域の陰影特徴を計算することで、異なる時刻間の小領域の対応付けを行っている⁵⁾。以上のように、マーカレス顔モーションキャプチャ手法では、人工的なマーカの代替となるような形状変化に頑健な特徴を抽出することで、変形を追跡する方法が主流である。一方で、形状を構成する全ての3次元点の情報から、直接形状同士の位置合わせを行う、Non-rigid Registrationと呼ばれる手法も提案されている。

2.2 Non-rigid Registration アルゴリズムに関する研究

Non-rigid Registration 手法は、異なる姿勢をもった剛体同士の位置合わせを行うための手法である、Registration 手法を非剛体 (Non-rigid object) への位置合わせに拡張した手法である。この手法は、MRI 画像から臓器の3次元形状を取得する場合等の、形状の自由度が高く、特徴が少ない対象物の位置合わせに適している。Jianらの研究では、欠落やノイズを含む形状同士の位置合わせを実現するために、顔形状を構成する頂点の分布を Gaussian Mixture Model (GMM) により表現し、形状変形を2つの GMM 同士の L2 Distance を最小化する問題に置き換えることで位置合わせを実現している⁶⁾。この研究の中では、形状の移動を滑らかに補間するために Thin-Plate Spline (TPS)⁷⁾ および Gaussian Radial Basis Function (GRBF)⁸⁾ を基底関数に用いている。また、その他の代表的な GMM に基づく Non-rigid Registration 手法としては、Coherent Point Drift (CPD) と呼ばれる Myronenkoらの研究⁹⁾ および Expectation-Maximization (EM) 法に基づき GMM 同士の距離を最小化する、Chuiらの研究¹⁰⁾ がある。Tevsらの研究では、ポリゴンメッシュをグラフとみなし、異なる形状間のグラフの対応関係を求めることで精度よく顔の変形を追跡している¹¹⁾。

これらの手法では、変形の際に生じる歪みを最小化するよう

にアルゴリズムが設計されており、ゴムボールや臓器のような単一の特性をもった対象物に対しては有効に働く。一方で、第1章でも述べたように顔のような下層に骨格があり、脂肪層の厚みが部位により異なる対象物では、厚みの変化が大きな箇所において変形の補間が過剰に行われてしまうという問題が生じやすい。そこで、本研究では顔形状を性質の異なる部位毎に分けることでこの問題を解決する。次節では人体形状から様々な部位を認識するための研究について述べる。

2.3 人体形状の識別手法

画像や3次元形状情報から顔の特定部位を識別する手法は、表情や姿勢、視線等の人間の行動を認識するために有用であり、これまでに数多くの研究が行われている^{12)~15)}。Caoらの研究では、目や口等の顔を構成する部位を画像から検出する手法を提案している¹²⁾。この手法では、顔画像と各部位の配置をペアにした教師データを約1,000件与えることで、部位を識別するための回帰木を生成している。Caoらによる実験では、顔の部位を約95%の識別率で認識することに成功した。しかし、識別する箇所は口や目等の色彩や陰影の特徴のある場所に限られており、頬や額等の平坦な部位への適用は難しい。Dantoneらの研究では、Random Forest¹⁶⁾ を用いて顔の3次元形状データを学習することで、リアルタイムに顔の部位を識別することを可能にした¹³⁾。Shottonらは、同様に Random Forest を用いることで全身の姿勢推定を実現した¹⁵⁾。Random Forest は多次元かつ大量のデータを学習する際に適した手法であり、画像特徴の1つである SIFT や3次元形状特徴の1つである Fast point feature histograms (FPFH)¹⁷⁾ のような特徴量に基づき形状を識別する際に有用である。また、学習後は高速に対象物を識別することができるので、数万点の点群から構成される顔形状データを短時間で処理できるという利点もある。その一方で、対象物とは無関係の情報を棄却できずに、いずれかの対象物として識別してしまうという問題もある。本研究では、一般物体認識ではなく、顔の3次元形状のみを対象としているので、この問題は生じにくいと考え、これまでに述べた利点から Random Forest を用いた顔の部位識別を行う。次節では、本稿で提案する手法の概要について述べる。

3. 提案手法の概要

提案手法は次に示す4段階の処理から構成される。

(a) 顔形状の取得 (図1(a)) : 既存の3次元形状計測手法により、顔形状を表す3次元点群データを取得する。また、その点群データに対応する3角形メッシュおよび深度マップを取得する。取得した点群データは各時刻毎に独立した点群データから構成されている。以降は、開始から n 番目に取得した顔形状を示す表現として、“第 n フレーム”という呼称を用いる。

(b) 顔形状の部位識別 (図1(b)) : (a) で取得した顔形状データから、鼻、口、目、頬の4種の部位およびノイズ領域を識別する。顔形状データには、顔形状とは無関係なノイズや他の物体が含まれている可能性があるため、これを排除するために、ノイズ領域を積極的に識別する。このように、顔を $4+1$ 種の領域に分類した理由については第4章で述べる。また、各部位の識別には Random Forest を用いる。Random Forest に与える教師データは入力データから6フレーム分の顔形状を取り出し、これに含まれる点群を手作業により5種の領域に分類することで作成する。

(c) 各部位の変形追跡 (図1(c)) : (b) で識別した各部位

を表す領域の時系列上での変化を **Non-rigid Registration** 手法により追跡する。追跡の際には、入力データの第 1 フレームにおける形状データを他のフレームの形状にフィッティングさせる。これにより、第 1 フレームの形状が各フレームにおいてどの位置に移動したかを知ることができる。

(d) 変形の統合 (図 1 (d)) : (c) では、各部位毎に独立して形状変形を追跡しているので、部位の境界付近では動きが不連続となる。そこで、**Radial Basis Function** に基づく 3 次元形状変形手法¹⁸⁾ に基づき各部位の動きを統合する。

以上の処理により、時間的に連続ではない 3 次元形状データから、第 1 フレームの連続的な変形という形式で、顔の動きを取得できる。(b), (c) および (d) の詳細については、以降の第 4 章から第 6 章において順に述べる。

4. 顔の部位識別

本章では、3 次元点群データから顔を構成する部位を識別する手法について述べる。顔は、鼻や目のように表皮が薄く動きの少ない部位および、口や頬のように表皮が厚く動きの大きい部位の組み合わせにより構成されている。この特徴が顔形状の変形追跡をより困難にする要因の 1 つとなっている。そこで本研究では、顔は鼻、口、目、頬の 4 種類の部位およびノイズ領域から構成されていると定義し、これを 3 次元点群データから識別するための手法を提案する。

4.1 顔の部位分類

顔を 4 種類の部位に分類した理由は以下である。表情認識を行うための研究において広く用いられている、顔の特徴点セットの 1 つである **Active Appearance Model**²⁾ では、目、眉、鼻、口および顔全体の輪郭 (顔を正面から見た際の、耳から顎にかけてのライン) を個別に扱っている。このことから、これらの部位には顔に固有の特徴がよく表れており、これを正確に捉えることが顔を精度よく追跡するために重要であると考えた。ただし、**AAM** における特徴点のうち眉に相当する部位は個別に識別しないものとする。その理由は、本研究では顔の 3 次元形状を処理の対象としているので、画像上では強い特徴をもつ眉であっても、形状としては僅かな凹凸でしかなく、識別することが困難であるからである。そして、本研究では顔を構成する全ての 3 次元点をいずれかの部位に識別するので、眉周辺の形状は「目」に相当する部位に、顔全体の輪郭から、目・鼻・口に至るまでの領域を「頬」に相当する部位として扱うこととする。

4.2 Random Forest による部位識別の特徴

Random Forest による識別では、入力として識別対象の特徴を表したベクトル (特徴ベクトル) を与えることで、各識別対象への該当率を得ることができる。さらに詳しく述べると、**Random Forest** は教師データを与えることで多数の回帰木を構築し、各回帰木の返す結果を投票のように集計することで、入力データが識別対象のうちどれに該当するのかわかり、得票率 (該当率) として表現する。**Random Forest** による識別を行う場合、識別対象の特徴を十分に表すことのできる特徴ベクトルを定義することで、識別率を高めることができる。本節ではまず、本研究で用いる特徴ベクトルについて述べる。

4.3 顔の特徴ベクトル

顔の部位識別に用いる特徴ベクトルとしては、3 次元形状特徴の 1 つである **Fast point feature histograms (FPFH)** および鼻の先端を原点とした、各頂点の相対座標を組み合わせる。FPFH は 3 次元点群データに適した特徴記述子の 1 つであり、着

目する点とその近傍点との平均局率をヒストグラムの形式で記述したものである。本研究で扱う顔形状は、鼻や口のように部位ごとに異なる特徴的な曲率変化をもつので、FPFH が顔に適した特徴記述であると考え、これを特徴ベクトルの一部として利用する。FPFH を計算する際には各点の法線情報が必要となるので、深度マップからこれを取得する。また、予備実験により FPFH を用いた顔の識別を試みたところ、曲率の似通った領域で誤認識が生じる結果となったので、これを修正するために、鼻からの相対座標を特徴として加える工夫を行う。顔の各部位は動作によって相対的な位置関係がほとんど変化しないという特徴があるので、鼻を原点とし、3 次元座標値が -1.0 から 1.0 に収まるように正規化した値を特徴ベクトルに追加する。この相対座標値の追加により、頬および目のように本来は左右の区別のある領域が 1 つの部位として扱われている領域に対して、ある次元の正負が反転しているという左右を区別する情報が与えられる。これにより、左右の形状で異なる変形を生じている教師データからは、片側のみに生じた固有の変形を学習させることができると考える。顔の特徴ベクトルについてまとめると、本研究では 14 次元のベクトルから構成される FPFH と、鼻からの相対位置により求める 3 次元座標の、計 17 次元の情報を顔の特徴ベクトルとして定義する。

4.4 顔部位の学習と識別

前節で定義した特徴ベクトルに基づき、顔の部位を識別する方法について述べる。本研究では、顔の部位を識別するための教師データとして、顔の点群データから任意の 6 フレームを選択し、手作業により顔を構成する 4 種の部位およびノイズに相当する領域を指定する。ここで指定するノイズ領域とは、3 次元形状計測の失敗による、本来顔に存在しない凹凸を含む領域、および図 1 の上段中央にあるような手等の顔以外の物体の領域である。これらの領域は時間とともに出現位置が大きく変化するので、追跡対象から除外する目的で識別を行う。この作業は各部位に該当する部分をマウスで描いた軌跡により囲むことで行い、この作業に要する時間は 10 分未満である。**Random Forest** の学習を行う際にはまず、上記の方法により選択した各部位に該当する点群データに対して特徴ベクトルを計算する。そして、各点群データの特徴ベクトルとそれが属する部位の番号を 1 組の教師データとして **Random Forest** に与えることで学習を行う。**Random Forest** による顔の部位認識に関する研究では、複数人の顔形状を学習することで未知の顔の部位を識別できるという報告もあるが¹³⁾、本研究では入力データ特有のノイズ領域を排除し、追跡を確実にを行うために上記のような作業を行う。

識別処理では、学習後の **Random Forest** に対して顔形状 (頂点の特徴ベクトル) を入力することで、その頂点がどの部位に該当するのかわかり、該当率を得ることができる。本研究では、ある 1 種の部位が 90% 以上の該当率をもつ場合のみ、その頂点がある部位に属するものとする。この時、どの部位にも特定されない点が生じるが、第 6 章で述べる、追跡結果の統合処理において、これらの点も含めて顔全体の変形を求めるので、問題とはならない。

5. 顔形状の変形追跡

本章では、顔の部位毎に形状の時間変化を追跡するための手法について述べる。以降では、ある時刻 t における顔の点群データの集合を $F^t = (v_0^t, \dots, v_i^t, \dots, v_N^t)^T$ 、識別後の部位 c に属する点群データを $P_c^t = (v_x^t, v_y^t, \dots)^T$ と表記する。 P_c^t は F^t

Fig. 2 Flow diagram of the motion tracking and deformation of a face

の部分集合の関係にある。本研究では、各部位の変形を追跡するために Non-rigid Registration の 1 手法である L2+TPS⁶⁾ を用いるとともに、顔形状の特徴に合わせて次の 2 点の工夫を加える。L2+TPS は、位置合わせを行う 2 つの点群データを Gaussian Mixture Model によりモデル化し、両モデル間の L2 ノルムを最小化することで位置合わせを行う手法である。その際に、点群の移動を Thin-plate Spline(TPS) により補完することから、L2+TPS と記述する。

- 追跡点数を削減する。
- 隣接するフレーム間の比較ではなく、第 1 フレームと他のフレームを比較することで追跡を行う。

両者の工夫について、以下で詳しく述べる。

5.1 追跡点数の削減

L2+TPS は数千点規模の位置合わせ問題を想定した手法であり、本稿で想定する数十万点の点群データからなる顔形状を扱うことは困難である。しかし、L2+TPS は顔の表皮のような自由変形する物体を追跡する目的において有効な手法であることから、入力点数を削減する方法をとる (図 2 (a))。点数の削減には、対象物体のある領域を一定間隔のボクセルに区切り、各ボクセルを代表する点だけを残すことで点数の削減を行う、Approximate Voxel Grid¹⁹⁾ を用いる。これにより、各パーツの点群データ P_c^t を 2000 点以下まで削減することで形状のサブセット $\hat{P}_c^t$ を求め、L2+TPS による変形追跡を可能にする。点数の削減を行えば、変形を追跡されない点が生じるが、残された点は前章で述べた、どの部位にも識別されなかった点と同様に第 6 章で述べる手法により変形を与える。

5.2 変形追跡の対象フレームに関する考察

顔形状の変形を追跡する手順として、第 1 フレームから順に隣り合うフレーム間の変形を連続的に求める方法と第 1 フレームから他のフレームへの変形を直接求める方法がある。前者の方法は変形を追跡する際の形状間の差が小さいので、誤りが生じにくいという利点がある一方で、総フレーム数が多い場合には連続的な変化により誤差が蓄積するという問題がある。後者の方法は形状間の差が大きいので、誤りの生じる可能性が高い一方で、各フレームでの変形は 1 回のみなので誤差の蓄積は無く、対象物体の運動が変形をして元に戻るような動作の場合、前者の方法よりも誤りの発生を抑えることができるという利点がある。顔の運動を対象とする場合、身体動作などと比べて変形の範囲は限定的であることに加え、部位識別によりフレーム間の形状同士の同一性はすでに与えられているので、本研究では

Fig. 3 RBF based deformation

後者の、第 1 フレームから他のフレームへの変形を求める方法をとる (図 2 (b))。

5.3 Non-rigid Registration 手法に基づく変形追跡

部位 c において、第 1 フレームの形状 $\hat{P}_c^1$ を L2+TPS により第 t フレームの形状 $\hat{P}_c^t$ に変形するための処理を次式で表す。

$$L2 + TPS_{\hat{P}_c^1 \rightarrow \hat{P}_c^t}(\hat{P}_c^0) = Q_c^t \quad (1)$$

第 t フレームに対する変形結果は、時刻を表す添え字 t を用いて Q_c^t と表現する。 Q_c^t はいずれの時刻においても $\hat{P}_c^1$ と同数の点数をもち、この時点で顔の点群データの一部 $\hat{P}_c^1$ の追跡結果を得ることができる。次章では、この Q_c^t を用いて、顔全体の形状 F^t を変形するための処理について述べる。

6. 追跡結果の統合と顔形状の変形

本研究では、前章までに述べた手法により顔上のいくつかの代表点の変形情報 Q_c^t が得られているので、顔形状の変形に適した手法を適用することで、顔全体の変形を行う。なお、本章では顔形状の部位の差を考慮しないので、部位毎に追跡した点の情報を統合し Q^t と表現する。 Q^t を構成する全ての点は、各時刻の入力顔形状に一致する座標値をもっているため、変形に用いる手法は変形後の形状が特定の点を通るような性質のある手法が望ましい。そこで、Radial Basis Function (RBF) に基づく形状変形手法^{20, 18, 21)}を用いる。本手法の概要を以下で述べる。

6.1 RBF に基づく形状変形

3 次元形状を変形する典型的な方法に、特定の点の移動ベクトルが与えられた際に、周囲の点をその移動ベクトルに基づき点の移動量を補間しながら変形する手法がある。例えば、今、平面状の 1 点に移動ベクトル T_1 を与え、基底関数にガウス関数を用いてこの変形を行った場合、基底関数と同様の形状をした起伏が生じる (図 3)。距離がゼロであった場合のガウス関数の値は 1 であるので、変形の中心位置は T_1 の表す位置に必ず移動する。しかし、2 点以上の移動ベクトル T_1, T_2 が与えられた場合では、移動ベクトルが互いに影響を与えるので、変形結果が特定の点を通らないという問題が生じる。この問題を解決するために、移動ベクトルの与えられている点同士の距離から、お互いに与えあう影響度を考慮することで、新たな移動ベクトルを求める。なお、本研究においては、初期の移動ベクトルは $T^t = Q^t - Q^1$ となるので、以降の説明ではこれらの記号を用いる。各点の移動ベクトルは、移動ベクトルの起点 Q^1 を構成する点同士の距離に応じて影響を受ける。よって、各点同士の距離を次式で示す行列の形で表現する。 Q_i^1 の添え字 i は i 番目の移動点を表す。この行列 $D(Q^1, Q^1)$ の各要素に RBF を適用したものが、 Q^1 を構成する各点が他の点の移動ベクトルから受ける影響となる。そこで、影響を受けた後に各点の移動が T^t となるような、新たなベクトル (制御ベクトルとよぶ) $C^t = (C_0^t, \dots, C_m^t)^T$ を用いて、両者の関係を示すと式 (3) となる。

$$D(Q^1, Q^1) = \begin{bmatrix} \|Q_0^1 - Q_0^1\| \cdots \|Q_0^1 - Q_i^1\| \cdots \|Q_0^1 - Q_m^1\| \\ \vdots \\ \|Q_j^1 - Q_0^1\| \cdots \|Q_j^1 - Q_i^1\| \cdots \|Q_j^1 - Q_m^1\| \\ \vdots \\ \|Q_m^1 - Q_0^1\| \cdots \|Q_m^1 - Q_i^1\| \cdots \|Q_m^1 - Q_m^1\| \end{bmatrix} \quad (2)$$

$$T^t = RBF(D(Q^1, Q^1))C^t \quad (3)$$

そして、 T^t および Q^1 は既知であるので、点同士の影響を与えあう行列の逆行列を求めることで、制御ベクトル C^t を求めることができる。

$$RBF(D(Q^1, Q^1))^{-1}T^t = C^t \quad (4)$$

この制御ベクトル C^t を用いることで、任意の点 x における時刻 t での移動ベクトル X^t は次の式により与えられる。

$$X^t = RBF(\hat{D}(x, Q^1))C^t \quad (5)$$

$$\hat{D}(x, Q^1) = (\|x - Q_0^1\|, \cdots, \|x - Q_m^1\|) \quad (6)$$

式 6 を顔を構成する全点に適用することで、顔形状を任意の時刻における形状に変形することができる。

6.2 基底関数に関する考察

RBF に基づく形状変形では、移動ベクトルの与えられた点（代表点）以外は RBF に基づく補間によってその移動量が決定する。よって、RBF に対して具体的にどのような関数を用いるかが、変形結果に大きな影響を与える。本研究では、顔の部位毎に独立して変形の追跡を行っているので、部位間の領域に動きを代表する点がとられにくいという性質がある。このような点の配置に対して、ガウス関数のような局所的に影響を与える基底関数に基づき変形を行うと、点の密度に応じて移動量が増減するので、各部位が独立して動いているような結果となる。そこで、代表点が全体の動きに影響を与えられるように、代表点間の 3 次元空間上での距離を基底関数として与える。

$$RBF(x) = \|x\| \quad (7)$$

この基底関数により、局所的な点の偏りに左右されない形状変形を実現する。同様の基底関数による顔の変形は Noh らの研究にも見られる¹⁸⁾。

7. 実験と評価

提案手法の精度を評価するために、顔を手で叩く動作 (Slap)、笑顔の表情を作る動作 (Smile)、縦方向に縮めた顔を大きく伸ばす動作 (Stretch)、および一般物体に対する適用例としてボールを握る動作 (Grip) の 4 種の動作を 3 次元形状計測手法²²⁾により取得し、実験を行う。なお、顔以外の対象物として、ボールを握る動作を追加する理由としては、本研究で主張する、大きく弾性変形する部位 (ボール) と平行移動が主体となる部位 (指) が混在する対象に対する変形追跡の有効性を示す一例として評価に加える。ただし、この部位識別に関しては多くの手作業により 4 本の指およびボールの識別を実施しており、提案手法の一般物体に対する適用可能性を主張するものではない。各動作に含まれるフレーム数および第 1 フレームの点数を表 7. に示す。また、図 4 に各動作における追跡後の形状の一部を示す。

7.1 追跡精度の評価および考察

提案手法の評価を行うための指標として、変形追跡後の形状に含まれる点から入力形状への最短距離を合計し、点の数および

Fig. 4 Examples of the input motions and tracked ones (upper row:input, bottom row:tracked) (a)Slap (b)Smile (c)Stretch (d)Grip

Table 1 Size of frames and points of input data

Data set	Slap	Smile	Stretch	Grip
Size of frames	46	55	48	48
Number of points	115169	190599	321142	161517

フレーム数で平均したものを平均誤差として定義する。そして、提案手法および 2.2 で述べた 4 種の Non-rigid Registration 手法 (L2+TPS⁶⁾, L2+GRBF⁶⁾, EM+TPS¹⁰⁾, CPD⁹⁾) を顔形状全体に適用した場合における平均誤差を比較する。図 5 に示す結果から、Slap, Stretch および Grip 動作では、従来手法により変形を追跡した結果に比べて、平均誤差を約 13% から 45% 削減することができた。特に、Slap 動作および Grip 動作のように、外部から力を加えることで対象物の一部のみが大きく変形する動作において誤差の減少量が多く、部位識別に基づく変形追跡が有効であることが確認できた。提案手法において精度が向上した理由としては次のような理由が考えられる。Non-rigid Registration 手法は、頑健に対象物体の変形を追跡できるようにするために、基底関数に基づき局所的な変形の偏りを無くするような処理が行われる (L2+TPS⁶⁾ におけるスプライン関数, L2+GRBF⁶⁾ におけるガウス関数等)。この処理は、均質な材質をもつ物体に対しては有効に働くものの、顔のような部位毎に変形の異なる対象物では問題となる。例えば、図 6 に示す追跡誤差を評価した図のうち、(b) および (c) の左側 2 列にある結果から、鼻の部位において誤差に大きな違いがあることが分かる。この時の顔の変形は、頬が大きく変形する一方で鼻は横方向に移動しているのみであり、両者の変形の様相は大きく異なる。しかし、Non-rigid Registration 手法では、頬の大きな変形を鼻や目の領域を含めて滑らかに拡散させた結果、周囲に誤差が広がるという現象が見られる。このように、Non-rigid Registration 手法の利点が顔を対象とした場合には有効に働かなかった点を、本研究で提案する手法が解決したことにより、精度が向上したものと考えられる。一方で、smile 動作では他の手法に比べて改善が見られなかった。これは、smile 動作の形状変形が小さく、従来手法でも十分に追跡が可能な対象であったためであり、従来手法の適用方法等に問題がないことを示している。

次に、特に変形および誤差の大きい Slap 動作に対する結果に

ついてより詳細に述べる。図 6 に、誤差の生じた領域を可視化した結果を示す。提案手法では、向かって左側の頬が叩かれることによって生じるう表皮のうねりによる誤差が減少していることが確認できる。また、鼻筋の部分に関しても誤差が減少している。これらのことから、部位毎に形状を追跡することにより、個々の部位の性質に適した変形の追跡が可能となったといえる。

7.2 追跡点の同一性に関する評価と考察

7.1 節では、入力形状と追跡後の形状間の距離を比較することで評価を行った。これは、追跡後の形状の凹凸が入力形状に一致していることを評価することはできるが、ある部位を表す点の時間変化を正しく追跡できているかどうかは明らかではない。例えば、口角上にある点が他のフレームでは頬上に移動していたとしても、頬にそのフレームにおける口角と同様の凹凸があれば誤差として現れない。そこで、特に変形が大きい Slap 動作の口元に着目し、図 7 上部に示す 5 点に関して手作業により、フレーム毎に同一点と思われる位置を指定し、提案手法により追跡した点との誤差を求める。フレーム毎の誤差を図 8 に示すとともに、図 7 に提案手法および最も誤差の少なかった L2+TPS 手法における結果を示す。図 8 に示す結果から、提案手法は従来手法に比べて同一点を精度よく追跡していることが分かる。図 7 に示した結果から、従来手法では手の形状の影響により向かって左側の口角を表す点が手の方へ移動しているが、提案手法では口角の位置をより少ない誤差で追跡できている。これは、提案手法が部位の識別を事前に行うことで、変形追跡の範囲を限定することができたからである。

8. おわりに

本稿では、顔の 3 次元点群データから連続的な動きを取得するためのマーカレス顔モーションキャプチャ手法を提案した。その成果を以下に纏める。

- (1) 顔形状の変形を追跡する際に、顔形状を構成する点群データを鼻、口、目、頬の 4 種類の部位およびノイズ領域に分けることで変形追跡を行う範囲を限定し、追跡精度を高めることができた。
- (2) RBF に基づく形状変形手法を利用することで、部位毎に独立して行った顔の変形追跡結果を統合し、顔全体の動きとして表現することができた。
- (3) 4 種類の異なる動作を用いて、入力データと変形追跡後の形状との差および特定の追跡点の精度評価を行い、従来手法に対する優位性を示した。

今後は、画像データから得られる特徴量を形状特徴量と組み合わせることで、追跡の際の誤対応を減らし、さらなる精度向上を目指す。

謝 辞

本研究の一部は、内閣府・最先端・次世代研究開発支援プログラム (LR030) の助成を受けて実施されたものである。

参 考 文 献

- 1) Weise, T., Bouaziz, S., Li, H., Pauly, M.: Realtime performance-based facial animation. In: ACM SIGGRAPH 2011 papers. SIGGRAPH '11, New York, NY, USA, ACM, 77 (2011) 1.
- 2) Zhang, W., Wang, Q., Tang, X.: Real time feature based 3-d deformable face tracking. In: Proceedings of the 10th European Conference on Computer Vision: Part II. ECCV '08, Berlin, Heidelberg, Springer-Verlag (2008) 720.
- 3) Huang, H., Chai, J., Tong, X., Wu, H.T.: Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. ACM Trans. Graph. 30, 74 (2011) 1.
- 4) Bickel, B., Lang, M., Botsch, M., Otaduy, M.A., Gross, M.: Pose-space animation and transfer of facial details. In: Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. SCA '08, Eurographics Association (2008) 57.
- 5) Sibbing, D., Habbeke, M., Kobbelt, L.: Markerless reconstruction and synthesis of dynamic facial expressions. Comput. Vis. Image Underst. 115 (2011) 668.
- 6) Jian, B., Vemuri, B.C.: Robust point set registration using gaussian mixture models. IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1633.
- 7) Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. Pattern Analysis and Machine Intelligence, IEEE Transactions on 11 (Jun) 567.
- 8) Broomhead, D.S., Lowe, D.: Multivariable Functional Interpolation and Adaptive Networks. Complex Systems 2 (1988) 321.
- 9) Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2010) 2262.
- 10) Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Comput. Vis. Image Underst. 89 (2003) 114.
- 11) Tevs, A., Berner, A., Wand, M., Ihrke, I., Bokeloh, M., Kerber, J., Seidel, H.P.: Animation cartography –intrinsic reconstruction of shape and motion. ACM Trans. Graph. 31, 12 (2012) 1.
- 12) Cao, X., Wei, Y., Wen, F., Sun, J.: Face alignment by explicit shape regression. In: CVPR, IEEE (2012) 2887.
- 13) Dantone, M., Gall, J., Fanelli, G., Gool, L.V.: Real-time facial feature detection using conditional regression forests. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. (2012).
- 14) Fanelli, G., Gall, J., Van Gool, L.: Real time head pose estimation with random regression forests. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. CVPR '11, Washington, DC, USA, IEEE Computer Society (2011) 617.
- 15) Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. CVPR '11, Washington, DC, USA, IEEE Computer Society (2011) 1297.
- 16) Breiman, L.: Random forests. Mach. Learn. 45 (2001) 5.
- 17) Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. In: Proceedings of the 2009 IEEE international conference on Robotics and Automation. ICRA'09, Piscataway, NJ, USA, IEEE Press (2009) 1848.
- 18) Noh, J.y., Fidaleo, D., Neumann, U.: Animated deformations with radial basis functions. In: Proceedings of the ACM symposium on Virtual reality software and technology. VRST '00, New York, NY, USA, ACM (2000) 166.
- 19) Rusu, R.B., Cousins, S.: 3D is here: Point Cloud Library (PCL). In: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China (2011).
- 20) Botsch, M., Kobbelt, L.: Real-time shape editing using radial basis functions. Comput. Graph. Forum 24 (2005) 611.
- 21) Sumner, R.W., Schmid, J., Pauly, M.: Embedded deformation for shape manipulation. In: ACM SIGGRAPH 2007 papers. SIGGRAPH '07, ACM (2007).
- 22) Sagawa, R., Kawasaki, H., Furukawa, R., Kiyota, S.: Dense one-shot 3d reconstruction by detecting continuous regions with parallel line projection. In: Proc. 13th IEEE International Conference on Computer Vision (ICCV 2011). (2011) 1911.

Fig. 5 Graph of the mean distances according to the algorithms

Fig. 6 Visualization of errors on the face (a)Input motion (b)Results tracked by proposed method (c)Results tracked by L2+TPS

Fig. 7 Movements of the key points on around the mouth (a)Ground truth (b)Proposed method (c)L2+TPS

Fig. 8 Compressions of the tracking error of the key points around the mouth