

顔の器官識別に基づくマーカレスモーションキャプチャの提案

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

○赤木康宏†, 古川亮‡, 佐川立昌††, 小川原光一‡‡, 清田祥太†, 沖佳憲†, 森永寛紀†, 川崎洋†
Yasuhiro AKAGI†, Ryo FURUKAWA‡, Ryusuke SAGAWA††, Koichi OGAWARA‡‡, Syota KIYOTA†,
Yoshinori OKI†, Hiroki MORINAGA† and Hiroshi KAWASAKI†

†: 鹿児島大学, {akagi,kawasaki}@ibe.kagoshima-u.ac.jp ‡: 広島市立大学, ryo-f@hiroshima-cu.ac.jp
††: 産業技術総合研究所, ryusuke.sagawa@aist.go.jp ‡‡: 和歌山大学, ogawara@sys.wakayama-u.ac.jp

概要: 本稿では, 顔の動きを密に計測した点群データに対して, フレーム間の連続性を定義するための, マーカレスモーションキャプチャ手法を提案する. 一般的な形状計測装置により得られた点群データには時間軸上の連続性が与えられておらず, これを利用する際には様々な不利益が生じる. 本稿では, 顔を4種の部位(鼻, 口, 目, 頬)に分類し, それぞれの部位の変形を **Non-rigid Registration** 手法に基づき追跡する手法を提案する. 顔形状の変形追跡を部位毎に行うことで, **Non-rigid Registration** のみを利用した場合に比べ, 形状追跡の平均誤差を 8.7mm から 6.1mm に低減させることができた.

<キーワード> マーカレスモーションキャプチャ, 顔, 器官識別

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では, 身体動作のように関節の動きを推定する問題に比べて表皮の自由変形を追跡する必要があり, 異なるフレーム間の形状を対応付けることが難しいという問題がある. そこで, 多く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設置する手法が用いられている. 近年では, マーカの設置に要する手間の削減やマーカの設置が困難な動作の取得を目的とした, マーカレス顔モーションキャプチャ手法[1][2]が提案されているが, 口元などの変形の自由度の高い場所では追跡が困難であるという問題が残されている. 一方で, **Non-rigid Registration** アルゴリズムに基づく形状変形を追跡手法[5-8]は自由変形する点群の追跡, 特に人体形状の変形追跡に有効である. 本稿ではこの特徴に着目し, 顔の形状追跡に特化することで, 追跡精度を向上させるための手法を提案する.

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これまでの研究について述べる.

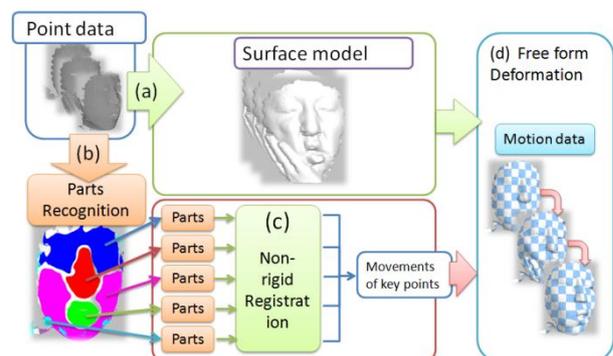


図1 提案手法の処理概要

2.1 顔のモーションキャプチャに関する研究

表情変化等の顔の動きを計測するための手法には, 顔に特徴的な位置(特徴点)を追跡するためのマーカを用いる手法と, 特徴点を画像等から検出して追跡するマーカレスの手法がある. 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[3]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. Bickel らの研究[4]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. これらのマーカを用いる方法では, 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利用できないという問題がある.

後者のマーカを用いずに顔の動きを取得する研究では、マーカに由来する撮影時の手間や制約条件が緩和されるものの、動作を正確に追跡するための工夫が重要となってくる。Weise らの研究^[1]では、撮影者の顔の 3 次元変形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、特徴点の位置からリアルタイムに 3 次元顔形状を合成することに可能にした。Sibbing らの研究^[2]では、顔の画像(陰影)情報から異なるフレーム間の対応関係を求め、その運動を追跡することで顔の変形を再現している。これら 2 つの方法では、3 次元形状と対になる画像情報を同時に取得する必要がある。

2.2 Non-Rigid Registration による物体の変形追跡

一方で、取得した顔の 3 次元形状のみを用いて、形状の変形を追跡する手法としては、Non-rigid Registration アルゴリズムを用いて、異なるフレーム間の顔形状の対応関係を直接求める方法[5-8]がある。Tevs らの研究[5]では、顔形状のポリゴンメッシュをグラフとみなし、異なるフレーム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。しかし、処理の過程で形状が平滑化され、細部の形状が失われる、数時間の計算時間を要するという問題もある。Jian らの研究[6]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) に基づき表現し、形状変形を 2 つの GMM 同士の L2 Distance (L2) を最小化する問題として表現している。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) および Gaussian Radial Basis Function (GBRF) を基底関数に用いている。また、その他の代表的な GMM に基づく手法としては、Coherent Point Drift (CPD) と呼ばれる Myronenko らの研究[7] および EM アルゴリズムに基づき GMM 同士の距離を最小化する、Chui らの研究[8]がある。以降は、ここに挙げた 4 種の手法を以下の略称で表現する。

- L2-TPS :Jian et. Al.[6]
- L2-GBRF:Jian et. Al. [6]
- CPD :Myronenko et. Al. [7]
- EM-TPS :Chui et. Al. [8]

3. 提案手法の概要

提案手法では、佐川らの手法^[9]により取得した顔形状の時系列点群データを用いて、第1フレームの形状から他フレームへの変形を求めることで、顔の動

きを追跡する(図1)。この時、変形の追跡には Non-rigid Registration アルゴリズムの1つである L2-TPS を用いる。ただし、単純に L2-TPS を用いた場合では顔の部位毎の皮膚の厚みによる変形の違いを考慮できず、特に口元などの変形が大きい個所では対応箇所を正しく発見できないという問題が生じる。そこで、本研究では顔を 5 種の部位(鼻、口、目、頬、その他)に分類し(図1(b))、変形の追跡を行う(図1(c))。その後、個々の部位の変形を統合することで、顔全体の変形を求める(図1(d))。本稿では、(b)の詳細について 4 章で、(c)および(d)の詳細について 5 章で述べる

4. 顔の部位識別

顔表面の動きは、鼻や目のように表皮が薄く骨格の影響を受けやすい部位と口や頬のように表皮が厚く、自由に動くことのできる部位がある。この顔に固有の特徴から、本研究では顔の動きを部位ごとに追跡し、その結果を統合することで顔全体の動きを取得する方法をとる。本章ではまず、顔の 3 次元形状特徴を用いた、各部位の識別手法について述べる。

4.1 顔の部位定義

本節ではまず、本研究で用いる顔の部位の分類について述べる。本研究では顔の部位を、鼻、口、目、頬の 4 種類に分類する(目、頬の左右は区別しない)と共に、顔に重なる障害物やノイズ等を分離するためのノイズ領域を加えた 5 種の領域を識別の対象とする。

4.2 顔の形状特徴

本節では、顔の各部位を識別するための形状特徴を表す特徴ベクトルについて述べる。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)[10]、および鼻の先端を原点とした、各頂点の相対座標を用いる。FPFH は鼻や口等の特徴的な凹凸をもつ形状を識別する際に有効であり、顔に適した特徴量である。また、顔の動作によって部位の相対的な位置関係が変化することはないので、鼻を中心とした相対座標を特徴量に加えた。本研究では、14 次元のベクトルから構成される FPFH と 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。

4.3 Random Forest[11]に基づく部位識別

4.2 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、各部位を識別するための教師データとして、入力された顔の

点群データから任意の 6 フレームを選択し、手作業により各部位の分類を与える。この作業は各部位に該当する部分をマウスに描いた軌跡により囲むことで行う。この時、ノイズに相当する領域も同時に指定することで、顔以外の領域を器官として誤認識することを防ぐ。上記の方法により作成した教師データおよび、4.2 で定義した特徴ベクトルの組から、部位識別用の Random Forest を構築する。この Random Forest に対して未知の顔形状を入力すると、各頂点の各部位に対する一致度が算出される。本研究では 90% 以上の一致度を持つ頂点のみを、その部位に対応した頂点として扱う。以降は、ある時刻 t における顔の点群データの集合を F^t 、識別後の部位 i を構成する点群データを P_i^t と表記する。Random Forest による識別では、式 1 に示すように F^t から P_i^t (鼻、口、目、頬) とノイズ領域 e^t に分離する。

$$\text{RandomForest}(F^t) \rightarrow P_i^t + e^t \quad \dots(1)$$

次章では、 P_i^t を用いて部位毎に形状の変形を追跡する手法について述べる。

5. Non-rigid Registration による変形追跡

4 章で述べた手法により得られた顔の各部位 P_i^t の時間軸上での対応関係を求めるために、CT 画像からの臓器の認識などにも用いられる、ロバスト性を有する Non-rigid Registration アルゴリズム^[2] (詳細は 2 章参照、以降 L2-TPS) に基づく形状変形の追跡を行う。そして、各部位の変形追跡を行った後に、顔全体の動きを得るために Free-Form Deformation による動きの統合を行う(図 2)。

本研究では、入力データの開始フレームにおける顔形状 F^0 から任意の時刻 t における形状 F^t への変形(各点の平行移動量)を求める方法をとる。この他にも、隣接するフレーム間の変形を順次求める方法もあるが、この方法では変形を繰り返すうちに形状の歪みが蓄積しやすいという問題が生じるので、前者の方法を用いる。

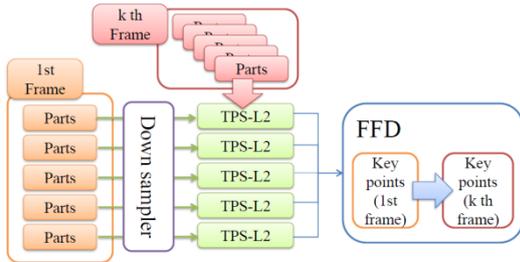


図 2 変形追跡処理

5.1 疎な部位変形の追跡

本稿で扱う顔の点群データ F^t は非常に密であり、 P_i^t に含まれる点群データであっても数万点に及ぶ。一方で L2-TPS は数千点の位置合わせを想定した手法であり、そのまま点群の移動追跡に用いた場合、計算に数時間を要する。そこで、点群データ P_i^t を Approximate Voxel Grid^[12] により 2000 点以下まで削減することで形状のサブセット p_i^t を求め、L2-TPS による変形追跡を可能にする。L2-TPS により、時刻 0 における形状 p_i^0 から時刻 t における形状 p_i^t への位置合わせを行い、 p_i^0 を変形した形状である \hat{p}_i^0 を得る(式 2)。

$$\text{L2-TPS}(p_i^0 \rightarrow p_i^t) = \hat{p}_i^0 \quad \dots(2)$$

ここで、 p_i^0 の各点の平行移動量を T_i^t とする(式 3)。

$$\hat{p}_i^0 - p_i^0 = T_i^t \quad \dots(3)$$

顔の各部位および各時刻での T_i^t を求めることで、個々の部位の動きを疎に得ることができる。次節では、 T_i^t を顔全体の動きとして統合するための手法について述べる。

5.2 Free-Form Deformation による動きの統合

顔の部位毎に得られた変形を表す T_i^t は顔を構成する点群データに対して疎であるので、顔全体の変形を得るためには T_i^t を補間するような処理が必要となる。そこで、少数の代表点の変形情報から、形状全体の大域的な変形を表現することのできる Free-Form Deformation(FFD)手法^[13] に基づき統合を行う。

FFD による変形を行う上で重要となる点は、代表点の動きを補間する際に用いる、各点の影響力を決定する基底関数をどのように与えるかにある。ガウス関数および Spline 関数のように、山型の形状をもつ基底関数を用いた場合、ある代表点の影響範囲はその周辺の限られる。本研究では、Random Forest により顔部位の分離を行っているので、部位間の中間領域に代表点がとられにくいという特徴がある。このような代表点に対して山型の基底関数を用いた FFD を行うと、部位間の補間が十分に行われず、各部位が独立して動いているような変形が行われる。そこで、代表点全体の影響を考慮できるように、代表点間の距離を基底関数とする FFD を行う。

FFD を行う際には、各部位の区別は行わないので、各部位の代表点 p_i^t および移動量 T_i^t を統合し、 K^t (K_i^t は代表点集合の i 番目の点) および T^t と表記する。これを用いて FFD による変形を表すと式 4 となる。式 4 のうち、 K^t および T^t は既知であるので、変形を制御するベクトル C^t を求める。

$$F^t = \left\| K^0_i - K^0_j \right\|_{i,j=0 \sim N} C^t \quad \dots (4)$$

最終的に、ある時刻 t における顔の点群データの F^t は式5により求まる。

$$F^t = \left\| K^0_i - F^0 \right\|_{i=0 \sim N} C^t \quad \dots (5)$$

上記の一連の処理により、顔モーションを連続したFFDの集合として取得することができる。

6. 実験と評価

提案手法の精度および実行速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)、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)を対象に、本手法による動作の取得を行った(図3, 7)。各データに含まれるフレーム数および頂点数を表1に示す。

表1 計測データ

	Slap	Smile	Grip
フレーム数	46	55	48
頂点数	115169	190599	161517

6.1 追跡精度の評価および考察

提案手法の評価としてまず、顔全体が大きく変形している動作である「Grip」を対象とした精度の検証を行う。精度評価を行うために、顔や表情認識を行う際に広く用いられる特徴点(20点)[14]を人間が手作業により指定することで、評価用の基準データとした。これらの点に対して、提案手法および第2章で述べた4種のNon-rigid Registrationアルゴリズムにより追跡された該当箇所との誤差を計測した(図4, 表2)。図4は、各手法のフレーム毎の平均誤差を示したグラフである。この結果から、提案手法は顔の部位別に対応点を追跡することで誤差の減少に成功していることがわかる。表2に示すように、単純にNon-rigid Registrationアルゴリズムを適用した場合に比べて、約30%以上誤差を減らすことができた。また、図4に示したグラフから、フレーム間の誤差の変動が小さいことも特徴である。これは、顔の動きを連続的なアニメーションとしてみた際に、フレーム間で大きく対応点を間違えることが少ないことを表している。この特徴は図3に示す結果にも表れており、従来手法では形状が大きく歪むフレームが含まれている。

次に、特に変形が大きい口元に着目し、図5に示す5点に着目した誤差の評価結果を図6および表2に示す。表2に示した平均誤差の値から、提案手法のみ全点での平均に対して誤差が減少しており、他

の手法では誤差が増大していることがわかる。この結果は、口や鼻のような特徴的な形状をもつ部位についてはRandom Forestにより正しい識別がしやすいので、変形の追跡もまた、より正確に行われていることを示している。他の手法では、手の形状からの影響を受け大きく歪んだり、フレーム毎に口を構成する頂点が増えたりする等の問題が生じやすい傾向にあった。よって、変形追跡の前処理として顔を部位毎に分類する方法は、形状が複雑で変形の大きな箇所に対して特に有効であるといえる。

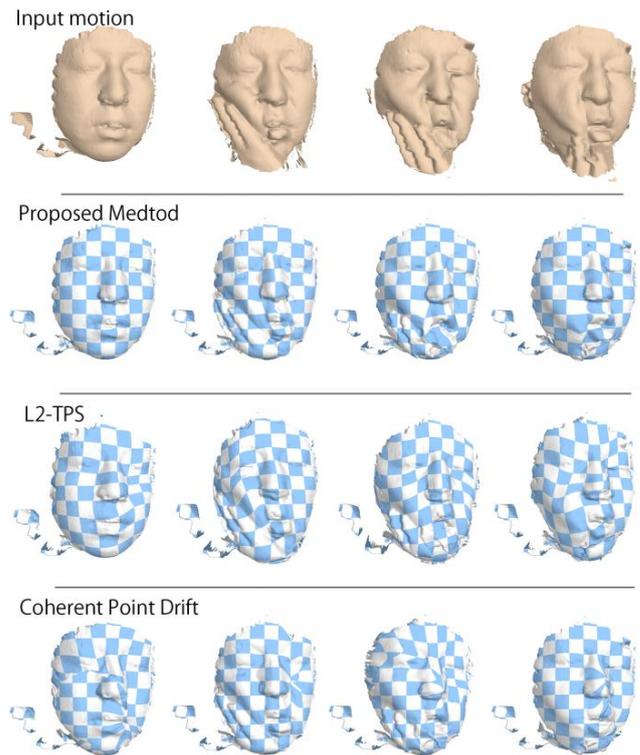


図3 Registration手法による変形の違い

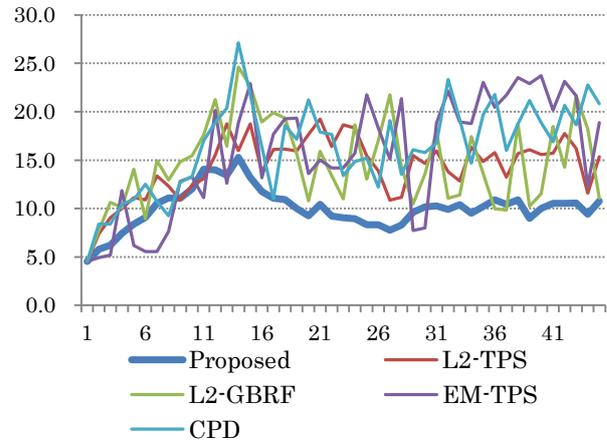
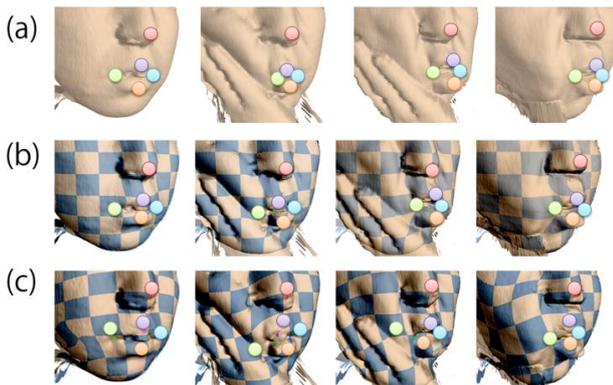


図4 Slap動作におけるフレームごとの誤差変化

表 2 Slap 動作における平均誤差および RMSE

	全点		口元のみ	
	平均誤差 (mm)	RMSE (mm)	平均誤差 (mm)	RMSE (mm)
Proposed	6.1	7.0	4.7	5.5
L2-TPS	8.7	9.5	10.1	10.7
L2-GBRF	9.0	9.6	11.3	11.8
EMTPS	9.5	10.4	11.1	11.8
CPD	9.9	11.0	11.9	12.7



(a) 評価基準点, (b) 提案手法, (c) L2-TPS
図 5 顔に特徴的な箇所の誤差評価

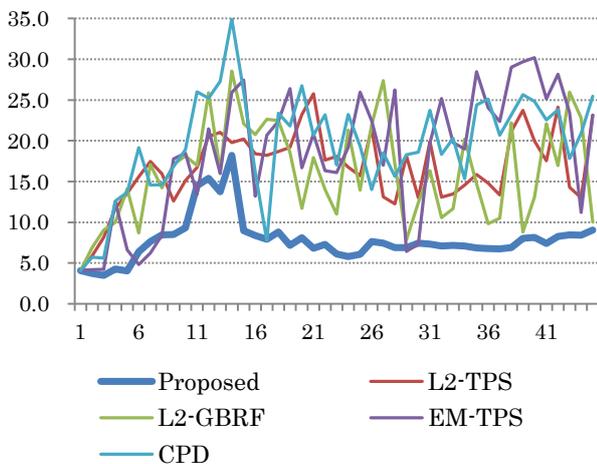
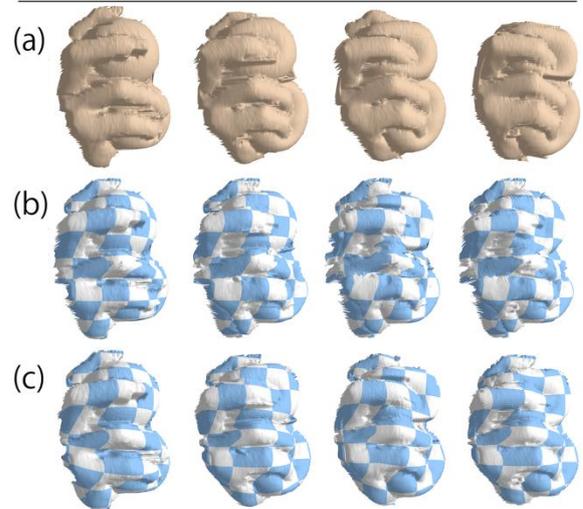
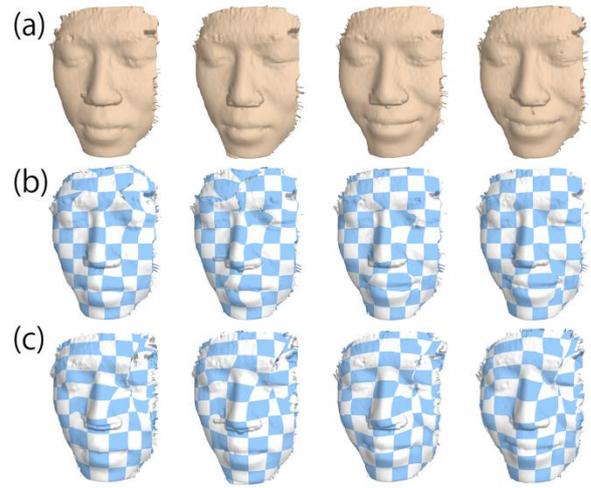


図 6 Slap 動作における口および鼻の追跡誤差

次に、図 7 の上部に Smile 動作の比較結果を示す。図 7 では、提案手法および他の手法で最も誤差の少なかった L2-TPS 手法による結果を示している。Smile 動作は無表情から笑顔を作る動作を緩やかに行ったものであり、手法による結果の違いは少なかった。しかし、L2-TPS 手法による結果では左目付近で形状が不自然に大きく歪む問題が生じている。このような形状が比較的単調で、



(a) 入力データ, (b) 提案手法, (c) L2-TPS
図 7 Smile および Grip 動作の比較

緩やかな動作を行う場面においては、提案手法の従来手法に対する優位性は少ないものの、変形の安定性が増していることが確認できた。

最後に、顔以外の複雑な形状変形への適用例として、ゴム製のボールを握りこむ Grip 動作の結果を図 7 の下部に示す。この動作では、2つのボールとそれぞれの指を異なる部位として学習・識別し提案手法を適用した。Grip 動作では、変形結果の差は少なく、どの手法においても安定して変形追跡を行うことができた。これは、Grip 動作のもつ、特徴的かつ大きな凹凸を多数含み、フレーム間の動作量が小さいという特徴から、フレーム間の点群同士の対応関係を発見しやすい問題であったことが原因と考えられる。この結果から、本稿で提案する手法は、起伏の多い形状となだらかな形状が混在する対象に対して有効な手法であることいえる。

6.2 処理速度の比較

本節では、提案手法および従来手法の処理速度の比較を行う。提案手法は顔の部位を識別し、部位毎に独立して変形を求めているので、識別のための処理が追加されるという特徴がある。そこで、処理全体に要する時間の計測を行い、従来手法との比較を行う(表3)。表3に示すように、提案手法は従来手法に対して同等かより短時間で処理が完了している。提案手法では、部位識別のための処理を必要とする一方で、位置合わせの際には比較する対象の点数を限定することができるので、位置合わせ手法自体の処理速度は短くなる。その結果、処理全体の時間で見た場合には、計算時間を短縮することができた。

表3 形状追跡処理に要する時間

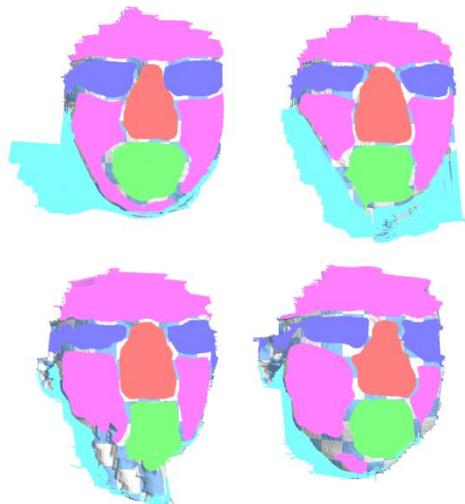
		Calc. Time (sec.)		
		Slap	Smile	Grip
Method	Proposed	811	831	1111
	L2-TPS	1722	867	1203
	L2-GBRF	1659	837	1167
	EM-TPS	4189	4521	5660
	CPD	26227	14805	7365

6.3 Random Forest による顔の部位識別

第4章で述べた Random Forest による顔の部位識別結果を図8に示す。図8は識別結果を5種の部位毎に色分けしたものである。各部位を表す領域は、概して顔の角度や変形に対応して正しく分離できているが、図8右下の頬に相当する領域(紫)の部分にあるような、領域の分断や大幅な領域の減少が見られる。このような領域の変形を求める際には、形状の歪みが生じる可能性もあるが、Gaussian Mixture Model に基づく位置合わせ手法は、欠損のある形状の位置合わせに対しても頑健性を有するので、今回行った実験では大きな問題は見られなかった。また、本研究ではエラー領域(シアン)以外にも、識別のための表価値が低い部位(図8で色分けのされていない領域)を位置合わせの対象に含めないことで、障害物(手)を含む領域を除外できていることが図7から確認できる。

7. 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。顔を5種類の部位に分けて形状



赤：鼻，緑：口，青：目，
紫：頬，シアン：ノイズ

図8 Random Forest による顔の部位識別

追跡を行うことで、点群間の対応関係を求める範囲を効果的に削減することができ、追跡精度の向上を可能にした。特に、口元のような形状の自由度が高く動きが複雑であるような部位において、精度を大幅に向上させることができた。また、顔の各部位を3次元点群データから識別するための特徴ベクトルを定義することで、顔の各部位を Random Forest により識別する手法を提案した。そして、部位毎に追跡した変形を顔の変形に適した FFD 手法により統合することで、顔全体の動きを追跡することができた。評価実験により、従来手法では 8.7mm あった平均誤差を、6.1mm に削減することができた。

本手法は、顔の各部位を識別するために、入力形状に対して6フレーム分の顔の部位の情報を手作業により与える必要があり、事前準備が必要である。今後は、より一般化した顔の部位識別手法を用いる等の工夫が必要である。また、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、位置合わせの際の誤対応を削減し、さらなる精度向上を目指す。

参考文献

- [1] Weise, T., Bouaziz, S., Li, H., Pauly, M.: Realtime performance-based facial animation. SIGGRAPH '11, pp.77:1-10, 2011.
- [2] Sibbing, D., Habbecke, M., Kobbelt, L.: Markerless reconstruction and synthesis of dynamic facial expressions. Comput. Vis. Image Underst., 115, pp. 668-680, 2011.
- [3] Huang, H., Chai, J., Tong, X., Wu, H.T.: Leveraging

motion capture and 3d scanning for high-delity facial performance acquisition. *ACM Trans. Graph.*, 30, pp. 74:1-74:10, 2011.

- [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*, pp. 57-66, 2008.
- [5] 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, pp.12:1-12:15, 2012.
- [6] Jian, B., Vemuri, B.C.: Robust point set registration using gaussian mixture models. *PAMI*, 33, pp.1633-1645, 2011
- [7] Myronenko, A., Song, X.: Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, pp. 2262-2275, 2010.
- [8] Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Underst.*, 89, pp. 114-141, 2003.
- [9] Sagawa, R., Kawasaki, H., Furukawa, R., Kiyota, S.: Dense one-shot 3d reconstruction by detecting continuous regions with parallel line projection. *ICCV*, pp.1911-1918, 2011.
- [10] Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. *ICRA'09*, pp.1848-1853, 2009.
- [11] Breiman, L.: Random forests. *Mach. Learn.*, 45, pp.5-32, 2001.
- [12] Rusu, R.B., Cousins, S.: 3D is here: Point Cloud Library (PCL). In: *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- [13] Botsch, M., Kobbelt, L.: Real-time shape editing using radial basis functions. *Comput. Graph. Forum*, 24, pp.611-621, 2005.
- [14] Frischholz, R., Dieckmann, U.: BioID: A Multimodal Biometric Identification System, In *IEEE Computer*, Vol. 33, No. 2, 2000.