

テクスチャ-デプスパッチタイリングに基づく 正面顔画像からの3次元形状推定

郷原裕明[†] 前島謙宣[†] 森島繁生[†]

本稿では、正面顔画像から3次元形状を推定する、新たな手法を提案する。提案手法では、3次元形状を持つ顔モデルからカラーマップとデプスマップを取得し、パッチに区切りデータベースを生成する。そして、入力顔画像のカラーとデータベース中のパッチのカラーマップと比較し、評価関数によりパッチを選択しパッチタイリングを行う。その際にデプスマップも同様に選択しタイリングをすることでテクスチャ情報と奥行情報の関係を利用した、テクスチャ-デプスパッチタイリングに基づく正面顔画像からの3次元顔形状推定なる新たな手法を提案する。

3D Face Reconstruction From A Facial Image based on Texture-Depth Patch Techniques

Hiroaki Gohara[†], Akinobu Maejima[†]
and Shigeo Morishima[†]

The paper presents an adaptation of the image quilting algorithm for 3D reconstruction and synthesis from 2D images. We build a DB of 3D faces that are normalized and converted into depth-maps. Next, given a 2D image, we compute its 3D depth-map by synthesizing texture-depth patches from the database using a minimization framework.

1. はじめに

近年、コンピュータグラフィクス(以下CG)技術の発達に伴い、多くの場面でCGが用いられる場が増加している。その中でも、映画やゲーム等の制作においてCG技術により実在人物の顔を表現することは、多くの需要があることで知られている[1]。しかしながら、それらを制作する際の顔の3次元形状を作りこむモデリング等の作業は非常に煩雑であり、作り手の負担とすることが大きい。そのため、実在人物を正確に、かつ作り手が容易に再現できるようにするために、様々な手法が提案されてきた。例えば、3Dレンジスキャナ[2, 3]のような機材を用いて、正確な顔の3次元形状を有する顔モデルを計測することで実在人物を表現する方法があげられる。しかし、これは正確にデータを得るために特殊かつ非常に高価な機材を必要とする。Blanzら[4]は、一枚の顔画像から、Morphable Model と呼ばれる形状・テクスチャ情報を有する汎用の顔モデルに基づき、顔の3次元形状を推定して実在人物を表現した。この手法は、推定する対象の顔画像の特徴点と特徴点と形状を関連付けたデータベースを用いることで、統計的に3次元形状を推定して実在人物を表現するため、非常に安価に実現可能である。しかしながら、推定に特徴点の位置関係を用いるため、形状の推定精度は特徴点検出器に強く依存する。

この問題を解決するために、我々は、一枚の正面顔画像とテクスチャと奥行を関連付けたデータベースを用い、パッチタイリングにより特徴点情報を直接用いることなく3次元形状を推定する手法を提案する。ここで言うパッチタイリングとは、画像をグリッドで区切り、区切られたパッチをデータベース中から選択し、それらを統合することにより新しい画像を作る手法のことである。パッチタイリングを顔に用いた代表的な手法にMohammedらのVisio-lization: Generating Novel Facial Images[5]がある。これは、顔画像のデータベースを用い、パッチタイリングにより新たな尤もらしい顔画像を生成する手法である。通常のパッチタイリングのみであると、大域的に見ないため右目と左目で形が大きく違うといったように顔全体として破綻することがある。しかしながら、この手法はパッチの選択時に大域的制約を用いることで、顔全体として破綻していないように見える、新たな尤もらしい顔を生成することができる。尚、この手法においては、大域的制約として、顔画像を主成分分析することにより算出される主成分軸に摂動を加えることで生成できる顔画像を用いている。しかし、この大域的制約は顔らしさを保つための制約であるので、顔の解像度や顔向きが顔画像のデータベースと同様の顔画像であればどの様なものでもよい。つまり、ある顔画像を大域的制約として入力すると、データベース中からそれに似たパーツが選ばれ、結果本人の顔画像に似た画像ができる(図1参照)。つまり、大域的制約として用いた顔画像に近

[†] 早稲田大学先進理工学研究科
Department of Advanced Science and Engineering

い顔画像をデータベース中のパッチをタイリングすることにより生成できるということである。また、事前調査により顔画像と顔の奥行情報には関係があることがわかった。そこで我々は、Mohammedらの手法と、顔画像と顔の奥行に関係があることを踏まえ、大域的制約として入力顔画像を使用し、顔画像と奥行情報が保存されたデータベースを利用することにより、1枚の正面顔画像から特徴点位置情報ではなくテクスチャ情報を利用した特徴点検出器に大きく依存しない新たな手法による3次元形状推定を提案する。提案手法の概要を図2に示す。

2. 顔におけるテクスチャと奥行の関係

顔画像から3次元形状を推定するためには、形状と見えとの間に関係がある必要がある。そこで、事前調査として顔画像と奥行の関係を確かめる実験を行った。実験の手順は以下の通りである。

1. 学習データとして、顔の3次元モデルを複数集める
2. モデル間の位置、回転、スケールを統一するために正規化を行う
3. 正規化された顔モデルに対し、画素ごとのRGBカラー情報(以後RGBカラーと呼ぶ)と奥行情報を取得する。
4. 3で取得したRGBカラー、奥行のマップを画素位置により矩形に区切りパッチを形成する。これをテクスチャ-デプスパッチと定義する。
5. 図3で示すように、テクスチャ-デプスパッチを一つ入力し、同じ画素位置の別の人物の顔のテクスチャ-デプスパッチと、RGBカラーの平均値、奥行平均値を合わせたうえで各々のRGB空間上での距離(3次元ノルム)と奥行値の差分の絶対値を算出する。
6. (RGB空間上での距離、奥行値の差分の絶対値)を300人のテクスチャ-デプスパッチに対してグラフにプロットする。

上記処理を複数の顔画像に行った。代表として、鼻の下部分 20×20 画素についてのRGB距離値と奥行差分値の関係を図4(a)に、口の中心部分 20×20 画素についてのRGB距離値と奥行差分値の関係を図4(b)に示す。図中直線は、分布に対する最少二乗直線でありその相関係数はそれぞれ0.596、0.562となった。図4からわかることは、RGB距離値が低い、つまり画素ごとにRGB空間上での距離が一定以上近いパッチの、奥行が近くなることを示している。また、顔の各場所に関してほぼ同様の傾向が見られた。以上のことから、画像とその見えには一定の関係性がみられた。故にRGBカラーから形状を推定できる可能性があることが分かった。

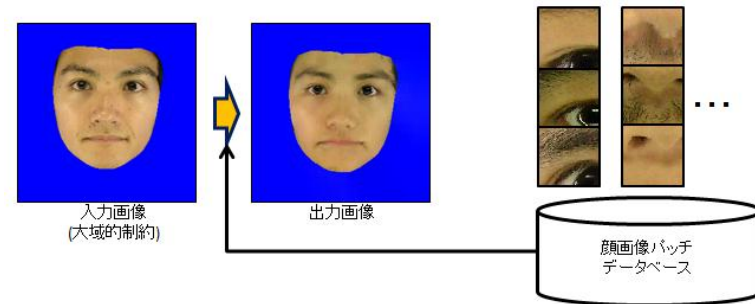


図1 Visio-lizationの概要

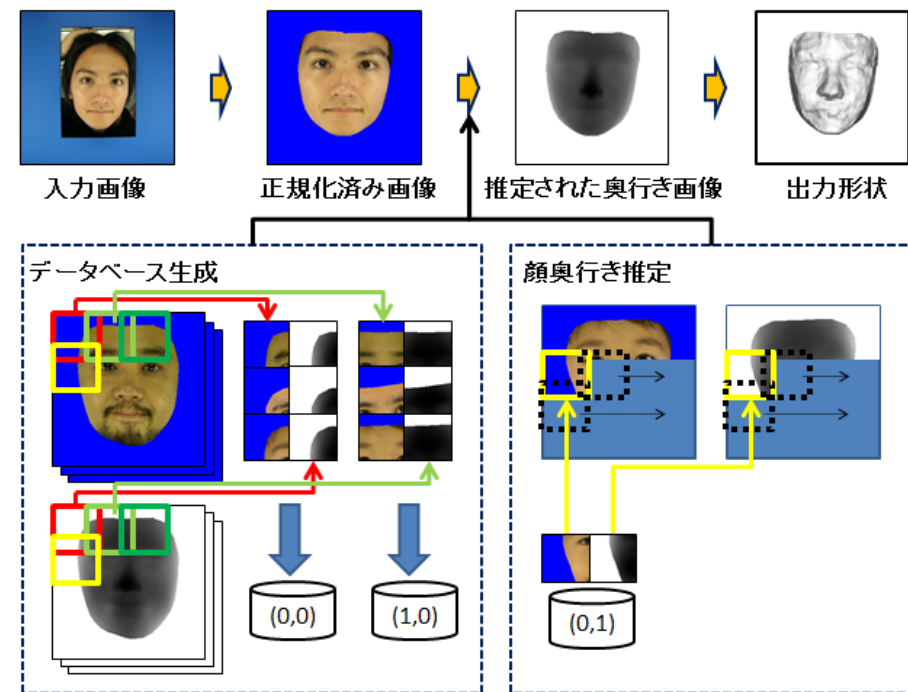


図2 形状推定概要

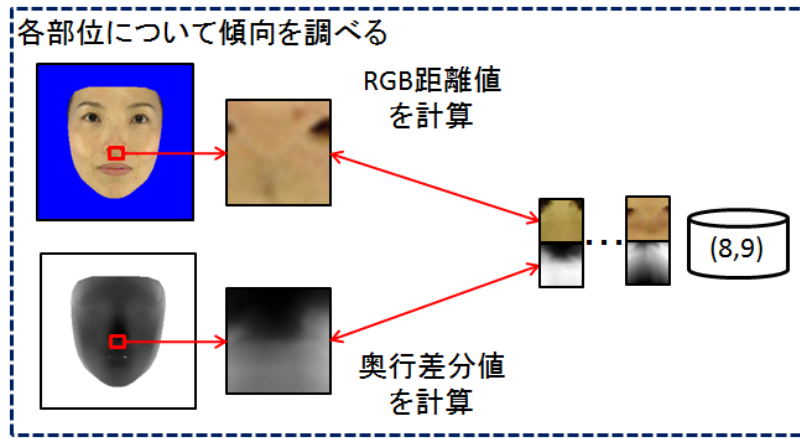
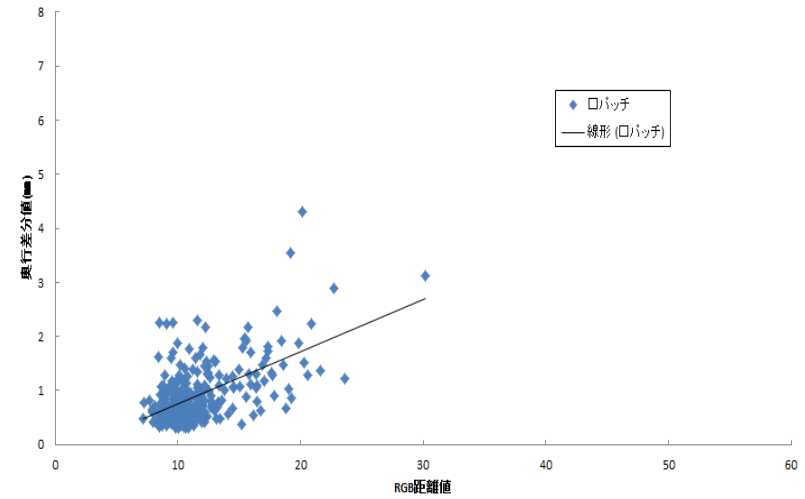
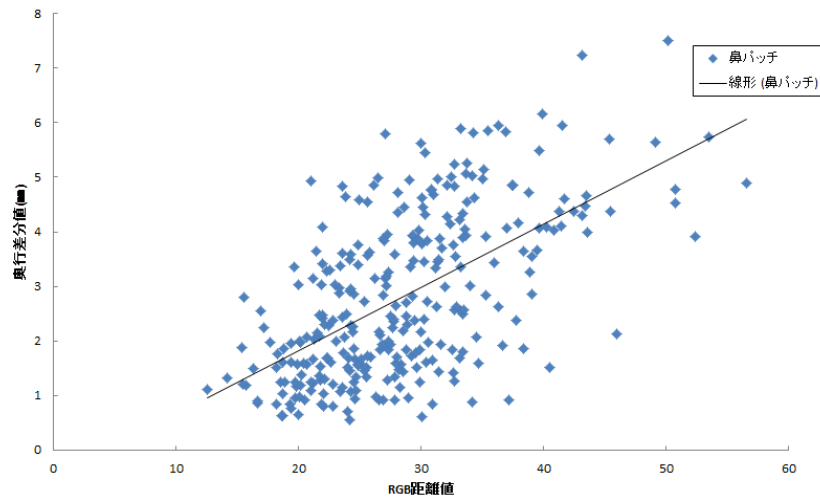


図3 各箇所の RGB 距離と奥行差分の関係



(b)

図4 各箇所の RGB 距離と奥行差分の関係



(a)

また、提案手法は最も RGB 距離が近いパッチの持つ奥行値が実際の奥行に限りなく近くなることが要求される。そこで、RGB 距離が最も近いパッチの奥行誤差が部位ごとにどれくらい異なるのか調べるため、20 人の顔に対し、最も RGB 距離が近いものを選択した際の奥行誤差をパッチの位置ごとに平均したものを求めた。結果を図 5 に示す。尚、今回 $16 \times 16 = 256$ 個のパッチに分割して行った。図 5 から、顔画像から顔の奥行を推定した時は輪郭付近の精度が悪く、中央に寄るほど精度がよいことが分かる。

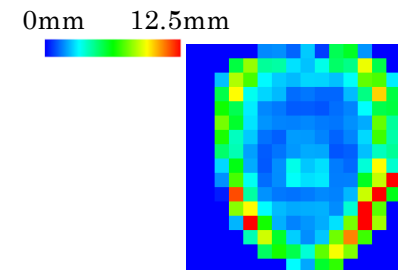


図5 位置による奥行誤差平均

3. データベースの構築

提案手法のデータベースは、400人分の男女のテクスチャ-デプスパッチから構成される。テクスチャ-デプスパッチデータベースの構築は以下の手順で行われる。

1. 前島らの手法によりレンジスキャンデータから3次元顔モデルを生成する[6].
2. 顔モデル間の位置、顔向き、スケールを統一するよう正規化する。この時各顔モデル間で同じ位置に同一の顔器官が重なるように正規化されていることが理想的である。
3. 各顔モデルを正面方向から見たときのRGBカラー情報と奥行情報をそれぞれテクスチャにレンダリングする。
4. 得られたRGBカラー情報と奥行情報を格納したテクスチャをパッチに切り分ける。これをテクスチャ-デプスパッチとする。尚、パッチの切り分けの際には、必ずパッチの一部が重複するように切り分け、パッチにはそれが存在していた位置のラベルが付けられる。切り分けられたパッチは、同じ位置のラベルが付けられたパッチ毎に分けられデータベースに保存される。今回は両目間隔約47画素の所、 20×20 画素で切り分け、重複部分は5画素とした。

テクスチャ-デプスパッチのデータベースは、各個人間で三重と形状は異なるものの、同じ位置にほぼ同じ顔器官が存在する。これは例えば、入力画像の鼻のパッチを同じ位置にラベル付された別人のパッチと入れ替えると、鼻が頬になるのではなく、鼻が別人の鼻の見えと形状に代わるといことである。これにより通常のパッチタイリングで起こり得る顔器官配置の矛盾を無くすることができる。

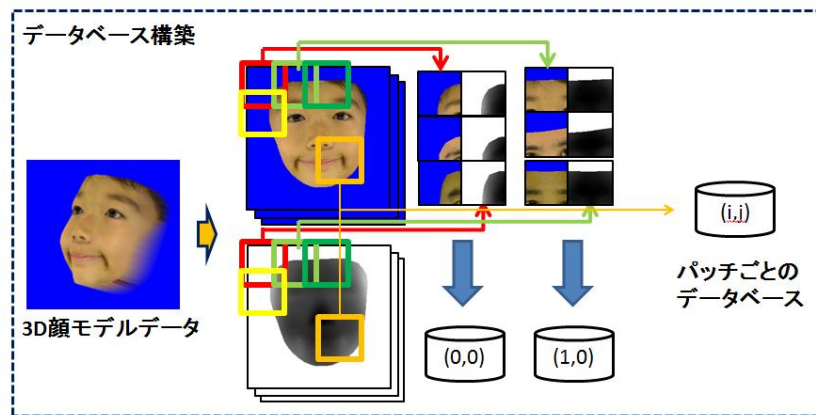


図6 データベース構築

4. 形状推定

形状推定は、パッチタイリングにより行う。パッチタイリングは、ラスタ上に左上の位置(0,0)から、右下まですべてのパッチがタイリングされるまで行われる。タイリングは以下を(i,j)に対し繰り返すことで行われる。

1. 位置(i,j)にタイリングされる際には、データベース中の位置(i,j)とラベリングされているすべてのテクスチャ-デプスパッチに対し評価関数Eを評価する。
 2. 評価関数Eが最も低い値を示したテクスチャ-デプスパッチを、(i-1,j)までタイリングされたパッチを統合したパッチ群 $I_{current}$ (テクスチャパッチ), $D_{current}$ (デプスパッチ)と、統合し $I_{current}$, $D_{current}$ を更新する。
- 1,2を右下まで行うことで、 $I_{current}$, $D_{current}$ は入力された顔画像に対する、データベースから生成した最も近いテクスチャと推定された奥行となる。以下に評価関数と統合を示す。

(1) 評価関数

評価関数Eは3つの項の線形和で表わされる。その3つとは、大域色特徴量、局所色特徴量、局所奥行特徴量である。

大域色特徴量 d_{gc} とは、入力された顔画像 I_{input} の(i,j)部分と、評価対象のテクスチャ-デプスパッチ $I_{selected}$ のRGBカラーに関して画素平均を合わせた上で、画素ごとにRGB色空間上の距離を測り、平均をとったものである。これにより、入力顔画像に近いパッチを選択しやすくなるような制約を与えることができる。

局所色特徴量 d_{lc} は、(i-1,j)までの選択されたパッチを統合したパッチ群のカラー情報である $I_{current}$ と、評価対象のパッチ間でオーバーラップさせて、その部分の画素平均を合わせた上で、画素ごとにRGB色空間での距離をとったものの平均である。局所色特徴量は、パッチ同士の連続性を保つような制約である。

局所奥行特徴量 d_{ld} は、(i-1,j)までの選択されたパッチを統合したパッチ群の奥行情報である $D_{current}$ と、隣接パッチとをオーバーラップさせ、その部分の奥行きを合わせた上で、奥行き値の距離差分の絶対値の画素平均である。これも局所色特徴量と同様に隣接間での連続性の制約である。

以下に式を示す。Nを評価対象であるテクスチャ-デプスパッチ $I_{selected}$ と入力顔画像をオーバーラップさせた際に重複した画素数、 $\overline{I_{selected}}$ を $I_{selected}$ の平均RGBカラーベクトル、nを(i-1,j)までの選択されたパッチを統合したパッチ群 $I_{current}$ とパッチをオーバーラップさせた際に重複した画素数、 $\overline{I_{current}}$ を $I_{current}$ に対し $I_{selected}$ をオーバーラップさせた際に重複した画素の平均、 $I[k]$ でオーバーラップした画素のk番目の画素のRGBを示す3次元ベクトル、 $D[k]$ でオーバーラップした画素のk番目の画素の奥行値、 $D_{selected}$ を評価するパッチの奥行情報、 $\overline{D_{selected}}$ を $D_{selected}$ の奥行

平均, $\overline{D_{current}}$ を, $D_{current}$ に対し $D_{selected}$ をオーバーラップさせた際に重複した画素の奥行平均, α , β , γ を各重みの係数とする.

$$d_{gc} = \frac{1}{N} \sum_{i=1}^N \|I_{input}[i] - (I_{selected}[i] - \overline{I_{selected}} + \overline{I_{input}})\| \quad (1)$$

$$d_{lc} = \frac{1}{n} \sum_{i=1}^n \|I_{current}[i] - (I_{selected}[i] - \overline{I_{selected}} + \overline{I_{current}})\| \quad (2)$$

$$d_{ld} = \frac{1}{n} \sum_{i=1}^n \|D_{current}[i] - (D_{selected}[i] - \overline{D_{selected}} + \overline{D_{current}})\| \quad (3)$$

$$E = \alpha d_{gc} + \beta d_{lc} + \gamma d_{ld} \quad (4)$$

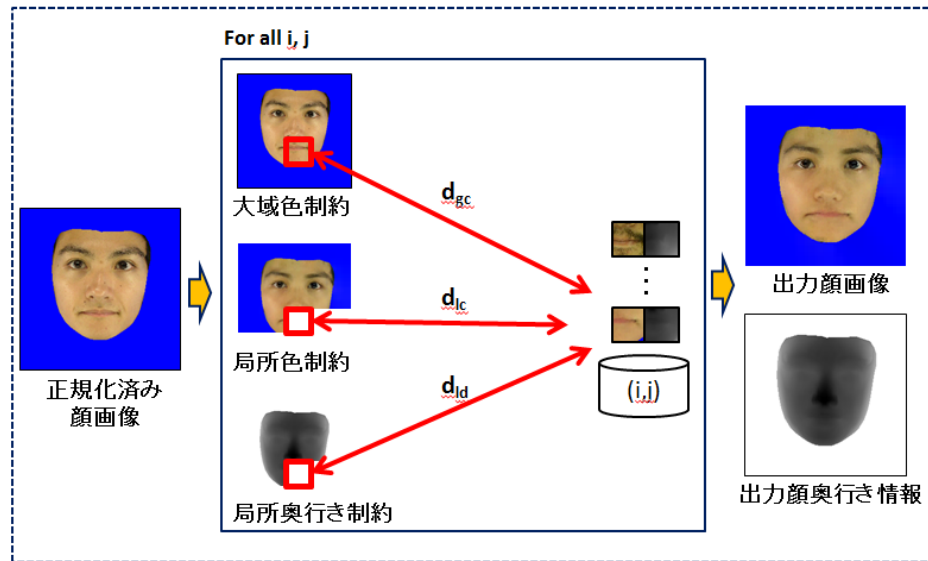


図 7 形状推定概要

(2) 統合

パッチをオーバーラップさせ, 評価関数 E からパッチを選択するが, 今まで選択してきたパッチ群を統合した $I_{current}$, $D_{current}$ と, 選択したテクスチャ-デプスパッチ $I_{selected}$ の間には, オーバーラップした部分をどのようにするかという問題がある. そこで, 重複した部分に適切な RGB カラー値と, 奥行を設定するために画素の位置

の距離比を利用し, 重みを決定した. オーバーラップさせていくと, パッチをラスタ状にタイリングするため, 以前にタイリングしたパッチと重複部分が生まれる. 具体的には, 現在のタイリングする位置を (i, j) とすると, $(i-1, j-1)$, $(i, j-1)$, $(i+1, j)$, $(i-1, j)$ の位置にある画素群と重複する部分ができる. そのうち, 2 つのパッチ間で重複する矩形部分と 4 つのパッチ間で重複する矩形部分ができる. 2 つのパッチ間で重複する部分は, 各々の重みは手前で 1, 反対側で 0 となるように線形補間を行う. 4 つのパッチ間で重複する矩形の場合は, 距離比を用いて重みを決定した. 尚, この処理を変えることは容易であるため, 2 次微分連続性を考慮した形で行うことでよりよい形状が得られると考える.

5. 精度評価

本稿では, 奥行の推定精度を評価するため, 理想的な正規化が行われた場合, どのぐらいの精度で奥行が推定できるか検証するため, 正規化済みの顔画像を入力として形状推定実験を行った. テクスチャ-デプスパッチデータベースを 400 人の顔から構成し, 入力顔画像はデータベース構築に使用していない 30 人の顔画像を用いた. また, 推定した形状を評価するために, 事前に計測済みの形状に対し画素ごとの誤差を求めた.

表 1 に, 各入力顔画像に対する推定形状と計測形状の誤差平均と標準偏差を示す. 表 1 から, 誤差平均の全入力顔画像における平均は 3.79(mm)であった. 図 8 に最も誤差の小さい例, 図 9 に 2 番目に誤差の小さい例, 図 10 に最も誤差の大きい例, 図 11 に誤差が誤差平均に最も近い例を示す.

表 1 各顔に対する誤差平均と標準偏差

順位	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
誤差平均(mm)	2.01	2.18	2.22	2.45	2.52	2.64	2.64	2.70	2.96	3.00	3.04	3.25	3.29	3.44	3.57
標準偏差(mm)	1.99	2.19	1.99	2.22	2.24	2.28	2.59	2.03	2.64	2.38	2.26	2.42	2.62	2.66	2.95

順位	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
誤差平均(mm)	3.75	3.89	3.94	3.95	4.00	4.18	4.19	4.47	4.57	4.69	5.13	5.15	6.19	6.77	7.06
標準偏差(mm)	3.21	2.32	2.85	3.52	2.96	3.36	3.25	3.49	3.27	3.01	4.30	4.81	3.98	3.65	5.11

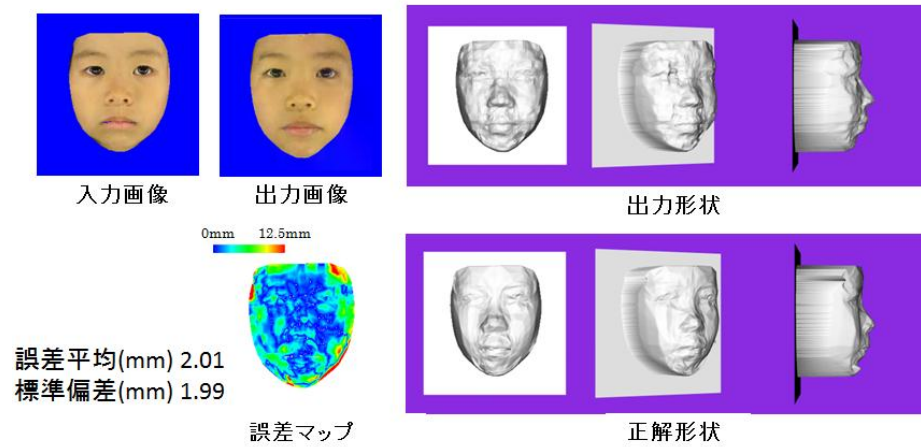


図 8 最も誤差の小さい例

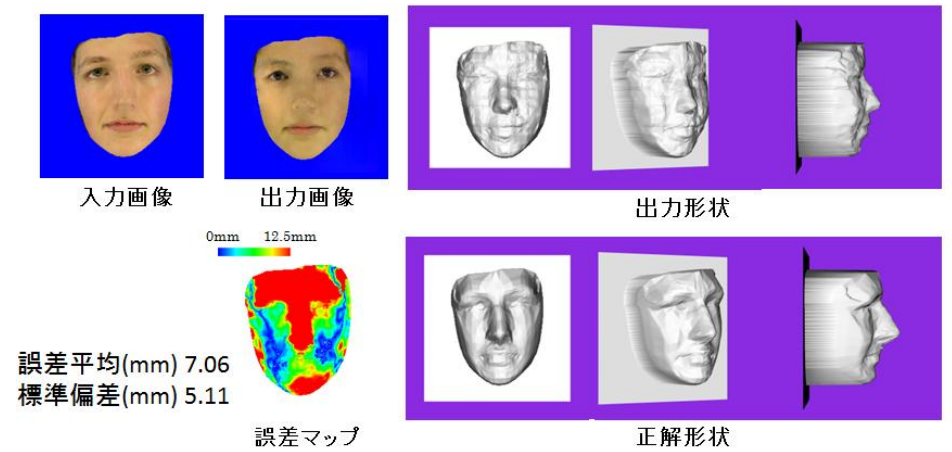


図 10 最も誤差の大きい例

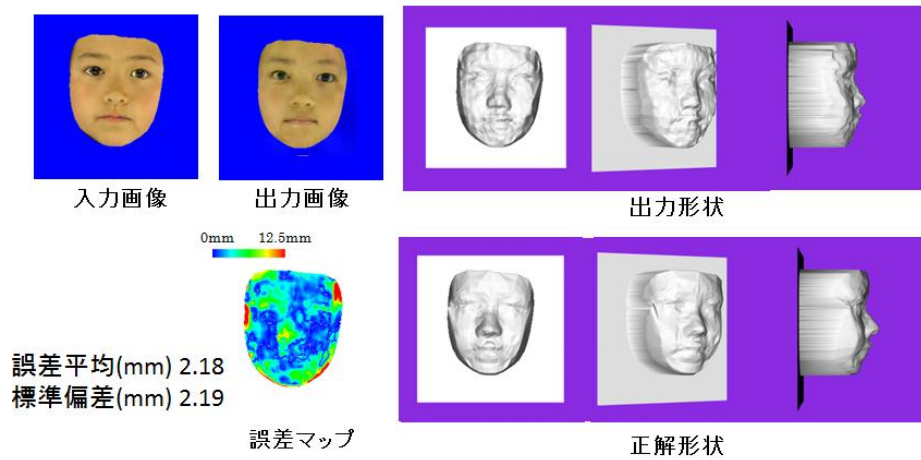


図 9 2 番目に誤差の小さい例

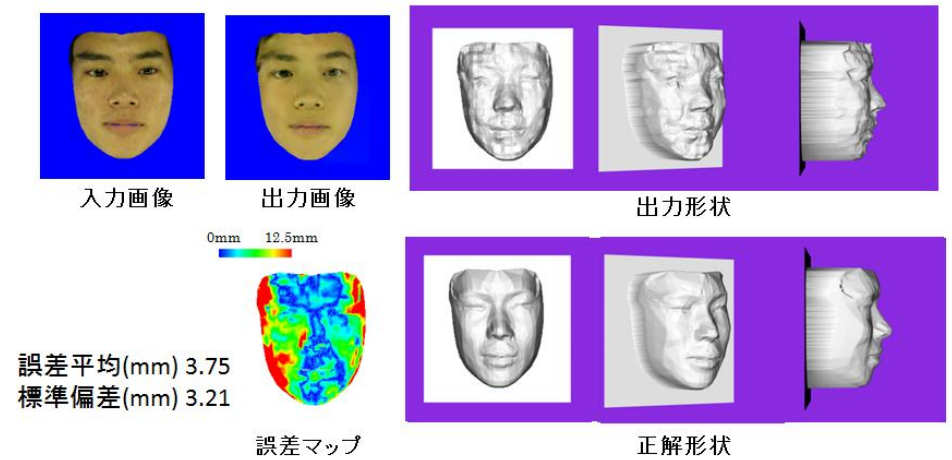


図 11 最も誤差が平均誤差に近い例

図 8, 9 は事前調査とほぼ同様に, 輪郭付近に誤差が表れている. また, 1 画素平均の誤差は 2 mm 程度となっている. 一方, 図 10 は最も誤差の大きかった外国人女性の例である. これは, テクスチャ-デプスパッチのデータベース中にこの様な比較的掘の深いデータが存在しないため, 表現しきれなかったと考えられる. 事前調査から, データ数が増えることにより精度が向上する傾向が見られたため, 図 10 のような例をより精度よく推定するためにデータベースの拡充は必須である. 図 11 は, 平均誤差に近い例であり, 事前調査と同様に輪郭部分が誤差が大きい傾向が出ている.

6. まとめと今後の課題

本稿では, 1 枚の正面顔画像から, テクスチャ-デプスパッチタイリングにより, 顔の奥行情報を推定する実験によりその効果を確認, 理想的に正規化された際の 3 次元形状推定に対する有効性を示した. 今後の課題は, 精度の向上のために, より多彩なデータを集めデータベースを拡充していくこと, パッチをタイリングする際はラスタ方式で埋めているがもっともテクスチャと形状の関係が強い箇所から埋めていくようにするなどの工夫をし, 評価し, 精度を向上させること, パッチ間の統合の処理をより正確に行うこと, 評価関数, テクスチャ-デプスパッチの大きさによる奥行の推定精度の違いを検証することが挙げられる.

参考文献

- 1) S.Morishima: Dive into the Movie -Audience-driven Immersive Experience in the Story-, IEICE TRANSACTIONS on Information and Systems, Special Section on Human Communication II Vol.E91-D, No.6, pp.1594-1603 (2008)
- 2) K.Fujimura, Y.Matsumoto, E.Tetuichi: Multi-camera 3D modeling system to digitize human head and body, Proc. of SPIE Vol. 4298, pp.40-47 (2001)
- 3) NEC エンジニアリング(株), “Danae”, <http://www.nec-eng.co.jp/pro/3d/introduction.html>
- 4) V.Blanz, T.Vetter: A Morphable Model For The Synthesis Of 3D Faces, *International Conference on Computer Graphics and Interactive Techniques, Proc. 26th annual conference on Computer graphics and interactive techniques*, pp.187-194(1999).
- 5) U.Mohammed, S. J. D. Prince, J.Kautz: Visio-lization: Generating Novel Facial Images, *ACM Transactions on Graphics (Proceedings SIGGRAPH 2009) 28(3)*, pages 57:1-57:8, (2009).
- 6) A.Maejima, S.Wemler, T.Machida, M.Takebayashi, S.Morishima: Instant Casting Movie Theater: The Future Cast System, *IEICE - Transactions on Information and Systems*, v.E91-D n.4, p.1135-1148, (2008).