

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 21 日現在

機関番号：32689

研究種目：基盤研究(B) (一般)

研究期間：2014～2016

課題番号：26280062

研究課題名(和文) 個人特徴を表現可能な動画像に基づくフォトリアルな顔の分析・合成の研究

研究課題名(英文) Photo-realistic Face Analysis and Synthesis in Image and Video to Generate Personal Characteristics

研究代表者

森島 繁生 (Morishima, Shigeo)

早稲田大学・理工学術院・教授

研究者番号：10200411

交付決定額(研究期間全体)：(直接経費) 12,500,000円

研究成果の概要(和文)：本研究課題では、主として次の6つの研究成果をあげることができた。  
 1) 犯罪捜査に威力を発揮する経年変化や肥瘦変形を考慮したモンタージュシステムの実現、2) 1枚の顔画像のみからの小規模なデータベースを利用する高精度な3次元顔形状復元システムの実現、3) エンタテインメントシステム応用としての他言語音声への3次元モデルを一切使用しないビデオリシャッフリングに基づく顔動画像リップシンクシステムの実現、4) 個性あるダンスキャラクタを自動生成可能なインスタントキャストシステムの実現。5) 表情変化時の個性を反映可能な経年変化ビデオ生成システムの実現。6) 似顔絵の個性を反映した実写映像復元の実現である。

研究成果の概要(英文)：This research project has succeeded to produce 6 systems concerning to face image analysis and synthesis considering personal characteristics as follows. 1) Face montage system effective to criminal investigation by considering an aging and slim/fat effect handling. 2) High quality pixel by pixel 3D face geometry reconstruction by single shot image with only small database. 3) Video lip synchronization system by video re-shuffling method applicable to translated voice without any 3D model for entertainment application. 4) High quality instant casting system for dance character reflecting personal characteristics in face and body. 5) Aging face video generation system reflecting personal characteristics in expression transition. 6) Automatic photo-realistic image generation from facial caricature keeping personal feature in original image.

研究分野：画像情報処理

キーワード：3次元顔形状復元 リップシンク合成 経年変化顔動画像合成 似顔絵からの実写復元 キャラクタ合成  
 肥瘦顔画像合成 犯罪捜査支援システム パッチベース画像合成

### 1. 研究開始当初の背景

バイOMETRICS個人認証は、指紋や虹彩、DNAなど高い精度が実現されている。しかし、いずれも被測定者に精神的な負担や装置を覗く動作などの制約を課すため、最終手段として用いられるのが一般的である。

一方、顔認証は負担や制約が少なく空港などの保安検査でも用いられているが、精度の面で上記の方法相当の性能が出せないばかりでなく、現在は静止画の認証が主流であるので、なりすましなどの問題点を含んでいる。また経年変化や髪型の変化、メイク、体格の変化等によって、異なる印象を与える点も顔画像による認証を困難な課題としている。

そこで、本研究課題では、動画もしくは複数方向からの静止画から頭部の3次元構造を復元すること、さらに表情変化時の3次元形状の動的な変化を復元すること、男女や年齢などの属性を表現し、それらの属性による顔テクスチャの変化をモデル化すること、痩せたり太ったりした場合の顔の形状およびテクスチャのバリエーションをモデル化すること、さらにこれらの個人特徴変化による影響を考慮して認証することにより顔認証の様々な問題点を解決することを目指している。このような視点での研究は過去に例を見ないものである。

### 2. 研究の目的

本研究は、対象者本人に撮影されているという意識を与えず、負担の全くない監視ビデオカメラ情報に基づき、本人同定あるいは感情状態の同定に必要な情報を安定に抽出するばかりでなく、観察者の判断を助けるための顔3次元合成や表情変化のリターゲットを実現するものである。これにより、顔表情の3次元形状とその動的変化を復元して観察することを可能とする。

同時に、個性という切り口でのフォトリアリスティックな顔合成技術に関する基礎的研究を遂行することを意味する。例えば顔認証精度を向上させるのに必須な個人のバリエーションを制御可能なフォトリアルな数億人単位での顔データの生成をも可能にする革新的な技術開発である。今回アプリケーションとして想定している顔認証、表情認識、デジタルサイネージ、キャラクターアニメーション以外にも、医学、心理学、工学、芸術等、様々な関連分野への波及効果が期待される。重要なことは、実写と変わらない顔合成のクオリティの追求にある。

### 3. 研究の方法

動画もしくは複数地点の監視カメラで撮影された顔写真からの3次元構造復元と表情変化時の個性表現モデルの実現に向け、平成26年度は主として、表情変化時の顔形状およびテクスチャのデータベース整備とその分析手法の検討を進める。平成26年度以降は、表情合成時のフォトリアリティ追求に

向けて、顔画像の表現方法について検討する。特に痩せたりや経年変化に対応可能な属性表現方法について検討し、様々な属性をもつ個性を表現可能な表情分析・合成システムの実現を行う。またこれらのシステムを、ビデオモニターやデジタルサイネージなどのアプリケーションシステムとして実装し、システム評価を進める。

### 4. 研究成果

本研究課題では、主として次の6つの研究成果をあげることができた。

#### （1）犯罪捜査に威力を発揮する経年変化や肥瘦変形を考慮したモニタージュシテムの実現

ビデオモニタージュシテムは、犯罪捜査等を目的とし、捜査員がモバイル環境において利用するシステムである。経年変化や肥瘦変形をインタラクティブに施し、犯人同定のためのリファレンスとなる顔画像合成を行うばかりではなく、髪型選択や皺の付加などを逐次行いながら合成画像を提示できるシステムである。個性を反映する目や鼻といった部位の画像はオリジナルを保持し、経年変化の影響が顕著に表れる皺やたるみの表現をパッチ単位でテクスチャ合成する手法を考案し、ポアソンイメージディティンクによって、頭髪部分や首から下の部分を他人の画像を置換することで実写レベルの高画質な経年変化顔画像合成を実現した。特に研究の独創性としては、若年層では表出しにくい皺の位置を、笑い顔をリファレンスとすることによって決定する方法を提案し、個性に関する再現性を実現した。その結果、図1に示すように、入力された原画像の年齢に対して、個人の印象を保持したまま、任意の年代に若返らせたり、歳をとらせることが可能となった。また処理時間も数十秒と高速に実行できる。アウトリーチ活動として、オープンキャンパスやデモ展示などを行い、鑑賞者に満足して頂けるクオリティを実現している。

また、Face and Gesture Recognition 国際会議において開催された顔認証コンペティションに参加し、競合する7つの顔経年変化アプリケーションに対して、最も高い成績で、経年変化させた顔画像が、ターゲット年齢のグランドトゥルースと認証精度を実現できることが証明された。



50代 30代 70代  
(a)原画像 (b)合成画像

図1 経年変化画像合成結果

(2) 1枚の顔画像のみからの小規模なデータベースを利用する高精度な3次元顔形状復元システムの実現

当初はビデオを複数報告から撮影された画像から顔復元を行う予定であったが、パッチ単位で画像から法線に直接変換する手法を独自に提案したことにより、1枚の顔画像から高精度に3次元形状を復元することが実現できた。

まず、データベースとして3次元顔形状の正解値(レンジスキャンデータ)とそれに対応する2次元画像データをおよそ10名~15名分用意し、画像の各ピクセル毎にRGB-D(画像三原色と奥行き)の値を格納しておく。本手法の特徴はデータベースサイズが比較的小さくても3次元復元精度が確保できる点にあり、撮影環境や条件ごとにデータベースを簡単に更新できる点が大きな特徴である。この点、深層学習に基づく手法では大量のデータベースを必要とするのに対して優位性が実現できる。

このデータベースから、まず正方パッチ領域に画像を分割し、そのパッチの中心に対して1つの法線ベクトルを計算し、これを顔の領域ラベルづけをして保存しておく。実行時には、入力された画像を同一のサイズで1画素シフトでパッチ分割し、誤差評価を行って最小距離のデータベース中のパッチに付与された法線ベクトルをその画素の法線として定義する。その後、この法線情報から顔表面のメッシュ構造を構築する。

図2にこの手法によって復元された顔の3次元構造を示す。左から入力原画像、本手法による復元結果、レンジデータで計測された正解形状、既存の最新手法による復元結果を示している。鼻や唇、目元など細部においてオリジナルの形状が忠実に再現されていることが分かる。また本手法は、用意する画像データベースの画質によって処理対象の画像を制御できるため、監視カメラ等の画質劣化画像や一部欠損した顔画像が入力された場合でも、再現精度の劣化を最小限に抑えることができるという際立った特徴がある。

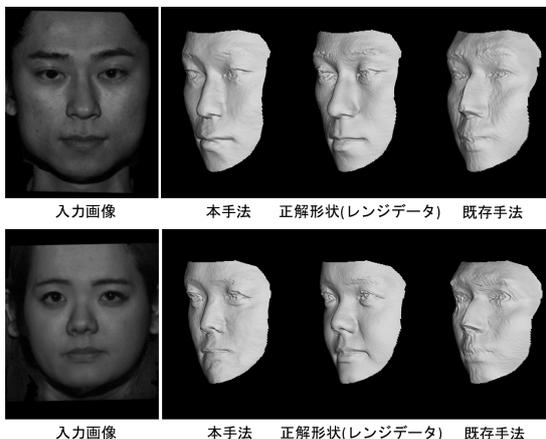


図2 顔の3次元構造復元結果

また従来のシェープフロムシェーディング手法がランバート反射を仮定することが一般的であるが、本手法では顔のパーツ毎の表面化散乱を含む反射特性が暗に考慮されていることも性能向上に寄与する点である。

3: エンタテインメントシステム応用としての他言語音声への3次元モデルを一切使用しないビデオリシャッフリングに基づく顔動画像リップシンクシステムの実現

オリジナルの言語からターゲット言語に吹替する場合に、3次元CGモデルを一切介さず、ターゲット言語毎に変換ルールを用意する必要のない全く新しい発想のリップシンク方式を提案した。

吹替対象映像(以下、俳優動画)と声優が吹替をしている様子を撮影した映像(以下、声優動画)の二つを入力し、声優と同じタイミングで同じ口形状が現れるように俳優動画の口周辺画像のフレームを並べ替えることでリップシンク映像を作成する。本手法は音素情報を全く用いないので多言語に適用可能であるという特徴を有する。また本手法のフレームワークは二次元アニメなどの顔形状モデルの取得が難しい映像にも適用できる点が独創的である。本手法の結果を図3、図4に示す。

俳優動画には英語を発話しているもの(長さ6分程度)を用い、声優動画には日本語を発話しているもの(僅か4秒程度)を用いた。図3からは正解動画(本人が日本語を喋るシーン)と比較して忠実に口形状だけではなく、歯や舌のポジションも再現されていることがわかる。また図4にRMSE値を用いて定量的に誤差評価をした結果を示す。図4からは提案手法が従来吹替(ボイスオーバーと呼ばれる従来のアフレコ手法)と比べて誤差が小さく、ほぼ全てのフレームで1mm未満に抑えられていることがわかる。

生成結果



正解動画



図3 生成結果と正解動画との比較

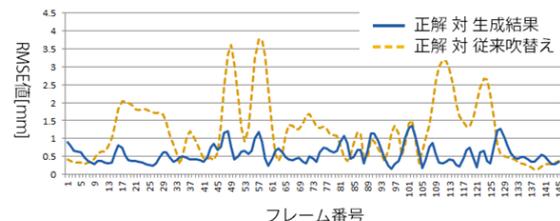


図4 RMSE値による誤差評価

(4) 個性あるキャラクタを自動生成可能なインスタントキャストシステムの実現

過去のプロジェクトにより実現されたダイブイントゥーザムービーの性能向上のための技術として、本人の顔形状モデルを手作業を介さず瞬時に反映させ、体形や衣装選択を行った後、自由にアニメーションさせることができるキャラクタ生成技術を開発した。

1枚の画像のみから、顔の3次元形状復元を行った後に、後頭部モデルを付加し、頭髪モデルを生成。これを本人の体系に合わせて生成したボーンおよびリグを配置した全身モデルに融合し、選択された衣装を体型にフィットさせて、テクスチャマッピングおよびシェーディングを行い、個性を再現したキャラクタモデルを生成する。

アニメーション時は、ボーンの制御によって動きを生成し、クロスシミュレーションおよびコリジョン判定・回避、ヘアシミュレーションを行う。技術的には、既存の最新技術の融合であるが、リアルタイムでのキャラクタアニメーションおよびパストレーシングによるオフラインレンダリングを施し、個性を反映したアニメーションがどの程度再現可能かについて様々な知見を得ることができ、新たな技術課題の発見に寄与する結果となった。特に、運動時の肉揺れ再現のための高速シミュレーション、高速レイトレーシング、表情合成・発話合成における個性を反映したアニメーション生成などが新たな課題として浮上した。

(5) 表情変化時の個性を反映可能な経年変化ビデオ生成システムの実現

静止画レベルで個性を保持したままの経年変化特徴付加に留まらず、表情変化ビデオ画像に対しても、ターゲット年齢への経年変化を実現する手法を提案した。

目標年代の別の人物の類似表情動画をリファレンスとして用いることにより、年齢変化時の顔動画の合成を行う。また、データベースの動画から得られる表情変化に伴う皺の濃さの変化を適用することで、皺形状の個性を維持した目標年代の表情変化を実現する。図5に提案手法のフローチャートを示す。

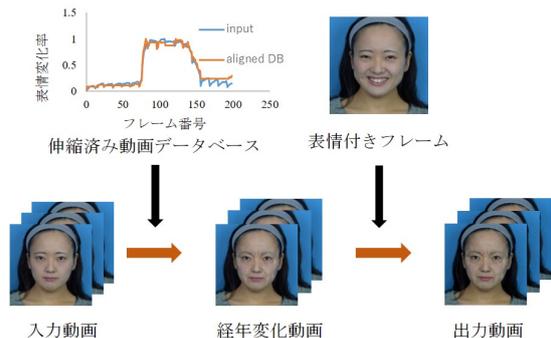


図5 経年変化動画生成のしくみ

入力動画と動画データベースの表情変化をフレーム単位で一致させるのは一般に困難である。そこで、各動画の表情変化の順番のみ統一し、動画データベースの伸縮を行うことで、表情変化のタイミングを入力動画におおよそ一致させる。この方法として既存研究では、両目、鼻及び口領域におけるLBP特徴量に基づき、入力動画のフレームを表情別にクラスタリングした。また、各表情クラスに対して表情が類似するフレームを1枚選択している。しかしこの方法では各表情クラス間での連続性が考慮されていない。そこで本研究では、動的時間伸縮法(DTW)を用いてデータベースの動画伸縮を行い、入力動画と動画データベースから選択されたサンプルとの表情変化のタイミングを一致させる。具体的には、各動画の全フレームに対して初期フレームとのLBP特徴量の類似度を算出し、その類似度変化が入力動画のそれと誤差最小化するように伸縮を行う。なおここで両目及び口領域のみに対してLBP特徴量を抽出した。次に入力動画及び伸縮済み動画の各フレームの顔特徴点を検出し動画データベースの顔形状を入力動画に一致させるために、検出した顔特徴点を用いてRBF補間により動画データベースの各フレームを入力動画の対応フレームに合わせて変形する。

この後、入力動画のテクスチャを目標年代の人物のテクスチャで再構築することにより、年齢変化要素を付加する。ここで顔画像を図6に示すように領域分割し、各領域ごとに合成を行う。

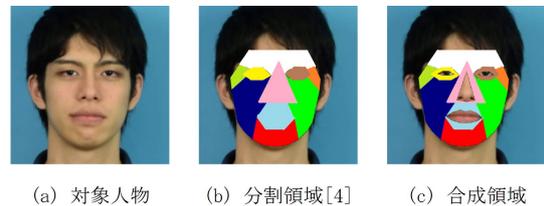


図6 顔の領域分割結果

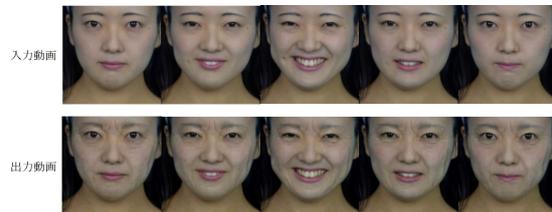


図7 加齢表情変形動画生成結果

老化時の皺の発生が表情皺に起因することから、表情付きフレームにおいて入力動画と出力動画の皺形状は一致することが望ましい。しかし、データベースに表情皺の形状が入力動画と完全に一致する人物が含まれる保証はない。そこで、入力動画の表情フレームにおける表情皺を合成動画に再合成することで皺形状を本人のものと一致させる。

UvA-NemoSmileDatabaseを用いて実験を

行った結果を図7に示す。各動画は無表情、笑顔、無表情という順に表情が変化する。なお顔向きが正面の動画のみを使用し、眼鏡などの遮蔽が含まれる動画に関しては、遮蔽が存在する領域の合成には用いないようにした。入力動画は20代女性、目標年代として60歳以上とした際の生成結果を図7に示す。これにより笑顔時の皺形状は保存されており、他の部分の皺の濃さは表情変化に応じて変化するという極めて自然な結果が得られた。

#### (6) 似顔絵の個性を反映した実写映像復元の実現。

捜査員や似顔絵師の描いた似顔絵画像からテクスチャトランスファーによって、オリジナルの個人の特徴を忠実に反映して、実写映像を復元する方式を実現した。

肖像画の特徴は、顔の陰影や形状は写実的であり、肌の質感(テクスチャ)や色味は描画材料で描かれているため非写実的な点である。昨今、様々な顔の研究・分析が行われてきたが、多くは写真を対象としている。肖像画を分析するためには、肌のテクスチャと色味を写実的に表現した画像が必要である。そこで入力した肖像画の陰影や形状は保持しつつ、実写真の肌テクスチャと色味を参照することで、肖像画の実写化を行う。

似顔絵から写実的な画像を生成する手法として、溝川らの手法がある。溝川らは、入力画像を平均顔形状に正規化してパッチ合成を行うことを提案した。この手法の問題点は変形に伴い局所的な陰影情報が損なわれる点と、斜め顔への対応が困難な点である。

そこで入力画像と実写真データベース画像において、それぞれ特徴点基準の座標を定義し、その座標変換によるパッチ対応・合成を行うことで、画像の変形を行わない合成手法を提案した。これにより形状・陰影情報を損ねない実写化が可能となった。

人物の二次元顔画像をデータベースとして使用し、顔向きは0度から左右45度まで(5度間隔)、年齢は20~70代の男女の無表情写真である。処理を行う際に入力画像と同様の顔向き角度のデータベースを使用する。画像サイズは入力画像の眉間から顎の長さで正規化しておく。

画像の生成手法としてMohammedらの手法を応用する。Mohammedらの手法は、顔画像をパッチに分割・再配置することによる新しい顔画像の生成である。Mohammedらの手法において、パッチの探索を行う際、入力画像とデータベース上の画像において顔の各部位の位置が異なるため、対応する座標を計算する必要がある。そこで、顔画像を特徴点に基づく二次元メッシュで分割し、対応するメッシュ上で対応座標を計算する。

次に対応付けられた各画像のパッチにおいて、入力画像とデータベース画像の輝度値

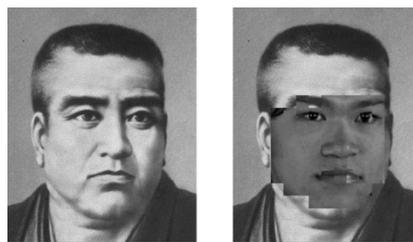
の比較を行い最適パッチを選択する。比較する際に、各データベース画像のパッチの輝度平均値を入力画像に揃えておくことで輝度値の正規化を行う。この処理により入力画像とデータベースで照明環境の影響を除去することができる。輝度値の正規化を行った後、データベースとの輝度値比較を行い最適パッチを選択する。

選択された最適パッチのRGB値を転写することで肌テクスチャの転写を行い、最後にポアソンイメージエディティングによってパッチ間の境界線を除去し自然な画像を生成する。



(a) 入力画像 (b) 提案手法 (c) 溝川ら

図8 似顔絵から生成された実写正面画像



(a) 入力画像 (b) 提案手法

図9 斜め顔(左15度)の生成結果

正面画像(0度)の結果画像を図8に示す。提案手法では溝川らの既存手法より、入力画像の頬や目周辺部分の陰影や、顔の各部位の位置・形状が再現されており、形状と陰影に忠実に実写化されていることがわかる。また、斜め画像(左15度)の結果画像を図9に示すが、本手法は顔向きによらず実写化できていることがわかる。ただし、このような個性ある例の場合には、よりデータベースを拡充して、多くの顔パーツのバリエーションをかばうことができるように工夫することが必要である。

## 5. 主な発表論文等

[雑誌論文] (計2件)

[1] 福里 司, 藤崎匡裕, 加藤卓哉, 森島 繁生, "頭蓋骨形状を考慮した肥瘦変化顔画像合成", 画像電子学会誌, Vol. 46, No. 1, pp. 197-205, 2017. 2. (査読有)

[2] サフキンパーベル, 加藤卓哉, 福里司, 森島 繁生, "老化時の皺の個性を考慮した経年変化顔画像合成", 情報処理学会論文誌 Vol. 57, No. 7, pp. 1627-1637, 2016. 7. (査読有)

〔学会発表〕（計 8 件）

- [1] 山本晋太郎, サフキンパーベル, 加藤卓哉, 佐藤優伍, 古川翔一, 森島繁生, “表情変化データベースを用いた経年変化顔動画合成”, Visual Computing / グラフィクスと CAD 合同シンポジウム 2017, 一橋講堂(東京都千代田区), No. 18, 2017. 6. (査読有)
- [2] Masanori Nakamura, Shugo Yamaguchi, Shigeo Morishima, “Live-action Portraits using Patch-based Texture Synthesis”, MIRU2016 第 19 回画像の認識・理解シンポジウム, インタラクティブ発表賞, アクトシティ浜松(静岡県浜松市), 2016. 8. (査読無)
- [3] Tsukasa Nozawa, Takuya Kato, Pavel A. Savkin, Naoki Nozawa, Shigeo Morishima, “3D Facial Geometry Reconstruction using Patch Database”, ACM SIGGRAPH 2016, Anaheim(USA), 2016. 7. (査読有)
- [4] Shoichi Furukawa, Takuya Kato, Pavel Savkin, Shigeo Morishima, “Video Reshuffling: Automatic Video Dubbing without Prior Knowledge”, ACM SIGGRAPH 2016, Student Research Competition Finalists (the 1st Award), Anaheim(USA), 2016. 7. (査読有)
- [5] Masanori Nakamura, Shugo Yamaguchi, Tsukasa Fukusato, Shigeo Morishima, “Creating a Realistic Face Image from a Cartoon Character”, ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA), Zurich(Switzerland), 2016. 7. (査読有)
- [6] 野沢 紘佐, 加藤卓哉, サフキンパーベル, 山口周悟, 森島繁生, “法線情報を含むパッチデータベースを用いた三次元顔形状復元”, Visual Computing / グラフィクスと CAD 合同シンポジウム 2016, 早稲田大学国際会議場(東京都新宿区), 2016. 6. (査読有)
- [7] 古川翔一, 加藤卓哉, サフキンパーベル, 森島繁生, “フレームリシャッフリングに基づく音素情報を用いない吹替え映像の生成”, Visual Computing / グラフィクスと CAD 合同シンポジウム 2016, 早稲田大学国際会議場(東京都新宿区), 2016. 6. (査読有)
- [8] Masahiro Fujisaki, Shigeo Morishima, “Facial Fattening and Slimming Simulation Based on Skull Structure”, ISVC 2015, Las Vegas(USA), 2015. 12. (査読有)

〔図書〕（計 0 件）

〔その他〕

ホームページ等

<http://www.mlab.phys.waseda.ac.jp/>

## 6. 研究組織

(1) 研究代表者

森島 繁生 (MORISHIMA, Shigeo)

早稲田大学・理工学術院・教授

研究者番号：10200411