

印象返還ベクトル法を用いた3次元顔表情画像の生成

堀井, 和也 / HORII, Kazuya

(出版者 / Publisher)

法政大学大学院理工学・工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学・工学研究科編 / 法政大学大学院紀要. 理工学・工学研究科編

(巻 / Volume)

58

(開始ページ / Start Page)

1

(終了ページ / End Page)

4

(発行年 / Year)

2017-03-31

(URL)

<https://doi.org/10.15002/00014233>

印象変換ベクトル法を用いた3次元顔表情動画の生成

Evaluating facial expression animations generated by impression transformation on the morphable 3D face model

堀井和也

Kazuya HORII

指導教員 赤松茂

法政大学大学院理工学研究科応用情報工学専攻修士課程

An approach to creating 3D animation of facial expression generation might be use of a motion capture system for dynamic 3D measurement of the target face, in which placing a lot of markers on the facial surface is a troublesome task. Another option is use of the morphable 3D face model which is built through principal component analysis of 3D data representing shape variations caused by facial expression generation. In this work, we adopted the latter approach, in which time series transformations of a novel 3D face is obtained by applying stepwise impression transfer vectors each of which gives displacement in the parametric space of 3D face model. In addition, the impression transfer vectors used in this experiment are designed by Support Vectors obtained by SVM learning.

Key Words : morphable 3D model , principal component analysis, support vector machine, animation

1. はじめに

コンピュータグラフィックスの発展により3次元アニメーションの制作[1]は発達してきた。また、映画、ビデオゲームなどの3Dアニメーションで人間の自然な動きを再現するときにはモーションキャプチャ[2]が用いられてきた。しかし、図1に示すようにモーションキャプチャの撮影には人の顔に多くのマーカをつける作業などが煩わしい。

また、印象変換[3][4]による顔画像の生成では、3次元顔画像からモーフィングモデルにもとづく顔印象の生成[5][6]がされてきた。

そこで、多数の顔の3次元顔形状データの主成分分析によってモーフィングモデルを作成し、サポートベクターマシン(SVM)[7]によって作られるパラメータ空間内での印象強調をすることで、新規3次元顔の表情生成の手法から表情動画の作成を試みた。



図1 モーションキャプチャ撮影時のマーカ貼り付け

2. 実験方法

(1) データベースの撮影

3次元形状を撮影するためにNECエンジニアリング社製のレンジファインダDanae-Rを用いた。撮影する表情は、「a」、「i」、「u」、「e」、「o」の発話表情と「真顔」、「閉口笑顔」、「開口笑顔」の8表情とし、3次元顔表情データベースの構築を行った。また、得られた形状データに対してRAPIDFORM社製のリバースモデリングソフトであるRapidForm2016を使用し必要となる領域の切り取り作業をし、3次元顔表情データベースの構築を行った。

(2) 正規化処理

3次元顔データは、撮影された画素の位置座標XYZと、その画素が持つRGB値を持つ点群データとして表現される。しかし、撮影した3次元顔データは、測定点の画素数と顔の傾きは人物や表情毎に異なっている。よって、特徴点を51点手動で打ち、そこから自動で73点抽出することにより特徴点数をそろえる。これをもとに、左右の瞳の特徴点が水平(x軸と平行)になるよう回転、さらに眉間の特徴点を原点として、鼻下部分と結んだ直線が垂直(y軸と平行)になるように回転、更に再サンプリングを行いサンプル点の数を統制する正規化処理[4][5]によってそれらの形態を整えた。

(3) 主成分分析

7種類の各表情について、特徴ベクトル \mathbf{X}_m を求め、無表情顔の平均からの差分ベクトル \mathbf{S}_m を求める。続いて各表情についての \mathbf{S}_m から共分散行列を求め、その固有値が大きいものから順に対応する固有ベクトルを並べた正規直交基底 $\mathbf{U}_k(k = 1, 2, \dots, K \leq 7)$ を求める。この \mathbf{U}_k を用いて各 \mathbf{S}_m と全差分ベクトルの平均ベクトル $\bar{\mathbf{S}}$ の差を式(1)によりK次元の主成分空間に投影することによって、 \mathbf{S}_m はK個のパラメータ $f_{m,k}(k = 1, 2, \dots, K)$ で表現される。

$$f_{m,k} = \mathbf{U}_k^t \cdot (\mathbf{S}_m - \bar{\mathbf{S}}) \quad (1)$$

(4) SVMを用いた印象変換ベクトル法

先行研究[6]では、(2)に示すように特徴ベクトル \mathbf{f} に印象変換ベクトル \mathbf{w} を足し合わせることで顔表情を生成していた。

$$\hat{\mathbf{f}}_m = \mathbf{f}_m + q_c \delta \cdot \mathbf{w} \quad (2)$$

印象変換ベクトルの求め方はFisherの線形判別法[5]、SVMに基づく非線形判別法[6]の2種類が比較されていたが、SVMに基づく印象変換ベクトルを用いた方が良い結果が得られていたので今回の動画の作成にはSVMを用いた。

(5) 新規人物の表情動画生成

上の過程で変換した主成分パラメータ $\hat{\mathbf{f}}_m$ と、正規直交基底 \mathbf{U}_k との線形和をとり、各表情の差分平均 $\bar{\mathbf{S}}$ との和によって無表情からの表情変化をあらゆる変位ベクトルを復元する。これを無表情顔 $\mathbf{X}_{neutral}$ に足しこむことにより、新規人物の表情 $\hat{\mathbf{X}}_m$ を生成する。この時パラメータの動きを $frame$ の数で分割することにより動画の途中となる表情を生成した。この過程において用いた式を(3)に示す。

$$\hat{\mathbf{X}}_m = \mathbf{X}_{neutral} + \frac{i}{frame} \left\{ \bar{\mathbf{S}} + \sum_{k=1}^K \hat{f}_{m,k} \cdot \mathbf{U}_k \right\} \quad (3)$$

$(i = 0 \sim frame)$

また、実際の動画の動きは線形的な速度ではないので、(4)に示すシグモイド関数を用いて加速度をつけてつなぎ合わせるにより描写した。

$$f(x) = \frac{1}{1 + \exp(-ax)} \quad (4)$$

表情パラメータの変化過程を図で表したものを図2に示す。赤矢印の部分が今回作成した動画となっている。

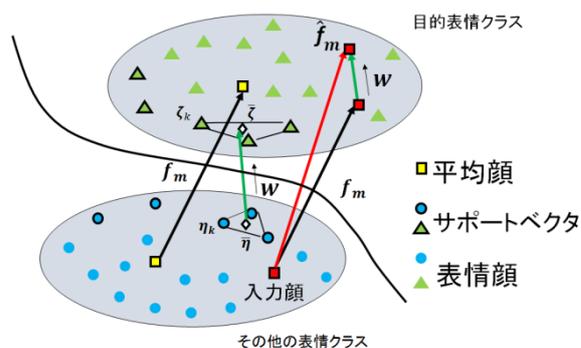


図2 表情パラメータの変化過程

生成された動画では口回りの領域がぼやけてしまっていた。これは先行研究[5][6]で使われていたデータベースにおいて、口回りの特徴点打ち方に問題があると考えられた。図3のように特徴点を打ち直し、口の中の領域を切り取った動画を新たに生成した。

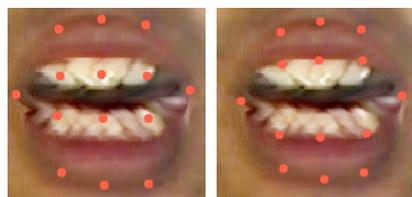


図3 修正前の特徴点(左)と修正後の特徴点(右)

さらに、生成された母音の最終フレームを新たな入力顔とし、母音から母音への発話動画も作成する。それらをつなぎ合わせるにより単語の発話動画を作成した。

3. 実験結果

(1) 生成された動画

図4に今回入力顔に使用したデータベース外の平均顔を示す。図5に「a」, 「i」, 「u」, 「e」, 「o」の発話動画の変化過程を示す。図6に「閉口笑顔」, 「開口笑顔」の表情動画の変化過程を示す。図5, 図6の上段, 中段は経過フレーム, 下段は最終フレームである。



図4 入力顔 (データベース外の平均顔)



図 5 発話動画像の変化過程



図 6 表情動画像の変化過程

図 7 に母音をつなぎ合わせた「大学」の発話動画像の変化過程を示す。



図 7 「大学」の発話動画像の変化過程

(2) 評価結果

モーショントラッキングを使用して作成された動画像[2]と精度比較を行うため主観評価実験を行った。20名の被験者に1つの動画像をランダムに提示し、動画像がどの表情に見えるかを判断してもらった。表1にモーショントラッキングを用いて作成された動画像の識別率の結果を示す。表2に本研究で作成された動画像の識別率の結果を示す。

表1 モーショントラッキングを用いた動画像の識別率

モーショントラッキング 識別率(%)	被験者が選択した表情						
	あ	い	う	え	お	閉口笑顔	開口笑顔
提示した表情	あ	82	0	0	18	0	0
	い	0	45	0	0	0	9
	う	0	0	82	9	9	0
	え	0	0	55	0	45	0
	お	0	0	0	0	0	100
	閉口笑顔	0	0	0	0	0	0
	開口笑顔	0	18	0	0	0	9

表2 本研究で作成された動画像の識別率

印象変換 識別率(%)	被験者が選択した表情						
	あ	い	う	え	お	閉口笑顔	開口笑顔
提示した表情	あ	91	0	0	0	9	0
	い	0	45	0	9	0	45
	う	0	0	65	15	20	0
	え	60	0	0	40	0	0
	お	0	0	48	0	52	0
	閉口笑顔	0	0	0	0	0	100
	開口笑顔	0	9	0	0	0	91

10種類の単語の発話動画像の精度評価をするために主観評価実験を行った。提示した単語の動画像は「大学」, 「アメリカ」, 「イタリア」, 「キネクト」, 「ウミガメ」, 「ブラジル」, 「映像」, 「エジプト」, 「ソビエト」, 「法政」である。被験者は25名とし上記の主観評価と同様な実験をした。表3に単語の発話動画像の識別率の結果を示す。

表3 単語の発話動画像の識別率

発話動画像 識別率(%)	被験者が選択した単語									
	ダイガク	アメリカ	イタリア	キネクト	ウミガメ	ブラジル	エイゾウ	エジプト	ソビエト	ホウセイ
提示した単語動画像	ダイガク	96	0	0	0	0	0	0	4	0
	アメリカ	6	84	0	4	0	0	0	6	0
	イタリア	0	4	82	0	0	8	0	0	6
	キネクト	10	0	0	74	0	0	2	14	0
	ウミガメ	4	0	10	0	82	4	0	0	0
	ブラジル	0	4	0	4	0	68	12	0	12
	エイゾウ	0	0	0	4	0	6	86	0	4
	エジプト	0	4	0	2	0	0	0	94	0
	ソビエト	8	0	0	20	0	0	0	62	10
	ホウセイ	12	0	2	8	0	0	4	4	62

4. 考察

提案手法により、表情動画像を作成することが出来た。これにより、顔の動きを計測するためのモーショントラッキングの撮影なしで新規人物の表情顔を作ることが出来る。今回の手法では、物理的なマーカを人に張り付けて撮影するという煩わしい作業がなくなったので従来研究

[2]に比べると撮影にかかる時間が短縮できた。

精度比較の評定結果から、モーションキャプチャを使って作られた動画像より識別率の高くなった表情もあれば、低くなった表情もあることがわかった。識別率が高くなった表情に関しては本手法で用いた印象変換ベクトルが生成顔をモーフィングモデル上で表情の境界から引き離れたことにより表情を強調している役割があるからだと考えられる。精度が低くなった表情に関しては、表情の変化過程が実際のモーションと違うということが考えられる。これは、印象変換ベクトルの重みとシグモイド関数の係数を変えて試行錯誤を繰り返し、主観評定を行い精度の良い動画像のパラメータを求めることで改善ができる。

単語の発話動画像の主観評定から、作成した10種類の発話動画像がある程度正確に識別されることがわかった。今回は母音をつなぎ合わせただけなので子音について考えていない。例えば「アメリカ」、「ソビエト」の「メ」や「ビ」は本来一度上唇と下唇を閉じてから発話する。この問題は、子音の口を閉じてから発する発話の部分は真顔への変化過程を挟むことで違和感がないものが作成できると考えられる。

今回の技術の精度を高め応用することで映画やアニメーションのCGに適用できると考えられる。

謝辞

本研究に際して日頃からご指導を頂いた赤松茂 教授に深く御礼申し上げます。また、本研究に共同研究者として協力して頂いた理工学部4年の田中ひかるさんを始めとする赤松研究室の皆様、そして実験を快く引き受けてくださった関係者の皆様に感謝いたします。

参考文献

- 1) 増田隆,高橋時市郎,師井聡子, “Facial Action Coding System に基づく表情アニメーションの評価,” 情報科学技術フォーラム一般講演論文集 6(3), pp. 313-314,2007-08
- 2) S. Nagata, Y. Arai, S. Takagi, Y. Inaba, and S. Akamatsu, “Synthesis of dynamic 3D facial expressions based on 3D facial movement measured by motion capture system,” Proc. of the 4th IEEE International Workshop on Image Electronics and Visual Computing, Koh Samui, Thailand, Oct. 7-10, 2014
- 3) 安善女主,小沢慎治, “顔画像からの筋肉パラメータの推定とそれに基づく他人の表情生成,” 電子情報通信学会論文誌 D-II, Vol. J8-D-II, No. 10, pp. 2081-2088, 2005
- 4) Y. Inaba, J. Kochi, H. Ishi, J. Gyoba, and S. Akamatsu, “Impression-driven Design Scheme for a Class of 3D Objects Based on Morphable 3D Shape Model, and its Automatic Buildup by Supplementary Feature Sampling,” Proc. of IWAIT2009, Jan. 2009
- 5) 天木毬江, 渡邊ちさ, 永田俊介, 山本俊太, 稲葉善典, 赤松茂, “多数の表情顔の3D形状に対する主成分分析に基づいた新規人物の表情生成の試み-” HCG シンポジウム 2012 2012-12
- 6) Y. Arai, K. Horii, Y. Inaba, and S. Akamatsu, “Creating more comprehensive facial expressions by morphable 3D face model: Transformation of 3D faces through impression transfer vectors defined by SVM,” Proc. of IWAIT 2016, 4C-2, Jan. 2016
- 7) K. Tsuda, “Overview of Support Vector Machine,” The Journal of the Institute of Electronics, Information, and Communication Engineers, Vol. 83, No. 6 pp. 460-466, June 2000