

画像生成 AI による顔表情映像の印象評価 Impression Evaluation of AI-Generated Facial Expression Videos

對馬 理向¹⁾ 原田 翔太¹⁾ 満上 育久²⁾
Riko Tsushima Shota Harada Ikuhisa Mitsugami

1 はじめに

近年、AI 技術の発展によって、映像中の人物から得られた骨格情報とスコアの対応関係から、体操やフィギュアスケートなどのスポーツの評価の定量化（自動採点）ができるシステムが開発されている [1]. しかし、スポーツとは違い、店員の接客態度は、接客態度と印象の対応関係が明らかではないため定量化が容易では無い。この対応関係を得るには、複数の店員が指定されたとおりの複数の異なる表情で接客行動をとらなければならない。しかもそれを複数の評価者に対して差異なく繰り返し行わなければならない。これを生身の人間が行うことは不可能である。この問題に対して、Sumino らは、表情と視線を任意に制御した顔表情映像を生成する手法を提案している [2, 3].

本研究では、Sumino らの手法を利用して多様な顔表情映像を生成し、それらの映像から表情が与える印象を調査する。具体的には、笑顔の度合いが異なる顔表情映像と、目の動き方が異なる顔表情映像を生成し、それぞれにおいてどの笑顔度、どの目の動き方が最も良い印象を与えるかに関して評価実験を行う。

実験参加者として、広島市立大学の学生と Amazon Mechanical Turk [4] のワーカーの 2 種類を用意し、それぞれの結果から両者の共通点・相違点などについても調査する。実験の結果、目の動き方が異なる顔表情映像においては、両者に同じ傾向が見られ、目線の揺れも瞬きも少ないことが良い印象に繋がるという事が分かった。一方で、笑顔度の異なる映像に関しては、それぞれの結果において異なる傾向が見られた。

2 表情変化映像の生成手法

2.1 First Order Motion Model

Siarohin らが提案した First Order Motion Model (FOMM) [5] は、1 枚の静止画を参照動画（ソース）をもとに動画化する手法である。入力するのはソース、ターゲット画像、ソースの基準画像で、ソースの基準画像とはソースのフレームのうちの 1 枚である。FOMM では、入力した顔画像から複数のキーポイントを抽出して、各キーポイントの位置と周辺の変形を表す特徴ベクトルを与える。ソースの基準画像とソースの各フレームに与えられた特徴ベクトルを \hat{k} , k_i とすると、これらの特徴ベクトルの差分ベクトル $\hat{k} - k_i$ をターゲット画像に付与することで、ターゲット画像にこの表情変化を転移することができる。

2.2 表情制御可能な顔映像生成手法

Sumino らは、前節で述べた FOMM をベースに、任意の特徴ベクトルを付加することで任意の顔表情映像を生成する手法を提案している [2, 3]. その概要を図 1 に

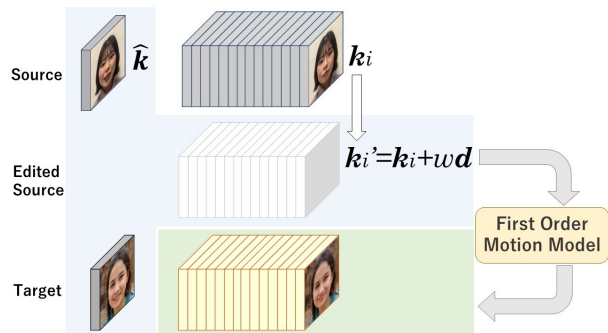


図 1: Sumino らの手法の概要 ([3] より引用)



図 2: 笑顔度が異なる映像



図 3: 目の動き方が異なる映像

示す。例えば、笑顔の度合いを制御する場合には、人物の真顔の画像と笑顔の画像の 2 枚の画像から特徴ベクトル d_{smile} を抽出し、それをソースの特徴ベクトル k_i に付加することで、ソース映像の顔動作をターゲット画像に転移しつつ任意の笑顔度に制御した映像を生成することができる。

3 表情変化映像の生成

本研究では、前章で述べた Sumino らの手法を用いて 5 段階で笑顔の度合いが異なる映像と、4 種類の目の動かし方が異なる映像を生成した。ソースは男性 3 名と女性 1 名の計 4 名、ターゲット画像は各ソースに対して 6 枚ずつ使用し、笑顔度が異なる映像と目の動き方が異なる映像をそれぞれ 24 個ずつ生成した。笑顔度の異なる映像の例を図 2 に示す。笑顔度の異なる映像の生成においては、笑顔度を 0% (真顔)、25%、50%、75%、100% の 5 段階で設定した。それぞれの笑顔度の顔画像を撮影し、各画像から特徴ベクトルを抽出して笑顔度 0% (真顔) の画像の特徴ベクトルとの差分ベクトルを取得した。ソースにこの差分ベクトルを付加してターゲット画像へ表情転移を行うことで、笑顔度を 5 段階に制御した動画を生成した。

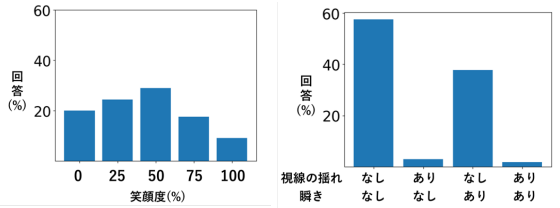
目の動かし方に対する印象調査のために、図 3 に示す

1) 広島市立大学

Hiroshima City University

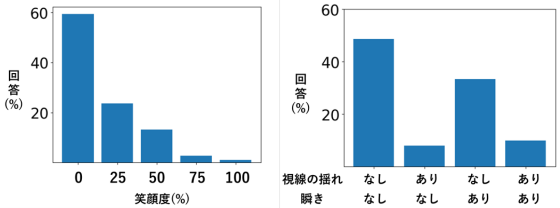
2) 立命館大学

Ritsumeikan University



(a) 笑顔度が異なる映像の結果 (b) 目の動き方が異なる映像の結果

図 4: 学生グループの結果



(a) 笑顔度が異なる映像の結果 (b) 目の動き方が異なる映像の結果

図 5: MTurk グループの結果

ように、視線の揺れと瞬きの有無を制御した 4 種類の映像を生成した。ソースのフレーム毎に異なる特徴ベクトルを付加することで、視線の揺れや瞬きを制御する。ここでは、視線が揺れる動画、瞬きをする動画、視線が泳いでかつ瞬きをする動画の 3 つを撮影して、それぞれの動画のフレームの特徴ベクトルと真顔の特徴ベクトルの差分を取得した。その差分ベクトルをフレーム毎にソースに付加することで、視線の揺れと瞬きの制御を実現した。

4 印象評価実験

4.1 実験環境

本研究では、前章にて説明した映像を用いて、2 つの異なるグループに対して評価実験を行った。第一は広島市立大学の学生 (学生グループ)、第二はクラウドソーシングサービスである Amazon Mechanical Turk のワーカー (MTurk グループ) である。学生グループから得た回答と、多様な背景を持った MTurk グループから得た回答を比較し、両者の違いを調査した。

どちらに対しても、図 2、図 3 に示したように、笑顔度が異なる映像を 0%~100% の 5 択で、目の動き方が異なる映像を目線の揺れと瞬きの有無の組み合わせの 4 択で提示し、それぞれ最も印象の良かった一つを選択させた。映像は何度でも視聴可能で、印象評価は動画の発言内容に関わらず行うように指示した。回答人数はそれぞれ学生グループが 20 名、MTurk グループが 33 名で、学生グループからは 960 個、MTurk グループからは 1,165 個の回答を得た。

4.2 結果

学生グループの結果を図 4 に示す。縦軸は回答割合、横軸は映像の種類を示している。図 4 (a) から、広島市立大学の学生は笑顔度 50% の映像を最も多く選択したことが分かる。笑顔度 50% を中心に、笑顔度の増減に従っ

て選択された割合が少なくなる。特に最も笑顔度の高い笑顔度 100% が選択された割合が一番少なかった。図 4 (b) からは、視線の揺れがない映像が多く選択されたことが分かる。

MTurk グループの結果を図 5 に示す。図 4 と同様に縦軸は回答割合、横軸は映像の種類を示している。MTurk グループの結果では、図 5 (a) から笑顔度 0% の真顔が最も多く選択されたことが分かる。図 5 (a) では図 4 (a) の傾向とは異なり、全体の半分以上が笑顔度 0% (真顔) を選択しており、笑顔度の増加に従い選択割合が減少した。笑顔度 100% が選択された割合は学生グループの結果と同様で一番少ないが、その割合は MTurk グループの方がはるかに少なかった。一方、図 5 (b) では、図 4 (b) と同様に、視線の揺れがない映像が多く選択される傾向が見られた。

4.3 考察

本実験において、目線の動きが異なる映像の評価においては広島市立大学の学生と Amazon Mechanical Turk の 2 つのグループで同じ傾向が見られたが、笑顔度の異なる映像の評価では両者に異なる傾向の結果が得られた。したがって、同じ表情であっても、受け取り手の属性によって違う印象をもたらす場合があると考えられる。

この違いが生じた要因として、印象を評価する人の年齢層・国籍・文化などの様々な違いが考えられる。今後はこれらの点に着目した調査も検討したい。

5 おわりに

本研究では、Sumino らの提案した表情変化を制御できる顔映像生成手法を用いて、笑顔度が異なる映像と目の動き方が異なる映像を生成し、それぞれにおいて最も良い印象を与えた映像を調査した。その結果、目の動き方が異なる映像では、異なる実験参加者グループ間で同じ傾向の結果が得られ、笑顔度の異なる映像では両者に異なる傾向が見られた。グループによって違いが生じた要因としては、印象を評価する人の年齢層・国籍・文化などの様々な違いが考えられ、今後はこれらの点に着目した調査も検討している。

謝辞

本研究は、JST、未来社会創造事業、JPMJMI22J1 の支援を受けたものである。

参考文献

- [1] 樹井昇一, 手塚耕一, 矢吹彰彦, 佐々木和雄, 「3D センシング・技認識技術による体操採点支援システムの実用化」, 情報処理, Vol.61, No.11, 2020 年.
- [2] K. Sumino, I. Mitsugami, and R. Sagawa, "Expression-controllable Facial Video Generation for Impression Quantification," Global Conference on Consumer Electronics, 2022.
- [3] 住野奏, 満上育久, 佐川立昌, 「対面時の印象評価に向けた多様な顔表情映像の生成」, 情報処理学会研究報告, Vol.2022-CVIM-230, No.15, 2022 年.
- [4] "Amazon Mechanical Turk," <https://www.mturk.com/>
- [5] A.Siarohin, S.Lathuilière, S.Tulyakov E.Ricci, N.Sebe, "First Order Motion Model for Image Animation," Conference on Neural Information Processing Systems, 2019.