Vol.2022-CVIM-230 No.15
2022/5/12

対面時の印象評価に向けた多様な顔表情映像の生成

住野 奏[1,a)]　満上 育久[1,b)]　佐川 立昌[2,c)]

概要：本研究では，対面時の相手に対する印象評価に向けて，ある人物の発話中の顔映像をもとに，人物の変更や表情の調整を反映した映像を生成できる手法を提案する．提案手法は，1 枚の静止画と参照動画（ソース）を元にリアルタイムで動画化する手法である First Order Motion Model (FOMM) をベースとしつつ，ソースに笑顔・視線変化を施すことのできる機構を組み込むことで，生成動画の表情・視線を任意に制御することを実現している．FOMM では，ソースの動画をターゲットの動画に転移させる手法であるのに対して，提案手法では，ソースの特徴ベクトルに笑顔・視線変化を表す差分ベクトルを付与した上で FOMM を適用する．また，笑顔の度合いと視線方向を制御する重みを時系列信号として表現しておくと動画に適切に反映させるように実装し，表情・視線を同時に変化させることも可能である．提案手法により生成した多様な顔表情映像から，表情・視線の変化が制御されつつも違和感や破綻のない自然な動画が生成できていることを確認する．

キーワード：顔画像生成，ディープフェイク，First Order Motion Model，印象評価

1. はじめに

近年の AI 技術の進展により，映像中の人物の動きをリアルタイムかつ正確に推定できるようになり，体操やフィギュアスケートなど，選手の動作とスコアの対応が定められている種目では，自動採点すなわち「良さの数値化」ができるシステムが開発されている [1]．一方，店員の接客に対する印象などのように，対面コミュニケーションにおける人の印象の良さを数値化するのは容易ではない．これは，印象の良さが感覚的・抽象的であり，店員のどのような表情・所作が「良い」のかが明らかではないためである．接客の印象の良さを数値化するためには，まずどのような店員が接客時にどのような表情・所作をとると印象がよく感じるのかを調査しなければならない．しかし，実際の人物が笑顔の度合いや視線などのバリエーションを変えながら同じ接客動作を多数の実験参加者に対して繰り返し提示することは現実的ではない．また，異なる人物が同一の笑顔度合い・視線の動きを行うことも不可能である．一方，このような人物・表情のバリエーションは CG を使えば実現されるものの，写実的でない CG 映像での心象評価は，実際の人物に対する心象評価とは必ずしも一致しないであろうという問題が生じる．

そこで本研究では，さまざまな人物がさまざまな表情・所作をとっている映像セットを用意するために，ある人物の発話中の顔映像をもとに，人物の変更や表情の調整を反映した映像を生成できる手法を提案する．この提案手法は，1 枚の静止画を参照動画（ソース）を元にリアルタイムで動画化する手法である First Order Motion Model (FOMM) [2] をベースとしつつ，ソースに笑顔・視線変化を施すことのできる機構を組み込むことで，生成動画の表情・視線を任意に制御することを実現している．FOMM では，ソース・ターゲットの基準画像のペアとソースの動画を入力とし，ソースの動画をターゲットの動画に転移させる手法であるのに対して，提案手法では，ソースの特徴ベクトルに笑顔・視線変化を表す差分ベクトルを付与した上で FOMM を適用する．この差分ベクトルを付与する重みを調整することにより，本来の FOMM で生成される動画に対して，笑顔の度合いや視線方向を制御した映像が生成される．なお，笑顔・視線の表情変化を表現する差分ベクトルは，表情や視線方向が異なる 2 枚の顔画像の特徴ベクトルの差から得られる．また，差分ベクトルを付与する重みの時間推移を時系列信号として入力することで，表情・視線を同時に制御することも可能である．提案手法により生成した多様な顔表情映像の評価において，表情・視線の変化が制御されつつも違和感や破綻のない自然な動画が生成できていることから，本システムの有用性を示す．

[1] 広島市立大学
[2] 産業技術総合研究所
[a)] sumino@sys.info.hiroshima-cu.ac.jp
[b)] mitsugami@hiroshima-cu.ac.jp
[c)] ryusuke.sagawa@aist.go.jp

2. 関連研究

対面時の人間の表情や視線と，相手の印象や満足度の関係性に関する研究は，社会心理学分野で盛んに行われている．例えば，土屋は，笑顔の有無と視線の有無を組み合わせた 4 パターンの人間の上半身を映した動画を被験者に提示し印象評価を行った結果，アイコンタクトを取り，かつ笑顔を表出している方が印象評価が高まったことを報告している [4]．ただし，この実験において動画内の人物は常に被験者を注視したものになっているが，実際の対人コミュニケーションにおいて相手に視線を向ける時間割合は約 50 %程度であることが知られていることから [3]，これは自然な対人コミュニケーションとは言い難い実験条件である．また菅原らは，眼と口に関する形状や位置を変えた多様な笑顔タイプの画像に対する印象を SD 法で評点化し主成分分析を行うことで笑顔における眼と口の重要性を確認している [5]．

これらの先行研究より，人の表情や視線，またそれらの動的な変化の仕方は，受け手の印象に大きな影響を与える．このことを踏まえると，人間の表情・視線の変化と印象評価の関係性を調査するには，多様な表情・視線変化で対面コミュニケーションを行った際の印象を多数の実験参加者から聴取する必要がある．しかし，生身の人間が顔表情を微妙に変化させ，その顔表情を多数の実験参加者に対して毎回完全同一に表現することは現実的に不可能である．さらに，人物に対する印象評価になることを避けるために，複数人の人間が同じ動作をする顔表情に対して印象評価を行いたいが，生身の人間では，このような実験は困難である．そこで本研究では，さまざまな人物がさまざまな表情・所作をとっている映像を自動生成できる手法を提案する．

この目的のための映像の生成のために用いられる技術として，ディープフェイクが挙げられる．ディープフェイクという言葉は本来，「ディープラーニングによって 2 つ以上の動画や画像の一部を交換する技術」を指す．従来，このような合成技術は，映画業界などの特別な機材やスタジオがある環境でしか実現できていなかったが，近年では，敵対的生成ネットワーク (GAN) の発展により，個人の PC だけでも，一枚の顔画像に対して，別の動画内の顔動作を反映した映像を生成することができる手法が提案されている．例を挙げると，Aliaksandr Siarohin らによって提案された First Order Motion Model [2] や，Ting-Chun Wang らによって提案された Few-shot Video-to-Video Synthesis [6] 等がある．また，顔に対する合成技術だけでなく，喋らせたい内容を入力すると，任意の人間があたかも本当に話している映像とその口の動きに合わせた音声を生成することができる Synthesia [7] や，映像中の人物や背景に対して，漫画のようなタッチにしたものをマッピングする技術 [8] など，ディープラーニングを活用した手法が数多く存在する．

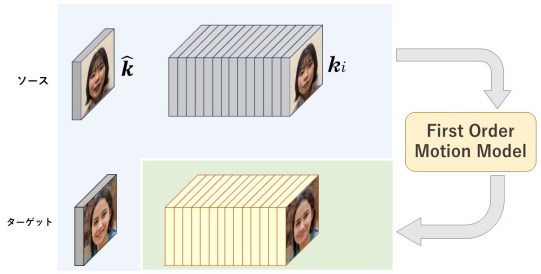


図 1: FOMM の概要

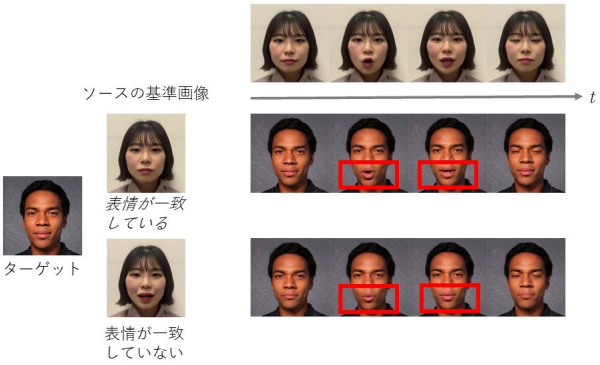


図 2: ソースの基準画像の違いによる生成結果の変化

本研究では，ディープフェイク技術の一手法である First Order Motion Model をベースとしながら，笑顔の度合いや視線動作の制御を施した顔表情映像を生成する手法を提案する．

3. 表情変化制御可能な顔映像生成手法

3.1 顔映像生成手法の概要

提案手法は，1 枚の静止画を参照動画（ソース）を元にリアルタイムで動画化することのできる手法のひとつである First Order Motion Model (FOMM) [2] をベースとしている．FOMM は，ソース・ターゲットの基準画像のペアとソースの動画を入力とし，ソースの動画をターゲットの動画に転移させる手法である（図 1）．この手法では，入力した顔画像から複数のキーポイントを抽出し，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動画の各フレームの特徴ベクトルをそれぞれ $\hat{\boldsymbol{k}}$，$\boldsymbol{k}_i$ とすると，これらの特徴ベクトルから算出される差分ベクトルは $\boldsymbol{k}_i - \hat{\boldsymbol{k}}$ である．この差分ベクトル $\boldsymbol{k}_i - \hat{\boldsymbol{k}}$ の動きがターゲットの画像と生成するターゲットの動画の各フレームの間にも生じていると考え，ターゲットの画像に $\boldsymbol{k}_i - \hat{\boldsymbol{k}}$ を転移する．そのため，ソースとターゲットの基準画像のペアは，顔表情が似たものを用いる必要がある．

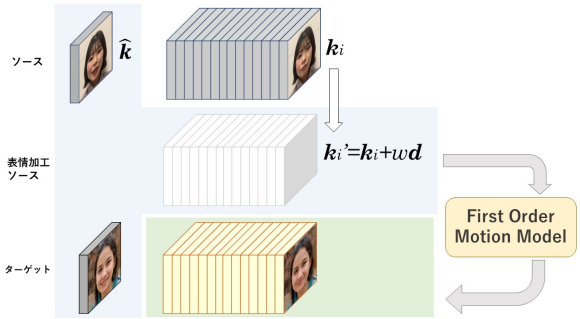


図 3: 提案手法の概要

図 4: 無表情顔

図 5: 笑顔

図 2 の下側に示したように，表情が大きく異なる基準画像ペア（ターゲットは口を閉じているのに，ソースの画像は口を開いている）を用いた場合は，生成される動画に破綻や違和感が生じる．

提案手法では，ソースの特徴ベクトル $\boldsymbol{k}_i$ に笑顔や視線変化に関する成分を付与した上で FOMM を適用することで，ターゲット動画の笑顔の度合いと視線方向を制御することを実現する（図 3）．笑顔や視線変化に関する成分について以降で説明する．

3.2 笑顔の度合い・視線動作の制御方法

笑顔の度合いを制御するためには，ソースの人物の無表情顔と笑顔の 2 種類の顔画像を用意する．一例として，今回の実験で使用した無表情顔と笑顔の画像をそれぞれ図 4，図 5 に示す．これらの画像から抽出された特徴ベクトルの差分ベクトルを $\boldsymbol{d}_{\mathbf{smile}}$ とし，その差分ベクトルをソースの特徴ベクトル $\boldsymbol{k}_i$ に付与することで，ソースの映像の顔動作をしながら，顔表情が笑顔になったターゲットの映像を生成することができる．また，$\boldsymbol{d}_{\mathbf{smile}}$ を $\boldsymbol{k}_i$ に付与する重みを w_{smile} $(0 \leq w_{\mathrm{smile}} \leq 1)$ とすると，w_{smile} を調整することで，笑顔の度合いを制御することが可能になる．また，視線方向の制御も同様に，正面注視と側方注視の 2 種類の顔画像から求めた差分ベクトル $\boldsymbol{d}_{\mathbf{eye}}$ とその重み w_{eye} $(-1 \leq w_{\mathrm{eye}} \leq 1)$ により算出する．

3.3 時系列信号による顔表情変化の制御

前節の処理は，1 枚 1 枚の静止画に対して行われるものであり，動画を生成するためには，差分ベクトルに付与する重み（w_{smile}，w_{eye}）の時間推移を時系列信号として入力する必要がある．しかし，この時系列信号を，ターゲットの動画の各フレームに対して手作業で設定するのは大変手間を要する．

表 1: 視線動作に関する重みの時間推移を時系列信号で表した例

t	w_{eye}
0	0
1	0
2	1
3	0
4	0

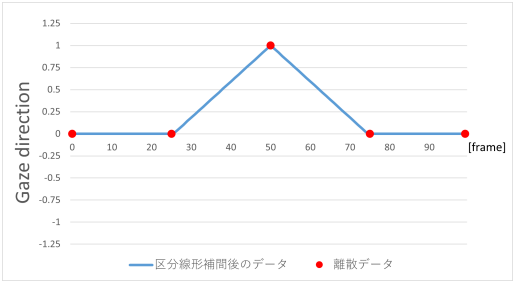


図 6: 平滑化処理後の時系列信号

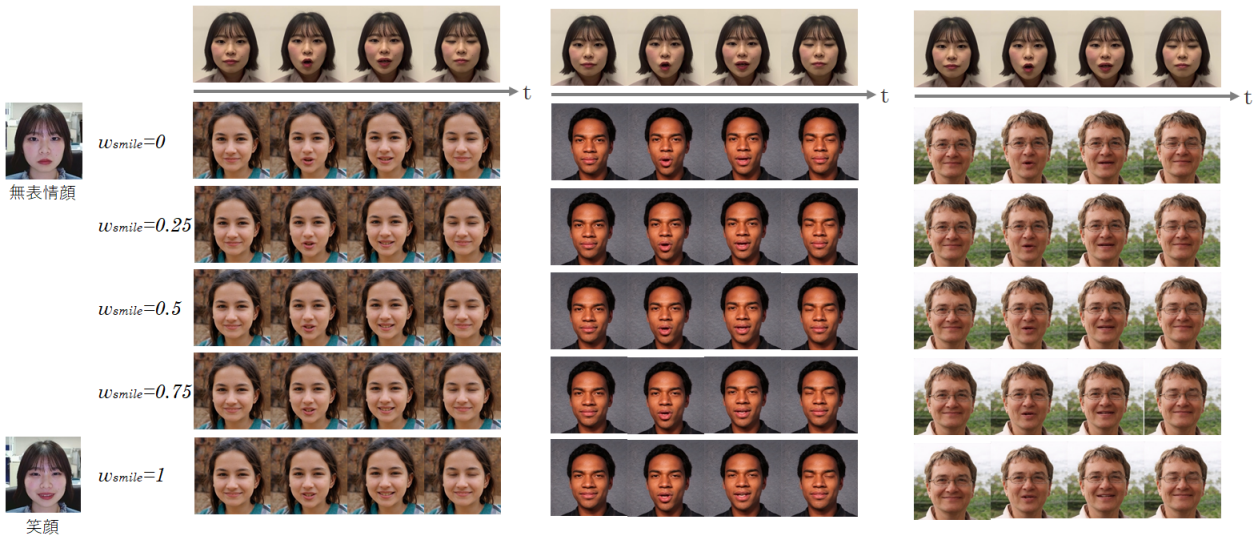
図 7: 正面注視

図 8: 右方向注視

そこで本手法では，粗い時間間隔での重みのみ与えれば，動画の各フレームにおける重みを密に算出できる機能を実装した．例えば，動画中のターゲットの視線動作が序盤と終盤には正面方向注視（$w_{\mathrm{eye}} = 0$）を行い，中盤は右方向注視（$w_{\mathrm{eye}} = 1$）を行うような動画を生成したい場合は，表 1 に示すような簡易な数字列のみ記述すれば，図 6 に示すような重みの時系列信号を生成する機能である．これにはまず，与えられた数字列の長さと動画の長さ（フレーム数）を線形に対応付け，図 6 の赤点のように離散データを生成する．そして，この離散データに区分線形補間処理を施し，同図の青線のような連続系列を生成した後，動画の各フレームに対応づけている．なお，笑顔の度合いも同様の制御方法を利用でき，笑顔の度合いと視線動作を同時に制御することも可能である．

4. 実験

提案した表情変化制御可能な顔映像生成手法により生成した結果を以降で示す．実験で使用したソースの映像には，接客応対の場面を想定して，「本日は，どのようなご用件でしょうか」と話しているものを用いた．笑顔の度合い



(a) 生成例 1　(b) 生成例 2　(c) 生成例 3

図 9: 生成動画の笑顔度調整

の制御のための差分ベクトル $\boldsymbol{d}_{\mathbf{smile}}$ を図 4 と図 5 から，視線動作の制御のための差分ベクトル $\boldsymbol{d}_{\mathbf{eye}}$ を図 7 と図 8 から算出する．今回，ターゲットの画像として用いる顔画像は，StyleGAN [9] により生成された実在しない人物の顔画像である．

4.1　笑顔の度合いの制御による生成結果

StyleGAN で生成された 3 人の顔画像に対して提案手法によりさまざまな笑顔の度合いで接客応対している動画を生成した結果を以下で示す．別のターゲットの画像を利用した生成結果を図 9 に示す．

4.2　視線動作の制御による生成結果

4.1 節の図 9 で使用した少女の画像をターゲット画像とし，視線動作を制御しつつ接客応対している動画を生成した結果を図 10 で示す．

4.3　時系列信号による顔表情変化の制御の結果

笑顔の度合いと視線を同時に変化させた結果を図 11 に示す．前節と同じく，4.1 節の図 9 で使用した少女の画像をターゲット画像として使用している．各図の上側のグラフは，w_{smile} と w_{eye} の時系列信号を表しており，指定された笑顔と視線の推移を適切に表現した生成結果を示している．

提案手法により，多様な人物に対して笑顔の度合い・視線動作が制御されつつも違和感や破綻のない自然な動画が生成できていることが確認できる．また，差分ベクトルの重みの時間推移を時系列信号として入力することで，指定された笑顔と視線の推移を適切に表現した映像が生成できていることが確認できた．

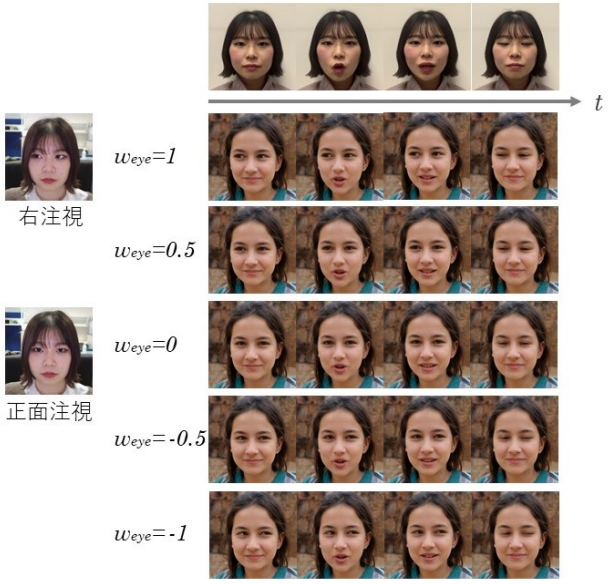
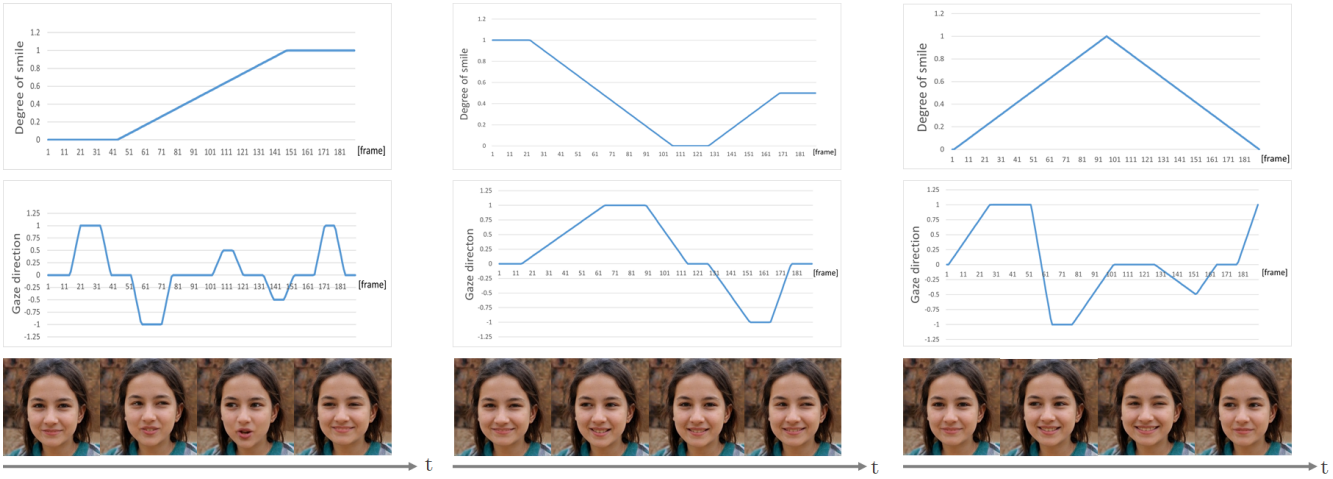


図 10: 生成動画の視線動作制御

5.　おわりに

本研究では，店員の表情・所作と店員に対する印象評価の関係の調査に向けたさまざまな人物がさまざまな表情・所作をとっている映像を自動生成できる手法を提案した．提案手法は，1 枚の静止画を参照動画（ソース）を元にリアルタイムで動画化する手法である FOMM をベースとしつつ，ソースの特徴ベクトルに笑顔・視線変化を施すことのできる差分ベクトルを付与することにより，生成動画の表情・視線を任意に制御することを実現した．また，差分ベクトルを付与する重みを変化させることで，笑顔の度合いと視線方向を任意に制御できるよう実装し，重みの時間



(a) 生成例 1　(b) 生成例 2　(c) 生成例 3

図 11: 笑顔度・視線の時系列信号を指定した動画生成

推移を時系列信号として表現しておくと動画に適切に反映させ，表情・視線を同時に変化させることも可能である．提案手法により生成した多様な顔表情映像から，表情・視線の変化が制御されつつも違和感や破綻のない自然な動画が生成できていることを確認した．

今後は，本研究の技術を利用した印象評価実験に取り組む．印象評価実験では，表情・視線の制御に加えて，瞬きの制御も行った多様な顔表情映像を生成し，それらの顔表情映像に対して，多数の実験参加者による印象評価を行う．その評価から，それぞれの顔表情映像に対する印象と，表情・所作の関係の分析を行いたい．

謝辞

本研究の一部は，内閣府総合科学技術・イノベーション会議の「SIP/ビッグデータ・AI を活用したサイバー空間基盤技術」（管理法人：NEDO）によって実施されました．また，本研究は JSPS 科研費 JP18H03312 の助成を受けました．

参考文献

[1] 桝井昇一, 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] A. Siarohin, S. Lathuilière, S. Tulyakov E. Ricci, N. Sebe,“First Order Motion Model for Image Animatio,” Conference on Neural Information Processing Systems (NeurIPS),December,2019.
[3] Kendon,A, “Some Functions of gaze direction in social encounters,” Acta psychologia, 26:1-47
[4] 土屋裕希乃,「会話場面における視線行動と満足度および印象評価の検討」, 国際経営・文化研究, Vol.21 No.1, December,2016.
[5] 菅原徹, 笠井直子, 佐渡山亜兵, 上條正義, 細谷聡, 井口竹喜,「笑顔の多様性と印象の関係性分析」, 日本感性工学会研究論文集, Vol.7 No2, pp.401-407, 2007 年.
[6] Ting-Chun Wang and Ming-Yu Liu and Andrew Tao and Guilin Liu and Jan Kautz and Bryan Catanzaro, “Few-shot Video-to-Video Synthesis,” Advances in Neural Information Processing Systems (NeurIPS),2019.
[7] https://www.synthesia.io/
[8] NAVER,https://webtoon.github.io/WebtoonMe/
[9] Tero Karras, Samuli Laine, Timo Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” CVPR2019, 2019.