複雑な変容を呈示する顔への追従投影のための画像変換の高速化の検討

野﨑 亮汰[*1]　　渡辺 義浩[*1]

Abstract – 従来の顔への追従投影では事前に作成したテクスチャ画像を顔の形状に合わせて投影し対象の見た目を操作していたため，複雑な変容を呈示することが困難だった．そこで，**Image-to-Image Translation (I2I)** を顔への追従投影に組み合わせることでモデル化が困難な変換をリアルタイムに実行し，複雑な変容を呈示可能になると期待できる．しかし，**I2I** は時間を要するため，顔の動きに対して投影像が遅れる問題があった．そのため本稿では，高フレームレート動画を入力とする **I2I** を対象として，時系列情報を基に画像変換領域を適応的にスパース化することによる **I2I** の高速化手法を提案する．また，顔への追従投影において，複雑な変容を呈示する応用例として年齢の変容がある．本稿では，年齢の変容を **I2I** に学習させ，変換された画像の画質や処理時間について評価を行った．

Keywords : **Dynamic Facial Projection Mapping, Image-to-Image Translation,** 年齢変容

1　はじめに

動的対象に映像を追従させて投影することで対象の見た目を操作する Dynamic Projection Mapping (DPM) が注目されている．なかでも，顔を対象として追従投影を行う DPM は Dynamic Facial Projection Mapping (DFPM) と呼ばれ，舞台演出やバーチャルメイクなどに応用されている [1, 2]．

DFPM では対象を撮像して特徴点の検出を行い，これを基に顔の 3 次元形状を推定し形状に合わせて投影画像を生成して追従投影を行っている．このため DFPM では撮像から投影までの処理時間が長い場合，対象の動きに対して投影像が遅延することで鑑賞者がずれを知覚し，没入感の低下につながる問題がある．また，DFPM において人間が上記のずれを知覚する遅延時間の閾値は，単純な顔の動きの場合に平均 3.87 ms になる実験結果が報告されている [3]．そのため，DFPM では一連の処理の高速性が求められる．

また，既存の DFPM では，事前に用意したテクスチャ画像を顔の形状に合わせて変形することで投影画像生成を行うシステムが提案されている [1, 2, 4]．このため，既存の DFPM における見た目の操作は事前に用意したテクスチャ画像の表現力に依存しており，顔の表情の変化に応じて現れる新たな特徴を呈示することができない課題がある．また，テクスチャ画像は事前に手動で用意する必要があるため，投影対象に合わせた個別の変容を呈示する画像を作成したい場合，手軽に体験を提供することが困難である．

一方，2 つの画像ドメイン間の対応関係を学習することで，入力画像を異なる見た目の画像に変換できる Image-to-Image Translation (I2I) と呼ばれる技術がある．I2I を用いることで対象を基にした複雑な変容を呈示する投影画像の生成が可能になる．例えば，人の顔の年齢感を変化させる複雑な変容を，U-Net を生成器に用いる FRAN と呼ばれる I2I によって実現した研究がある [5]．上記の変容は年齢変容と呼ばれ，対象の表情に合わせて目尻やほうれい線などの位置に大小さまざまな皺を自然に生成することで，写実的な年齢の変化を表現した．同手法を用いてリアルタイムに投影画像を生成し DFPM に用いることで，複雑な変容を投影によって呈示することが可能になる．しかし，同手法を含めた I2I 全般において，DFPM に用いるためには処理の高速性が不十分である課題がある．

上記に対し，I2I を DFPM へ導入するため，処理時間が短い軽量な CNN を用いて構成された LPTN [6] と呼ばれる I2I を，視覚における輝度と色に対する感度の違いと画像の周波数分解を用いて画質の劣化を軽微にとどめながら高速化した手法がある [7]．しかし，同手法は LPTN のネットワークの計算量が小さく処理時間が短い代わりに，過学習しやすく一部の領域で過度な変容が生成される課題や変容感が不足する課題があった．そのため，より計算量の大きいネットワークを用いる I2I を高速化することで，処理時間を短縮しつつ複雑な変換を呈示可能な I2I の提案が求められる．

ここで，画像の編集モデルにおいて，画像と画像に対しての編集指示がある場合に，編集される領域を選択的に処理し，条件付き GAN や拡散モデルを含む様々な変換モデルを高速化する SIGE と呼ばれる手法が提案されている [8]．SIGE は生成器のアーキテクチャを変更しないため，複雑な変容を呈示できる I2I の精度

[*1] 東京科学大学

を維持しながら高速化が図れると考えられる．

そこで，本稿では複雑な変容を呈示する顔への追従投影を実現するため，先述の FRAN に SIGE を適用して高速化を図る．ここでは，SIGE の編集領域のみを処理することで高速化する考えを応用し，高速カメラを用いて撮像したフレーム間の時系列情報を基に変換を行う領域を適応的に選択することで，DFPM に用いる I2I を高速化する手法を提案する．また，高フレームレート撮像を行うと，入力画像に多くのノイズが含まれるため，ノイズの影響を低減し変換領域が小さくなるように決定する．さらに，DFPM において，複雑な変容を呈示する応用例として年齢変容がある．提案手法の有効性を検証するため，年齢変容を I2I に学習させ，変換された画像の画質や処理時間について評価を行う．

2 関連研究

2.1 顔への追従投影

DFPM における対象と投影像の間の遅延について，人間が知覚できる位置ずれを回避するための最小の遅延時間を意味する，遅延時間の弁別閾を調査した研究がある [3]．同研究では，顔と投影像の遅延をシミュレートした動画を作成し，被験者実験を行うことで，顔の移動速度が 0.5 m/s のときに，遅延時間の弁別閾が平均 3.87 ms であることを明らかにした．このような要請のもと，顔の特徴点検出，3 次元形状推定，画像変形の流れで処理をする DFPM において，撮像・処理・投影までを 4.869 ms で実現する手法が提案されている [1]．

一方，DFPM において複雑な変容を呈示することに注力した，年齢変容 DFPM が提案されている [4, 9]．具体的には，変容目標から年齢感を表す成分を抽出したテクスチャ画像を事前に生成し変形して投影する手法 [4] や，カメラフィードバックを用いて投影によって対象の見た目を変容目標を表すテクスチャ画像に近づける手法 [9] がある．前者では対象が表情を変えた場合に，対象の皺と投影像の皺がずれる課題，後者ではフィードバック計算の処理内部でガウシアンブラーを用いていたために，投影結果にぼけが生じる課題があった．また，2 手法に共通してテクスチャ画像を事前に生成していたため，表情の変化に伴う変容感の呈示が十分ではない課題があった．上記に対し，I2I を用いてリアルタイムに投影画像を生成することで，皺のずれや手法を原因とするぼけが無い，表情の変化に伴う変容感の呈示が可能になると考えられる．

2.2 Image-to-Image Translation

I2I は 2 つの画像ドメイン間の対応関係を学習し，入力画像を別ドメインの画像に変換する技術である．このため，解析的なモデル化をしにくい複雑な変容を表現することが可能である．また，データセットを作成し学習に用いることで，様々な変容が実現でき，例えば年齢変容を写実的に実現した手法がある [5]．

ここで，I2I は StyleGAN2 [10] のような大規模モデルをベースとする画像変換手法と比較すると 1 枚の画像の変換に要する処理時間が短いが，DFPM に用いるには処理時間が長い課題がある．これに対し，I2I の高速化はモバイル端末や組み込みシステムで動作させる目的から注目されている．例えば畳み込み層の層数を削減する手法や，学習済みモデルの出力結果に近づくように層数の少ない生成器を学習させる手法などが提案されている [11, 12]．しかし，これらの手法はアーキテクチャの削減に伴い変換の精度が低下する課題がある．

一方，編集を行う領域に対して選択的に変換を実行し，条件付き GAN や拡散モデルを含む様々な変換モデルを高速化する SIGE と呼ばれる手法がある [8]．SIGE では入力画像上において編集を行う領域から，生成器の各畳み込み層で変換を行う領域を適応的に決定し，同変換領域のみを変換することで高速化を図る．そのため，SIGE では生成器のアーキテクチャを変更しないので，先述の高速化手法と比較して変換の精度が劣化しづらいと考えられる．一方，同手法はユーザの編集指示を基に画像を変換する手法の高速化に向けて設計されているため，1 枚の画像に対し編集領域が変化する画像変換手法に特化している．これに対し DFPM に用いる I2I では，入力が高フレームレートである点と時系列情報からフレーム間の画像の変化を得られる点から，同手法の編集領域のみを変換することで高速化を行う考えを応用できると考えられる．しかし，SIGE では 512 × 512 pixels の画像に対して，編集指示を基に各畳み込み層の変換領域を決定する処理に数ミリ秒要し，同決定処理と変換処理を逐次的に行うと処理時間が延びる課題がある．

3 フレーム間の時系列情報を用いた画像変換領域の適応的スパース化による高速化

3.1 手法概要

本稿では，U-Net ベースの生成器を持つ FRAN [5] に対し，SIGE [8] をフレーム間の時系列情報に基づいて画像変換領域を適応的にスパース化する構成に拡張することで高速化を図る．図 1 に提案手法の概要図を示す．DFPM では低遅延化のため，撮像には高フレームレートの高速カメラを用いる．このとき，高速撮像によりフレーム間の画像の変化が小さい特徴がある．そこで，フレーム間の差分を基に変換領域を決定することで，対象の動作に合わせて変換が必要な領域のみ

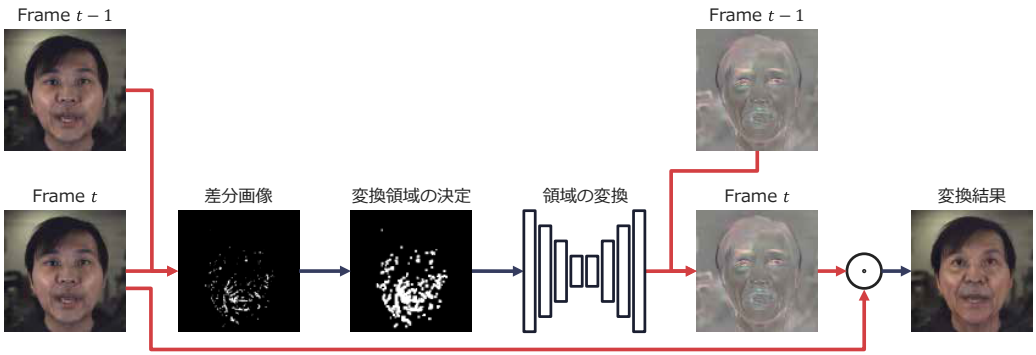

図 1　顔への追従投影における I2I に対して SIGE [8] を用いる高速化手法.
Fig. 1　Acceleration method using SIGE [8] for I2I in dynamic facial projection mapping.

を適応的に処理することが可能となる．SIGE は変換領域が小さいほど高速化の効果が大きいため，本条件下では処理の高速化が期待できる．ただし，SIGE および本手法のいずれにおいても，1 フレーム目は参照フレームが存在せず変換領域を決定できないため，画像全体を変換する必要がある．

また先述の通り，SIGE では編集指示に基づく変換領域の決定処理に数ミリ秒要し，逐次的に処理を行うと処理時間が延びる課題がある．そこで，本手法ではフレーム間の差分を基に変換領域を決定する処理と I2I の変換処理を並行化し，その時点で利用可能な最新の変換領域を用いて変換処理を行う．図 2 に同処理の概要を示す．上記では，変換領域の決定処理と I2I の変換処理で異なる入力画像を用いるため，対象の動作によって領域がずれ，適切な変換が行えない可能性がある．しかし，高フレームレート撮像下では，フレーム間における画像変化が小さい特徴がある．加えて，SIGE ではフレーム間の差分を含むようにブロック単位で変換領域を決定するため，領域のずれがブロック内に収まると期待される．

ここで，FRAN は畳み込み層，プーリング層，バッチ正規化，アップサンプリング処理などによって構成されるが，SIGE では処理のスパース化が畳み込み層にのみ適用されている．そこで本手法でも FRAN の畳み込み層のみにスパース化を適用する．そのため，

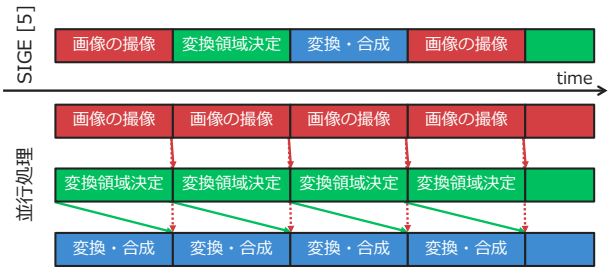

図 2　変換領域の決定処理と I2I の変換処理の並行化.
Fig. 2　Parallelization of transformation region determination and I2I transformation processing.

畳み込み層以外の処理は画像全体に対して行われる．さらに，FRAN では畳み込み層で処理を行った後，常にバッチ正規化を用いている．処理の削減のため，変換時において，バッチ正規化の処理を畳み込み層の処理と統合する．

3.2　差分の作成と変換領域の決定処理

SIGE では，2 画像間の差分領域を含むようにブロック単位の領域を生成し，これを基に畳み込み層の各解像度で変換領域が決定される．動的対象を撮像して変換を行う I2I では，フレーム間で対象の動きによる変化以外にノイズが含まれるため，意図せず変換領域が拡大する問題がある．このノイズは主にカメラと環境光の変化に起因する．そのため，変換領域を削減し高速化を図るために，ノイズを除去し差分画像から対象の動き領域を特定し，変換領域として決定する必要がある．そこで，差分画像，変換領域の決定処理を行う前に撮像画像にガウシアンフィルタを適用する．

まず，画素 p における 1 フレーム前の画像との差分 $d_{t,1}(p)$ と，それに基づく変換領域 $b_{t,1}(p)$ を下式で求める．

$$\begin{aligned} d_{t,1}(p) &= \left\| \tilde{I}_t(p) - \tilde{I}_{t-1}(p) \right\|_1 \\ b_{t,1}(p) &= [d_{t,1}(p) > \tau_1] \end{aligned} \tag{1}$$

ここで，$I_t \in [0,1]^{H \times W \times C}$ はフレーム t の画像，$\tilde{I}_t = G(I_t)$ はガウシアンフィルタを適用した画像，τ_1 は閾値である．

また，上記の連続フレーム間の差分のみに基づく変換領域決定では，緩やかに画素値が変化する領域を変換領域として検出できない問題がある．そこで，画素 p において，現在フレーム t と $\tilde{n}$ フレーム前の画素値の差分を用いる．さらに，ある画素が一度変換領域として判定された後は，変換領域の過度な拡大を抑えるため，当該画素の差分計算で参照する過去フレームをリセットする．このとき，画素 p における直近のリセットからの経過フレーム数を n' とし，差分を計算するフレームの間隔を $\tilde{n} = \min(n, n')$ と定める．以上よ

り，差分 $d'_{t,\tilde{n}}(p)$ と，それに基づく変換領域 $d'_{t,\tilde{n}}(p)$ を次式で定義する．

$$d'_{t,\tilde{n}}(p) = \left\| \tilde{I}_t(p) - \tilde{I}_{t-\tilde{n}}(p) \right\|_1$$
$$b_{t,\tilde{n}}(p) = \left[d'_{t,\tilde{n}}(p) > \tau_{\tilde{n}} \right] \quad (2)$$

ここで，$\tau_{\tilde{n}}$ は閾値である．

最後に，式 (1),(2) の 2 処理を組み合わせて得られる変換領域 $b_t(p)$ は下式によって決定される．

$$b_t(p) = [b_{t,1}(p) \vee b_{t,\tilde{n}}(p)] \quad (3)$$

さらに，ノイズ低減効果を高めるため，周囲に変換領域に含まれている画素が少ない画素は変換領域に含めないように決定する処理を加える．これにより，ノイズのような周囲に変化している領域が無い画素が変換領域として選ばれないようになる．

4　実験

4.1　概要

本手法の評価を行うため，顔の動作についての 512×512 pixels，500 fps，5 秒間の動画を 4 種類用意し，入力として用いた．これらの動画では 5 秒間の間にそれぞれ平行移動 (tr)，回転運動 (rot)，複数の表情への変化 (emo)，発話 (talk) の動きと静止した状態を含んでいる．上記の動画に対し，本手法を導入した FRAN を用いて変換を行った．ここで，3.1 節で述べた処理の並行化と利用可能な最新の変換領域を用いる処理をシミュレートするため，3 フレーム前の入力画像から決定した変換領域を用いて変換を行った．3 フレームという値は変換領域の決定の処理時間を基に決定した．また，LPTN [6] や LPTN を高速化した LPTN' [7] を用いて変換を行った結果と比較し評価を行った．

4.2　学習条件

変換によって年齢変容を呈示するように I2I の学習を行うためには同一人物の異なる年齢における画像を含んだ顔画像データセットが必要である．そのため，まず StyleGAN2 [10] を用いて 512 × 512 pixels の 2,000 人分の合成顔画像を作成した．次に，作成した顔画像に対し SAM [13] を用いて 18 歳から 85 歳の間の 14 年齢に顔の年齢感を変容させた顔画像を作成した．FRAN は画像と元の年齢，目標とする年齢を入力として変換を行うため，異なる年齢感の変容を 1 度の学習で表現が可能である．一方，LPTN は画像が入力されると変換が行われ，1 度の学習で単一の変容のみ表現できる．そのため，FRAN にはデータセット全体を用いて同一の年齢も含む 14 年齢間の変換を学習させ，LPTN の学習では 2 年齢のみを抽出して 23 歳から 75 歳への変換を学習させた．また，投影における変容結果は，対象の顔の反射率と投影画像の乗算で表現できると仮定した．この仮定に基づき，入力画像に出力画像を乗算した結果が年齢感の変化した画像となるように，FRAN および LPTN の生成器を学習させた．

さらに，FRAN, LPTN の学習において共通の損失関数を用いた．損失関数は Zoss らが FRAN の学習に用いたもの [5] を用い，その内容を式 (4) に示す．

$$\mathcal{L} = \lambda_{L1}\mathcal{L}_{L1} + \lambda_{LPIPS}\mathcal{L}_{LPIPS} + \lambda_{adv}\mathcal{L}_{adv} \quad (4)$$

上式の $\mathcal{L}_{L1}$ は目標画像と生成画像の $L1$ ノルム，$\mathcal{L}_{LPIPS}$ は同じく目標画像と生成画像を入力とした LPIPS，$\mathcal{L}_{adv}$ は識別器を用いた敵対的損失を表す．また，λ_{L1}，λ_{LPIPS}，λ_{adv} はそれぞれの損失関数の重みを表しており，FRAN の学習では $\lambda_{L1} : \lambda_{LPIPS} : \lambda_{adv} = 1 : 1 : 0.05$，LPTN の学習では $\lambda_{L1} : \lambda_{LPIPS} : \lambda_{adv} = 0.5 : 1 : 0.2$ と設定した．さらに，学習において識別器は PatchGAN Discriminator を用い，Adam Optimizer を使用して学習を行った．

上記以外の学習の条件について，FRAN の学習では，年齢間の変換 196 組を 2,000 枚の顔画像に対しそれぞれ 1 度ずつ学習させることを 1 epoch とし，20 epoch の学習を学習率 $1e^{-5}$ で行った．一方，LPTN の学習では 300,000 iterations 分の学習を学習率 $1e^{-5}$ で行った．

4.3　定性評価

まず，動画の変換を行う際にいくつかのパラメータを指定する必要がある．ここで，ガウシアンフィルタを過度に強くかけるとノイズだけでなく，対象の動きに伴う変化もボケる．そのため，ガウシアンフィルタのカーネルを 7×7, σ を 2 とした．また，式 (1) において $\tau_1 = 1e^{-2}$，式 (2) において $n = 5$，$\tau_{\tilde{n}} = 5e^{-3}$ と設定した．さらに，変換領域の決定において，周囲の 24 画素の内 6 割以上が変換領域でない場合は，その画素が変換領域に含まれないようにした．

ここで動画は 500 fps であるため，フレームごとに評価すると人間の知覚とは異なる条件で評価される．そのため，500 fps の出力動画を基に 60 fps の動画を作成した．上記の条件下で本手法を用いて表情変化，平行移動の動画を変換した結果と，FRAN や LPTN，LPTN' を用いて変換した結果を図 3 に示す．ただし，同図では 60 fps の動画の 1 フレームを載せた．

まず，図 3 において FRAN を用いた変換結果では，表情の変化に伴って生成される目元の皺の密度の変化や頬の皺の深さの変化が確認された．一方，LPTN を用いた変換結果では，白唇部に髭が生成され過度に青緑色に変化する課題や，目の周辺に細かな皺が生成されるのではなく黒く変色する課題があり，写実性が十

表 1　動画の変換結果に対し，PSNR，SSIM，LPIPS，FovVideoVDP [14] を用いた評価．FRAN [5] による変換結果を Reference として用いた．
Table 1　Evaluation using PSNR, SSIM, LPIPS and FovVideoVDP [14] for the video transformation results. The transformation results using FRAN [5] are used as a reference.

		Ours	LPTN [6]	LPTN' [7]
tr	PSNR (↑)	**39.312**	30.645	30.642
	SSIM (↑)	**0.9826**	0.9366	0.9365
	LPIPS (↓)	**0.0409**	0.0733	0.0735
	FovVideoVDP (↑)	**7.6721**	6.3055	6.3015
rot	PSNR (↑)	**40.349**	30.205	30.209
	SSIM (↑)	**0.9822**	0.9396	0.9394
	LPIPS (↓)	**0.0396**	0.0796	0.0799
	FovVideoVDP (↑)	**7.8069**	6.2055	6.2033
emo	PSNR (↑)	**44.059**	30.237	30.234
	SSIM (↑)	**0.9880**	0.9249	0.9248
	LPIPS (↓)	**0.0132**	0.0791	0.0792
	FovVideoVDP (↑)	**8.6359**	6.0832	6.0773
talk	PSNR (↑)	**42.813**	30.244	30.243
	SSIM (↑)	**0.9844**	0.9186	0.9185
	LPIPS (↓)	**0.0182**	0.0829	0.0830
	FovVideoVDP (↑)	**8.3843**	5.9848	5.9823

分ではない．これは，生成器のアーキテクチャの違いにより，LPTN と比較して FRAN の方が表現力が高いことに起因すると考えられる．また，FRAN と本手法の結果を比較すると顔領域において細かな違いは確認できない．一方，平行移動の変換結果では背景において黒いノイズ状のアーティファクトが確認された．これは，本来変換領域とすべき部分が領域決定処理の段階で見落とされたことに起因すると考えられる．

4.4　手法導入に伴う変換結果の変化の評価

まず，60 fps の動画について，各フレームの画像全体に対して PSNR，SSIM，LPIPS を用いた評価を行った．さらに，人間の視覚特性をモデル化した動画評価指標である FovVideoVDP [14] を用いて，500 fps の出力動画を評価した．ここで，FovVideoVDP は，参照動画と比較対象動画を入力として，最大値が 10 の Just-Objectionable-Difference (JOD) を出力する指標である．評価にあたって，FRAN によって変換した動画を Reference として用い，表 1 に結果を示す．同表より，提案手法は LPTN および LPTN' よりも各指標で良好な値を示している．これは，提案手法が FRAN の変換結果を Reference として設計されているためであり，妥当な結果であると考えられる．

また，表 1 の結果を動作の種類ごとに比較すると，対象の動きが比較的大きく，顔全体が剛体運動を行う平行移動および回転運動では，各指標の結果が悪いことが確認できる．一方，対象の動きが比較的小さく，顔内部が非剛体運動を行う表情変化や発話では，各指標の値が良いことが読み取れる．以上より本手法は，画像内における対象の動きが小さい動作に対して，画質の劣化が比較的抑えられる傾向があると考えられる．

表 2　提案手法を用いる変換の処理時間と変換率．FRAN [5]，FRAN' を比較対象とする．
Table 2　Processing time and transformation rate of the proposed method. FRAN [5] and FRAN' are used as a comparison.

		FRAN [5]	FRAN'	Ours
tr	処理時間 [ms]	7.067	**6.465**	6.745
	変換率 [%]	-	-	30.96
rot	処理時間 [ms]	7.085	**6.463**	6.613
	変換率 [%]	-	-	30.48
emo	処理時間 [ms]	7.120	6.464	**5.293**
	変換率 [%]	-	-	11.98
talk	処理時間 [ms]	7.025	6.456	**5.481**
	変換率 [%]	-	-	14.71

4.5　処理時間

4.4 節と同様のパラメータを用いて顔の 4 種の動作の動画に対して変換を行い，処理時間の計測を行った．本稿では，変換領域の決定処理と画像の変換処理のうち，後者のみを処理時間として計測した．また，画像における変換領域の割合を変換率として表に示す．計測には CPU に Intel Core Ultra 9 285K，GPU に RTX PRO 6000 Blackwell Workstation Edition を搭載した計算機を用いた．比較のため，FRAN と FRAN のバッチ正規化の処理を畳み込み層の処理と統合した FRAN' の処理時間を計測した．表 2 に各動画に対する計測結果の平均値を示す．同表より，本手法は FRAN と比較して 4 種すべての動作の動画に対する変換において処理時間の削減が確認できる．特に，対象の動きが小さい動作では変換率が小さく，FRAN と比較して処理時間は 1.5 ms 以上短縮された．また，表情変化の動画では約 1.345 倍の高速化を達成した．一方，対象の動きが大きい平行移動と回転運動では FRAN' よりも処理時間が長いことが確認された．

5　考察

表 1 に示される FovVideoVDP を用いた本手法に対する結果は 9 を下回っていた．JOD が 9 以下の場合，比較対象動画は参照動画と比較して劣化していると見なされる．つまり，表 1 は本手法の変換結果が FRAN と比較して劣化していることを示している．しかし，図 3 を確認すると，顔領域では明確な劣化を確認できない．そのため，表 1 の結果は平行移動の変換結果のような背景におけるアーティファクトによって JOD が低く算出されていると考えられる．ゆえに，今後は顔領域に限定した評価を行う必要がある．

また，表 2 から，対象の動作によって変換領域の大

図 3　表情変化，平行移動の動画に対する変換結果を 60 fps に変換した画像.
Fig. 3　Images translated from 500 fps video of facial expressions and parallel movement, converted to 60 fps.

きさが異なり，変換率の低下に応じて処理時間が削減できる傾向が読み取れる．このことから，画質を維持しつつ変換領域を可能な限り小さくする決定処理が重要である．しかし，本手法では対象の動作に対する差分の作成および変換領域の決定処理について，十分に最適化されておらず，さらなる改善が必要だと考えられる．加えて，提案手法では環境や投影距離に応じたパラメータ調整が必要である．そのため，この調整は変換領域の決定に大きな影響を及ぼす要因である．このパラメータ調整の自動化が可能になれば，DFPM への導入時に変換領域が対象の動きに合わせて適切に決定され効果的な高速化が期待できる．

6　まとめ

本稿では，複雑な変容を呈示する顔への追従投影を実現するため，時系列情報を基に画像変換領域を適応的にスパース化することにより FRAN を高速化する手法を提案した．また，提案手法の有効性を検証するため，変換結果の定性評価に加え，変換結果の画質および処理時間の定量評価を行った．その結果，対象の表情に合った変容感を呈示する I2I に対し，顔の動きが小さい動作では約 1.345 倍の高速化が達成されたことを確認した．

一方，本稿の評価は，画像上での変換結果に関する定性評価と，画質・処理時間の定量評価にとどまっており，変容感そのものについては十分に検証できていない．そのため，今後は被験者実験などを通じて，変容感を評価する必要がある．加えて，本稿では画像上の評価に限定しているため，実投影環境において呈示される変容の効果についても検証が必要である．

参考文献

[1] Hao-Lun Peng, et al. Perceptually-Aligned Dynamic Facial Projection Mapping by High-Speed Face-Tracking Method and Lens-Shift Co-Axial Setup. *TVCG*, 2025.
[2] Nao Tsurumi, et al. Rediscovering your own beauty through a highly realistic 3D digital makeup system based on projection mapping technology. *IFSCC*, 2023.
[3] Hao-Lun Peng, et al. Studying User Perceptible Misalignment in Simulated Dynamic Facial Projection Mapping. *ISMAR*, pp. 493–502, 2023.
[4] 袁璐ほか．プロジェクションマッピングによる顔の年齢変容に関する検証．第 68 回複合現実感研究会, MR2023-4, 2023.
[5] Gaspard Zoss, et al. Production-Ready Face Re-Aging for Visual Effects. *TOG*, Vol. 41, No. 6, 2022.
[6] Jie Liang, et al. High-Resolution Photorealistic Image Translation in Real-Time: A Laplacian Pyramid Translation Network. *CVPR*, pp. 9387–9395, 2021.
[7] 野崎亮汰, 渡辺義浩. 動的対象への追従投影のための image-to-image translation の高速化. 第 30 回日本バーチャルリアリティ学会大会, 2E1-09, 2025.
[8] Muyang Li, et al. Efficient Spatially Sparse Inference for Conditional GANs and Diffusion Models. *TPAMI*, Vol. 45, No. 12, pp. 14465–14480, 2023.
[9] 袁璐ほか．色補償を用いたプロジェクションマッピングによる顔の年齢変容に関する検討．第 28 回日本バーチャルリアリティ学会大会, 2B2-08, 2023.
[10] Tero Karras, et al. Analyzing and Improving the Image Quality of StyleGAN. *CVPR*, pp. 8107–8116, 2020.
[11] Yihui He, et al. Channel Pruning for Accelerating Very Deep Neural Networks. *ICCV*, pp. 1398–1406, 2017.
[12] Geoffrey Hinton, et al. Distilling the knowledge in a neural network. *arXiv*, pp. 1–9, 2015.
[13] Yuval Alaluf, et al. Only a matter of style: age transformation using a style-based regression model. *TOG*, Vol. 40, No. 4, 2021.
[14] Rafał K. Mantiuk, et al. FovVideoVDP: a visible difference predictor for wide field-of-view video. *TOG*, Vol. 40, No. 4, 2021.

© 2026 by the Virtual Reality Society of Japan (VRSJ)