

三次元表現に基づく照明効果を考慮した化粧の抽出と応用

楊 興超^{1,2,a)} 武富 貴史^{1,b)} 金森 由博^{2,c)}

概要: 映画や広告などの制作現場において、バーチャルアバターなどの三次元顔モデルに対する化粧のニーズが高まっている。しかし、三次元の顔に直接化粧を施すことは手間がかかり、二次元顔画像から化粧パターンを抽出・転写する既存手法は照明による陰影や遮蔽の影響を受けやすい、という問題がある。本研究では一枚の化粧付き顔画像から、三次元顔復元、逆レンダリング、化粧転写の技術を活用することによって、化粧パターンを三次元顔モデルの UV マップとして抽出する手法を提案する。提案手法で抽出した化粧パターンは、三次元顔モデルだけでなく、二次元の顔画像にも応用可能である。例として、大規模な化粧パターンデータセットとそれを用いたパラメトリックな化粧モデルの構築や、顔写真から化粧のレイヤーを分離することで、照明効果を考慮した化粧の転写や補間・除去を実現できることを示す。

1. はじめに

化粧は見た目の印象を操作する有力な手段であり、VR/AR、ゲーム、オンラインコマースなどのアプリケーションを制作する際にも重要な要素となっている。特に、映画や広告業界では、写実的なデジタルヒューマンの活用とともに三次元顔モデルに対する化粧の需要が増加しつつある。

三次元顔モデルに化粧を施す方法として、テクスチャを手動で修正する方法や特別な計測機器を利用する方法 [1] があるが、これらは時間やコストがかかり、インターネット広告などで大量の化粧パターンを作成する必要がある場合に問題となる。一方、二次元顔画像に対する化粧転写技術は、手軽に化粧パターンを置き換え、見た目のシミュレーションを行うことができることから多くの研究がなされている [2]。しかし既存手法では、照明による陰影や自己遮蔽の影響を受け易い、三次元顔モデルへの応用は難しい、といった問題がある。

本研究では、一枚の顔画像から、三次元顔モデルで利用可能な形式で化粧パターンを抽出する手法を提案する。図 1 に提案手法による処理結果の一例を示す。提案手法では、まず、三次元顔変形モデル (3D Morphable Model: 3DMM) [3] を利用し、形状、拡散アルベド、鏡面アルベド



図 1 化粧写真の逆レンダリングと復元。上段は入力写真とレンダリング画像の重畳表示 (左から素肌のみ、素肌+化粧、素肌と拡散シェーディングの乗算、さらに鏡面復元成分を加算)、下段は分解で得られた各マテリアルの UV テクスチャ表現。

などの粗いマテリアル情報を、三次元モデルで一般的に用いられている UV テクスチャとして抽出する。その後、逆レンダリング技術に基づく最適化により詳細な情報を復元する。最後に、化粧転写技術を活用することで、復元された高精細な拡散アルベドから素肌、化粧パターンを抽出する。図 2 に示すように、提案手法は、再照明可能な化粧の転写、化粧パターンの補間および除去、大規模な化粧画像データセットの構築とそれに基づくパラメトリック化粧モデルの構築、などへの応用が可能である。

2. 関連研究

提案手法は、三次元顔復元、逆レンダリングによる画像分解、化粧転写、3つの技術要素から構成される。以下、それぞれの技術要素について、関連手法を概観する。

2.1 単一画像からの三次元顔復元

画像からの三次元顔モデルの復元では、遮蔽や複雑な照

¹ 株式会社サイバーエージェント
CyberAgent, Inc.

² 筑波大学
University of Tsukuba

a) you_koutyo@cyberagent.co.jp

b) taketomi_takafumi@cyberagent.co.jp

c) kanamori@cs.tsukuba.ac.jp



図 2 提案手法の応用例。左から順に、化粧が制御可能な三次元顔復元，照明を考慮した化粧の補間・除去および再照明，化粧転写，化粧の抽出とデータセットの作成。

明効果の影響があるため，事前知識なしに単一画像からモデルを復元することは困難である．既存の三次元顔復元手法は，一般に，事前知識として 3DMM [3,4] を利用している [5-8]．三次元顔復元は，レンダリングした 3DMM と入力画像の間で顔ランドマークの位置誤差などを最小化することによって実現されている．3DMM の構築や三次元顔復元に関する既存研究については，文献 [9,10] を参照されたい．提案手法では，文献 [7] と同様に，光源環境として仮想的なライトステージ環境を想定した顔の三次元形状復元を行う．また，顔の遮蔽による顔テクスチャの欠損の問題を解決するために，UV テクスチャ補完手法 DSD-GAN [11] を採用する．

2.2 逆レンダリングによる画像分解

人物画像に関連する逆レンダリングによる画像分解手法について概観する．SfSNet [12] は，実画像と合成画像を用いてネットワークを学習することで，入力顔画像を形状，反射率，照度に分解する手法を提案した．Relighting Humans [13] は，光の遮蔽を考慮することで，より写実的な再照明を実現した．これらの手法は，照明を 2 次の球面調和関数 (Spherical Harmonics; SH) [14] で表現しているため，拡散反射以外への対応が困難であった．一方で，ライトステージデータを用いて，光沢や環境マップなどを推定する手法が提案されている [15,16]．本研究では，化粧に着目し，ライトステージデータを用いずに顔画像を素肌，化粧，照明効果などの情報に分解する．

2.3 化粧転写

一般に，化粧転写技術では，参照画像の化粧パターンを対象画像に転写することを目的としている．BeautyGAN [17] は化粧転写技術のベースラインとして用いられる手法であり，化粧前後の画像ペアを必要としない方法である．後続のアプローチでは，濃い化粧の転写 [18]，顔の位置や角度の違い [19-21]，陰影や遮蔽の影響 [22]，編集可能性 [23] などの点において改善がなされている．しかし，既存手法

には，照明効果が考慮されていない，二次元画像のみを対象としている，という問題が残っている．本研究では，三次元モデルの表現として一般に用いられている UV テクスチャ上で処理を行うことでこれらの問題を解決する．

3. アプローチ

図 3 に示すように，提案手法は，1) 顔画像への 3DMM 当てはめによる粗いマテリアルの推定，2) 逆レンダリングによるマテリアルの高精細化，3) 化粧転写技術を活用した素肌成分と化粧成分の分離，の 3 つのステップで構成されている．ステップ 1 では，三次元顔復元手法 [6] を拡張し，鏡面反射の取り扱いを可能とする．次にステップ 2 では，ステップ 1 で得られた粗いマテリアルを初期値とし，逆レンダリングに基づく最適化によって高精細化する．最後にステップ 3 では，高精細な拡散アルベドを入力として，素肌成分と化粧成分を抽出する．以下，各ステップの詳細を述べる．

3.1 粗いマテリアルの復元

逆レンダリングに基づく 3DMM の当てはめにより粗いマテリアル情報を取得する．鏡面反射を推定するために，FLAME モデル [3] の拡張であり，鏡面アルベドを取り扱い可能な AlbedoMM [24] を利用する．これにより，文献 [6] の顔復元ネットワークを，拡散反射だけでなく鏡面反射も推定できるように拡張し，新しい三次元顔復元ネットワーク FRN を構築した．本研究では，仮想ライトステージを用いた顔復元手法 [7] を参考に，正二十面体上に 20 個の平行光源を設定して推定に用いる．仮想ライトステージは鏡面反射の推定のみで使用され，光源の強度および方向は調整できるようにする．また，FLAME が限られたデータから構築されていることによる実画像との間に生じる差異を補償するため，拡散アルベドを調整するパラメータを導入する．

3DMM の形状 G_c ，拡散アルベド D_c ，鏡面アルベド S_c は以下のように定義される．

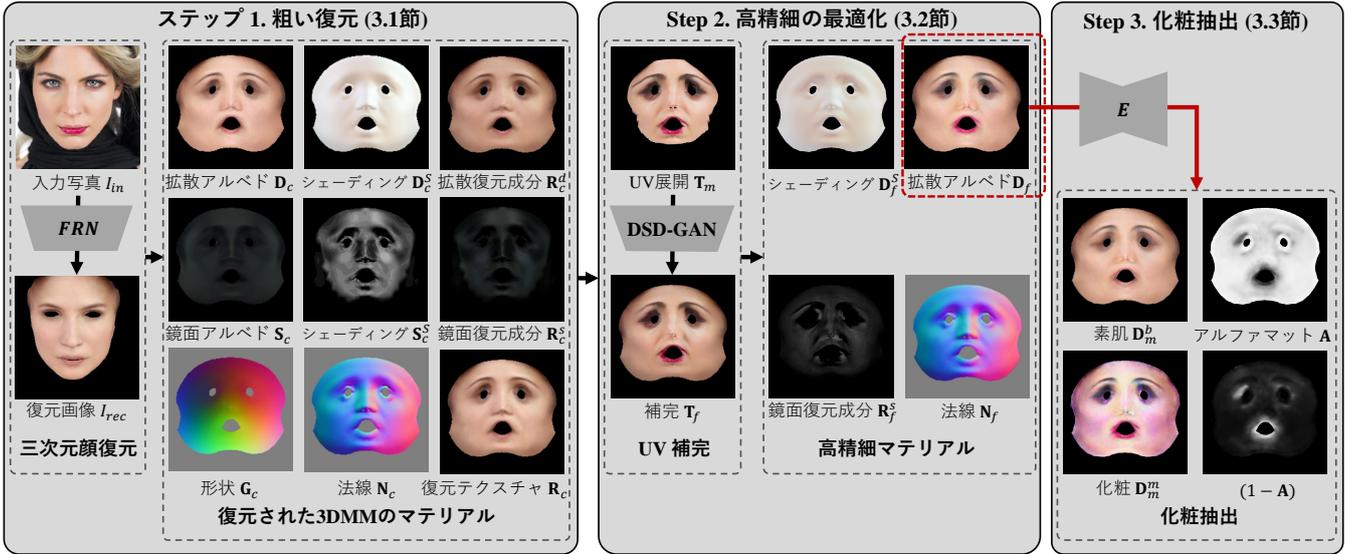


図3 提案手法の概要. 3つのステップによって, 化粧写真から素肌と化粧パターンを抽出する. まず, 三次元顔復元ネットワーク FRN を用いて, 粗いマテリアルを推定する (3.1節). 次に, 展開したテクスチャが欠損する問題を解決するため, UV テクスチャ補完ネットワーク $DSD-GAN$ [11] を適用する. その後, 粗いマテリアルに対して最適化を行う (3.2節). 最後に, 化粧抽出ネットワーク E を用いて, 高精細な拡散アルベドから素肌, 化粧, アルファマットを抽出する (3.3節).

$$\begin{aligned} \mathbf{G}_c &= \bar{\mathbf{G}} + \mathbf{B}_{id}\alpha + \mathbf{B}_{ex}\beta, \\ \mathbf{D}_c &= \bar{\mathbf{D}} + \mathbf{B}_d\gamma \odot \mathbf{C}_{gain} + \mathbf{C}_{bias}, \\ \mathbf{S}_c &= \bar{\mathbf{S}} + \mathbf{B}_s\delta, \end{aligned} \quad (1)$$

ここで, $\bar{\mathbf{G}}, \bar{\mathbf{D}}, \bar{\mathbf{S}}$ はそれぞれ形状, 拡散アルベド, 鏡面アルベドの平均を表す. \odot はアダマール積を表す. 添え字 c は粗いマテリアル情報を示す. また, $\mathbf{B}_{id}, \mathbf{B}_{ex}, \mathbf{B}_d, \mathbf{B}_s$ はそれぞれアイデンティティ, 表情, 拡散アルベド, 鏡面アルベドの PCA 基底ベクトル列である. $\alpha \in \mathbb{R}^{200}, \beta \in \mathbb{R}^{100}, \gamma \in \mathbb{R}^{100}, \delta \in \mathbb{R}^{100}$ は三次元顔の形状や反射率を制御するためのパラメータである. 最後に, \mathbf{C}_{gain} と \mathbf{C}_{bias} は拡散アルベド調整パラメータである.

復元した三次元顔モデルのテクスチャ \mathbf{R}_c を以下のように計算する.

$$\begin{aligned} \mathbf{R}_c &= \mathbf{R}_c^d + \mathbf{R}_c^s \\ &= \mathbf{D}_c \odot \mathbf{D}_c^s + \mathbf{S}_c \odot \mathbf{S}_c^s, \end{aligned} \quad (2)$$

ここで, $\mathbf{R}_c^d, \mathbf{R}_c^s, \mathbf{D}_c^s, \mathbf{S}_c^s$ はそれぞれ, 拡散復元成分, 鏡面復元成分, 拡散シェーディング, 鏡面シェーディングを表す. 本研究では, 幾何的に計算された陰影成分を「シェーディング」と呼称し, シェーディングと反射率であるアルベドとの乗算結果を「復元成分」と呼称して区別する. 拡散シェーディングは, 法線 \mathbf{N}_c と SH 照明 $\mathbf{L}_c^{sh} \in \mathbb{R}^{27}$ を用いて Lambert 反射モデル [25] に従い算出できる. ここで用いる法線 \mathbf{N}_c は, 形状 \mathbf{G}_c から算出する. 鏡面シェーディングについては, Blinn-Phong 反射モデル [26] に従い算出する. 算出において, 仮想ライトステージに用いた平行光源 20

個分の光強度を $\mathbf{L}_i \in \mathbb{R}^{20}$, 光源方向を $\mathbf{L}_d \in \mathbb{R}^{60}$ とし, 鏡面反射指数を $\rho \in \mathbb{R}^{20}$ とする. 三次元顔復元に必要なパラメータを $\chi = (\alpha, \beta, \gamma, \delta, \mathbf{C}_{gain}, \mathbf{C}_{bias}, \mathbf{r}, \mathbf{t}, \mathbf{L}_c^{sh}, \mathbf{L}_i, \mathbf{L}_d, \rho)$ とする. ここでパラメータ $\mathbf{r} \in \mathbb{R}^3, \mathbf{t} \in \mathbb{R}^3$ はそれぞれ頭部の回転および平行移動を表す. χ と復元したテクスチャ \mathbf{R}_c を用いてレンダリングした結果画像を I_{rec} とする. 損失関数は, 文献 [6] を参考に以下のように設定する.

$$\begin{aligned} \mathcal{L}_c &= \omega_{photo} \mathcal{L}_{photo} + \omega_{lan} \mathcal{L}_{lan} \\ &\quad + \omega_{skin} \mathcal{L}_{skin} + \omega_{reg} \mathcal{L}_{reg}, \end{aligned} \quad (3)$$

ここで \mathcal{L}_{photo} は入力画像 I_{in} と復元した画像 I_{rec} の L1 損失である. \mathcal{L}_{lan} は画像から検出したランドマークと 3DMM を投影して得られるランドマーク間の L2 損失である. \mathcal{L}_{skin} は, I_{in} の皮膚領域と拡散アルベド \mathbf{D}_c それぞれの平均色の間の L1 損失である. \mathcal{L}_{skin} は, FLAME と実画像との間に生じる拡散アルベドの差異を補償するために本研究で新たに導入した損失項である. また \mathcal{L}_{reg} は, 推定される 3DMM パラメータが平均顔から大きく外れることを抑制するための正則化項である. 既存手法 [6] とは異なり, 本研究では, 鏡面反射に関連する係数 $\delta, \mathbf{L}_i, \mathbf{L}_d$ の項を新たに追加する.

$$\begin{aligned} \mathcal{L}_{reg} &= \omega_\alpha \|\alpha\|_2^2 + \omega_\beta \|\beta\|_2^2 + \omega_\gamma \|\gamma\|_2^2 \\ &\quad + \omega_\delta \|\delta\|_2^2 + \omega_L \|\mathbf{L}_i\|_2^2 + \omega_L \|\mathbf{L}_d\|_2^2, \end{aligned} \quad (4)$$

ここで $\|\cdot\|_2$ は L2 ノルムを表す. それぞれの重みは $\omega_{photo} = 19.2, \omega_{lan} = 5, \omega_{skin} = 3, \omega_{reg} = 3 \times 10^{-4}, \omega_\alpha = 1.0, \omega_\beta = 0.8, \omega_\gamma = 1.7 \times 10^{-2}, \omega_\delta = 1.0, \omega_L = 1.0$

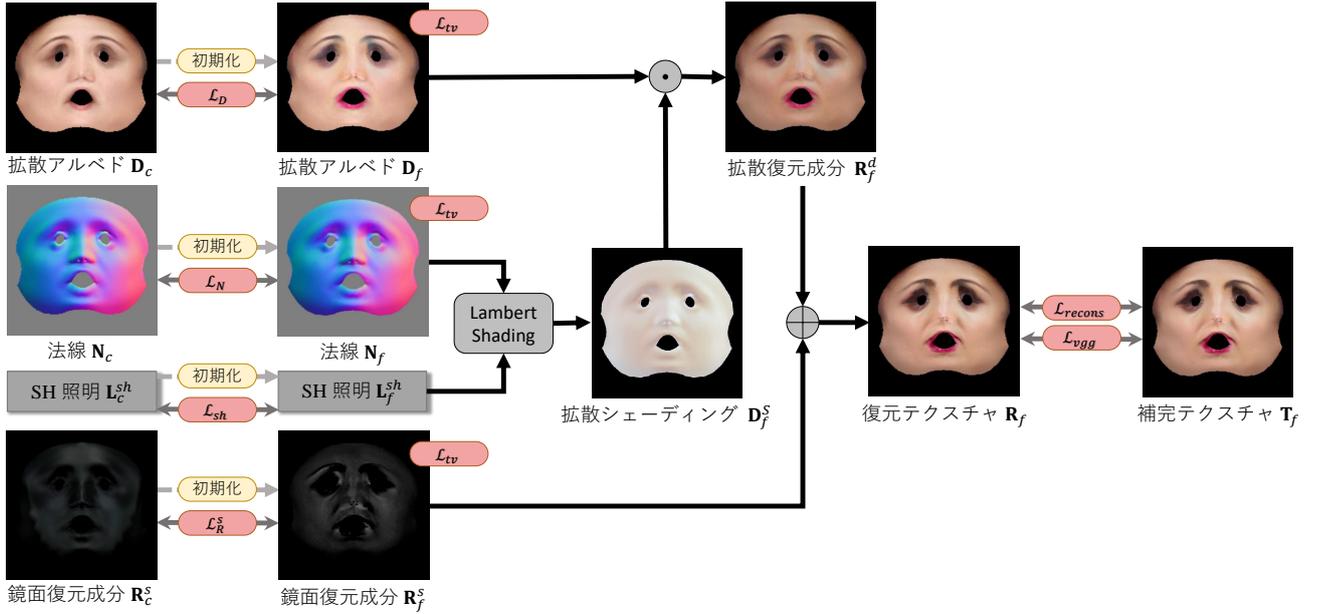


図 4 高精細なマテリアル生成のための最適化モジュール (3.2 節). 粗いマテリアル $\mathbf{D}_c, \mathbf{N}_c, \mathbf{R}_c^s$ と SH 照明 \mathbf{L}_c^{sh} を初期値として, 復元テクスチャ \mathbf{R}_f が補完テクスチャ \mathbf{T}_f に一致するように各マテリアルを最適化する. 高精細な拡散アルベド \mathbf{D}_f , 法線 \mathbf{N}_f , 鏡面復元成分 \mathbf{R}_f^s が最終出力である. \oplus と \odot はそれぞれピクセル毎の加算と乗算を表す.

に設定した.

3.2 欠損補完と最適化による高精細なマテリアル生成

図 3 に示すように, まず, 入力画像 I_{in} を UV テクスチャ \mathbf{T}_u に展開する. \mathbf{T}_u には遮蔽による欠損が含まれており, そのままでは後段の処理で不具合が生じる. そこで顔テクスチャの補完手法である DSD-GAN [11] を適用することで, 欠損のない UV テクスチャ \mathbf{T}_f を生成する. なお, 添え字 f は詳細なテクスチャ情報を含む UV テクスチャであることを示す.

図 4 に示すように, 復元されたテクスチャ \mathbf{R}_f が UV テクスチャ \mathbf{T}_f と一致するように, 各種マテリアルの最適化を行う. 最適化処理では, 3DMM の当てはめによって得られた $\mathbf{D}_c, \mathbf{N}_c, \mathbf{L}_c^{sh}, \mathbf{R}_c^s$ を初期値として用い, 高精細な情報 $\mathbf{D}_f, \mathbf{N}_f, \mathbf{L}_f^{sh}$ を復元する. 鏡面復元成分 \mathbf{R}_f^s については, 最適化の自由度を高めるため, 仮想ライトステージの光源に関する変数は用いず, \mathbf{R}_f^s を直接最適化する. 最適化に用いる損失関数を以下に示す.

$$\mathcal{L}_f = \omega_{recons} \mathcal{L}_{recons} + \omega_{vgg} \mathcal{L}_{vgg} + \omega_{tv} \mathcal{L}_{tv} + \omega_{prior} \mathcal{L}_{prior}, \quad (5)$$

ここで \mathcal{L}_{recons} は \mathbf{R}_f と \mathbf{T}_f の間の L1 損失である. \mathcal{L}_{vgg} は詳細情報を保持するための知覚的損失 [27] である. また, \mathcal{L}_{tv} は最適化されたテクスチャ $\mathbf{D}_f, \mathbf{N}_f, \mathbf{R}_f^s$ 中のノイズを抑制するための Total Variation (TV) 損失 [28] である. \mathcal{L}_{prior} は, 最適化されたマテリアルと粗いマテリアルとの類似度を測る L1 損失である.

$$\mathcal{L}_{prior} = \omega_D \mathcal{L}_D + \omega_N \mathcal{L}_N + \omega_R^s \mathcal{L}_R^s + \omega_{sh} \mathcal{L}_{sh}, \quad (6)$$

ここで $\mathcal{L}_D, \mathcal{L}_N, \mathcal{L}_R^s$ はそれぞれ拡散アルベド, 法線, 鏡面復元成分に対する L1 損失を表し, \mathcal{L}_{sh} は SH 照明に対する L2 損失を表している. 各パラメータの重みは $\omega_{recons} = 40$, $\omega_{vgg} = 5$, $\omega_{tv} = 10$, $\omega_{prior} = 1.0$, $\omega_D = 4$, $\omega_N = 1.0$, $\omega_R^s = 1.0$, $\omega_{sh} = 1.0$ とした.

3.3 化粧パターンの抽出

最適化によって得られた高精細な拡散アルベド \mathbf{D}_f を, 素肌, 化粧パターン, アルファマップに分解する. 図 5 に示すように, 化粧パターンの抽出ネットワークは, 化粧パターン抽出器 E , 化粧画像識別器 M , 素肌画像識別器 B から構成される. ネットワークの訓練のために, 顔画像を化粧ありアルベド \mathbf{D}_m と化粧なしアルベド \mathbf{D}_n の 2 種類に分類した拡散アルベドデータセットを作成して用いる. 素肌 \mathbf{D}_m^b , 化粧パターン \mathbf{D}_m^m , アルファマップ \mathbf{A} は E によって抽出される. \mathbf{A} を用いて, 抽出された化粧パターン \mathbf{D}_m^m と化粧なしアルベド \mathbf{D}_n をブレンドすることによって新しい化粧ありアルベド \mathbf{D}_n^m を生成する. 復元される化粧ありアルベド \mathbf{D}_m^r と生成される化粧ありアルベド \mathbf{D}_n^m は, 以下の式により定式化する.

$$\begin{aligned} \mathbf{D}_m^r &= \mathbf{A} \odot \mathbf{D}_m^b + (1 - \mathbf{A}) \odot \mathbf{D}_m^m, \\ \mathbf{D}_n^m &= \mathbf{A} \odot \mathbf{D}_n + (1 - \mathbf{A}) \odot \mathbf{D}_m^m, \end{aligned} \quad (7)$$

ここで, $(1 - \mathbf{A})$ は \mathbf{A} の顔領域内での反転を表し, 各画素は $[0, 1]$ の値をとる. 識別器は生成された素肌 \mathbf{D}_m^b と化粧パターン \mathbf{D}_n^m を判別する. 損失関数は次式で定義する.

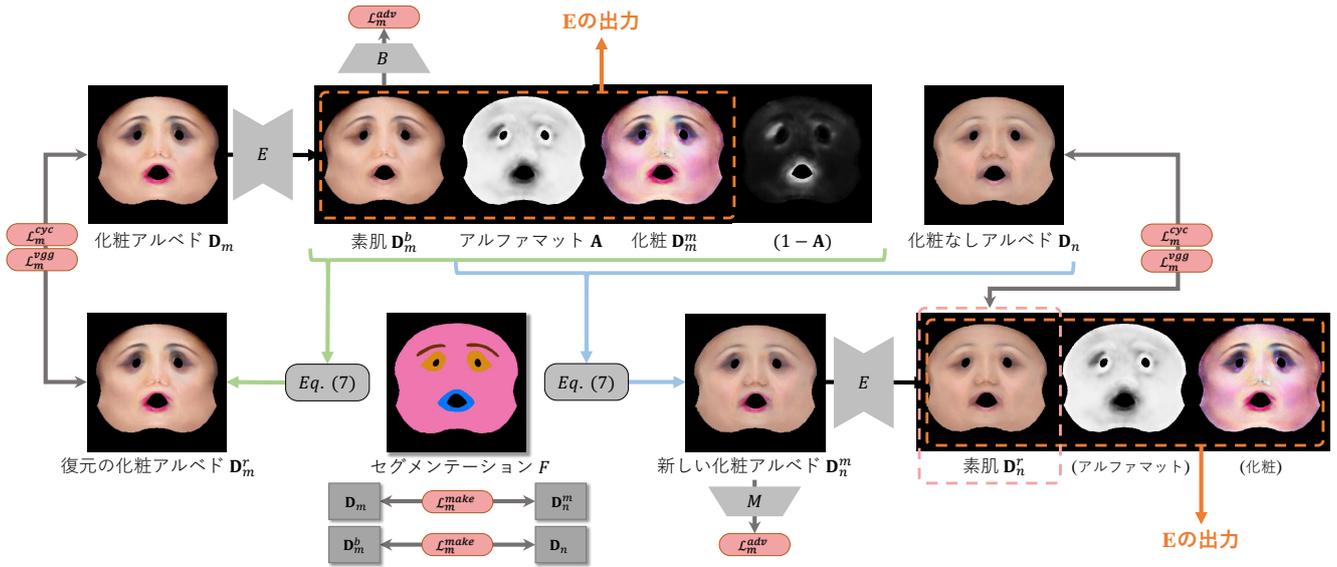


図5 化粧の抽出ネットワーク (3.3節). ネットワーク E はアルファブレンドに基づき, 高精細な拡散アルベドを素肌, 化粧, アルファマットに分解する.

$$\mathcal{L}_m = \omega_m^{cyc} \mathcal{L}_m^{cyc} + \omega_m^{vgg} \mathcal{L}_m^{vgg} + \omega_m^{adv} \mathcal{L}_m^{adv} + \omega_m^{tv} \mathcal{L}_m^{tv} + \omega_m^{make} \mathcal{L}_m^{make}, \quad (8)$$

ここで \mathcal{L}_m^{cyc} , \mathcal{L}_m^{vgg} はそれぞれ L1 再構成損失, 知覚的損失を表す. また, \mathcal{L}_m^{adv} は識別器と生成器に対する敵対的損失を表し, \mathcal{L}_m^{tv} は滑らかなテクスチャ D_m^b と D_m^m を生成するための TV 損失を表す. さらに, \mathcal{L}_m^{make} は BeautyGAN [17] で用いられている化粧損失を表す. 化粧損失は, UV テクスチャ上に定義された眉, 目, 唇の領域 F を用い, D_m と D_m^r , D_n と D_n^m の間で算出する. 各項目の重みは, それぞれ $\omega_m^{cyc} = 20$, $\omega_m^{vgg} = 2$, $\omega_m^{adv} = 5$, $\omega_m^{tv} = 8$, $\omega_m^{make} = 1$ とした.

4. 実装の詳細

4.1 粗いマテリアルの復元

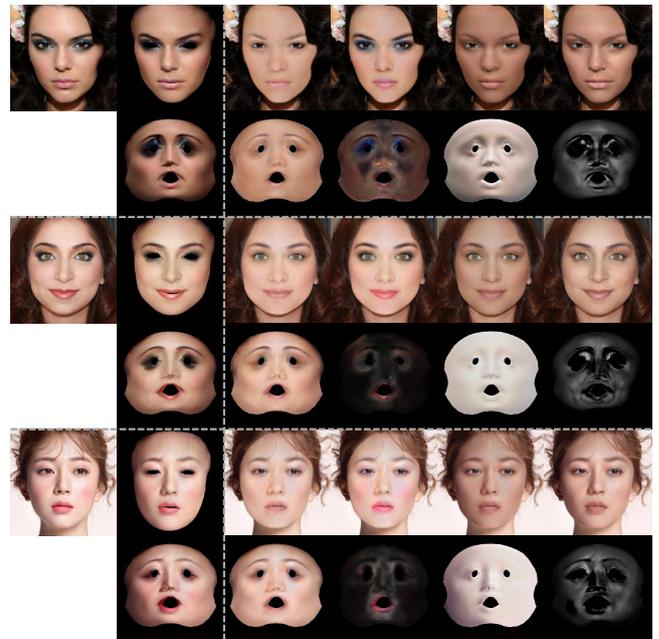
提案手法の 3DMM パラメータを推定するために, 文献 [6] のネットワークの最終層を修正した. 学習には, 文献 [6] と同様に 260,000 枚の画像を用いた. バッチサイズは 8 とした. 最適化には Adam を用い, 学習レートは 1×10^{-4} に設定して, 20 エポックの学習を行った.

4.2 高精細なマテリアルの最適化

高精細なマテリアルを生成する際の最適化処理では, 各テクスチャに対して 500 回の繰り返し処理を行った. Adam を用いた最適化での学習率は 1×10^{-2} に設定し, 学習率の減衰は 0.1 とした.

4.3 化粧パターンの抽出

学習用データとして, MT データセット [17] と LADN データセット [18] の 2 種類の化粧データセットより, 3,070



(a) 入力 (b) レンダリング (c) 素肌 (d) 化粧パターン (e) 拡散 (f) 鏡面

図6 提案手法の出力結果. (a) 入力の化粧写真, (b) レンダリング画像の重畳表示とそれに対応するテクスチャ. (c) から (f) までは図1と同様.

枚の化粧あり画像と 1,449 枚の化粧なし画像を抽出しデータセットを構築した. ネットワーク訓練時のバッチサイズは 1 とし, 40 エポックの学習を行った. 最適化には Adam を用い, 学習率は 1×10^{-4} とした. 化粧抽出器 E は DSD-GAN の生成器と同じ構造であり, 識別器 M と B には PatchGAN [29] を使用した.

学習には NVIDIA RTX 2080 Ti を用いた. 学習時間は, 粗いマテリアルを復元するネットワークに約 3 日, 化粧パターン抽出ネットワークに約 1 日であった. また, 高精細

表 1 化粧転写手法の機能比較. 各項目の意味は次の通り. “顔姿勢”: ポーズの異なる顔を扱える, “補間”: 化粧の濃淡を補間可能, “部分転写”: 転写する顔領域を選択可能, “編集”: 手で化粧を編集可能, “遮蔽”: 遮蔽に頑健, “照明効果”: 照明効果を制御可能.

Method	顔姿勢	補間	部分転写	編集	遮蔽	照明効果
BeautyGAN [17]						
LADN [18]		✓				
PSGAN [19]	✓	✓	✓			
SCGAN [21]	✓	✓	✓			
CPM [20]	✓	✓	✓	✓		
SOGAN [22]	✓	✓	✓	✓	✓	
EleGANt [23]	✓	✓	✓	✓	✓	
提案手法	✓	✓	✓	✓	✓	✓



図 7 化粧転写を複雑化させる要素を含む化粧転写の結果. 提案手法は, 遮蔽, 照明条件および顔姿勢の違いに対応できる.

のマテリアル最適化処理では, 1 枚のテクスチャの処理時間は約 1 分であった.

5. 実験結果と応用例

5.1 化粧パターンの抽出結果

化粧画像の分解結果を図 6 に示す. 3 名分の各結果画像の上段の (c)~(f) は, 入力画像にレンダリング結果画像を重畳表示したものである. レンダリングしたのはそれぞれ, 素肌, 化粧+素肌, 素肌と拡散シェーディングの乗算, さらにそれに鏡面還元成分を加算した結果である. また (b) は, (c)~(f) の各マテリアルを用いて入力写真と同等なテクスチャを還元した結果である. これらの結果より, 入力画像が素肌, 化粧パターン, 拡散, 鏡面のテクスチャに分解できていることが分かる.

5.2 照明効果を考慮した化粧転写

提案手法は抽出された化粧パターンを用いて, 式 (7) に基づき新しい化粧のテクスチャを合成する. この合成テクスチャをレンダリングし, 元画像に重畳表示することによって, 化粧転写を実現できる.

提案手法と従来の化粧転写手法の機能に関する比較を表 1 に示す. 提案手法では, UV テクスチャの化粧を抽出することによって, 顔姿勢の違いを解消できる. また, 化粧を転写する領域の指定・編集が可能となり, 遮蔽にも頑健である. さらに既存手法と異なり, 提案手法では照明効果にも対応できる.

化粧転写の結果を図 7 に示す. 参照写真には, 遮蔽, 顔姿勢の違い, 複雑な照明効果など, 化粧転写を複雑化させる要素が含まれている. 提案手法では, 参照写真から化粧を抽出し, 元写真の照明効果を維持しながら化粧転写を実現できている.

図 8 に, 元写真と参照写真とで照明条件が異なる場合の化粧転写の結果を示す. 従来手法と比べて, 提案手法では照明効果の影響を受けずに, 化粧の転写を実現できている.

5.3 照明効果を考慮した化粧の補間と除去

既存の化粧転写手法では, 化粧補間を行う際に参考写真が必要となる. 図 9 に示すように, 提案手法は参照写真なしで化粧の補間と除去を実現できる. また, 提案手法により, 照明効果を維持しながら化粧を自然に補間できる. 一番上の行は, 照明効果を含む顔写真での化粧の補間・除去結果を示している. 照明効果を変化させずに化粧の濃淡を調整できている. 二行目は, 化粧の補間時のアルファマットの変化の様子を示している. アルファマット \mathbf{A} は次式によって \mathbf{A}_σ に調整される.

$$\mathbf{A}_\sigma = \text{clamp}(\mathbf{A} + \sigma, 0, 1), \quad (9)$$

ここで $\sigma \in [0, 1]$ であり, \mathbf{A}_σ の値は $[0, 1]$ の範囲に切り詰められる. \mathbf{A}_σ を変化させることによって, 化粧パターンの濃淡が変化し, 化粧を補間できる. また, 照明効果と化粧を分離したことにより, 三行目に示している通り, 化粧を調整しながら再照明が可能となっている.

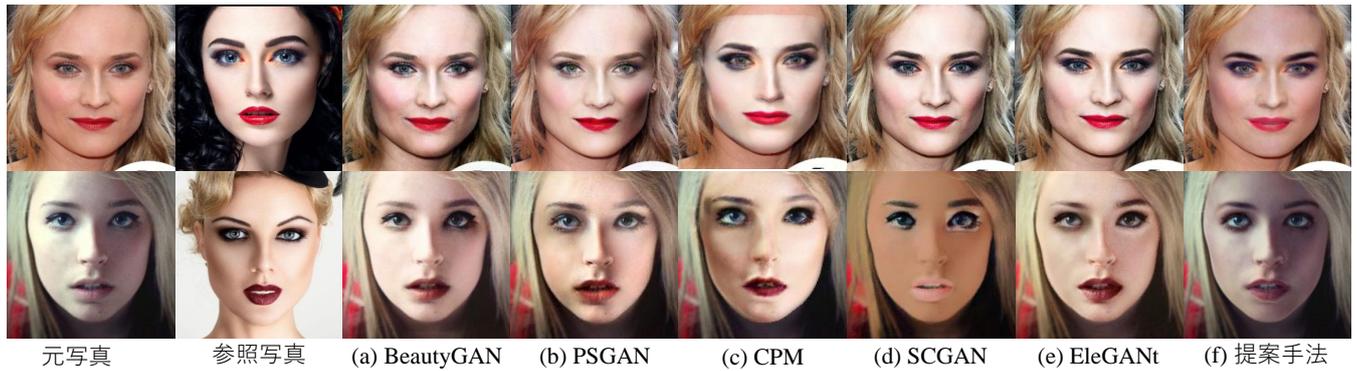


図 8 照明条件が異なる場合での化粧転写の定性評価. 元写真の照明効果を維持しながら, 参照写真の化粧を転写している. 左から順に, (a) BeautyGAN [17], (b) PSGAN [19], (c) CPM [20], (d) SCGAN [21], (e) EleGANt [23], (f) 提案手法の結果.

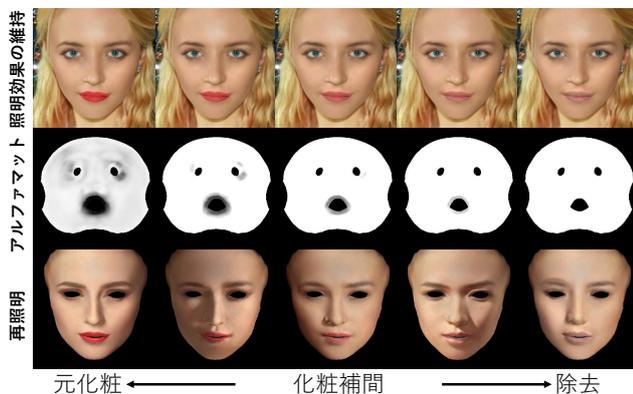


図 9 照明を考慮した化粧の補間と除去. 一番左から徐々に濃い化粧が除去される様子. 一行目は照明効果を保持した化粧の補間結果, 二行目は素肌と化粧パターンの混合比率を表すアルファマット A の変化の様子, 三行目は復元した三次元顔の化粧を補間しながら, 再照明を行った結果.

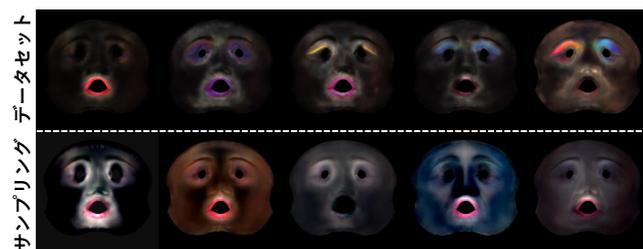


図 10 本研究で構築した化粧データセットと PCA に基づく化粧モデル. 上段は顔写真から抽出した化粧パターン, 下段は化粧モデルから主成分に沿ってランダムにサンプリングした化粧パターン.

5.4 化粧写真の三次元顔復元

抽出した化粧パターンを収集してデータセットを構築し, 化粧付き顔写真の三次元顔復元に応用する. 本研究では FLAME [3] の拡散アルベドテクスチャモデルの構築手法と同様な手順を適用した. まず, 収集した化粧テクスチャ $(1 - A) \odot D_c^m$ を用いて, PCA に基づく化粧パターンの統計モデルを構築した. 化粧パターンの PCA モデルから, 主成分の軸に沿ってランダムにサンプリングされたテクス



図 11 化粧の PCA モデルを用いて 3DMM の拡散アルベドモデルを拡張し, 三次元復元に適用した結果. 復元した三次元顔モデルの忠実度が大幅に向上した.

チャを図 10 に示す. さらに, 化粧の PCA モデル D_c^m を用いて FLAME の拡散アルベドモデル D_c を拡張した, 新しい化粧付き拡散アルベドモデル D_c' を次式で定義した.

$$D_c' = D_c + D_c^m, \quad (10)$$

2つのモデル D_c, D_c' を用いて化粧付き顔写真の三次元復元を行った結果の比較を図 11 に示す. D_c' を利用することによって, 口紅とアイシャドウを復元でき, また, 唇や目の形状をより正確にマッチングできている.

6. まとめと今後の課題

本研究では, 一枚の化粧写真から三次元表現に基づく化粧を抽出する初めての手法を提案した. 併せて, 照明効果を考慮した化粧の転写, 補間, 除去といった新しい応用を実現した. また, 抽出した化粧パターンが UV テクスチャとしてアラインメントされていることを利用し, 大規模な化粧テクスチャデータセットと PCA に基づく化粧モデルを構築した. この化粧モデルを活用することによって, 化粧付き顔写真のより正確な三次元復元を実現できた.

提案手法では拡散アルベドから化粧パターンを抽出できたが, 現実世界では化粧パターンに光沢が含まれることが考慮されていない. 今後は, BRDF を考慮することで, より物理的な化粧の抽出を目指す.

参考文献

- [1] Huang, C.-G., Lin, W.-C., Huang, T.-S. and Chuang, J.-H.: Physically-Based Cosmetic Rendering, *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, p. 190 (2013).
- [2] Ma, X., Zhang, F., Wei, H. and Xu, L.: Deep learning method for makeup style transfer: A survey, *Cognitive Robotics*, Vol. 1, pp. 182–187 (2021).
- [3] Li, T., Bolkart, T., Black, M. J., Li, H. and Romero, J.: Learning a model of facial shape and expression from 4D scans, *Transactions on Graphics*, Vol. 36, No. 6, pp. 194:1–194:17 (2017).
- [4] Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Lüthi, M., Schönborn, S. and Vetter, T.: Morphable Face Models - An Open Framework, *Proceedings of International Conference on Automatic Face & Gesture Recognition*, pp. 75–82 (2018).
- [5] Blanz, V. and Vetter, T.: Face Recognition Based on Fitting a 3D Morphable Model, *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, pp. 1063–1074 (2003).
- [6] Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y. and Tong, X.: Accurate 3D Face Reconstruction With Weakly-Supervised Learning: From Single Image to Image Set, *Proceedings of Conference on Computer Vision and Pattern Recognition Workshops*, pp. 285–295 (2019).
- [7] Dib, A., Bharaj, G., Ahn, J., Thébaud, C., Gosselin, P. H., Romeo, M. and Chevallier, L.: Practical Face Reconstruction via Differentiable Ray Tracing, *Computer Graphics Forum*, Vol. 40, No. 2, pp. 153–164 (2021).
- [8] Danecek, R., Black, M. J. and Bolkart, T.: EMOCA: Emotion Driven Monocular Face Capture and Animation, *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 20311–20322 (2022).
- [9] Egger, B., Smith, W. A. P., Tewari, A., Wuhler, S., Zollhöfer, M., Beeler, T., Bernard, F., Bolkart, T., Kortylewski, A., Romdhani, S., Theobalt, C., Blanz, V. and Vetter, T.: 3D Morphable Face Models - Past, Present, and Future, *Transactions on Graphics*, Vol. 39, No. 5, pp. 157:1–157:38 (2020).
- [10] Stylianou, G. and Lanitis, A.: Image Based 3D Face Reconstruction: a Survey, *International Journal of Image and Graphics*, Vol. 9, No. 2, pp. 217–250 (2009).
- [11] Kim, J., Yang, J. and Tong, X.: Learning High-Fidelity Face Texture Completion without Complete Face Texture, *Proceedings of International Conference on Computer Vision*, pp. 13970–13979 (2021).
- [12] Sengupta, S., Kanazawa, A., Castillo, C. D. and Jacobs, D. W.: SfSNet: Learning Shape, Reflectance and Illuminance of Faces 'in the Wild', *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 6296–6305 (2018).
- [13] Kanamori, Y. and Endo, Y.: Relighting humans: occlusion-aware inverse rendering for full-body human images, *Transactions on Graphics*, Vol. 37, No. 6, p. 270 (2018).
- [14] Ramamoorthi, R. and Hanrahan, P.: An efficient representation for irradiance environment maps, *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 497–500 (2001).
- [15] Pandey, R., Orts-Escolano, S., LeGendre, C., Häne, C., Bouaziz, S., Rhemann, C., Debevec, P. E. and Fanello, S. R.: Total relighting: learning to relight portraits for background replacement, *Transactions on Graphics*, Vol. 40, No. 4, pp. 43:1–43:21 (2021).
- [16] Tajima, D., Kanamori, Y. and Endo, Y.: Relighting Humans in the Wild: Monocular Full-Body Human Relighting with Domain Adaptation, *Computer Graphics Forum*, Vol. 40, No. 7, pp. 205–216 (2021).
- [17] Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W. and Lin, L.: BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network, *Proceedings of International Conference on Multimedia*, pp. 645–653 (2018).
- [18] Gu, Q., Wang, G., Chiu, M. T., Tai, Y. and Tang, C.: LADN: Local Adversarial Disentangling Network for Facial Makeup and De-Makeup, *Proceedings of International Conference on Computer Vision*, pp. 10480–10489 (2019).
- [19] Jiang, W., Liu, S., Gao, C., Cao, J., He, R., Feng, J. and Yan, S.: PSGAN: Pose and Expression Robust Spatial-Aware GAN for Customizable Makeup Transfer, *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 5193–5201 (2020).
- [20] Nguyen, T., Tran, A. T. and Hoai, M.: Lipstick Ain't Enough: Beyond Color Matching for In-the-Wild Makeup Transfer, *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 13305–13314 (2021).
- [21] Deng, H., Han, C., Cai, H., Han, G. and He, S.: Spatially-Invariant Style-Codes Controlled Makeup Transfer, *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 6549–6557 (2021).
- [22] Lyu, Y., Dong, J., Peng, B., Wang, W. and Tan, T.: SO-GAN: 3D-Aware Shadow and Occlusion Robust GAN for Makeup Transfer, *Proceedings of International Conference on Multimedia*, pp. 3601–3609 (2021).
- [23] Yang, C., He, W., Xu, Y. and Gao, Y.: EleGANt: Exquisite and Locally Editable GAN for Makeup Transfer, *Proceedings of European Conference on Computer Vision* (2022).
- [24] Smith, W. A. P., Seck, A., Dee, H., Tiddeman, B., Tenenbaum, J. B. and Egger, B.: A Morphable Face Albedo Model, *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 5010–5019 (2020).
- [25] Basri, R. and Jacobs, D. W.: Lambertian Reflectance and Linear Subspaces, *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 2, pp. 218–233 (2003).
- [26] Blinn, J. F.: Models of light reflection for computer synthesized pictures, *Proceedings of the 4th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 192–198 (1977).
- [27] Johnson, J., Alahi, A. and Fei-Fei, L.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution, *Proceedings of European Conference on Computer Vision*, Vol. 9906, pp. 694–711 (2016).
- [28] Gatys, L. A., Ecker, A. S. and Bethge, M.: Image Style Transfer Using Convolutional Neural Networks, *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423 (2016).
- [29] Isola, P., Zhu, J., Zhou, T. and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 5967–5976 (2017).