論文 / 著書情報
Article / Book Information

論題	部位の領域分割画像を入力とした微分可能レンダラによる人体の三次元再構成
Title	3D human pose and shape reconstruction with differentiable renderer from body part segmentation
著者	櫻井凜太郎, 宇都有昭, 篠田浩一
Authors	Rintaro SAKURAI, Kuniaki UTO, Koichi SHINODA
出典	電子情報通信学会 技術研究報告, vol. 121, no. 23, pp. 31-36
Citation	IEICE technical report, vol. 121, no. 23, pp. 31-36
発行日 / Pub. date	2021, 5
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright(c) 2021 IEICE

一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
PRMU2021-6(2021-05)

部位の領域分割画像を入力とした微分可能レンダラによる人体の三次元再構成

櫻井凜太郎† 　宇都　有昭† 　篠田　浩一†

† 東京工業大学 情報理工学院
〒152-8552 東京都目黒区大岡山 2-12-1
E-mail: †{sakurai,uto}@ks.c.titech.ac.jp, ††shinoda@c.titech.ac.jp

あらまし　三次元人体再構成のための深層学習の訓練データとして，二次元信号から復元された三次元人体形状がしばしば用いられる．本研究では部位の領域分割画像が与えられる場合に，部位ごとの相互隠蔽を考慮したシルエットを微分可能レンダラによって作成し，勾配法によって人体形状を精確に推定する手法を提案する．部位の領域分割画像には，全身のシルエットが持たない姿勢の情報と，二次元関節座標がもたない体型の情報が含まれる．3D Poses in the Wild dataset (3DPW) を用いた実験では部位の領域分割画像と二次元関節座標を併用した場合に，三次元関節座標誤差と三次元頂点座標誤差について，従来手法 SMPLify による最適解の誤差が 0.224 と 0.191 であったのに対して，提案手法では 0.210 と 0.184 であった．

キーワード　三次元人体再構成，微分可能レンダラ，部位の領域分割

3D human pose and shape reconstruction with differentiable renderer from body part segmentation

Rintaro SAKURAI†, Kuniaki UTO†, and Koichi SHINODA†

† School of Computing, Tokyo Institute of Technology
2-12-1 , Ookayama, Meguro-ku, Tokyo 152-8552 Japan
E-mail: †{sakurai,uto}@ks.c.titech.ac.jp, ††shinoda@c.titech.ac.jp

1. はじめに

コンピュータビジョンにおける重要なタスクの一つに人体の三次元再構成があり，ヴァーチャルリアリティや自動運転などの多くの領域で応用が期待される．本稿がいう人体の三次元姿勢推定とは関節の三次元座標の推定であり，三次元再構成とは姿勢と体型を含む人体の完全な形状を推定することを指す．また入力として単 RGB 画像が与えられるとする．

人体形状の表現としてはパラメトリックなモデルがよく使われる．Loper *et al.* [1] は人体の姿勢，体型を表現するパラメトリックなモデル，Skinned Multi-Person Linear model (SMPL) を提案した．SMPL は与えられたパラメータから関節を持つ骨格とそれを覆う皮膚をスキニングして，人体を表すメッシュを出力する．

再構成の手法には，パラメトリックな人体形状モデルを用いて RGB 画像から得られる信号に人体形状の投影信号をフィッティングさせるものがある．例えば，Bogo *et al.* [2] が提案した SMPLify は，与えられた二次元関節座標とモデルの関節座標が画像空間上で一致するように，勾配法により SMPL パラメータをフィッティングする．最適化の損失関数に実データで訓練した混合ガウス分布，相互浸透誤差項を加えることにより，姿勢だけでなく体型についても考慮に入れた推定ができることを示した．

近年，他のコンピュータビジョンの分野と同様に，人体の三次元再構成に深層ニューラルネットワークを用いる研究が高いパフォーマンスを出している．深層学習を適用するにあたっては教師データとして入力と出力の組が大量に必要となるが，実世界の人体の形状を取得するコストは大きく，さらに撮影環境も限定される．例えば，Ionescu *et al.* [3] はマーカによるモーションキャプチャシステムを用いて人体の三次元姿勢を計測し，Human3.6M データセットを構築したが，そこに含まれる画像は室内で撮影されたものに限定されている．

この問題を解決するために三次元のアノテーションを必要としない手法が提案されてきた．三次元情報の代替信号とし

ては，画像上の関節座標，全身のシルエット，腕や胴体などの部位ごとの領域分割画像などがあり，そうした二次元信号を含むデータセットも提供されている．例えば，Johonson *et al.* [4] は単画像から二次元関節座標を推定する方法を提案し，LSP データセットを構築した．Lassner *et al.* [5] は LSP で提供される二次元関節座標に SMPLify と微分可能レンダラを用いた最適化を適用して，三次元形状を生成することで部位の領域分割画像を得ている．Tsung-Yi *et al.* [6] では二次元関節座標や部位の領域分割画像を人手で作成している．

こうした二次元の代替信号を深層ニューラルネットワークの訓練に利用する研究のアプローチとしては，代替信号を中間表現や再投影を適用した教師信号として用いる方法がある．その種の手法は，コンピュータグラフィックスによって大量に生成された三次元形状と代替信号の組を用いてニューラルネットワークの訓練を効果的に実行できることに依拠している [7]〜[10]．これらは，RGB 画像から二次元信号の推論と，二次元信号から三次元形状の推論を行う多段階のネットワークを組み，さらに三次元形状を二次元空間に投影することで，教師信号として扱う手法である．

一方で Kolotouros *et al.* [11] は，三次元形状やモデルパラメータなど直接的な教師信号をもって訓練を行うべきと主張する．彼らはニューラルネットワークにより RGB 画像より推定された三次元形状を初期値として SMPLify による最適化を行い，その最適解を当該画像と対となる教師信号としてネットワークを訓練をする手法を提案し，そのネットワークが多段階の推定を行う手法よりも高い性能を持つことを示した．代替信号に対する最適化を用いて教師データを強化するという点では Lassner *et al.* [5] も共通性を持つと言えるであろう．

最適化手法の課題として，SMPLify における再構成は二次元関節座標にのみ依拠しているために，精確に体型を推定することはできない．一方で Lassner *et al.* [5] は微分可能レンダラとシルエットによる最適化を用いたが，全身のシルエットに依拠するものであったため，局所最適解に陥りやすくなる．

本研究では与えられた部位の領域分割画像に一致する三次元形状を反復法により計算する手法を提案する．全身のシルエットより詳細な姿勢の情報を持ち，関節座標よりも体型の情報を強く持つ部位の領域分割画像を用いることにより，より精確な推定が可能となった．

2. 提 案 手 法

2.1 動　　機

部位の領域分割画像には，各部位の位置，人体の体型，相互隠蔽による奥行き方向の情報が強く反映されており，これを利用すれば二次元関節座標や全身のシルエットを用いた場合よりも精確な再構成が実現できると予想される．予測された人体の形状から領域分割画像を生成するにあたっての問題は，部位同士による相互隠蔽の再現である．本研究では，部位同士の相互隠蔽を考慮した各部位の領域分割画像の生成法と微分可能レンダラを用いた勾配法による SMPL パラメータの最適化手法を提案する．

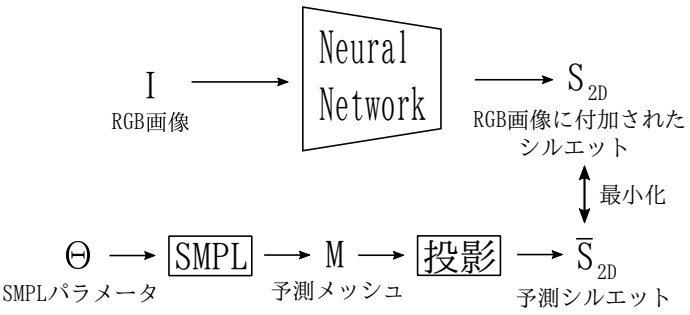

図 1　単 RGB 画像から生成されたシルエットへの最適化

図 1 に示すように，単 RGB 画像からニューラルネットワークを用いてシルエットを生成するなど，単 RGB 画像とシルエットの組が与えられる場合にも本手法を応用できる．前例としては，Lassner *et al.* [5] のようにシルエットが含まれるデータセットに三次元形状を付加する，あるいは Kolotouros *et al.* [11] のようにニューラルネットワークの訓練に組み込むなどが挙げられる．RGB 画像はシルエットに含まれない情報を持ち，例えば，服や皮膚に現れる陰影などは人体形状と影響を及ぼし合う．しかし，そういった光学現象を反映してレンダリングをする際には光源や人体モデルの光学的特性などを推定する必要があり，これ自体が難問である．したがって，本研究では SMPL によって生成される人体モデルの形状を推定の範囲とし，光学現象やより詳細な形状変形は考慮しない．

2.2 パラメトリックな人体形状モデル

本手法では人体形状の表現として SMPL を用いる．23 個の関節の回転を表す回転ベクトル $\theta \in \mathbb{R}^{23\times3}$，体型を表すパラメータ $\beta \in \mathbb{R}^{10}$，大域的な回転ベクトル $\theta_g \in \mathbb{R}^3$ と並進移動ベクトル $\boldsymbol{t} \in \mathbb{R}^3$ (以下，全パラメータを総称して $\Theta = (\theta, \beta, \theta_g, \boldsymbol{t})$ と書く)，出力される固定長の頂点ベクトルの集合 $\boldsymbol{V} \in \mathbb{R}^{6890\times3}$ とすると，$\boldsymbol{V}(\Theta)$ と表される．またスキニングを適用する骨格の表現として，骨格を構成する各頂点の三次元座標 $\boldsymbol{J}_{3D} \in \mathbb{R}^{45\times3}$ を用いると，$\boldsymbol{J}_{3D}(\Theta)$ と表される．後の議論のために $\boldsymbol{V}$ と $\boldsymbol{J}_{3D}$ はすべての入力で微分可能であることに注意が必要である．

2.3 レンダリング

勾配法の損失関数として予測された SMPL モデルのレンダリング結果と Ground Truth の残差を用いるが，これが SMPL パラメータで微分可能である必要がある．SMPL パラメータからメッシュへの変換は微分可能である一方で，これをシルエットに変換する際に通常のレンダリングでは微分不可能な処理が含まれる．そこで，Lie *et al.* [12] が提案した微分可能レンダラ Soft Rasterizer(SoftRas) を用いる．

画像に付加される Ground Truth のシルエットが，Lassner *et al.* [5] で扱われた条件と同様に全身のシルエットである場合，さらに詳細な解析結果として人体を N_p 個の部位に分割した領域が与えられる場合を考える．前者は従来手法と同様に，各画素にアルファ値が格納され，これに対応する予測シルエットは全身のメッシュをレンダリングして得られるシルエットである．後者の場合には各画素に N_p 次元ベクトルが格納される．N_p 次元ベクトルの各要素は，それぞれ各部位に相当す

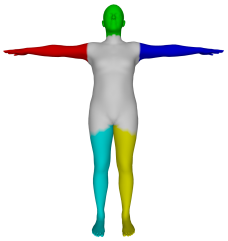

図 2　分割された人体

図 3　Ground Truth のシルエット

図 4　Ground Truth の各部位のシルエット

図 5　Ground Truth の各部位のマスク画像

図 6　再構成されたメッシュ

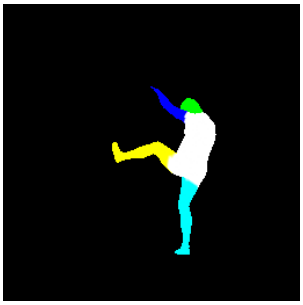

図 7　再構成されたメッシュの投影

図 8　再構成された人体の各部位のシルエット

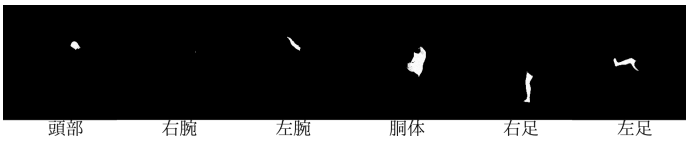

図 9　マスク演算を適用した再構成された人体の各部位のシルエット

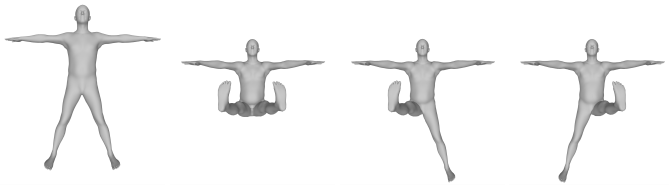

図 10　初 期 姿 勢

る部分が投影される確率に対応する．この画像はニューラルネットワークや人の手によって与えられ，人体がどのように分割されるかについては既知とする．これに対応する予測シルエットを生成するとして，N_p 個の部位ごとに人体のメッシュを分割し，各部位ごとにシルエットをレンダリングする．そうしたシルエットは部位同士の隠蔽が考慮されていないため，各部位に対して Ground Truth の画像から当該部位以外のシルエットを抜き出して，これを用いてマスク演算をする．

レンダリングはメッシュの座標だけでなくカメラのパラメータ (カメラの座標，向き，画角) も引数としてとるが，各 RGB 画像について撮影されたカメラの情報が付与されているとは限らない．さらにメッシュの幾何的な情報とカメラの幾何的な情報が変動することは冗長である．そこで本手法では SMPLify [2] と同様に，一律にカメラのパラメータを固定し最適化の対象としない．したがって最適化の結果は固定されたカメラパラメータの下での結果である．

部位の領域分割画像が与えられる場合の例として，図 2 に示すように頭部・胴体・右手・左手・右足・左足の $N_p = 6$ で分割する場合をとりあげ，具体的な手続きを説明する．Ground Truth のシルエットは，図 3 に示すように各画素についてどの部位が投影されているかを可視化したものである．これを基に図 4 に示す各部位のシルエットと図 5 に示す各部位のマスク画像を作成する．次に再構成された人体の形状について，Ground Truth と比較可能なシルエットを生成する．最適化の解として図 6 に示すような結果が得られたとしよう．このメッシュを投影した結果は図 7 である．まず人体のメッシュを部位ごとに分割し各部位ごとに SoftRas を用いて図 8 に示すようなシルエットを生成する．この各部位のメッシュのシルエットは部位同士の隠蔽が考慮されていない．そこで図 5 に示したマスクを適用した結果を図 9 に示すが，これを予測シルエットとする．

2.4　初　期　値

反復法による最適化においては，初期値の設定が重要となる．SMPLify [2] では胴体を構成する二次元関節座標の間の距離で初期位置が決定された．本手法ではシルエットのみが与えられた場合も対象とするため，シルエットを用いて初期値を決定する．

勾配法の初期値姿勢のうち大域的な並進運動 $\boldsymbol{t}$ は全身のシルエットから決定される．シルエットのアルファ値の総和をシルエットの面積とみなす．これは人体と投影面の距離の二乗に反比例すると仮定して，並進運動の z 成分を決定する．並進運動の xy 成分については，腰の関節がシルエットの重心に投影されるように決定する．

初期姿勢 θ としては 4 つの姿勢を用意する．図 10 はそれぞれの初期姿勢を示す．それぞれの姿勢では各脚を挙上させているが，脚については投影面積が大きく腕の運動に比べてフィッティングが難しいため，複数の初期値を用いて最適化をするのが望ましいためである．各姿勢のパラメータについて，大腿の関節について左右ともに 0.3 π 開かせる．脚の挙上については 0.45 π 持ち上げるように設定する．

大域的な回転運動 θ_g については各初期姿勢について y 軸まわりの $0\pi, 0.5\pi, 1.0\pi, 1.5\pi$ 回転の 4 つの値を初期値として用いる．ここで仮定として人体は y 軸の正の方向に頭部を向けているとした．

以上より，Θ は 16 通りの初期値をもつ．

2.5 損失関数

部位の領域分割画像が与えられる場合の勾配法で用いる損失関数は次式で定義される．

$$\mathcal{L} = \mathcal{L}_s(\bar{\mathcal{I}}_s(\Theta); \mathcal{I}_s, \mu_0) + \lambda_0 \mathcal{L}_p(\bar{\mathcal{I}}_p(\Theta); \mathcal{I}_p, \mu_1) + \lambda_1 \mathcal{L}_\theta(\theta) + \lambda_2 \mathcal{L}_a(\theta) + \lambda_3 \mathcal{L}_\beta(\beta) \tag{1}$$

ここで $\bar{\mathcal{I}}_s \in \mathbb{R}^{255\times255\times1}$，$\bar{\mathcal{I}}_p \in \mathbb{R}^{255\times255\times N_P}$ は Θ により生成された人体のメッシュをレンダリングした全身のシルエットと部位ごとのシルエットである．$\mathcal{I}_s$，$\mathcal{I}_p$ は Ground Truth の全身のシルエットと部位のシルエットであり，予測されたシルエットと同じサイズのテンソルである．テンソルについて，上付添字 i を用いて各要素を表現する．$\lambda_0, \lambda_1, \lambda_2$ は各項の重み付けをする変数である．

$\mathcal{L}_s$ は Ground Truth のシルエットと予測シルエットに関する損失関数である．これは単純にテンソル間の距離を測るのではなく次式で定義されるように本来メッシュが投影されるべき画素に投影されない場合，すなわち False Negative に対して強くペナルティを課す指標である．

$$\mathcal{L}_s = \frac{1}{255^2} \sum_i (\Delta \mathcal{I}_s^i)^2 \tag{2}$$

$$\Delta \mathcal{I}_s'^i = \begin{cases} \mathcal{I}_s^i - \bar{\mathcal{I}}_s^i & \text{if} \quad \mathcal{I}_s^i > \bar{\mathcal{I}}_s^i \\ \mu_0(\mathcal{I}_s^i - \bar{\mathcal{I}}_s^i) & \text{otherwise} \end{cases} \tag{3}$$

$\Delta \mathcal{I}_s$ は差分画像であり，μ_0 は False Negative である場合の重みをなすパラメータである．

$\mathcal{L}_p$ は Ground Truth の部位のシルエットと予測された部位のシルエットに関する損失関数である．これも同様に False Negative の場合に強くペナルティを課す．

$$\mathcal{L}_p = \frac{1}{255^2} \sum_i (\Delta \mathcal{I}_p^i)^2 \tag{4}$$

$$\Delta \mathcal{I}_p'^i = \begin{cases} \mathcal{I}_p^i - \bar{\mathcal{I}}_p^i & \text{if} \quad \mathcal{I}_p^i > \bar{\mathcal{I}}_p^i \\ \mu_1(\mathcal{I}_p^i - \bar{\mathcal{I}}_p^i) & \text{otherwise} \end{cases} \tag{5}$$

$\Delta \mathcal{I}_p$ は差分画像であり，μ_1 は False Negative である場合の重みをなすパラメータである．

$\mathcal{L}_\theta$ は SMPLify で利用された θ についての混合ガウス分布で表現される事前分布についての負の対数である．混合ガウス分布のパラメータは公開されているものを利用する．

$\mathcal{L}_a$ は膝・肘に対する回転ベクトルの x 成分にかかる損失項であり，次式で定義される．

$$\mathcal{L}_a = \sum_i \exp\left(\theta_x^i\right) \tag{6}$$

i は右肘・左肘・右膝・左膝を特定する添字であり，θ_x^i はそれらに対応する回転ベクトルの x 成分である．出力を指数関数としているため，負の方向すなわち自然な回転方向にはほとんどペナルティを課さないが，正の方向すなわち不自然な回転方向には大きくペナルティを課す．

$\mathcal{L}_\beta$ は β に対する正則項であり，β の L2 ノルムである．$\mathcal{L}_\theta, \mathcal{L}_a, \mathcal{L}_\beta$ については Bogo *et al.* [2] が提案したものを用いた．一方で相互浸透誤差項については Kolotouros *et al.* [11] の議論に従い計算コストの削減のため省略した．

2.6 最適化

本手法では，Ground Truth のシルエットが与えられたときに，目的関数 (1) を最小化する入力 Θ を勾配法を用いて数値的に求める．反復の回数は 100 とする．目的関数は Θ の他に諸パラメータを持つ．各反復過程において目的関数に含まれる諸パラメータ，$\lambda_0, \lambda_1, \lambda_2, \lambda_3$ を減衰，μ_0, μ_1 を増幅させる．

各反復過程において，反復過程を通して変動するパラメータの影響を受けない評価関数として，全身のシルエットと部位の領域分割画像の両方に対して，Ground Truth と予測シルエットの間の L2 ノルムを計算する．全反復過程において評価関数を最小にする入力 Θ を最適解とする．

本手法では，2.4 節で述べた 16 通りの初期値について最適解を求め，それらのうち最小のものを最終的な最適化とする．

3. 実験

3.1 データセット

本章で報告される実験は 3D Poses in the Wild dataset (3DPW) [13] 上で実行された．3DPW は人物を撮影した動画とフレームに付加されるアノテーションからなる．アノテーションには画像に映る人物の形状を表す SMPL パラメータが含まれる．本研究ではこの連続する SMPL パラメータの列 (シークエンス) のみを用いる．人体の分割方法については図 2 に示した分割を使用する．

シークエンス中の各フレームの Ground Truth の SMPL パラメータ，Θ_{GT} による人体のメッシュ $\boldsymbol{V}(\Theta_{GT})$ に基づき，SoftRas を用いてレンダリングを行いシルエット $\mathcal{I}_s, \mathcal{I}_p$ を生成し，$(\Theta_{GT}, \mathcal{I}_s, \mathcal{I}_p)$ の組を実験に用いる．また，SMPLify との比較実験においては Θ_{GT} から生成される三次元関節座標 $\boldsymbol{J}_{3D}(\Theta_{GT})$ を画像空間に投影し，二次元関節座標 $\boldsymbol{J}_{2D}$ を得る．

3.2 実験の設定

SoftRas は Lie *et al.* [12] による実装を利用した．SMPL は Vassilis Choutas による実装 [14] を用いた．SMPL モデルは Bogo *et al.* [2] が配布するモデルに Vassilis Choutas のツール [14] を適用したものを使用する．混合ガウス分布のパラメータは Kolotouros *et al.* [11] が公開したものを用いる．混合ガウス分布による損失関数 $\mathcal{L}_\theta$ は，Pavlakos *et al.* [15] による実装を利用した．

最適化のアルゴリズムは AdaGrad を用いる．学習率の初期値は特に言及しない限り θ に 0.06，θ_g に 0.08，$\boldsymbol{t}$ に 0.06，β に

表 1　3DPW における三次元関節座標誤差 (J_{err}) と頂点座標誤差 (V_{err})

	本手法 (全身)	本手法 (部位)	本手法 (部位 + 関節)	SMPLify
J_{err}	0.444	0.262	0.210	0.224
V_{err}	0.395	0.229	0.184	0.191

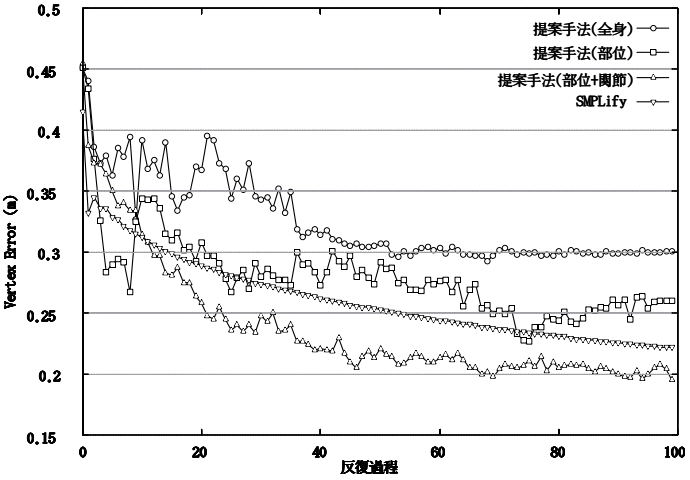

図 11　V_{err} の推移

図 12　Ground Truth の人体形状

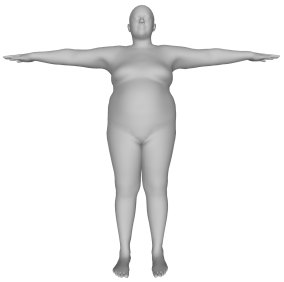

図 13　Ground Truth の T 姿勢

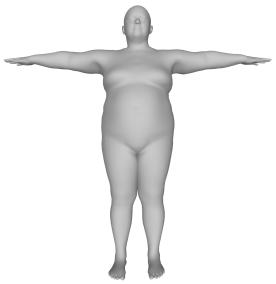

図 14　提案手法 (部位) による最適解　図 15　SMPLify による最適解

表 2　三次元頂点座標誤差 (V_{err})

	提案手法 (部位)	SMPLify
V_{err}	0.0595	0.0697

0.04 を用いる．

3.3　評 価 指 標

本手法によって数値的に求められた最適解 Θ_{opt} と Ground Truth の SMPL パラメータ Θ_{GT} の間の再構成誤差を評価するために，三次元関節座標誤差 J_{err} と三次元頂点座標誤差 V_{err} を用いる．これは再構成された骨格，メッシュに含まれる頂点の誤差の平均であり，単位はメートルである．大域的な並進移動については評価せず，姿勢と体型のみを評価するために，最適解と Ground Truth を $\boldsymbol{t} = 0$ で正規化して指標を計算する．

3.4　SMPLify との定量的評価

3DPW の train セットに含まれる全シークエンスについて，本手法と SMPLify を比較する．10 フレームごとに最適化を実行し，評価指標の全シーケンスでの平均値を表 1 に示す．

平均の結果を見ると，全身のシルエットによる最適化は再構成の誤差は大きく，部位の領域画像を与えることで最適化の誤差が小さくなることが分かった．部位の領域画像による最適化は SMPLify より誤差が大きく，部位と関節座標による最適化では SMPLify より誤差が小さくなる．全体の傾向として，関節座標の誤差が頂点座標の誤差より大きいのは，推定が困難な腕と比較して推定が容易な胴体や脚などに頂点が多くあるためであると考えられる．

capoeira の 100 フレーム目に付加される Θ_{GT} から生成された $\boldsymbol{V}(\Theta_{GT}), \boldsymbol{J}(\Theta_{GT})$ を投影して得られた二次元信号に対して提案手法と SMPLify について最適化を実施し，各反復過程における評価指標 V_{err} の推移を図 11 に示す．シルエットのみに依拠する場合の評価指標は二次元関節座標を用いる場合と比較して変動が激しく，シルエットにより構成される損失関数が複雑な形状をしていること，シルエットを二次元関節座標と併用することで変動が緩和されることが示された．

3.5　体型の復元

Ground Truth として図 12 に示すモデルを出力する SMPL パラメータを用いる．このモデルについて姿勢パラメータを零ベクトルとして T 姿勢を取らせたものを図 13 に示す．提案手法と SMPLify の最適化の結果について同じく T 姿勢に正規化をして Ground Truth のモデルと比較する．この実験では β の学習率の初期値として 1.00 を用いる．

最適化の結果として得られた解を T 姿勢に正規化したものを図 14，図 15 に示す．実験の結果を表 2 に示す．本手法は SMPLify と比較して，人体の体型についての復元力が高いこと，また二次元関節座標と比較して，シルエットが人体の体型についての情報を多く含むことが示された．

3.6　考　　察

シルエットのみによる最適化の精度として，SMPLify に劣る理由としては Ground Truth と予測された関節座標の各点が互いに対となる点を持つ一方で，シルエットを構成する画素については，ランドマーク間の対応がないことが挙げられる．部位の領域分割画像を与えて各部位に属する画素についての対応を持たせることで精度が上がるという実験の結果が，これを示唆するであろう．また，本実験では二次元関節座標について，RGB 画像に付加されるアノテーションとしてはありえない場合が含まれ得る．例えば，腕が胴体によって完全に

隠蔽されている場合に RGB 画像から腕の関節座標を完全に推定することはできない．しかし実験ではそのような場合にも腕の関節座標を SMPLify に伝えている．一方で，シルエットについてはレンダラによって投影される段階で隠蔽された部分は遮断され，画像には含まれない．この点で本実験は二次元関節座標を用いた場合に有利に働くと言える．

4. 結　　論

本研究では部位の領域分割画像に対する SMPL パラメータの最適化手法を考案した．従来手法を用いた場合の関節誤差 (m) と頂点誤差 (m) が，それぞれ 0.224 と 0.191 であったが，部位の領域分割画像のみを用いた場合では，0.262 と 0.229 であり，より良い最適解を得られなかった．二次元関節座標と組み合わせて最適化を行うことにより，誤差は 0.210 と 0.184 となり，従来手法より良い最適解を得た．また，体型の推定に関する実験で，提案手法の最適解の頂点誤差 (m) が 0.0595 であり，従来手法の 0.0697 より小さく，提案手法が人体の体型に対して従来手法より高い復元力を有することを示した．

5. 今後の課題

本稿では人体の分割として $N_p = 6$ の場合のみを扱った．分割の粒度による精度の影響についても議論の余地はある．

本手法の応用例としては部位の領域分割画像が含まれるデータセットに対して三次元形状を新たに付加し拡張する，あるいは訓練に組み込むことでより強力な教師信号を提供させる，などが挙げられる．実験では精確な二次元信号が与えられるとしたが，実際の RGB 画像からニューラルネットワークにより部位の領域分割画像を推定する場合，あるいは人力で与えられる場合に，その精度が本手法の解に与える影響を調べることが重要になる．訓練に組み込む研究としては，SMPLify による応用例は研究されているが，本手法を組み込んだ場合のネットワークの性能について比較する研究が必要である．

部位の領域分割画像を単独で使用した場合では SMPLify より精度の良い最適解を出すことができない，したがって二次元関節座標と併用して最適化を行うことが望ましいが，本稿では，すべての反復過程で両方の信号を扱うべきかどうかの議論には立ち入らなかった．二次元関節座標が姿勢について強力な情報を持つ一方で，体型についてはシルエットが優位性を持つことを踏まえると，大域的には二次元関節座標で最適化をして局所的にはシルエットを用いることが妥当であると考えられる．すなわち反復過程の最初の方においては前者を用い，ある程度収束したら後者に切り替えるなどの方法をとるのが望ましい．

本手法では光学現象を考慮しなかったが，これは微分可能レンダラが実装するシェーディングの機能が貧弱であり，古典的なラスタライゼーションが大域イルミネーションを苦手とするためである．人間の服や皮膚は多様性と複雑な反射特性を持つが，これを再現するために幾何的な計算はラスタライズベースの微分可能レンダラが行い，シェーディングについては複雑な光学現象のシミュレーションを RGB 画像やメッシュの形状を入力にとるニューラルネットワークで行うなどの方法が考えられる．

文　　献

[1] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M.J. Black, “SMPL: A skinned multi-person linear model,” ACM Trans. Graphics (Proc. SIGGRAPH Asia), vol.34, no.6, pp.248:1–248:16, Oct. 2015.

[2] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M.J. Black, “Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image,” Computer Vision – ECCV 2016, pp.••–••, Lecture Notes in Computer Science, Springer International Publishing, Oct. 2016. http://smplify.is.tuebingen.mpg.de/.

[3] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.36, no.7, pp.1325–1339, jul 2014.

[4] S. Johnson and M. Everingham, “Clustered pose and nonlinear appearance models for human pose estimation,” Proc. BMVC, pp.12.1–11, 2010. doi:10.5244/C.24.12.

[5] C. Lassner, J. Romero, M. Kiefel, F. Bogo, M.J. Black, and P.V. Gehler, “Unite the people: Closing the loop between 3d and 2d human representations,” IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.••–••, July 2017. http://up.is.tuebingen.mpg.de

[6] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, PiotrDollár, C.L. Zitnick, “Microsoft coco: Common objects in context,” European Conference on Computer Vision (ECCV), pp.••–••, Zürich, 2014. Oral.

[7] I.B. Vince Tan and R. Cipolla, “Indirect deep structured learning for 3d human body shape and pose prediction,” Proceedings of the British Machine Vision Conference (BMVC), eds. by G.B. Tae-Kyun Kim, Stefanos Zafeiriou and K. Mikolajczyk, pp.15.1–15.11, BMVA Press, Sept. 2017. https://dx.doi.org/10.5244/C.31.15

[8] G. Varol, D. Ceylan, B. Russell, J. Yang, E. Yumer, I. Laptev, and C. Schmid, “BodyNet: Volumetric inference of 3D human body shapes,” ECCV, pp.••–••, 2018.

[9] M. Omran, C. Lassner, G. Pons-Moll, P. Gehler, and B. Schiele, “Neural body fitting: Unifying deep learning and model based human pose and shape estimation,” 3DV 2018 , International Conference on 3D Vision, pp.484–494, IEEE, Verona, Italy, 2018.

[10] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis, “Learning to estimate 3D human pose and shape from a single color image,” CVPR, pp.••–••, 2018.

[11] N. Kolotouros, G. Pavlakos, M.J. Black, and K. Daniilidis, “Learning to reconstruct 3d human pose and shape via model-fitting in the loop,” Proceedings of the IEEE International Conference on Computer Vision, pp.••–••, 2019. http://visiondata.cis.upenn.edu/spin/data.tar.gz.

[12] S. Liu, T. Li, W. Chen, and H. Li, “Soft rasterizer: A differentiable renderer for image-based 3d reasoning,” The IEEE International Conference on Computer Vision (ICCV), pp.••–••, Oct. 2019. https://github.com/ShichenLiu/SoftRas.

[13] T. vonMarcard, R. Henschel, M. Black, B. Rosenhahn, and G. Pons-Moll, “Recovering accurate 3d human pose in the wild using imus and a moving camera,” European Conference on Computer Vision (ECCV), pp.••–••, sep 2018.

[14] https://github.com/vchoutas/smplx,https://github.com/vchoutas/smplx/tree/master/tools.

[15] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A.A.A. Osman, D. Tzionas, and M.J. Black, “Expressive body capture: 3d hands, face, and body from a single image,” Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.10975–10985, June 2019. https://github.com/vchoutas/smplify-x/blob/master/smplifyx/prior.py. http://smpl-x.is.tue.mpg.de