

単眼着衣画像からの非着衣人体三次元モデル推定

繁木 結衣^{1,a)} 大倉 史生^{1,b)} 満上 育久^{2,c)} 八木 康史^{1,d)}

概要: 衣服の下に隠れた人体形状の計測は非常に困難である一方、様々な用途において必要とされている技術である。例えば近年、衣服のオンラインショッピングにおいて試着ができないという欠点を補うために、仮想的に試着を行う仮想試着が考案されてきている。これまでに提案されてきた仮想試着では、利用者の全身画像に衣服の画像を重ね合わせるなどの方法で、主にデザインやコーディネートを可視化することが目的とされてきた。しかし、仮想試着を通じて衣服のサイズやフィット感を知るためには、利用者の正確な体型情報が必要になる。本研究では、単眼着衣画像から非着衣時の三次元人体モデルを生成し、非着衣体型を推定する手法を提案する。従来の体型推定手法は多視点動画画像や三次元データから推定する方法が主流であり、単眼画像から衣服の下に隠された真の体型を推定することは困難であった。提案手法では、単眼人物画像から得られた服装のカテゴリに基づく領域分割に従い、あらかじめ推定しておいた衣服の厚さを考慮しながら人体モデルの最当ではめを行い、単眼着衣画像からの非着衣時の三次元体型推定を実現する。

1. はじめに

コンピュータビジョン (CV) 分野、特に人物映像解析分野において、人体形状が用いられる応用は非常に多岐にわたる。例えば、バイオメトリクスにおいては歩行者形状を特徴量とした歩容認証 [1] が試みられ、エンタテインメント用途においては RGB-D カメラの普及により三次元形状計測が容易になったこともあり、人物動画への仮想物体の重畳などに広く活用されている。しかし、映像から人物シルエットとして直接得られるのは、見た目の人物形状、すなわち多くの場合「服を着た状態の人物形状」である。よって、服装変化等によって起こる見た目の形状変化は、個人の人物形状の不変性を仮定する応用 (例えば歩容認証 [2]) において重要な問題となる。本来、着衣時の人物形状と人体形状は異なる。つまり、真の人体形状は、服の下に隠れている。

非着衣人体形状の重要な応用の一つが仮想試着である。近年、衣服のネットショッピングが普及し、実際に店頭で衣服を見ずに購入する顧客が増えている。しかし、オンラインショッピングでは衣服の試着ができないため、届いた商品が自分の想像したデザインやサイズと異なる、というようなことが多く発生する。このような問題を解消するために、利用者の全身画像に衣服の画像を重ね合わせるなど

の方法で仮想的に試着を行う「仮想試着」システムが提案されている [3]。しかし、衣服モデルの重畳に基づく仮想試着はデザインやコーディネートの参考にすることはできても、衣服の正確なサイズやフィット感を知るためには利用者の正確な体型情報が必要になる。

従来、非着衣時の体型を推定する研究は行われてきた。しかし、三次元スキャナや、同期された多数のカメラによって得られた着衣人物の三次元データを用いて体型を推定する手法が主流であり、自宅での仮想試着等に必要とされる、簡便な入力による体型取得は困難であった。本研究は、特殊な撮影環境・センサを必要としない体型取得手法として、一枚の着衣時の全身画像から非着衣時の三次元人体モデルを推定する手法を提案する。

2. 関連研究

2.1 三次元人体形状推定

衣服も含めた (着衣時の) 人体形状を推定する手法は、これまでも CV 分野における主要な問題の一つであった。多視点画像を入力する手法 [4] のほか、RGB-D 画像を用いたアプローチ [5] が提案されている。これらの手法は、入力画像 (群) に対し人体のパラメトリックモデルを当てはめることで三次元形状を推定するが、人物が着ている服も体型に含まれるため、服装によって推定される体型が変化することが問題となる。単眼画像からの三次元体型推定手法として、Guan らの手法が挙げられる [6]。本手法は、シルエット、影情報、関節位置から人体モデルを推定する。しかし、関節位置は手動で入力する必要があるとともに、他の手法と同様、服に隠された真の体型の推定については

¹ 大阪大学

² 広島市立大学

a) shigeki@am.sanken.osaka-u.ac.jp

b) okura@am.sanken.osaka-u.ac.jp

c) mitsugami@hiroshima-cu.ac.jp

d) yagi@sanken.osaka-u.ac.jp

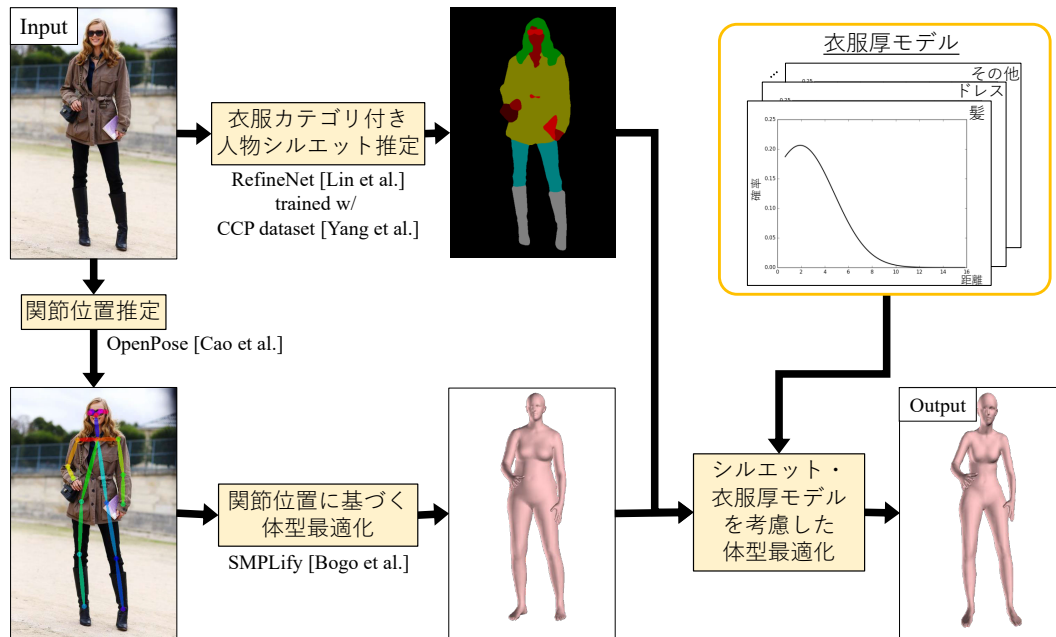


図 1 提案手法の流れ：関節位置に基づき最適化された人体モデルパラメータを初期値として、シルエットおよび衣服カテゴリに基づく衣服厚モデルを考慮した最適化を行う。

考慮されていない。

一方、ここ 10 年ほどの間に、着衣人物データから非着衣体型を推定する手法がいくつか提案されている。Balanchandraらは着衣人物の多視点映像から得られた三次元形状データと肌検出器を用い、肌が可視である部分の形状に当てはまるように人体モデルのパラメータを推定することで、非着衣体型を推定した [7]。また、Wuhrerらは、比較的体にフィットした薄い服を着た人物の三次元時系列データから非着衣体型の推定を行っている [8]。Yangら、Zhangらは、スカートなど厚みや広がりのある服装を着た人物についても非着衣体型の推定を実現した [9], [10]。しかし、これらの手法では同期された多視点映像や、三次元スキャナによる入力が必要とする。本研究は、上記のような非着衣体型の三次元モデルを、二次元単眼画像から推定する。

単眼 RGB 画像からの非着衣体型推定手法として、Bogoinらによって SMPLify と呼ばれる手法が提案されている [11]。SMPLify では、関節位置推定手法 [12] によって二次元画像上で検出された関節位置情報が、対象人物の体型情報を含んでいる、という仮説のもと、関節位置に合うようにパラメトリック人体モデルの姿勢・体型を最適化することで人体モデルを推定する。しかしこの手法は、シルエット形状等、画像に含まれる見た目の体型を用いることなく関節位置のみを用いて推定を行うため、画像に含まれるノイズや服装等による関節位置の変化等の影響で、推定される体型が大きく変化する。本研究で提案する手法は、SMPLify で得られた人体パラメータを初期値として、さらにシルエットおよび衣服の厚みモデルを考慮した体型の最適化を行うことで、より正確な体型推定を目指す。

2.2 人体三次元形状のパラメトリックモデル

三次元人体形状の推定には、しばしば人体形状のパラメトリックモデルが用いられる。これらのモデルは、主にコンピュータグラフィックス等の応用のために、三次元スキャナ等で取得された多くの人物形状データをパラメータ化したものである。これまでに、人物スケルトンに対して筋肉量等に基づく太さを付加した人体モデル [13] や、人物の形状のみならず姿勢情報を付加した Dyna[14]、SCAPE[15]、SMPL[16] 等のパラメトリックモデルが提案されてきた。本研究では、姿勢が変化した際にもより自然なモデルを生成可能な SMPL[16] のパラメータを推定することにより、人体形状推定を行う。SMPL は、24 関節（各 3 自由度）に対応する 72 の姿勢パラメータおよび、人体の形状変化の統計に基づき算出された主成分である 10 の体型パラメータによって体型・姿勢を表現する。

3. 提案手法：非着衣三次元モデルの推定

3.1 概要

提案手法は、図 1 に示すように、着衣単眼画像を入力とし、非着衣体型の人体モデル（SMPL モデル [16]）パラメータを出力する。SMPLify[11] により関節位置に基づく最適化が行われたモデルパラメータを初期値として、シルエット形状および衣服の厚みを考慮したさらなる最適化を行う。

本研究における重要な挑戦は、着衣時と非着衣時の人物の形状変化のモデル化にある。着衣画像から非着衣体型を推定するためには、まず衣服の厚みによって、着衣時と非着衣時でどれほど形状に差があるかを知る必要がある。ま

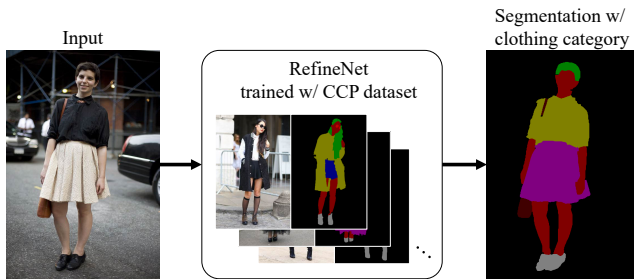


図 2 衣服カテゴリの領域分割

た、衣服の種類によって形状の変化量が異なると考えられるため、衣服のカテゴリごとに衣服の厚みをモデル化する必要がある。しかし、様々な衣服カテゴリを含む着衣・非着衣人物画像の大規模データセットは未だ存在せず、実画像群に基づくモデル化は難しい。そこで本研究では、人工的に非着衣体型モデルから様々な衣服の着衣画像群を生成し、それぞれの衣服カテゴリについて衣服の厚さをモデル化する。入力画像から推定された衣服カテゴリ付きの人物シルエットおよび、本研究で構築された衣服厚モデルを用いて、人体モデルをの当てはめを行う。

以下に、各処理について詳述する。

3.2 衣服カテゴリ付き人物シルエット推定

衣服カテゴリにより異なる衣服厚に基づく最適化を実現するため、提案手法では入力画像に含まれる人物シルエットだけでなく、服装カテゴリを出力する。本研究では、セマンティックセグメンテーション手法の一つであり、高解像度画像の入出力が可能で、人物パーツの領域分割等で高い精度を達成している RefineNet[17] を利用する。図 2 に示すように着衣人物画像中の衣服カテゴリの学習を行い、着衣画像の領域分割を行う。学習に用いたデータセットは、衣服の種類によってラベル付けされた Clothing-Co-Parsing (CCP) データセット [18] である。CCP データセットには、画素単位で 59 種類の衣服カテゴリがラベル付けされたおよそ 1000 枚の画像が含まれる。本研究では、服装カテゴリを“背景”、“肌”、“髪”、“トップス”、“スカート”、“ドレス”、“ズボン”、“靴”、“鞄”、“その他”の 10 種類に分類しなおし、データセットに含まれる 800 枚の画像を用いて学習を行った。

3.3 衣服厚モデルの構築

衣服厚のモデル化は、着衣画像から非着衣体型を推定する提案手法における核となる部分である。しかし、前述のように多くの衣服カテゴリの着衣・非着衣画像群を含むデータセットは存在しない。よって本研究では、深層学習に基づく画像生成アプローチを応用し、パラメトリック人体モデルの領域情報から、非着衣人物形状着衣人物シルエットの生成を行う。具体的には、Lassner らによる人物画像生成アプローチ [19] における着衣人物スケッチの生成

手法を用いる。

[19] は、3.2 節で述べた 10 種類の衣服カテゴリ付き人物シルエットおよび、SMPLify で推定された体型シルエットで学習が行われている。本研究では、図 3 に示すように、様々な体型・姿勢の SMPL モデルおよび体の部位を二次元画像に投影し、同じ体型・姿勢をとる着衣人物の服装ラベルを出力する。得られた着衣・非着衣の人物シルエット対について、非着衣シルエット輪郭上の各点から法線方向にある着衣シルエットの輪郭点までの距離 d_m (シルエットの外側方向を正、内側方向を負とする) を算出し、衣服カテゴリごとにシルエット変化量の統計を得る。非着衣シルエットの輪郭は必ず着衣シルエットと一致するか、内側にあるべきである。そのため、衣服カテゴリごとに $d_m \geq 0$ となるような切断正規分布のパラメータを最尤推定することで、シルエット変化量の確率密度関数を計算する。人体モデル形状の最適化には、衣服厚モデルおよび推定される形状輪郭点から計算される対数尤度に基づくコスト関数が用いられる。

3.4 パラメトリック人体モデルの当てはめ

提案手法は、パラメトリック人体モデルである SMPL を関節位置・シルエットおよび衣服厚モデルを考慮して当てはめることで、非着衣三次元人体モデルを推定する。本研究では、単眼画像からの関節位置推定手法である OpenPose[20] を用いて画像上での二次元関節位置を推定し、二次元関節位置に基づく体型パラメータの最適化手法である SMPLify によって、最適化の初期値となる SMPL パラメータを得る。

初期パラメータを入力として、3.2 節の領域分割によって得られた衣服カテゴリ付き人物シルエットと、3.3 節で作成された衣服厚モデルを用いてパラメータの最適化を行い、提案手法の出力を得る。ここで、SMPL のパラメータには体型および姿勢 (関節角) が含まれるが、本研究では、体型パラメータを最適化することによって、最適な姿勢が変化することが考えられるため、姿勢パラメータの最適化も併せて行う。

3.4.1 最適化

SMPL の体型パラメータ (10 次元) を β 、姿勢パラメータ (24 関節 72 次元) を θ と置くと、提案手法で用いる最適化関数は以下のように定義される。

$$E_{shape}(\beta) = \lambda_s E_s(\beta, \theta) + \lambda_c E_c(\beta, \theta), \quad (1)$$

$$E_{pose}(\theta) = \lambda_s E_s(\beta, \theta) + \lambda_c E_c(\beta, \theta) + \lambda_j E_j(\beta, \theta) + \lambda_a E_a(\theta) + \lambda_{sp} E_{sp}(\beta, \theta) + \lambda_\theta E_\theta(\theta). \quad (2)$$

E_{shape} 、 E_{pose} はそれぞれ体型パラメータと姿勢パラメータの最適化関数であり、 λ_s 、 λ_c 、 λ_j 、 λ_a 、 λ_{sp} 、 λ_θ は各項の重みで、定数である。 E_{shape} 最小化時には姿勢パラメータ θ を固定し、 E_{pose} 最小化時は体型パラメータ β を固定しながら、2 つの最小化を交互に繰り返すことで、モデル

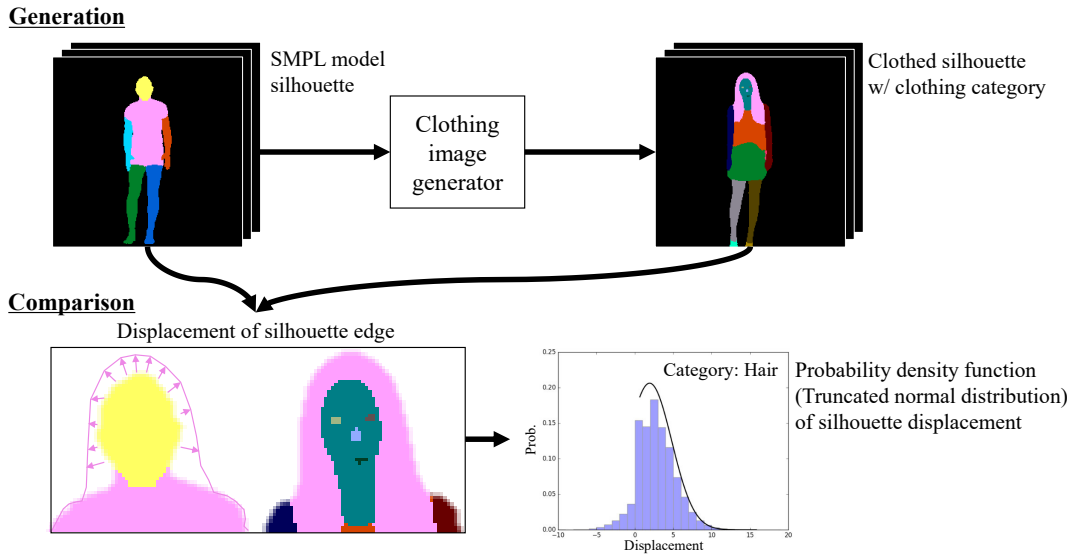


図3 衣服厚のモデル化：人物画像生成器によって得られた非着衣・着衣シルエット間の距離を服装カテゴリごとに計算し、確率密度関数（切断正規分布）を最尤推定に基づき当てはめることで、非着衣・着衣シルエット間の形状変化を確率モデルとして表現する。

全体の最適化を行う。

$E_j(\beta, \theta)$, $E_a(\theta)$, $E_{sp}(\beta, \theta)$, $E_\theta(\theta)$ は SMPLify における最適化でも用いられている項である。 $E_j(\beta, \theta)$ は SMPL モデルの関節位置と入力画像から推定された関節位置のユークリッド距離、 $E_a(\theta)$ は肘と膝が不自然な方向に回転しないようにする制約、 $E_{sp}(\beta, \theta)$ は人体モデルの相互侵入のペナルティ、 $E_\theta(\theta)$ はあらかじめ計算しておいたポーズの確からしさに関する項である。

提案手法で新たに追加する項は、肌領域のシルエットに関する項 $E_s(\beta, \theta)$ および衣服領域のシルエットに関する項 $E_c(\beta, \theta)$ である。以下に、各項の詳細について述べる。

肌シルエット項 $E_s(\beta, \theta)$: 入力画像領域が“肌”とラベル付けされたシルエット輪郭点に関する項であり、入力画像中の人物シルエット（以下、入力シルエットと呼ぶ）の輪郭点のうち“肌”とラベル付けされた点と、SMPL 人体モデルを投影したシルエット（以下、SMPL シルエットと呼ぶ）の輪郭点が一致するように振る舞う。

$$E_s(\beta, \theta) = \frac{\text{dist}(\beta, \theta, \mathbf{d}_{\text{skin}})}{n_{\text{skin}}}. \quad (3)$$

dist は、入力シルエットの“肌”とラベル付けされたエッジ点から、その最近傍の SMPL シルエットのエッジ点までの距離で、 n_{skin} は入力シルエットの“肌”とラベル付けされたエッジ点の数である。ここで、距離 dist は以下のように定義される。

$$\text{dist}(\beta, \theta, \mathbf{d}_{\text{skin}}) = \sum_i \begin{cases} d_i(\beta, \theta) & (d(\beta, \theta) \geq 0) \\ d_i(\beta, \theta)^2 & (d(\beta, \theta) < 0). \end{cases} \quad (4)$$

\mathbf{d}_{skin} は、入力シルエットの肌ラベルを持つ輪郭点と SMPL シルエットの最近傍輪郭点間のユークリッド距離を格納

するベクトルであり、その各要素は、SMPL シルエットの各点が入力シルエット内部に存在するときに正、入力シルエット外部に存在するときに負の値をとるようにする。また、非着衣の身体は必ず入力シルエットの内側に位置するので、生成される人体モデルが入力シルエットの内側に位置するようにするために、SMPL シルエットの輪郭点が入力シルエットの外部に存在するときのペナルティを大きくする。

衣服シルエット項 $E_c(\beta, \theta)$: “肌”および“背景”以外のラベルの輪郭点に関する項である。SMPL シルエットの輪郭点のうち、“肌”以外に相当する点を服の厚さだけ入力シルエットの内部に入るようにする。この項は、3.3 節で作成した、衣服厚の確率密度関数との対数尤度を用いて計算される。

$$E_c(\beta, \theta) = \sum_{c \in C} \frac{f(\beta, \theta, \mathbf{d}_c)}{n_c}, \quad (5)$$

$$f(\beta, \theta, \mathbf{d}_c) = \left(-\sum_i \log(\text{pdf}_c(d_{c_i}(\beta, \theta)) + \epsilon)\right). \quad (6)$$

ここで C は“肌”、“背景”以外のラベル集合（各衣服カテゴリ、“髪”、“その他”の計 8 種類）を表し、 $\mathbf{d}_{c \in C}$ は、これらのラベルを持つ入力シルエットの輪郭点から、最近傍の SMPL シルエットの輪郭点までのユークリッド距離を格納するベクトルである。3.3 節で作成した衣服カテゴリごとの衣服厚の確率密度関数 $\text{pdf}_{c \in C}$ を使い、各衣服カテゴリごとに輪郭点距離に基づく負の対数尤度を求め、輪郭点数 n_c で正規化したものをコストとして用いる。つまり、事前に構築された確率モデルに対し、推定された SMPL シルエットの輪郭点距離が合致するほどコストが小さくなるように振る舞う。

3.5 実装の詳細

提案手法は、公開されている SMPLify の実装と同様 Python の Champy ライブラリを用いて実装し、SMPL パラメータの最適化には非線形最小化手法である Dogleg 法を用いた。従来手法と異なり、提案手法で導入する E_s, E_c の計算には、シルエット輪郭点間の最近傍点の探索が必要となる。本稿で述べる実験では、入力シルエットの各輪郭点について対応する SMPL シルエットの最近傍輪郭点を探索した。しかし、繰り返し行われる非線形最小化の更新処理に際して、毎回対応点を探索することは計算量の点から現実的ではない。そのため、アルゴリズム 1 に示すように、対応点探索と非線形最小化を繰り返すことで漸近的に最小化を行った。なお、繰り返し回数 n は、 $n = 4$ 程度で十分に収束することを確認した。

Algorithm 1 SMPL optimization w/ silhouette

Input: Initial SMPL parameters β_0, θ_0 from SMPLify, image silhouette edge points I .

Output: Refined SMPL parameters β, θ .

```

 $\beta \leftarrow \beta_0, \theta \leftarrow \theta_0$     ▷ Initialize by joint-based optimization
for  $i = 0$  to  $n$  do
    Calculate SMPL silhouette edge points  $S$ .
    for each  $i \in I$  do
        Find nearest neighbor point  $s \in S$ .
    end for
    Minimize  $E_{\text{shape}}(\beta)$  using Dogleg method.
    Minimize  $E_{\text{pose}}(\theta)$  using Dogleg method.
end for

```

4. 実験

提案手法の有効性を検証するため、真値付きデータセットを用いた定量評価および、ファッションスナップ写真群を用いた目視比較を行った。

4.1 定量評価

ここでは、三次元スキャナによる高精度な計測手法によって得られた非着衣体型および、当該人物の着衣時の三次元動画像を含むデータセット [10] を用いた、非着衣三次元モデリングの定量評価について述べる。具体的には、関節位置のみ用いた従来手法、関節位置・シルエット形状を考慮した手法、および関節位置・シルエット形状・衣服カテゴリごとの衣服厚モデルを用いた提案手法を用いて、データセットから生成された単着着衣画像を入力とした体型推定を行い、真値との誤差を評価する。

4.1.1 実験方法

実験には、BUFF データセット [10] を用いた。BUFF データセットは、三次元スキャナで取得されたテクスチャ付き三次元モデルの時系列データからなり、5名の被験者が、各々2種類の服装 (Tシャツ長ズボン・サッカーウェア

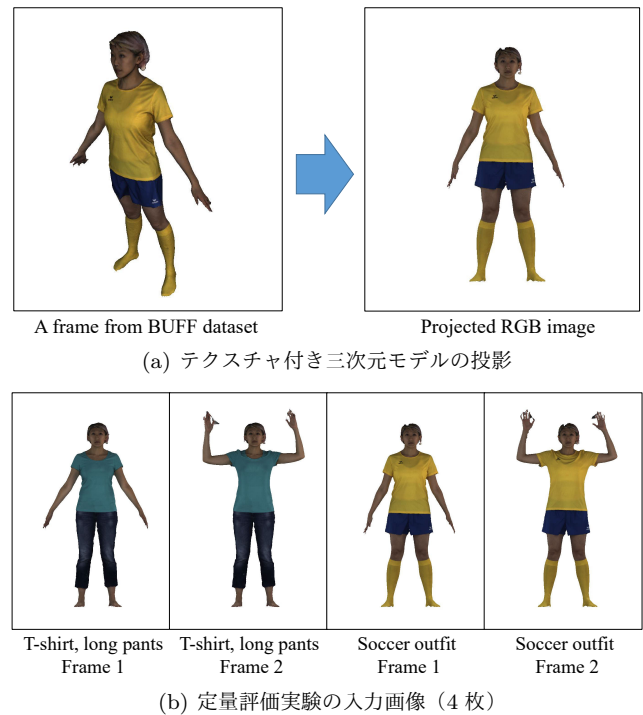


図 4 BUFF データセット [10] を用いた定量評価実験の入力画像生成

ア) で動作を行う様子が撮影されている。また、三次元スキャナで取得された各被験者の非着衣時体型が与えられている。本実験では、データセットに含まれるテクスチャ付き三次元モデルに含まれる 1 名の被験者 (ID: 3223) について、各服装について比較的服装・関節推定に適した直立時の 2 フレームずつを選択し、人物に正対する方向の仮想カメラに投影することで入力画像を作成した (図 4 参照)。

本実験では、以下の 3 手法の体型推定精度を比較した。

- (1) 関節位置のみを用いた手法 [11] (*SMPLify*)
- (2) 関節位置およびシルエット形状を用いる手法 (*Silhouette*)
- (3) 関節位置・シルエットおよび衣服厚による形状変化を考慮する提案手法 (*Proposed*)

手法 1 (*SMPLify*) には、SMPLify[11] の著者らにより公開されている Python+Champy による実装を用いた。手法 *Silhouette* は提案手法における肌シルエット項 $E_s(\beta, \theta)$ を、入力シルエット全体に適用することによって得た。つまり、入力シルエット全体に“肌”ラベルを与えた場合と同様の結果となる。

それぞれの手法で得られた SMPL 人体モデルパラメータについて、姿勢誤差の影響を排除して体型の推定精度を比較するため、特定の姿勢パラメータを代入することによって得られる基準姿勢に統一した上で得られる人体三次元モデルの各頂点と、真値モデルの各頂点間で対応する点間の平均ユークリッド距離を算出し、比較した。姿勢を統一した比較方法は、[10] に記載された体型推定の精度評価方法と同様である。BUFF データセットに含まれる三次元モデ

ルにはスケール情報が含まれる。しかし、本実験で比較する3手法はいずれもメトリック情報の推定はできない。そのため、本実験での精度評価においては、各手法で出力された三次元モデルのスケールを微小変化させながら探索し、最も真値との誤差が小さくなるスケールを採用した。

4.1.2 実験結果と考察

BUFF データセットから体型推定を行った結果を表1に示す。4画像の平均誤差は、提案手法が14.29mmとなり、比較手法の中で最良であった。入力異なるため単純比較はできないが、三次元ビデオ系列を入力とする従来手法であるYangらの手法[9]の平均誤差と比較しても高い精度が得られている。

比較的タイトな服装であるTシャツ+長ズボン(“T-shirt, long pants”)では、関節位置+シルエットによる最適化手法(*Silhouette*)で最も高い精度が得られた。これは、見た目の形状と非着衣形状間の変形が少ないため、見た目の形状(シルエット)による最適化が効果的であったことによるものであると考えられる。見た目の形状と非着衣形状の変化が比較的大きいサッカーウェア(“Soccer outfit”)画像を入力とした場合、手法*Silhouette*の誤差が大きくなり、提案手法(*Proposed*)による精度向上が見られた。

4.2 目視評価

既存のデータセットに含まれるデータは服装や体型のバリエーションが少なく、また、大きく広がりを持つ服装(スカートなど)は含まれない。提案手法の特長は、大きく広がる服装や、体型のバリエーションへの適応性にあると考えられる。そこで、ここでは様々な厚みが含まれた衣服を着用する人物画像の体型推定結果を視覚的に比較する。具体的には、ファッションスナップデータセットであるCCPデータセットおよびコミュニティフォトから収集した画像を入力として、4.1節で比較した3手法で体型推定を行った。図5に、5枚の画像を入力とした体型推定結果を示す。

図6に、関節位置を用いた手法(*SMPLify*)と提案手法(*Proposed*)による推定結果の比較を示す。関節位置のみを用いた最適化では、シルエットに合った体型が表現できていないことがわかる。一方、図7に示すような、コート

などの厚みのある服においては、提案手法で推定された人体モデルの輪郭が入力シルエットの内部に存在するのに対し、シルエット輪郭を用いた手法(*Silhouette*)は本来よりも太った体型を出力していることがわかる。

4.3 制限

実験結果より、提案手法が一枚の画像から非着衣三次元形状を推定可能であることが示された。しかし一方で、*SMPLify*[11]等、単眼画像による一般的な三次元モデル推定手法と同様に、カメラに対する奥行方向の推定精度は比較的低くなる傾向にある。本稿で述べた実験においても、図8に示すように、奥行方向の関節位置に誤差が生じることが多いことが明らかになっている。本研究で用いたパラメトリック人体モデル[16]では姿勢と体型が異なるパラメータで表現されるため、姿勢の推定誤差が直接的に体型の推定誤差として現れるわけではない。しかし、姿勢誤差によって変化したシルエット形状が影響することにより、本来とは異なる体型パラメータが出力される可能性がある。また、4.1節で述べたように、提案手法はスケール情報を出力しない。そのため、仮想試着等に直接的に応用するためには、スケールが既知の物体等を同時に撮影する必要がある。

5. おわりに

本稿では、単眼着衣画像から非着衣の体型を推定する手法を提案した。本研究では、画像生成のアプローチによる非着衣体型からの着衣画像生成に基づき、衣服カテゴリごとに、衣服厚を表現する着衣時・非着衣時のシルエット変化モデルを構築した。提案手法は、人物の関節位置・シルエットだけでなくシルエット変化モデルに基づく対数尤度を用いて、人体のパラメトリックモデルであるSMPLのパラメータを最適化することで体型推定を行う。定量評価および目視比較の結果、関節位置のみを用いた既存の非着衣体型推定手法および、関節位置・シルエット形状のみを用いた手法と比較し、高精度な体型推定が可能であることを示した。今後は、さらに詳細な着衣・非着衣形状のモデル化や、複数視点画像を入力とした高精度化を検討する。

表1 定量評価結果: BUFF データセット [10] を入力とした真値との平均誤差 [mm]。Yang *et al.* [9] および Zhang *et al.* [10] は、本実験で用いたフレームを含めた多数のフレームからなる三次元ビデオ系列を入力とする非着衣三次元モデル推定手法であり、参考として精度を記載する。

Input image	<i>SMPLify</i> [11]	<i>Silhouette</i>	<i>Proposed</i>	Yang <i>et al.</i> [9] (4D input)	Zhang <i>et al.</i> [10] (4D input)
T-shirt, long pants: Frame 1	15.52	14.61	15.99	18.15	2.26
T-shirt, long pants: Frame 2	16.06	13.21	13.26	(using whole sequence)	(using whole sequence)
Soccer outfit: Frame 1	14.01	15.72	14.70	19.47	2.33
Soccer outfit: Frame 2	14.61	15.98	13.21	(using whole sequence)	(using whole sequence)
平均誤差	15.05	14.88	14.29	18.81	2.30



図 5 CCP データセット [18] およびコミュニティフォトを入力とした非着衣体型推定結果。左から入力画像、RefineNet[17] による衣服カテゴリ付け結果、関節位置を用いた推定結果 [11] (*SMPLify*)、関節位置とシルエットを用いた推定結果 (*Silhouette*)、関節位置・シルエットおよび衣服厚モデルを用いた推定結果 (*Proposed*) を示す。

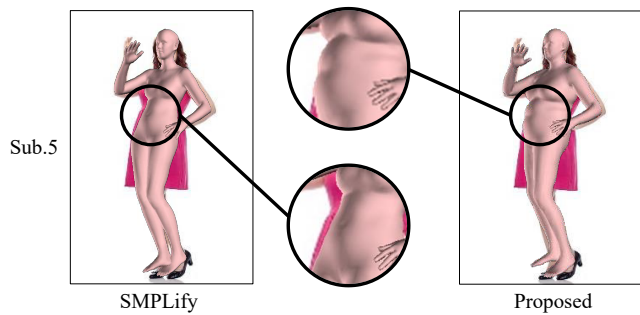


図 6 関節位置による推定結果 (SMPLify) と提案手法 (Proposed) の比較

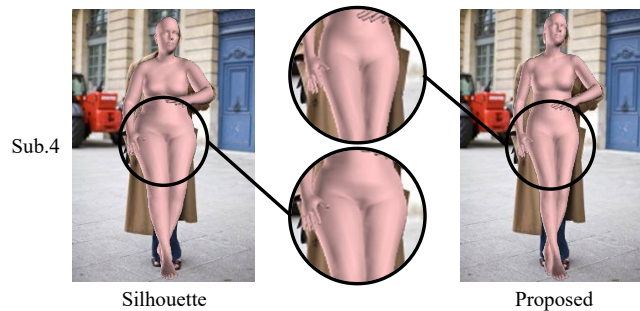


図 7 関節位置＋シルエットによる推定結果 (Silhouette) と提案手法 (Proposed) の比較

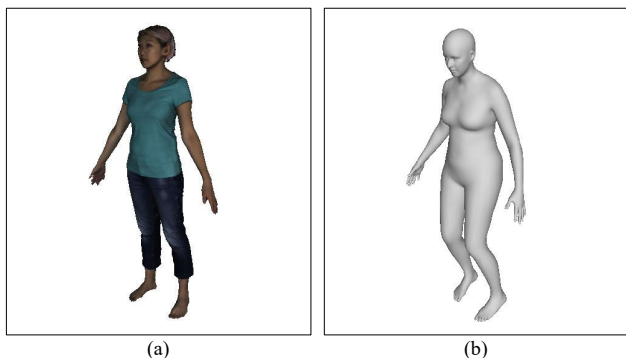


図 8 提案手法の制限：(a) テクスチャ付き三次元モデルおよび、(b) 人体前面から撮影された画像を用いた体型推定結果を (a) とほぼ同じ位置から撮影したもの。単眼画像による推定であるため、奥行き方向の関節位置に誤差が生じる。

参考文献

[1] M. S. Nixon, T. Tan, and R. Chellappa. *Human Identification Based on Gait*. Springer, 2006.

[2] X. Li, Y. Makihara, C. Xu, D. Muramatsu, Y. Yagi, and M. Ren. Gait energy response function for clothing-invariant gait recognition. In *Proc. Asian Conf. on Computer Vision (ACCV'16)*, pp. 257–272. Springer, 2016.

[3] M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar, and M-H. Foo. A mixed reality virtual clothes try-on system. *IEEE Trans. on Multimedia*, Vol. 15, No. 8, pp. 1958–1968, 2013.

[4] A. Hilton, D. Beresford, T. Gentils, R. Smith, W. Sun, and J. Illingworth. Whole-body modelling of people from multiview images to populate virtual worlds. *The Visual Computer*, Vol. 16, No. 7, pp. 411–436, 2000.

[5] A. Weiss, D. Hirshberg, and M. J. Black. Home 3D body

scans from noisy image and range data. In *Proc. IEEE Int'l Conf. on Computer Vision (ICCV'11)*, pp. 1951–1958. IEEE, 2011.

[6] P. Guan, A. Weiss, A. O. Balan, and M. J. Black. Estimating human shape and pose from a single image. In *Proc. IEEE Int'l Conf. on Computer Vision (ICCV'09)*, pp. 1381–1388. IEEE, 2009.

[7] A. Balan and M. J. Black. The naked truth: Estimating body shape under clothing. In *Proc. European Conf. on Computer Vision (ECCV'08)*, Vol. 5304, pp. 15–29, 2008.

[8] S. Wuhrer, L. Pishchulin, A. Brunton, C. Shu, and J. Lang. Estimation of human body shape and posture under clothing. *Computer Vision and Image Understanding*, Vol. 127, pp. 31–42, 2014.

[9] J. Yang, J.-S. Franco, F. Hétyroy-Wheeler, and S. Wuhrer. Estimation of human body shape in motion with wide clothing. In *Proc. European Conf. on Computer Vision (ECCV'16)*, pp. 439–454. Springer, 2016.

[10] C. Zhang, S. Pujades, M. J. Black, and G. Pons-Moll. Detailed, accurate, human shape estimation from clothed 3D scan sequences. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'17)*, 2017.

[11] 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. In *Proc. European Conf. on Computer Vision (ECCV'16)*, pp. 561–578. Springer, 2016.

[12] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V. Gehler, and B. Schiele. DeepCut: Joint subset partition and labeling for multi person pose estimation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'16)*, pp. 4929–4937, 2016.

[13] S. I. Park and J. K. Hodgins. Data-driven modeling of skin and muscle deformation. *ACM Trans. on Graphics (TOG)*, Vol. 27, No. 3, p. 96, 2008.

[14] G. Pons-Moll, J. Romero, N. Mahmood, and M. J. Black. Dyna: A model of dynamic human shape in motion. *ACM Trans. on Graphics (TOG)*, Vol. 34, No. 4, p. 120, 2015.

[15] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape completion and animation of people. *ACM Trans. on Graphics (TOG)*, Vol. 24, No. 3, pp. 408–416, 2005.

[16] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. on Graphics (TOG)*, Vol. 34, No. 6, p. 248, 2015.

[17] G. Lin, A. Milan, C. Shen, and I. Reid. RefineNet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'17)*, 2017.

[18] W. Yang, P. Luo, and L. Lin. Clothing co-parsing by joint image segmentation and labeling. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'14)*, pp. 3182–3189, 2014.

[19] C. Lassner, G. Pons-Moll, and P. V. Gehler. A generative model for people in clothing. In *Proc. IEEE Int'l Conf. on Computer Vision (ICCV'17)*, 2017.

[20] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'17)*, 2017.