身体と被服のサイズ関係に基づく着用シルエットの印象推定

池田　宗也† 　桂井麻里衣†† 　真木　勇人††† 　後藤　亮介†††

† 同志社大学大学院理工学研究科　〒 610–0394 京田辺市多々羅都谷 1-3
†† 同志社大学理工学部　〒 610–0394 京田辺市多々羅都谷 1-3
††† ZOZO Research　〒 150–0001 東京都渋谷区神宮前 5 丁目 52-2 青山オーバルビル 3F
E-mail: †{ikeda,katsurai}@mm.doshisha.ac.jp, ††{hayato.maki,ryosuke.goto}@zozo.com

あらまし　画像認識技術によるファッションの意味理解や分析を目的として，衣服を着用した人物の画像（以下，スナップ画像）内の衣服の認識に関する研究が活発化している．過去に，スナップ画像に写った衣服の種類や領域，色を抽出する研究が報告されている．しかし，衣服の着こなし（コーディネート）やスタイルは衣服の種類や柄，色の組合せのみで表現されるわけではない．衣服着用時のシルエットが与える印象（タイト，ルーズなど）もコーディネートを構成する上で重要な要素である．そこで本稿は，衣服着用時のシルエットが与える印象の推定手法を提案する．提案手法では，スナップ画像から推定した衣服の領域と人体形状の領域を比較することで，衣服着用時のシルエットが与える印象を推定する．ファッションコーディネートアプリ WEAR 上に投稿されたスナップ画像を用いた実験では，提案手法によって下半身の衣服をルーズまたはタイトの 2 クラスを判別するタスクにおける Accuracy は 0.734 となった．

キーワード　ファッション, 深層学習, 衣服の領域分割, 人体形状推定

1 はじめに

ファッションに関する市場は，現在全世界で約 3 兆ドルの市場規模 [1] をもつ．画像認識技術を使ったファッション分析やサービス開発が新たな市場価値の創出や顧客体験につながることから，スナップ画像を対象とした研究が活発化している．特に，スナップ画像に写った衣服の認識に関する研究は，EC サイトにおける類似商品の推薦や Amazon Echo Look[2] などのコーディネート評価サービスなどに応用され，実用化が進んでいる．近年では，衣服に関するラベルが付与された大規模なファッション画像データセットが構築されており [15, 19, 31]，研究対象のタスクは多岐に及ぶ．具体的には，ファッションスタイルの分類 [9, 14, 26]，類似商品検索 [11]，トレンド予測 [29] などが挙げられる．中でも，スナップ画像から衣服の種類や色，柄などの特徴を推定する手法 [6, 26] は，画像中の衣服の組合せからファッションスタイルを分析する研究に用いられてきた．

しかし，ファッションスタイルやコーディネートは衣服の種類や柄，色の組合せのみで表現されるわけではない．衣服にはサイズがあり，着用する人物の体型との関係によって生まれる着用時のシルエットの印象（タイト，ルーズなど）もコーディネートを構成する上で重要な要素である [2]．そこで本研究では，衣服の着用シルエットの印象という新たな属性の推定に向け，衣服下の体型と衣服の関係性に着目する．過去に，衣服下の人体の情報を姿勢推定や人体形状推定によって得ることで，体型と衣服の関係性について調査した研究が報告されている [5, 9]．本研究においても，人体形状推定によって得られた 3D 人体モデルと画像内の衣服領域を比較することが，衣服着用時のシルエットの印象推定に対して有効と考えられる．

提案手法では，入力されたスナップ画像に対して衣服領域，衣服下の人体形状を推定する．得られた各衣服領域と 人体領域の画素集合を比較することで，人体が衣服に対して占める割合を算出する．算出した数値は，着用したファッションアイテムと体型の関係性を示すシルエットの印象指標として用いる．算出したシルエットの印象指標を特徴量とした Gaussian Naive Bayes により，衣服着用時のシルエットの印象を推定する．ファッションコーディネートアプリ WEAR[3] に投稿されたスナップ画像を用いた実験では，シルエットの印象指標が衣服の着用時シルエットの印象推定に対して有効であることが明らかになった．

本稿の構成は以下のとおりである．2 章では，衣服領域推定に関する過去の研究，人体形状の推定に関する過去の研究，および体型と衣服の関連性に着目した過去の研究を紹介する．3 章では，スナップ画像の着用シルエットの印象推定手法を提案する．4 章では，実験結果に基づき提案手法の有効性を検証する．最後に 5 章では，本研究のまとめと今後の課題について述べる．

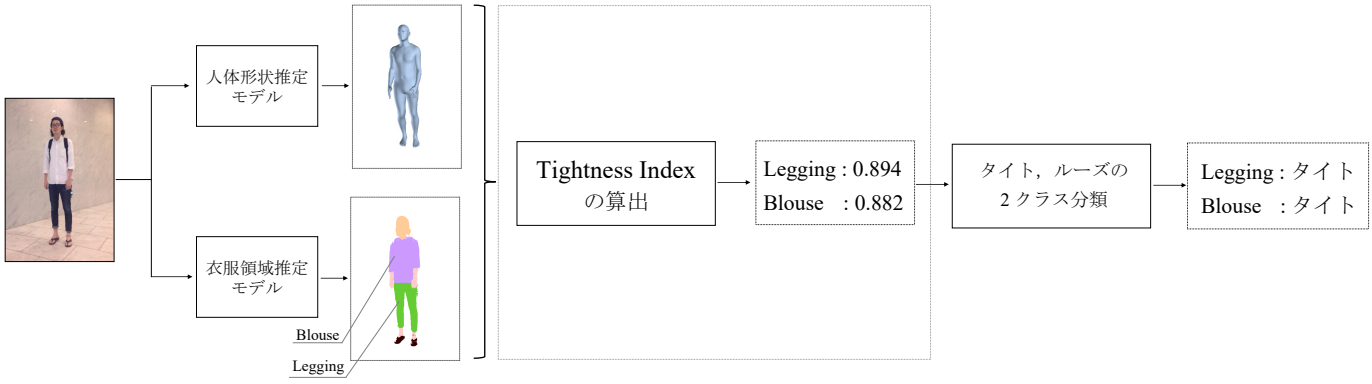
2 関連研究

本章では，衣服領域の推定に関する先行研究，人体形状の推定に関する先行研究，および体型と衣服の関連性に着目した先行研究についてそれぞれ述べる．

2.1 衣服領域の推定に関する研究

与えられた画像内の衣服領域を推定するタスクは Clothing Parsing と呼ばれ，セマンティックセグメンテーションの一分

1：https://www.mckinsey.com/industries/retail/our-insights/
2：https://www.amazon.com/Amazon-Echo-Look-Camera-Style-Assistant/dp/B0186JAEWK
3：https://wear.jp/



(a) 入力画像.　(b) 人体形状推定，服装領域推定.　(c) 各衣服の領域と 3D人体モデルの領域の重なりを提案指標 Tightness Index として算出.　(d) Tightness Index を用いた各衣服着用時のシルエットの印象推定.

図 1: 提案手法の概要図．(a) スナップ画像を人体形状推定モデルと衣服領域推定モデルに入力として与える．(b) 人体形状推定モデルは入力画像中の人物を 3D 人体モデルによってレンダリングし，出力する．衣服領域推定モデルは入力画像中の衣服の種類，およびその領域を推定する．(c) 得られた各衣服の画素集合において人体モデルの画素集合が占める割合を提案指標 Tightness Index として算出する．(d) Tightness Index を基に，分類器によって各衣服の着用時シルエットの印象をタイト，もしくはルーズの 2 クラスに分類する．

野として研究されている [6, 15, 16, 22, 27]．Clothing Parsing では，スナップ画像に対して T シャツやスカートなどの衣服に関するラベルが付与されており，入力画像内の各画素に対し適切なラベルを推定するようモデルを学習する．近年では，セマンティックセグメンテーションに関する他のタスクにおいて成果を上げた Fully-CNN [12] を用いて精度向上を目指した研究がある [22, 28]．また，Fully-CNN を用いた衣服領域推定によって画像内の衣服の色と柄，種類を関連付けて推定した研究 [26] が報告されている．本研究においても，スナップ画像内の衣服領域の推定のために Fully-CNN をベースとしたモデルを用いる．

2.2 人体形状推定

人体形状推定とは，画像内の人物の姿勢や体型に 3D 人体モデルを当てはめるタスクである．推定に用いる 3D 人体モデルは，パラメータで人物の体型や姿勢を表現する Skinned Multi-Person Linear Model (SMPL) [20] が主流となっている．人体形状推定タスクにおける SMPL を用いたアプローチは二つに大別される．一つ目が，人体形状の推定を二段階に分ける方法である [7]．Bogo ら [7] は，姿勢推定モデルによって入力画像における人体の二次元関節座標を抽出し，SMPL の関節座標との誤差を最適化することで入力画像内の人物の形状を推定した．人体形状推定を二段階に分けるアプローチの欠点として，人体形状推定の精度が姿勢推定モデルに大きく依存すること，二次元関節座標と人体モデルの関節座標間を最適化する際に生まれる誤差が大きいこと，入力画像を二次元骨格座標のデータに変換するため入力画像の画素情報を推定に使えないことが挙げられる．

二つ目が，入力画像から直接 3D の人体形状を推定する方法である [1, 8]．Kanazawa ら [1] は，CNN によって入力画像から画像特徴量を抽出し，カメラ角度や 3D 人体モデルのパラメータを生成する Human Mesh Recovery (HMR) を提案した．HMR では，敵対的生成ネットワーク構造のアーキテクチャによって単一画像から 3D の人体形状の推定を可能とした．また，HMR は入力画像から抽出した画像特徴量を用いて 3D の人体モデルを推測することで，従来の手法よりも優れた性能を示した．そこで本研究では，スナップ画像から衣服下の人体形状を推定するために HMR を用いる．

2.3 体型と衣服の関係性に着目した研究

体型と衣服の関係性に着目した研究では，体型と着用衣服の嗜好の関係性を調査した研究 [9] やスナップ画像から画像内の人物の適切な衣服サイズを推定する研究 [5] などが挙げられる．Sattar ら [9] は衣服下の人体形状を 3D 人体モデルによって推定し，3D 人体モデルの形状と画像内から推定した衣服を分析することで，体型とファッションスタイルの嗜好を調査した．結果として，体型の違いによってファッションスタイルや着用する衣服に変化が現れることを示した．Song ら [5] は，ポーズや衣服などの状態が統一されたスナップ画像を入力として，画像内の人物の適切な衣服のサイズを推定する方法を検討した．具体的には，入力画像内の人物から推定した二次元関節座標を用いて，人物の体型とそれに適応する衣服のサイズを推定した．

本研究は体型と衣服の関係性に着目した中でも，スナップ画像内の衣服領域と人体形状推定によって得られた人体モデルの領域を比較することで，衣服着用時のシルエットの印象を推定するという点で先行研究と異なる．

3 身体と衣服のサイズ関係に基づく着用シルエットの印象推定

本章では，衣服着用時のシルエットの印象推定手法を提案する．提案手法の概要を図 1 に示す．提案手法では，スナップ画像に対して衣服領域を推定し（3.1 節），衣服下の人体形状を 3D 人体モデルによって推定する（3.2 節）．画像中に含まれ

る各衣服の画素集合と人体モデルの画素集合を比較することで，提案指標 Tightness Index を算出する (3.3 節)．最後に，Tightness Index を用いて，画像内の各衣服の着用シルエットの印象をタイト，ルーズの 2 クラスに分類する．(3.4 節)．

3.1 入力画像の衣服領域推定

本研究ではスナップ画像内の衣服領域推定のために，Zhao ら [10] が提案した Fully-CNN モデル Pyramid Scene Parsing Network (PSPNet) を用いる．PSPNet とは，CNN から抽出した画像特徴量とその画像特徴量に対して様々なフィルタサイズのプーリング処理を適用した結果を結合させ，ラベル領域を推定するモデルである．フィルタサイズの異なるプーリング処理を画像特徴量に対して適用することで，周辺画素との関係性（コンテキスト）を考慮した領域推定が可能となっている．また，本研究では PSPNet の事前学習のためにスナップ画像データセット Colorful Fashion Parsing Data (CFPD) [24] を用いた．CFPD は， 23 種類の衣服の領域に関するラベルが付与された 2,682 枚のスナップ画像から成るデータセットである．本研究で用いる PSPNet は，入力画像内の各画素に対して CFPD で定義された 23 種類のラベルの中から適切なラベルを推定するよう学習を行った．画像特徴量抽出のための CNN には ResNet50 [13] を用いており，損失関数には交差エントロピー誤差，最適化には Adam を用いた．学習では，CFPD の 80 %を学習に，20 %を検証に用いた．学習時の設定は Epoch 数を 50 とし，バッチサイズ を 12 とした．テスト時の Accuracy は 0.851 であった．

PSPNet による衣服領域の推定結果に対する後処理として，本研究では DenseCRF [21] を適用する．DenseCRF とは無向グラフィカルモデルの一種であり，領域推定の分野で用いられている手法である．画像内の各画素の色特徴の関連性を考慮することで，画素の所属するクラスを推定する．DenseCRF を適用することによる領域推定精度の向上が先行研究によって確認されている [17, 22].

3.2 入力画像に含まれる人体の形状推定

本研究では，スナップ画像から人物の衣服下の体型を推定するために Kanazawa ら [1] が提案した Human Mesh Recovery (HMR) を用いる．HMR とは，入力画像内の人物を検出し，出力としてカメラ角度項と SMPL [20] のパラメータ項の 2 つを推定するモデルである．カメラ角度項とは，画像中の人物がどのような角度から画像に写されているかを推定する項であり， SMPL のパラメータ項は，3D 人体モデルの体型や姿勢を決定するパラメータを推定する項である．HMR から出力されたこれらのパラメータに基づき 3D 人体モデルをレンダリングすることで入力画像中の人体形状を推定した結果が得られる．本研究で用いる HMR は，2 次元の関節座標が付与された画像データセット LSP, LSP-extended [23] , MPII [18], MS COCO [25], および 3 次元人体モデルの正解ラベルが付与された画像データセット Human.3.6M [3], MPI-INF-3DHP [4] によって事前学習済みのモデルを用いた．

3.3 衣服と着用者の体型の関係性の数値化

本研究では，衣服着用時のシルエットの印象は衣服を着用した人物の体型と衣服サイズの関係によって生じており，それらを比較することで衣服着用時のシルエットの印象推定が可能であると考えた．そこで，衣服着用時シルエットの印象推定のために衣服領域の推定結果と人体形状推定結果を用いる．ここで，衣服領域の推定結果は，画像内の各画素の衣服領域ラベルが付与された画像である．人体形状の推定結果は入力スナップ画像内で検出された人体が 3D モデルとしてレンダリングされた画像である．衣服領域推定結果と人体形状推定結果を用いて，式 (1) によって衣服と着用者の体型の関係性を数値化する．また，3.1 節で述べたように本研究では，衣服領域推定モデルの学習のために CFPD を用いた．CFPD では衣服に関する 23 種類のラベルが付与されているが，その中には本研究において区別する必要のないラベルが含まれている（“Legging”と“Pants”と“Jeans”など）．衣服領域モデルによる推定で得られた画像には，そのような類似したラベルが同一のスナップ画像内に混在している場合がある．このような推定結果が得られた場合，類似ラベルの中で最も領域内画素集合が大きいラベルに対してのみ提案指標を算出する．

画像 i から検出された N 種類の衣服について，それらの領域に含まれる画素集合をそれぞれ $c_1(i), ..., c_N(i)$ とする．また，人体形状推定の結果として得られた身体領域に含まれる画素集合を $b(i)$ とする．このとき，n 番目の衣服に対するシルエット印象の指標 $Tightness\ Index(i, n)$ を以下の式で定義する．

$$Tightness\ Index(i, n) = \frac{|c_n(i)| \cap |b(i)|}{|c_n(i)|}. \tag{1}$$

上式において，$|\cdot|$ は集合の要素数を表す．Tightness Index が大きいほど人体と衣服がより密着している状態（タイト）であるとみなし，Tightness Index が小さいほど人体と衣服がより離れている状態（ルーズ）であるとみなす．

3.4 衣服着用時シルエットの印象推定

本研究では，衣服の着用時シルエットの印象推定を タイトまたはルーズの 2 クラス分類問題として解く．分類器には，Gaussian Naive Bayes を用いる．入力特徴量には，3.3 節において，式 (1) によって算出した画像中の衣服と体型の関係性を示す Tightness Index を用いる．Gaussian Naive Bayes は，ベイズの定理に基づいた教師あり分類器である．次式のように，入力された特徴量 x を尤度が最も大きいクラス C_k に分類する．

$$\underset{k \in \{1,\ldots,K\}}{\arg\max}\ (\ln P(C_k) + \ln P(x|C_k)). \tag{2}$$

Gaussian Naive Bayes 分類器では，入力となる特徴量の分布を正規分布として仮定する．式 (2) における $P(x|C_k)$ は，以下の式によって定義される．

$$P(x|C_k) = \frac{exp(\frac{(x-\mu_k)^2}{2\sigma_k^2})}{\sqrt{2\pi\sigma_k^2}}. \tag{3}$$

ここで，μ_k はクラス C_k に属する特徴量の平均， σ_k はクラス C_k に属する特徴量の分散を表す．

4 実　　験

4.1 データセット

3 章で提案した手法の有効性を検証するための実データとして，WEAR でユーザが 2016 年の 1 月から 2018 年の 10 月までに投稿したスナップ画像を用いた．投稿された各スナップ画像には，ユーザ情報，ユーザが自由記述できるタグ，およびキャプションが付与されている．

本実験では，ユーザ付与タグの中から衣服着用時のシルエットが与える印象を表すタグ（ワイドパンツ，ビッグシルエット T シャツなど）を着用シルエットの印象を表す擬似的な正解ラベルとして用いた．具体的には，着用時にルーズな印象を与えるパンツを示すタグである“ワイドパンツ”タグ，タイトな印象を与えるパンツを示すタグである“スキニーパンツ”タグが付与された画像を収集した．

WEAR に投稿されたスナップ画像の中には，衣服やアイテムのみが写された画像が存在している．本研究の目的は衣服着用時のシルエットの印象を推定することであり，人物以外の画像をデータセットから除去する必要がある．そこで，収集した画像データセットの各画像に対して姿勢推定モデル OpenPose [30] を適用した．得られた推定結果に対して，人体が含まれない場合や全ての二次元関節座標が正しく検出できなかった場合はその画像を除去した．加えて，ユーザ付与タグが示す衣服着用時のシルエットの印象とスナップ画像を正しく対応させるために，複数人の二次元関節座標が検出された画像を除去した．また，衣服領域推定モデルを適用した時に検出された衣服領域の総画素数が 100 以下の画像も除去した．最終的に，“スキニーパンツ”，“ワイドパンツ”タグが付与されたスナップ画像各 1,200 枚，計 2,400 枚に対して提案手法を適用し，着用時のシルエットの印象を推定した．Gaussian Naive Bayes による分類では，画像データの 90 % を学習に，10 % を検証に用いた．

4.2 スナップ画像に対する衣服領域推定，および人体形状推定の結果に対する定性的評価

本節では，HMR を用いて推定した 3D 人体モデルと PSPNet によって推定された衣服領域の結果を定性的に評価する．スナップ画像に対する提案手法の適用結果の例を図 2 に示す．結果から，HMR は衣服によって隠れた体のパーツに対しても高精度に推定できることがわかった（図 2 四行目）．その一方で，足の向きなどの細かい部分の推定が困難であることがわかった．この問題は，HMR が入力画像から得ている二次元関節座標のランドマークの少なさに起因している．HMR は人体に対して 17 種類の二次元関節座標のランドマーク (MS COCO [25] 方式) を推定している．しかし，足を示すランドマークは踵の部分にしか存在せず，つま先の向きや形を示す情報が存在しないことが，細かいつま先などの情報を再現できていない原因だと考えられる．また，複雑なポージングや特殊な衣服の着用方法によって推定が失敗するケースも確認された（図 2 右端のスナップ画像）．

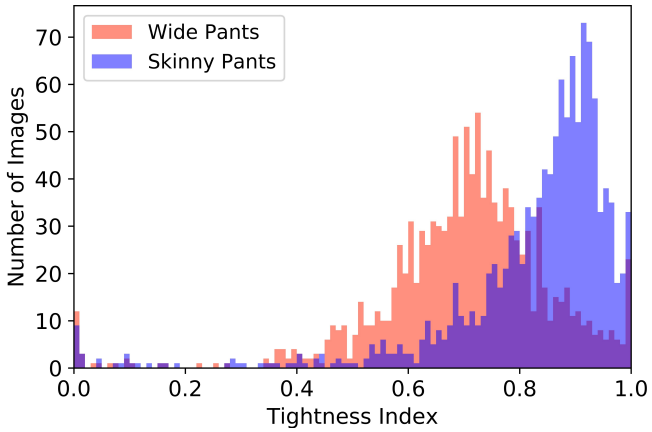


図 3: “スキニーパンツ”タグが付与された画像データセットと“ワイドパンツ”タグが付与された画像データセットに対する提案手法の適用結果のヒストグラム (x 軸が提案手法の数値．y 軸が画像の枚数．)．

表 1: “スキニーパンツ”タグが付与された画像データセットと“ワイドパンツ”タグが付与された画像データセットに対する提案手法の適用結果の比較

	平均値（標準偏差）	中央値
ワイドパンツ	0.701(±0.159)	0.716
スキニーパンツ	0.826(±0.165)	0.872

PSPNet は多くの衣服領域を正しく推定できている一方で，足元などの細かい領域や照明によって明度が変化した領域に対する推定が上手くいかないことがわかった（図 2 三行目）．他の誤推定の原因として，PSPNet の学習に用いた CFPD に含まれるスナップ画像と実験に用いたスナップ画像データの差異が挙げられる．CFPD は，主に海外で使用されているスナップ画像投稿 SNS サイト Chictopia[4]から収集したスナップ画像によって構成されている．海外と日本におけるファッションスタイルや着こなし，トレンドの違い，および体型の違いが推定に影響を与えていると考えられる．また，PSPNet の推定結果に対して，DenseCRF を適用することで誤った推定を低減できることがわかった（図 2 三行目）．しかし，DenseCRF の適用によって却って推定が悪化した例もある．それは，異なる衣服（パンツとシャツなど）の色彩，もしくは背景と衣服の色彩が類似している場合などに確認された（図 2 右端のスナップ画像）．このような問題に対する対策として，入力スナップ画像のコントラスト変更などが考えられる．

4.3 実 験 結 果

“ワイドパンツ”タグと“スキニーパンツ”タグが付与された各 1,200 枚の画像に対して，提案手法により得られた Tightness Index のヒストグラムを図 3 に示す．また，提案手法によって算出された各タグに対するシ Tightness Index の平均，標準偏差および中央値を表 1 に示す．表 1 に示した結果から，“スキ

4 : http://www.chictopia.com

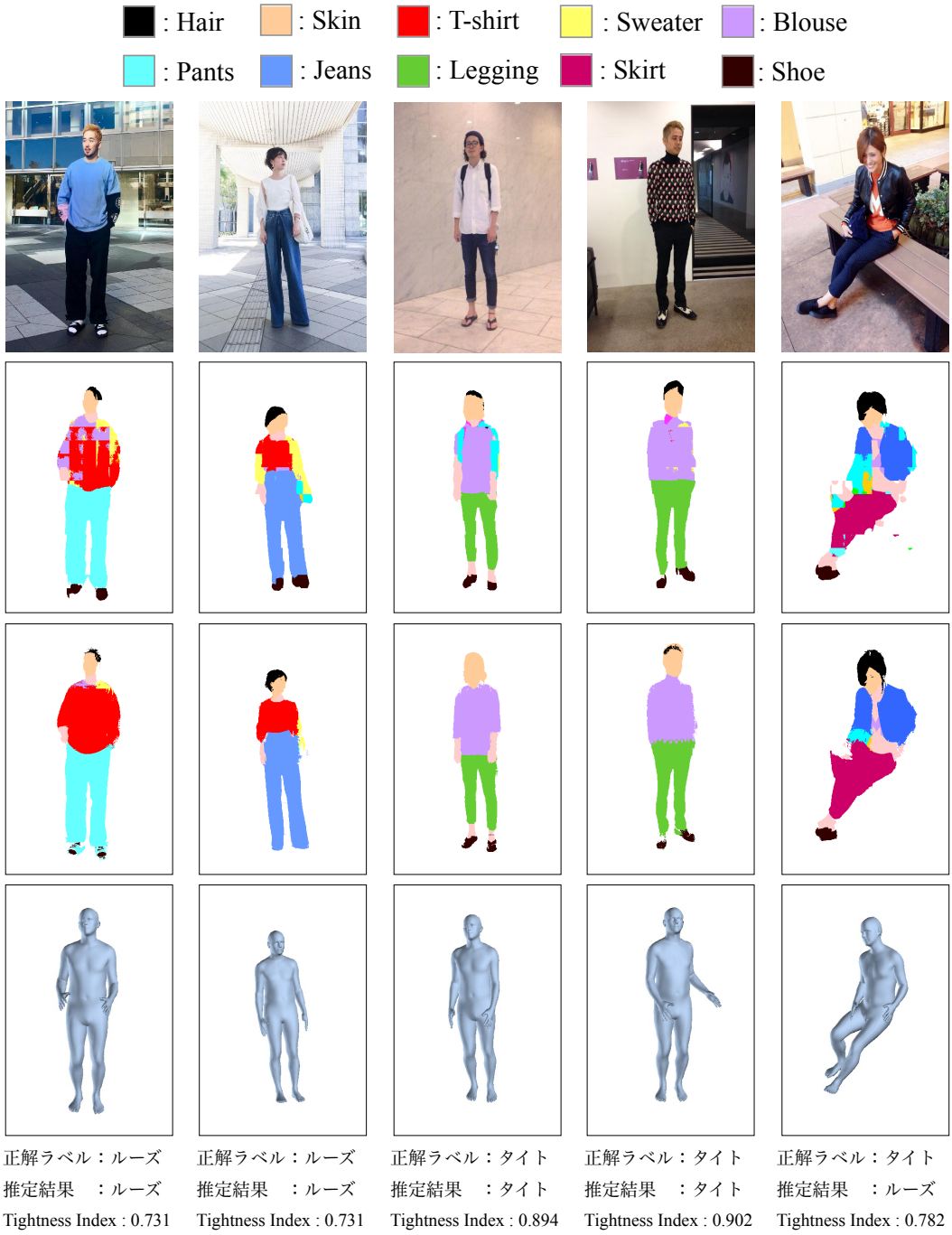


図 2: ファッションコーディネートアプリ WEAR から収集したスナップ画像に対する各推定手法の適用結果．一行目：入力として用いたスナップ画像．二行目：PSPNet による衣服領域推定の結果．三行目：DenseCRF の適用結果．四行目：HMR による人体形状推定結果．

ニーパンツ”タグが付与された画像データセットに対する提案手法の数値の平均が“ワイドパンツ”タグが付与されたものより大きいことが分かる．Gaussian Naive Bayes を用いた分類では，学習時の Accuracy は 0.742，テスト時の Accuracy は 0.734 となった．結果から，提案した指標は衣服の着用時シルエットの印象を推定に有効だと考えられる．

一方で，作成した二つの画像データセットにおいて，提案指標が 0 になる場合が確認された．これは 4.2 節で言及したように，特殊なポージングによって人体形状推定や衣服領域の推定が上手くいかなかったことが原因として挙げられる．提案手法は，衣服の推定と体型の推定の両方の結果が正しく出力されなければ指標の性能に影響を及ぼす．この問題に対する解決策として，推定モデルの構造の見直しなどによる精度改善が必要である．

5　まとめと今後の課題

本稿では，スナップ画像に対する衣服着用シルエットが与える印象を推定する手法を提案した．提案手法では，画像内の身体と衣服のサイズ関係に基づいて着用シルエットの印象を推定するアプローチをとり，スナップ画像の衣服領域の推定と衣

服下の人体形状推定した．得られた衣服領域と 3D 人体モデルの画素集合を比較することで，人体が衣服に対して占める割合を算出した．提案指標 Tightness Index を特徴量とした Gaussian Naive Bayes により画像内衣服の着用シルエットの印象を分類した．また，有効性検証を目的とした実験から得られた結果は，提案手法が衣服の着用シルエットを推定する上で有効であることを示した．

今後の課題として，今回の実験とは異なるカテゴリの衣服データセットの作成と評価が挙げられる．また，衣服領域推定モデルや人体形状推定モデルの精度改善を目指す．

文　献

[1] A.Kanazawa, M.J.Black, D.W.Jacobs, and J.Malik. End-to-end Recovery of Human Shape and Pose. In *IEEE Conference on Computer Vision and Pattern Regognition (CVPR)*, 2018.

[2] S.P. Ashdown. *Sizing in Clothing*. Woodhead Publishing, 2007.

[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] D.Mehta, S.Sridhar, O.Sotnychenko, H.Rhodin, M.Shafiei, H.-P.Seidel, W.Xu, D.Casas, and C.Theobalt. VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera. *ACM Transactions on Graphics(TOG)*, Vol. 36, No. 4, p. 44, July 2017.

[5] D.Song, R.Tong, J.Chang, T.Wang, J.Du, M.Tang, and J.J.Zhang. Clothes Size Prediction from Dressed-Human Silhouettes. In *International Workshop on Next Generation Computer Animation Techniques*, pp. 86–98. Springer, 2017.

[6] E.Simo-Serra, S.Fidler, F.Moreno-Noguer, and R.Urtasun. A High Performance CRF Model for Clothes Parsing. In *Proceedings of the Asian Conference on Computer Vision*, 2014.

[7] F.Bogo, A.Kanazawa, C.Lassner, P.Gehler, J.Romeror, and M.J.Black. Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image. In *European Conference on Computer Vision(ECCV)*, Lecture Notes in Computer Science. Springer International Publishing, October 2016.

[8] G.Pavlakos, X.Zhou, K.G.Derpanis, and K.Daniilidis. Coarse-to-Fine Volumetric Prediction for Single-Image 3D Human Pose. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1263–1272, 2017.

[9] H.Sattar, G.Pons-Moll, and M.Fritz. Fashion is Taking Shape: Understanding Clothing Preference Based on Body Shape From Online Sources. *arXiv preprint arXiv:1807.03235*, 2018.

[10] H.Zhao, J.Shi, X.Qiand, X.Wang, and J.Jia. Pyramid Scene Parsing Network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881–2890, 2017.

[11] J.Huang, R.S.Feris, Q.Chen, and S.Yan. Cross-Domain Image Retrieval with a Dual Attribute-aware Ranking Network. *IEEE International Conference on Computer Vision (ICCV)*, pp. 1062–1070, 2015.

[12] J.Long, E.Shelhamer, and T.Darrell. Fully Convolutional Networks for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[13] K.He, X.Zhang, S.Ren, and J.Sun. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[14] K.Matzen, K.Bala, and N.Snavely. StreetStyle: Exploring world-wide clothing styles from millions of photos. *arXiv preprint arXiv:1706.01869*, 2017.

[15] K.Yamaguchi, M.H.Kiapour, L.E.Ortiz, and T.L.Berg. Parsing Clothing in Fashion Photographs. pp. 3570–3577, 06 2012.

[16] K.Yamaguchi, M.H.Kiapour, L.E.Ortiz, and T.L.Berg. Retrieving Similar Styles to Parse Clothing. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 37, No. 5, pp. 1028–1040, 2015.

[17] L.C.Chen, G.Papandreou, I.Kokkinos, K.Murphy, and A.L.Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *arXiv:1606.00915*, 2016.

[18] M.Andriluka, L.Pishchulin, P.V.Gehler, and B.Schiele. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 3686–3693, 2014.

[19] M.H.Kiapour, X.Han, S.Lazebnik, A.C.Berg, and T.L.Berg. Where to Buy It:Matching Street Clothing Photos in Online Shops. In *International Conference on Computer Vision*, 2015.

[20] 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, October 2015.

[21] P.Krähenbühl and V.Koltun. Parameter Learning and convergent Inference for Dense Random Fields. In *International Conference on Machine Learning*, pp. 513–521, 2013.

[22] P.Tangseng, Z.Wu, and K.Yamaguchi. Looking at Outfit to Parse Clothing. *arXiv preprint arXiv:1703.01386*, 2017.

[23] S.Johnson and M.Everingham. Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2010.

[24] S.Liu, J.Feng, C.Domokos, H.Xu, J.Huang, Z.Hu, and S.Yan. Fashion Parsing With Weak Color-Category Labels. *IEEE Transactions on Multimedia*, Vol. 16, pp. 253–265, 2014.

[25] T.-Y.Lin, M.Maire, S.J.Belongie, L.D.Bourdev, R.B.Girshick, J.Hays, P.Perona, D.Ramanan, P.Dollár, and C.L.Zitnick. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision(ECCV)*, 2014.

[26] W.L.Hsiao and K.Grauman. Learning the Latent "Look" : Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images. *IEEE International Conference on Computer Vision (ICCV)*, pp. 4213–4222, 2017.

[27] W.Yang, P.Luo, and L.Lin. Clothing Co-Parsing by Joint Image Segmentation and Labeling. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3182–3189, 2014.

[28] X.Liang, C.Xu, X.Shen, J.Yang, S.Liu, J.Tang, L.Lin, and S.Yan. Human Parsing with Contextualized Convolutional Neural Network. *IEEE International Conference on Computer Vision (ICCV)*, pp. 1386–1394, 2015.

[29] Z.Al-Halah, R.Stiefelhagen, and K.Grauman. Fashion Forward: Forecasting Visual Style in Fashion. *IEEE International Conference on Computer Vision (ICCV)*, pp. 388–397, 2017.

[30] Z.Cao, T.Simon, S.Wei, and Y.Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 7291–7299, 2017.

[31] Z.Liu, P.Luo, S.Qiu, X.Wang, and X.Tang. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.