

Human Pose Guide Ontology を用いた 静止画像内人物のポーズ分類

田代 和浩 川村 隆浩 清 雄一 中川 博之 田原 康之

大須賀 昭彦

近年、多数のアイドルグループが人気を博しており、知名度や社会での影響力が大きくなっている。それに伴い、これらのアイドルの画像をポーズごとに分類し、鑑賞したいというニーズがアイドルファン内で高まっている。本研究は、アイドルのポーズを分析し、ユーザーに代わって分類してくれるエージェントを提案する。提案手法は大きくポーズ推定とポーズ分類の2つの処理に分かれている。ポーズ推定によって取得できる画像内人物の体のパーツ位置情報の特徴ベクトルとし、ポーズ分類によって8つのクラスへと画像を分類する。ポーズ分類の際、分類精度をより高めるため、Human Pose Guide Ontology(HPGO)を提案する。HPGOは人体構造に関する制約条件を内包し、ポーズ推定で得られた体のパーツ位置情報に補正を加えることで、分類精度を高める働きをする。評価実験により、提案手法の有用性を示し、実験結果について考察を行う。

The preference about poses of the idol varies and there is a demand for selection of idol images according to their pose. We present an agent for classifying by the pose of idol in still images. In this work, we focus on still images of idol wearing swimsuits. For each image, we create the feature vectors such as line segments indicating size and location, orientations of ten body parts(head, torso, upper/lower arms/legs) in reference to Eichner's Stickner Pose Estimation. Moreover, in order to raise the accuracy of the classification, we propose Human Pose Guide Ontology(HPGO) to guide and constrain the feature vectors. Finally, we evaluate our approach and show its efficiency of HPGO.

1 はじめに

近年、多数のアイドルグループが人気を博しており、知名度や社会での影響力が大きくなっている。このような状況はアイドル戦国時代とも呼ばれ、それに伴いWeb上に存在するアイドル画像の数も爆発的に増加している。また、これらのアイドル画像をポーズごとに分類し、鑑賞したいというニーズがアイドルファン内で高まっている。しかし、Web上のアイドル

画像の数は膨大であるため、人手での分類はユーザーにとって負担となっている。現在、主流の画像検索エンジン(Google画像検索^{†1}、Bing画像検索^{†2}など)では、「顔がアップで写った画像」、「青色の要素が多い画像」など、多少の絞り込み検索ができるとはいえ、ポーズの種類によって検索結果を分類することはできない。人物のポーズによって画像を自動的に分類するには画像の内容をコンピュータが解析し、理解する必要がある。本研究は、アイドルのポーズを分析し、ユーザーに代わって分類してくれるエージェントを提案する。本エージェントはポーズごとの鑑賞のニーズに応えることに加え、人体をデッサンする際の参考画像をポーズごとにweb上で検索する場合にも有用なシステムであると言える。このエージェント

Pose Estimation of Still Images using Human Pose Guide Ontology.

Kazuhiro Tashiro, Takahiro Kawamura, Yuichi Sei, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga, 電気通信大学大学院情報システム学研究所, Graduate School of Information Systems, University of Electro-Communications.

コンピュータソフトウェア, Vol.31, No.3(2014), pp.58-69.

[研究論文] 2013年7月16日受付.

†1 <http://www.google.co.jp/imghp>

†2 <http://www.bing.com/images>

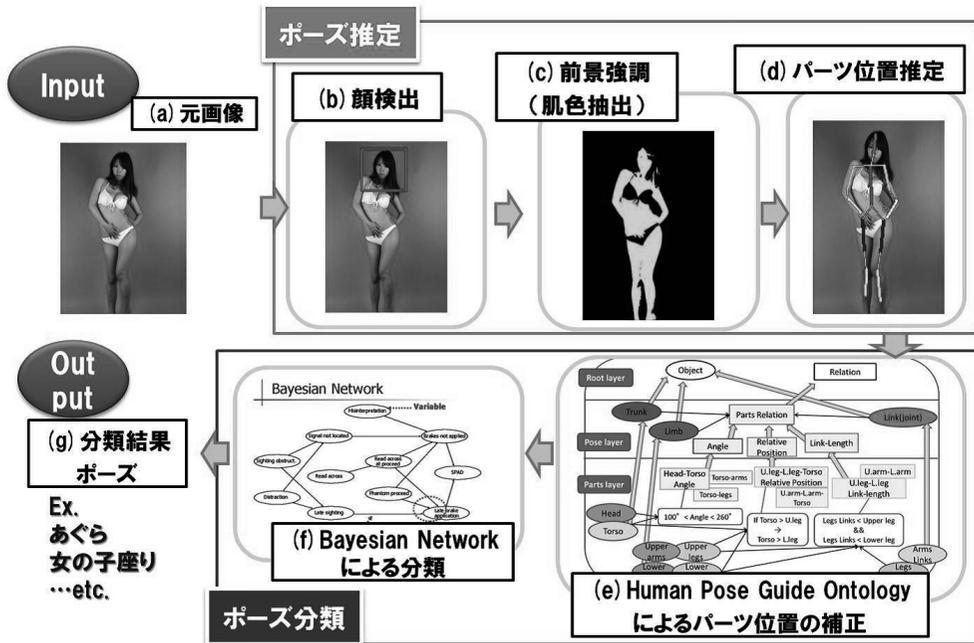


図1 提案システム概要図

システムは Google 画像検索の結果として表示されるアイドル画像をインプットとし、ポーズごとに分類することを目的としている。尚、本研究では対象を女性アイドル画像に限定し、男性アイドル画像は除外している。一般的に男性アイドルに比べて、女性アイドル画像はポーズの種類が多く、自動分類のニーズがより大きい分野であると言える。また、ここでは女性アイドル画像の中でも特に水着の画像を対象としている。これは、アイドルに関連する語彙で Google 画像検索を行った結果、半数以上の画像が水着画像であったためである。

本論文の構成を以下に示す。まず、2章において、提案システム全体の流れについて述べる。次に3章と4章で、提案システムの具体的な処理内容をポーズ推定とポーズ分類に分けて説明する。そして、5章では提案手法の有効性を示すための評価実験を行い、6章でその結果について考察する。最後に7章で本論文のまとめを行う。

本研究で実験・評価に用いた画像はすべて Web 上から収集したものであるが、本論文では著作権等の問

題から筆者らが独自に撮影した画像を掲載する。

2 提案アプローチの概要

本研究の提案システムの概要図を図1に示す。提案システムはポーズ推定とポーズ分類の2つの部分に大きく分けることができる。本研究におけるポーズ推定は、体のパーツ位置、角度、大きさを推定することと定義し、図1の(b)~(d)に相当する。ポーズ分類は図1の(e)~(f)に該当し、ポーズ推定で得られたパーツ情報をもとに、画像を特定のポーズごとに分類する処理を指す。ポーズ推定、ポーズ分類に関しては、それぞれ3章、4章で記述する。提案システムのインプットとなるのは処理がなされていないアイドル画像である(図1(a))。インプット画像に対して、まず顔検出(図1(b))を行い、画像内の人物のおおまかな位置と大きさを把握する(3.2.1節)。次に肌色情報を利用し、前景強調を行う(図1(c), 3.2.2節)。前景部分のみからエッジ情報を抽出した後、パーツ位置推定を行い(図1(d), 3.2.3節)、パーツ情報を取得する。得られたパーツ情報に対して、Human Pose Guide Ontology(以下、HPGO)による補正を行う(図1(e))。

HPGO は人体の各パーツを階層構造で表現し、パーツ間の位置関係や角度などの制約を内包する外部知識である。HPGO に関しては 4.2 節にて詳しく述べる。パーツ情報に補正を加えた上で Bayesian Network による分類を行い (図 1(f), 4.4 節), アウトプットとしてポーズの種類を出力する (図 1(g))。

3 ポーズ推定

3.1 既存研究:Stickman Pose Estimation

人物のポーズ推定に関する研究は数多くなされているが、それらの中でも Eichner ら [3] の研究は精度の高さから注目を集めている。この研究では、人体の上半身を 6 つのパーツ (head, torso, upper arm \times 2, lower arm \times 2) に分割し、それら位置、角度、大きさを推定して、棒人間を用いてポーズを推定している (図 2)。棒人間を用いてポーズを表現するところから Stickman Pose Estimation^{†3}と呼ばれている。Eichner らはこの手法をテレビドラマのシーンに映っている人物のポーズ推定に応用している。テレビドラマでは、登場する人物のポーズはほぼ直立であり、正面を向いていることを想定しているため、Eichner らの手法は正面向き、直立のポーズ推定に特化している。これは処理の過程で用いられるテンプレート情報が正面向き、直立用に作られているためである。本研究では Stickman Pose Estimation 手法を参考に画像内のアイドルのポーズを推定し、分類する。しか

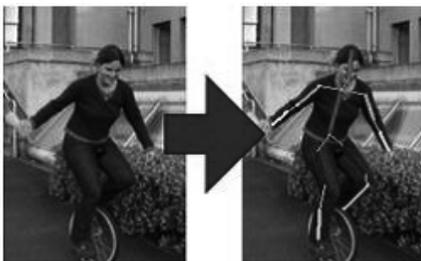


図 2 Stickman Pose Estimation の例 [3]

し、アイドルの画像は正面向き、直立のものばかりではなく、既存手法をそのまま応用することは難しい。そのため、本研究では、Stickman Pose Estimation に改良を加え、正面直立以外のポーズ推定に対応できるようにした。以下では Stickman Pose Estimation に対する改良について述べる。

3.2 ポーズ推定のアプローチ

3.2.1 顔検出

画像内の人物のおおまかな位置と大きさを特定するために、まず顔検出を行う。人物のおおまかな位置を把握することで、明らかに人物が存在し得ない場所を切り捨てて考えることができ、その後の処理においての計算量の削減、パーツ位置の推定精度の向上が見込める。また、検出された顔の大きさをもとに画像内の人物の大きさを推測することにより、被写体が遠くにいても近くにいても普遍的なポーズ推定が可能となっている。先行研究では、顔検出ではなく、頭部、肩のエッジ情報を用いた上半身検出を採用しているが [4]、これは正面直立以外のポーズには適用できない。そのため、本研究では顔検出を採用した。顔検出器は OpenCV^{†4}が提供しているものを使用した。顔検出器は検出した顔の検出窓の座標と大きさを返す。この検出窓を拡大することで人物が存在し得る場所を特定する。

3.2.2 肌色検出を用いた前景強調

人物のポーズ推定をすることは人体の各パーツの位置を推定することを意味する。人体のパーツ位置推定をする際、画像内における人体が存在している場所 (前景) と、存在していない場所 (背景) を分割することは精度向上に大きく貢献する。そのために前景強調処理を行う。Eichner らは、前景強調を行う際、GrabCut アルゴリズム [7] を使用している。このアルゴリズムは画像の 2 値分類に使用され、高い分類精度を持っているが、事前に前景と背景のピクセル情報を入力する必要がある。通常、前景と背景のピクセル情報はユーザーが入力するが、自動で前景強調を行うために前景強調テンプレートが使用されている [5]。

^{†3} Web サイトでは 10 のパーツ (head, torso, upper arm \times 2, lower arm \times 2, upper leg \times 2, lower leg \times 2) に分割してポーズ推定するコードが公開されている。 <http://groups.inf.ed.ac.uk/calvin/>

^{†4} <http://opencv.org/>

この前景強調テンプレートは正面直立に特化しているため、先行研究ではそれ以外のポーズに対する前景強調の精度は落ちる。

図3は先行研究での前景強調テンプレートの使用例である。図3中の凸型部分がテンプレートとなっており、このテンプレート部分に含まれるピクセル情報がGrabCutに対する前景情報の入力となる。通常、凸型の突出した部分と顔の位置が重なるように使用されるが、人物が直立の姿勢でなければ、効果的な前景情報を入力することは出来ない。実際に図3では人物が右に傾いて立っており、テンプレート内には多くの背景部分が含まれ、誤って前景情報としてGrabCutアルゴリズムに入力されている。

そのため、結果として得られた画像である図4では背景である部分も前景として強調されてしまっている(ハイライトされている部分が前景として強調されている部分)。

アイドル画像の多くは正面直立のポーズではなく、テンプレートによる前景強調を本研究の対象領域でそのまま使用することはできない。本研究では対象領域であるアイドルの水着画像の特性を活かして、肌色部分の抽出を行い、それを前景部分として強調する手法をとっている。これにより、直立以外のポーズにも対応が可能となる。

図5は画像から肌色の部分のみを抽出して描画したものである。図6は肌色部分を前景強調した結果であり、テンプレートによる前景強調(図4)よりも高い精度で前景を抽出できていることが分かる。

3.2.3 パーツ位置推定

人体をパーツごとに分割し、各パーツの位置、角度を推定する。本研究では人体を head, torso, upper arm × 2, lower arm × 2, upper leg × 2, lower leg × 2 の合計10個のパーツに分ける。Ramananが提案したImage Parsing[6]手法を使用する(図7)。Image Parsingは人体のエッジ情報と、あらかじめ用意しておいた各パーツのエッジ特徴量から得られるエッジテンプレートを用いて、各パーツ位置の尤度を推定する手法である。3.2.2節で得られた前景部分のみのエッジ(図7:b)を抽出し、エッジテンプレート(c)と組み合わせ、パーツ位置の尤度分布(d)を推定する。推



図3 テンプレート



図4 テンプレートを用いた前景強調



図5 肌色抽出画像



図6 肌色情報を用いた前景強調

定された尤度分布から各パーツの色情報を学習し、記録する。次は色情報を考慮に入れて[2]、より正確な各パーツ位置の尤度(e)を推定し、これを繰り返すことで精度を上げる。結果として人体は棒人間として表現され、10個のパーツの両端の座標が出力される。

4 ポーズ分類

3.2.3節で得られた体の各パーツ位置情報から、画像内の人物のポーズを推定し、分類する処理を行う。しかし、Eichnerら[3]の手法は正面直立のポーズ用に作られているため、それ以外のポーズの場合、棒人間のパーツ位置の推定精度は高くない。本研究では正面直立用の上半身検出の代わりに顔検出を採用し、正面直立用の前景強調テンプレートの代わりに肌色情報を用いた前景強調を行なっているが、アイドルのポーズは複雑なものが多く、精度の低い棒人間のパーツ位置情報をそのまま使ってポーズ分類を行うことはできない。

そこで本研究ではHuman Pose Guide Ontology(以下、HPGO)を用いて、棒人間のパーツの推定位置に補正をかけたうえで、ポーズ分類を行う手法を提案する。Guide OntologyはChenら[1]が提案した手法であり、このオントロジーを外部知識として使

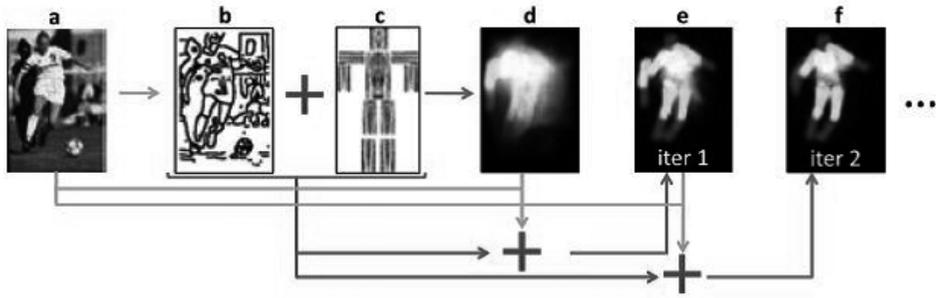


図 7 Ramanan's Image Parsing [6]

用し、画像内のオブジェクト・オブジェクト間のリレーションの推定結果に補正をかけている。本研究における HPGO は Chen らが提案したものに着想を得て、人体のポーズ推定用に独自に考案したものである。Chen らが提案した Guide Ontology と、本研究で提案する HPGO の違いについては、6.2 節で述べる。



図 8 Chen らのオブジェクト・リレーションのネットワーク [1]

4.1 既存研究:Guide Ontology

[1] において Chen らは画像のシーン認識をテーマとしている。テストセットとなる画像内には人やボールなどのオブジェクトが存在し、それらオブジェクト同士の関係性をリレーションとして扱う。これらオブジェクトとリレーションをそれぞれノードとして、関係のあるものはエッジで結合し、画像のシーンそのものをオブジェクト・リレーションのネットワークとして表現している。図 8 では 3 つの Person ノードに SoccerPlayer ラベル、Ball ノードに Soccerball ラベル、2 つの Relation ノードに Kick ラベルが付けられている。シーン認識の手順として、最終的なネットワーク構築の前に、ノードひとつひとつに対して視覚的な特徴量を用いて暫定的なラベリングを行なっている。このときのラベリングはオブジェクトの大きさや色、位置関係などの視覚的特徴をもとに SVM 分類によって行うため、オブジェクト間のセマンティックな関係性については考慮していない。よって、暫定ラベリングの精度次第では、“BasketBallPlayer が SoccerBall を kick している”というネットワークができることもあるこのシーン認識結果のセマンティックな誤りに補正をかける外部知識として Guide Ontol-

ogy が利用される。Guide Ontology は “SoccerBall を kick する Person ノードに付けられるべきラベルは SoccerPlayer である” といった制約条件を内包している。

4.2 Human Pose Guide Ontology

HPGO の概要図を図 9、HPGO が内包する制約を表 1 に示す。HPGO は Root layer, Pose layer, Parts layer に分かれている。Root layer はオブジェクトとオブジェクト間のリレーションが存在する。Pose layer は体幹、肢体、関節部分をオブジェクトとして含み、それらのリレーションがポーズを定義する層となっている。Parts layer がセマンティックなパーツ間のリレーションを定義する部分であり、HPGO における最重要層である。上層の体幹、肢体、関節部分のサブクラスとして各パーツが存在し、それらが相互に持つ関係をリレーションとして定義している。リレーションは各パーツの角度、各パーツ間の相対位置、リンクの長さの 3 つのリレーションに分けられ、それぞれがそれぞれの特徴を持つ。HPGO はこれらのリレーション情報と共に制約を内包し、これらを利用することでパーツ位置推定の結果に補正をかけ

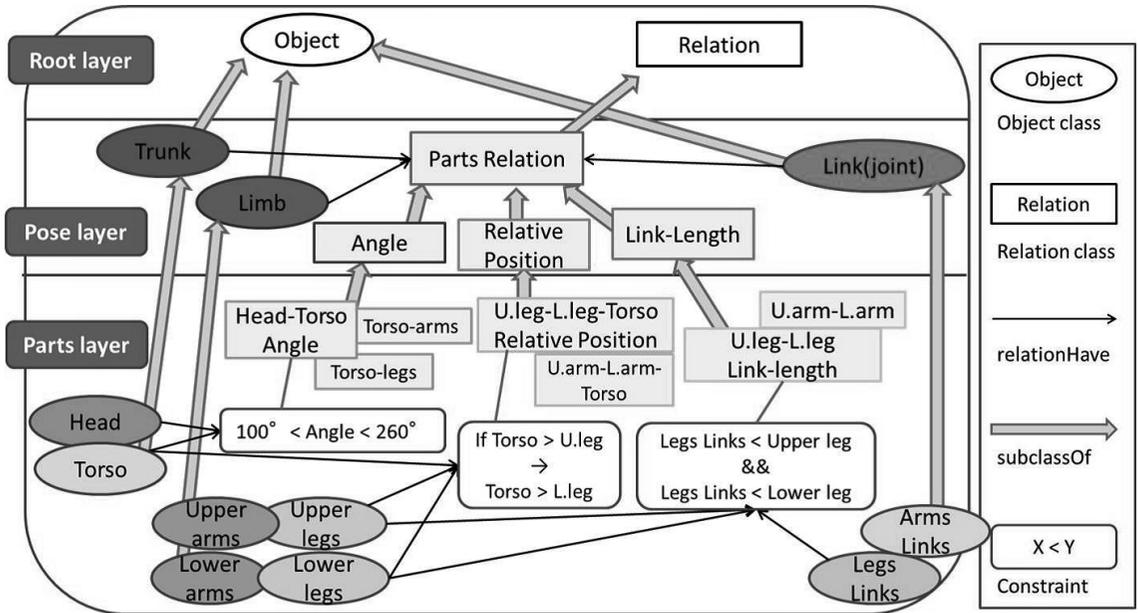


図 9 HPGO 概要図

表 1 HPGO が内包する制約

制約	備考
パーツ総数	パーツの総数は 10 である.
リンク総数	リンクの総数は 9 である.
パーツの繋がり方	torso 上端に繋がるパーツは head, upper arms である.
	torso 下端に繋がるパーツは upper legs である.
	upper arms に繋がるパーツは torso, lower arms である. upper legs に繋がるパーツは torso, lower legs である.
パーツ間の角度 (4.2.1 節)	head-torso 間の角度に関する制約.
	torso-upper arms 間の角度に関する制約.
	torso-upper legs 間の角度に関する制約.
	upper arms-lower arms 間の角度に関する制約.
	upper legs-lower legs 間の角度に関する制約.
パーツ間の相対位置 (4.2.2 節)	torso と upper arms, lower arms の相対位置に関する制約.
	torso と upper legs, lower legs の相対位置に関する制約.
リンクの長さ (4.2.3 節)	upper arms, lower arms の長さ と arms-link の長さに関する制約.
	upper legs, lower legs の長さ と legs-link の長さに関する制約.

る. 補正方法については 4.3 節にて記載する. 以下に HPGO が規定する制約条件について述べる.

4.2.1 各パーツ間の角度における制約

人体の各パーツ間には適切な角度範囲がある. たとえば head と torso はほとんど同じ方向を向っていることが多く, torso が上向き (直立) であるにも関わらず, head が下向きになっているということは考えにくい. そのため, HPGO は “torso から見て (真下方向を 0° として) head の角度は 100° から 260° の間である” という制約を設けている (図 9 内 Head-Torso Angle 以下).

4.2.2 各パーツ間の相対位置における制約

人体の各パーツの位置関係はそれぞれ関連性を持っている. 他のパーツとの相対的な位置と照らし合わせて, 考えにくい位置にパーツが存在するという推定結果が出たとき, その結果は補正されるべきである. たとえば, HPGO は “upper leg が torso よりも低い位置に存在するとき, lower leg が torso よりも高い位置に存在しない” という制約を規定している. 図 9 では, この制約を U.leg-L.leg-Torso Relative Position 以下に記載している.

4.2.3 各パーツを結ぶリンクにおける制約

人体の各パーツを結ぶ線をリンクと呼んでいる。リンクは人体の関節部分を表し、親パーツ (例:upper arm) の終点と、子パーツ (例:lower arm) の始点をつなぐものである。リンクの性質上、正常にパーツ位置推定が行われた場合、リンクの長さは比較的短くなる。よって、HPGO は“リンクの長さはパーツそのものの長さよりも短い”という制約を設けている。図9では例として、upper leg, lower leg, legs link間の制約を U.leg-L.leg Link-length 以下に記載している。

4.3 HPGO による補正

通常の Stickman Pose Estimation によるパーツ位置推定の結果を図10左、HPGOによるパーツ位置補正の結果を図10右に示す。補正は、誤ったパーツの推定位置を適切な位置に移動させることで行われる。パーツの適切な位置を計算するために、(1)肌色情報、(2)人体の左右対称性、(3)HPGOが規定する制約条件の3つを考慮する。これらを反映させるためエネルギー関数 E を導入した。

$$E = E_{skin} + E_{sym} + E_{con} \quad (式1)$$

式1の右辺の項はそれぞれ(1)肌色情報、(2)人体の左右対称性、(3)HPGOが規定する制約条件のエネルギーを表しており、それらの合計が最も低くなる時にパーツ位置が最適化するように定義した。以下ではそれぞれのエネルギー関数について記述した後、具体的な補正方法について示す。

4.3.1 エネルギー関数：肌色情報

パーツの位置として適した場所の条件として、単純に肌色であることが求められる。パーツ位置推定の精度が低く、補正の対象となりがちな upper arms, upper legs, lower arms, lower legs は高確率で肌色が露出しているため、肌色情報のエネルギーはパーツ位置補正において重視される。肌色情報のエネルギーを以下の式で表す。

$$E_{skin} = -\omega_{skin} \cdot \frac{Pixel_{skin}}{Pixel_{all}} \quad (式2)$$

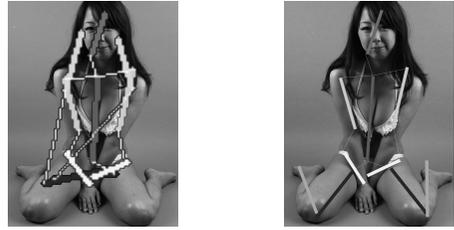


図10 左:補正前 右:補正後

ω_{skin} は重みを表す。 $Pixel_{all}$ はパーツ位置候補周辺のピクセルの全体数、 $Pixel_{skin}$ は、 $Pixel_{all}$ のうち、肌色を表すピクセルの数を表す。したがって、パーツ位置候補周辺のピクセル集合において、肌色ピクセルの数が多きほど、エネルギーは最適化へと向かう。

4.3.2 エネルギー関数：人体の左右対称性

人体の構造は一般的に左右対称である。ポーズ推定の結果としてパーツの位置が左右対称となることはほとんど無いが、torso と upper arms, upper legs 間の関節 (upper arms, upper legs の始点) は左右対称に近い位置をとることが多い。そのため、このエネルギーは upper arms, upper legs のパーツ位置補正時のみに使用される。推定結果の補正は、新しくパーツの始点と終点を求めることであり、始点のおおまかな位置を目安として推測しておくことが補正精度の向上につながる。このような人体の構造の左右対称性を考慮したエネルギーを以下の式で表す。

$$E_{sym} = \omega_{sym} \cdot d(x_i - X) \quad (式3)$$

ω_{sym} は重み、 X はパーツ位置の始点の目安を表す。また、 $d(x_i - X)$ は x_i と X の距離である。始点の目安は上述のように、左右で対となる反対側のパーツの始点と torso をはさんで左右対称の座標をとる。対となる反対側のパーツもまた HPGO の制約に違反している場合は、正常なパーツ位置推定結果の平均をとった座標に始点の目安を置くようにしている。式3はパーツ位置の始点候補 x_i と、始点の目安 X の距離が近いほど、エネルギーは最適化へと向かうことを示している。

4.3.3 エネルギー関数：HPGO が規定する 制約条件

補正を受けて新たに推定されたパーツ位置が制約条件を再び違反することが無いように、このエネルギーを設けている。

$$E_{con} = \begin{cases} 0 & \text{if } part_i \text{ satisfies constraints} \\ \infty & \text{otherwise} \end{cases} \quad (\text{式 4})$$

式 4 は、補正を受け、新たに推定されたパーツ ($part_i$) の位置が制約を違反する場合は、エネルギーを無限大にすることを表している。HPGO の制約を違反することは非常に厳しいペナルティを受けることになり、違反したパーツ位置が採用されることはない。制約を違反しない場合はこのエネルギーに関する影響は受けない。

4.3.4 補正方法

具体的な補正方法について本節で記述する。制約を違反したパーツは、まず始点の目安となる点を定められる。このとき、人体の左右対称性を考慮し、左右で対となる反対側のパーツの始点と torso をはさんで左右対称の座標を目安点とする。この目安点を中心として、5*5 の始点候補点を等間隔に配置する。間隔の大きさは顔検出窓の大きさをもとに決定する。その後、各始点候補点を中心として、仮のパーツ (棒) を 10° ずつ回転させ、すべての始点候補点、角度において仮パーツ周辺の肌色ピクセル数を取得する。回転させる仮のパーツの大きさは torso の大きさをもとに決定する。すべての仮パーツ始点位置と角度の情報が得られた後、式 2 と式 3 にそれぞれの情報を代入し、 E_{skin} と E_{sym} を足しあわせた数値が最も小さいときの仮パーツ始点位置、角度がパーツ位置候補となる。このパーツ位置候補が HPGO の制約を再び違反していた場合、式 4 により、 E_{con} の値が無限大となり、エネルギー関数全体 (式 1) もまた無限大となる。よって、 E_{skin} と E_{sym} の和が次に小さくなるときの仮パーツ始点位置、角度が新しく補正されたパーツ位置として採用される。基本的にパーツ始点位置よりも肌色情報のエネルギー関数を重視しているため、式 2、式 3 内で使用した重み ω_{skin} 、 ω_{sym} の値をそれぞれ 10.0、

0.2 とした。

4.4 Bayesian Network によるポーズ分類

3.2.3 節で得られたパーツの推定位置を 4.3 節で補正し、Bayesian Network に入力する素性を作成する。Bayesian Network に入力する素性は各パーツの位置、各パーツの相対角度の 2 種類に分けられる。各パーツの位置に関しては torso からの距離を利用する。head, upper arms, lower arms の 5 種類のパーツは torso 上端からの距離、upper legs, lower legs の 4 種類のパーツは torso 下端からの距離をもとに素性を作成する。次に各パーツの相対角度については、パーツ間のつながりがある場合、それらのパーツ間の角度をそのまま素性としている。

5 実験

本研究における 2 点の提案の貢献について確認を行った。1 点目は Eichner らの研究を改良し、正面直立以外でもポーズ推定できるようになったこと、2 点目は HPGO による補正を行い、ポーズ分類の精度向上を目指したことである。評価実験はテストセットを用意してパーツ位置を推定し、HPGO による補正をかけたものと、かけないもの両方に対して、Bayesian Network を用いたポーズ分類を行った。画像 1 枚につき、計算時間は 40 秒程度かかる。計算時間の内訳は、画像読み込み・顔検出に約 25%、前景強調に約 35%、パーツ位置推定に約 40% 程度である。パーツ位置推定結果が HPGO による補正を受ける場合はこの時間に加えて 20 秒から 40 秒程度の計算時間がかかる。

5.1 テストセット

アイドルのポーズとして多く見られる、8 つのポーズを分類対象として選定した (女の子座り・寝そべりは右向き・左向きでそれぞれ 1 つのポーズとしている)。テストセットは Web 上の画像を収集したものである。アイドルの名前を検索クエリとして、得られた検索結果の画像から水着画像を選び、ポーズごとに人手で分類した後、収集した画像の鏡像画像を作成し、サンプル数を増やした。今回は各ポーズごとにラン



図 11 全身



図 12 あぐら



図 13 ぺたん座り



図 14 両ひざ立ち



図 15 女の子座り:左



図 16 女の子座り:右



図 17 寝そべり:左



図 18 寝そべり:右

ダムに 20 枚を選出し、計 160 枚をテストセットとしている。Google 画像検索でアイドルに関連する語を検索クエリとして得られた結果画像群のうち、「顔がアップで写っているもの」、「上半身までしか写っていないもの」などを除くと、7 割以上の画像が以下の 8 つのポーズに該当する。残りの 3 割程度の画像はひとつのクラスとしてまとめてしまえるほど共通した特徴を持っていない姿勢が多かったため、今回はテストセットに含めなかった。

1. 全身 (図 11)
2. あぐら (図 12)
3. ぺたん座り (図 13)
4. 両ひざ立ち (図 14)
5. 女の子座り: 左向き (図 15)
6. 女の子座り: 右向き (図 16)
7. 寝そべり: 左向き (図 17)
8. 寝そべり: 右向き (図 18)

5.2 評価

全体の正解率を表 2、各ポーズにおける Precision, Recall, F 値を表 3 に示す。

表 2 テストセット全体に対する正解率

HPGO 補正	正解率
なし	66.3%
あり	78.1%

表 3 ポーズごとの Precision, Recall, F 値

補正	ポーズ	Precision	Recall	F 値
なし	全身	95.2%	100.0%	97.6%
	あぐら	56.3%	45.0%	50.0%
	ぺたん座り	34.4%	55.0%	42.3%
	両ひざ立ち	60.0%	75.0%	66.7%
	女の子座り:左	36.4%	20.0%	25.8%
	女の子座り:右	60.0%	45.0%	51.4%
	寝そべり:左	95.0%	95.0%	95.0%
	寝そべり:右	95.0%	95.0%	95.0%
あり	全身	100.0%	100.0%	100.0%
	あぐら	75.0%	75.0%	75.0%
	ぺたん座り	72.7%	80.0%	76.2%
	両ひざ立ち	61.5%	80.0%	69.6%
	女の子座り:左	58.8%	50.0%	54.1%
	女の子座り:右	81.8%	45.0%	58.1%
	寝そべり:左	86.4%	95.0%	90.5%
	寝そべり:右	90.9%	100.0%	95.2%

HPGO 補正がないものはテストセット全体の正解率が 66.3%、補正が加わったものは 78.1% となり、補正が約 12% の正解率向上につながった。

補正がない場合でも全身、寝そべり:左向き、寝そべり:右向きの 3 種類は精度が高い。それ以外のあぐら、ぺたん座り、女の子座りなどは精度が低く、テストセット全体の精度を下げているが、補正が加わることによって、Precision, Recall とともに向上する。

以下、ポーズごとに精度の高さ、精度向上率などの観点から評価を行う。

5.2.1 全身、寝そべり:左向き、寝そべり:右向き

これらの 3 種類のポーズは補正前から分類精度が高く、補正前と補正後で精度に変化はほとんど無かった。ポーズの特徴として挙げられるのは各パーツの見え方の明瞭さである。これらのポーズでは体のパーツが重なりあうことが少なく、画像内にパーツがしっかりと写っている。そのため、パーツの特徴量を正確に捉えることができ、パーツ位置推定の精度が高い。パーツ位置推定精度の高さが補正前、補正後ともに高い分類精度を実現したと考えられる。

5.2.2 女の子座り:左向き, 女の子座り:右向き

補正前, 補正後を通して精度が低くなったポーズである。補正前の分類精度が低い理由はパーツ位置推定の精度自体に問題があったからである。5.2.1 節の3種類のポーズとは反対に女の子座りは体のパーツが重なりあうことが非常に多く, 画像内にパーツが写っていないことも多い。このとき, パーツ位置推定の精度は著しく下がり, HPGO による補正も機能しづらい。よって精度は他のポーズよりも低い結果となった。

5.2.3 あぐら, ぺたん座り

この2種類のポーズは補正前と比べ, 補正後では大きく精度が向上したポーズである。あぐら, ぺたん座りともに脚部のパーツによって特徴づけられるポーズであるが, 5.2.1 節のポーズに比べると脚部パーツの見え方は明瞭ではない。そのため, 補正前ではパーツ位置の推定精度が低く, それぞれのポーズの特徴を分類器が正確に学習することができなかった。補正後の精度向上の理由としては, HPGO の補正方法と, この2種類のポーズの相性が良かったからだと考えられる。5.2.2 節のポーズとは異なり, パーツ自体が重なりあって一切見えないわけではないため, ピクセルの肌色情報をシンプルに利用した HPGO の補正方法が非常に有効に働いたと言える。

5.2.4 両ひざ立ち

両ひざ立ちは補正の有無に関わらず, 精度は相対的に見て中程度, 補正による精度向上率も小さい。両ひざ立ちのパーツ位置推定の精度は十分に高く, 制約を違反して補正を受けることが少ない。しかし画像によって lower legs の向きが異なっているなど, 一口に両ひざ立ちといっても, その特徴はひとつひとつの画像によって異なっていることが多いポーズであると言える。よって, 分類器にとってはひとつのポーズとしての特徴を学習することが難しく, 精度が上がらなかったと考えられる。

6 考察

本章では, 評価実験結果を受けて, 提案システムの改善案について考察を行う。その後, 既存の Guide Ontology と HPGO の違いについて述べる。

6.1 システム改善案

6.1.1 ポーズ種類の細分化

5.2.4 節で記述したように, 両ひざ立ちという1つのポーズ内でも, 画像によってパーツ位置の特徴が異なっていることがあった。結果として分類器が1つのポーズとしての特徴を捉えることができなかった。これはポーズクラスのカテゴリ粒度が大きすぎることを示唆している。両ひざ立ちに限らず, ポーズの種類をより細分化することで同様の特徴を持ったパーツ位置推定結果が得られ, 分類器の学習が容易となり, 精度が上がる事が考えられる。

6.1.2 画像内人物の肌の色に応じた前景強調

前景強調における肌色抽出はパーツ位置推定を行う前処理として重要であるが, 本研究では事前に肌色ピクセルのサンプルを用意し, 肌色抽出を行っている。この手法では, 画像内人物の肌の色が事前に用意したサンプルから大きく離れている場合, 抽出の精度が下がる可能性がある。この問題を防ぐため, 顔検出で抽出した領域内の肌の部分をベースとし, 動的に肌色サンプルを構築することで対応していくことを考えている。

6.1.3 複数情報を用いたパーツ位置推定のサポート

本研究では Eichner らの Stickman Pose Estimation に改良を加え, パーツ位置推定を行った。しかし, アイドルの多様なポーズに十分に対応できたとは言いがたい。特に, 5.2.2 節で記載した通り, 体のパーツが重なりあい, パーツそのものが画像内に写っていない場合に, パーツ位置推定の精度は非常に落ちる。また, 体のパーツが画像に収まらず, はみ出している場合に関しても同様に精度は著しく下がる。これら問題に対応するため, 今後はパーツの存在する場所に関する事前知識や, 画像の大きさと顔領域の大きさの比率, 肌色の連続性などの情報を導入することにより, 画像内に写っていないパーツの位置を推測するシステムを作ることで対応したい。

6.1.4 新しい制約条件の規定

現時点において, HPGO の制約条件は誤ったパーツ位置をすべて補正することはできていない。しかし, 制約条件を厳格にし過ぎると, 正しく推定された

パーツ位置にも補正をかけることになる。つまり制約条件をただ厳しくすることが補正機能を高めることにはつながらない。改善のためには、個々の制約条件を厳格化するのではなく、複数の制約条件を組み合わせることで更に新しく制約条件を作っていくことが必要である。パーツ位置推定のミス傾向を分析していくことで有効な制約条件を設定することができると考えている。

6.2 HPGO と Guide Ontology の違い

本研究の HPGO は Chen らの Guide Ontology に着想を得て作成したもののだが、目的が異なる。Chen らはノードに対して、SVM 分類によるラベリングを行ったあとに、Guide Ontology による補正を行っているが、本研究では Bayesian Network による分類を行う前に HPGO によるパーツ位置推定結果の補正を行う。つまり、Chen らは分類結果そのものに補正をかけるが、本研究はパーツ位置推定結果に補正をかけ、分類器に入力する素性を作成し、正しい分類結果に導くことを目的としている。分類結果そのものに補正をかける HPGO を作成するには、分類結果に応じて制約条件を規定する必要がある。この場合、ポーズの種類が増えるほど、制約条件もまた増え、対応が困難となる。本研究ではそれを防ぐため、分類結果そのものに補正をかけるのではなく、HPGO の補正によって分類器に入力する素性を作成している。

7 まとめ

本論文では、画像内のアイドルのポーズを推定し、ポーズの種類ごとに分類するエージェントを提案した。提案手法は、人体のパーツの位置、方向、大きさを推定するポーズ推定と、得られたパーツ情報を用いて分類を行うポーズ分類の2段階に分けられる。ポーズ推定は既存研究である [3] の手法を改良し、正面直立以外のポーズにも対応可能となった。加えて、人体構造に関する制約を含む Human Pose Guide Ontology を作成し、パーツ情報の補正を行うことで分類精度向上を目指した。評価実験によって HPGO による精度向上を確認し、本手法の有用性を示した。

今後はさらなるポーズ分類の精度向上と、対応でき

るポーズ種類の拡張に取り組む予定である。そのために、Stickman Pose Estimation, HPGO の両面からアプローチする必要がある。前者に関しては、肌色領域の連続性などを利用して、ポーズ推定をサポートすることを考えており、後者に関しては、新しい制約の規定や、既存の制約の組み合わせを考え、より HPGO の機能を高めていくことを考えている。

謝辞 本研究は JSPS 科研費 24300005, 23500039, 25730038 の助成を受けたものです。本研究を遂行するにあたり、議論・研鑽の場を提供して頂き、御指導頂いた国立情報学研究所/東京大学本位田真一教授をはじめ、研究グループの皆様へ感謝致します。モデルとして御協力頂いた中村有那樣、有限会社クリアツール様に感謝致します。

参考文献

- [1] Chen, N., Zhou, Q.Y. and Prasanna, V.: Understanding web images by object relation network, in *Proceedings of the 21st international conference on World Wide Web, WWW '12*, ACM, 2012, pp. 291–300.
- [2] Eichner, M. and Ferrari, V.: Better appearance models for pictorial structures, in *Proceedings of the British Machine Vision Conference*, BMVA Press, 2009, pp. 3.1–3.11, doi:10.5244/C.23.3
- [3] Eichner, M., Marin-Jimenez, M., Zisserman, A. and Ferrari, V.: 2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images, *Int. J. Comput. Vision*, Vol. 99, No. 2 (2012), pp. 190–214.
- [4] Felzenszwalb, P., McAllester, D. and Ramanan, D.: A Discriminatively Trained, Multiscale, Deformable Part Model, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [5] Ferrari, V., Marin-Jimenez, M. and Zisserman, A.: Progressive search space reduction for human pose estimation, in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.
- [6] Ramanan, D.: Learning to parse images of articulated bodies, in *NIPS'06*, 2006, pp. 1129–1136.
- [7] Rother, C., Kolmogorov, V. and Blake, A.: “GrabCut”: interactive foreground extraction using iterated graph cuts, *ACM Trans. Graph.*, Vol. 23, No. 3 (2003), pp. 309–314.



田代和浩

1989年生。2012年横浜国立大学教育人間科学部卒業。2014年電気通信大学大学院情報システム学研究科修士課程了。同年より(株)サイバーエージェント入社。



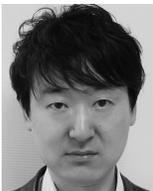
川村隆浩

1969年生。1994年早稲田大学大学院理工学研究科電気工学専攻修士課程了。同年(株)東芝入社。現在、同社研究開発センター主任研究員。2001-2002年米国カーネギ・メロン大学ロボット工学研究所客員研究員兼任。2003年より電気通信大学大学院情報システム学研究科客員准教授兼任。2007年より大阪大学大学院工学研究科非常勤講師兼任。工学博士(早稲田大学)。主としてセマンティック Web, エージェント技術の研究・開発に従事。人工知能学会理事。人工知能学会, 情報処理学会各会員。



清 雄一

1981年生。2009年東京大学大学院情報理工学系研究科博士後期課程修了。同年(株)三菱総合研究所入社。同社情報技術研究センター, 金融ソリューション本部等に所属。2013年より電気通信大学助教。現在に至る。分散コンピューティング, セキュリティ, プライバシ保護技術等の研究に従事。情報処理学会, 電子情報通信学会, IEEE Computer Society 各会員。



中川博之

1974年生。1997年大阪大学基礎工学部情報工学科卒業。同年鹿島建設(株)に入社。2007年東京大学大学院情報理工学系研究科修士課程修了。2008年同大学院博士課程中退。同年電気通信大学助

教, 2014年大阪大学大学院情報科学研究科准教授。現在に至る。工学博士(早稲田大学)。要求工学, 形式手法, エージェントおよび自己適応システム開発手法の研究に従事。情報処理学会, 電子情報通信学会, IEEE CS 各会員。



田原康之

1966年生。1991年東京大学大学院理学系研究科数学専攻修士課程修了。同年(株)東芝入社。1993-1996年情報処理振興事業協会に出向。1996-1997年英国 City 大学客員研究員。1997-1998年英国 Imperial College 客員研究員。2003年国立情報学研究所入所。2008年より電気通信大学准教授。博士(情報科学)(早稲田大学)。エージェント技術, およびソフトウェア工学などの研究に従事。情報処理学会, 日本ソフトウェア科学会会員。



大須賀昭彦

1958年生。1981年上智大学理工学部数学科卒。同年(株)東芝入社。同社研究開発センター, ソフトウェア技術センター等に所属。1985-1989年(財)新世代コンピュータ技術開発機構(ICOT)出向。2007年より, 電気通信大学大学院情報システム学研究科教授。2012年より, 国立情報学研究所客員教授兼任。工学博士(早稲田大学)。ソフトウェアのためのフォーマルメソッド, エージェント技術の研究に従事。1986年度情報処理学会論文賞受賞。IEEE Computer Society Japan Chapter Chair, 人工知能学会理事, 日本ソフトウェア科学会理事を歴任。情報処理学会, 電子情報通信学会, 人工知能学会, 日本ソフトウェア科学会, IEEE Computer Society 各会員。