

人の主観に基づく相対歩容属性の推定

林 優太¹ 楨原 靖¹ 村松 大吾¹ 八木 康史¹

概要: 人の歩行には様々な情報が含まれていることから、人の歩き方の個性（歩容）に基づいた個人認証や、病気の診断といった様々な応用がなされている。本研究では、人が見たときに感じる歩容の綺麗さといったような、人の主観に基づく歩容属性を推定する手法を提案する。大規模歩行映像データベースである OULP-Age より、30 代男性の歩行データを抽出し、その歩行データに適した人の主観に基づく歩容属性として八つの属性を選定した。これらの属性の絶対的な値を、データセットに含まれる全歩行映像に対して一貫性を持って付与することは一般に困難であることから、そのような一貫性の考慮を必要としない、歩行映像組に対する相対的な属性比較に基づくラベル付けを行うことで、歩行映像と相対属性ラベルからなるデータセットを構築した。歩容属性のランキング学習においては、まず、歩容認証の分野で標準的に用いられている歩容エネルギー画像を VGG16 のネットワークに入力して歩容特徴を抽出し、その特徴と相対属性ラベルを用いて、ランキング学習モデルの一つであるプライマルランクサポートベクターマシンを学習した。これにより、歩行映像から各属性のスコアを出力するモデルを構築した。実験では構築したデータセットを用いて属性推定の精度評価を行い、人の主観に基づく歩行による属性の推定の可能性を確認した。

1. はじめに

近年、計算機科学技術、特に、パターン認識・機械学習・画像処理といった人工知能の分野の発展により、人の生体特徴が様々な分野に利用されるようになってきている。生体特徴とは、人間の持つ個人ごとに異なる身体的特徴や行動的特徴を指す。生体特徴には、DNA や指紋、虹彩、顔、音声、歩容（人の歩いている姿）など様々なものが存在し、パスポートコントロールやアミューズメントパーク入場のための顔認証、科学捜査における DNA 鑑定や指紋認証など、幅広く活用されている。中でも人の歩容は他に比べて多様な特徴を内包しており、多方面の分野に利用され始めている。

人の歩容が利用されている分野の一つに、歩容認証が挙げられる [1][2]。歩容認証とは、人の歩き方から個人を特定する認証技術である。認証技術には顔認証 [3] や指紋認証 [4] などいくつかの生体特徴を活用したものが存在するが、いずれも距離が近い場合でないと高い精度を得ることが出来ない。しかし、人の歩容特徴は、他の生体特徴と異なり、対象の人物の協力を必要とせず、離れた位置からでも取得することが出来るという利点を持つことから、防犯カメラの映像のような解像度の低い映像への利用も可能となっている。そのような理由から、昨今、歩容認証の犯罪

捜査への利用が進められている。実際に 2008 年のイギリスにて歩容認証による認証結果が裁判所での証拠として初めて認められ、犯人逮捕へと繋がった事例がある [5]。また、日本でも 2009 年に犯罪捜査で歩容鑑定が使用された例があり、2016 年には裁判にて証拠として採用されたケースも報告されている [6]。

人の歩容が利用されている他の分野として、医療分野も挙げられる。例えば、脳梗塞後に起こる脳性麻痺などにより歩行が困難になる場合があり、そのような歩行障害を改善するためにリハビリテーションを行うが、その際に行う歩容解析の結果を利用して、診断や介助方針の決定を行い、将来的な回復に役立てる試みもなされている [7] [8]。更に、歩容解析は、認知症の早期発見にも活用されている。認知症は初期段階において症状が目に見える形で現れにくいいため、知らずの内に症状が進行するケースが多い。そこで、認知症患者は運動タスクと認知タスクといったような二つのタスクを同時に行うことが困難になるといった報告に注目し、認知機能の評価に歩容の変化の解析を用いることで診断が容易になることが期待されるシステムの開発も行われている [9] [10] [11]。

また、歩き方の持つ特徴の中には、上記のような特定の分野に焦点を当てたもの以外にも、一般大衆が興味を持つような属性といったものも含まれている。例えば、自分が人からどのように見られているのか、つまりある種の審美

¹ 大阪大学 産業科学研究所

的側面は大多数の人にとって一つの大きな関心事である。特に生活する上で欠かせない動作である歩行は、常に他人の目につくため、その良し悪しは、より関心の高いものであると思われる。実際に、近年、ポスチュアウォーキング協会といった、各シーンに応じた綺麗な姿勢の歩き方の指導を行う団体も設立されており、関心の一層の高まりを裏付けている [12]。したがって、歩行映像を解析することで、歩き方の綺麗さなどの人の主観に基づく属性を推定することが可能になると、歩容の美しさの自動判定システムや、綺麗な歩容を身につけるための自主練習システムへの発展が期待されるなど、その意義は大きいものと考えられる。

以上を踏まえ、本研究では、歩容に対して綺麗さなどの人の主観によって定まる見た目の属性推定する手法を提案する。これらの人の主観に基づく属性の絶対的な値を、データセットに含まれる全歩行映像に対して一貫性を持って付与することは一般に困難であることから、そのような一貫性の考慮を必要としない、相対属性によるラベル付けの枠組み [13] を利用した。それにより得られた歩行映像のペアとそれに対する相対属性のデータを用いて、属性のランキングモデルを学習する。これにより、歩行映像から人の主観に基づく属性を推定することが可能となる。

本論文の構成を以下に示す。まず 2 章で本研究の関連研究を紹介する。次に 3 章にてデータ収集のためのアノテーション実験について解説する。4 章では得られたデータを学習する際の手法について説明をし、5 章にて推定実験とその結果を提示する。最後の 6 章で本研究のまとめを述べる。

2. 関連研究

2.1 歩行映像解析

歩容を機械学習の入力データとする場合、歩行映像から歩容特徴を抽出して用いるのが一般的である。そのような歩容特徴表現は数多く考案されており、モデルに基づく表現と見えに基づく表現の大きく二つに分類される。

モデルに基づく表現には、体の各パーツ間の長さの特徴として用いるものや [14]、関節点の位置情報を周期的に取得し、観測方向の変化に対して影響を受けない関節角度の抽出を行うもの [15] などが考案されている。しかし、これらモデルに基づく表現は、低解像度画像に対し精度の良い結果を得ることが困難であることから、見えに基づく表現に分類される手法が主流となっている。

見えに基づく表現には、1 周期のシルエット画像列から時間的なエントロピーを計算した Gait Entropy Image (GEnI) [16] や、時間的に連続する 2 つのシルエットの差から特徴を得る Active Energy Image (AEI) [17]、時間的に連続する 2 枚のシルエット輪郭上のオプティカルフローを算出し 1 歩行周期で平均化する Gait Flow Image (GFI) [18] などが提案されている。上記以外にも、歩行者

のシルエット 1 周期分の画像列を平均化することで得られる Gait Energy Image (以下、GEI と表記する) [19] と呼ばれる手法があり、最も広く用いられている。

これらの歩容特徴表現は歩行映像解析の研究に幅広く使用されている。歩容を行動的生体特徴とみなして個人を識別する歩容認証や、それを防犯カメラ映像を用いた科学捜査に応用する試みもなされている [5] [20] [21]。また、歩容から年齢を推定する研究や、歩容から性別を識別する研究などにも用いられるなど、歩行研究の様々な分野で歩容特徴表現が用いられている [22] [23]。

2.2 属性推定

属性とはその事物が持つ性質のことを指し、画像から機械学習などによって属性を推定する研究や、また、属性を入力としてクラス識別を行う研究等、属性に関連した研究が数多くなされている。そのような属性は、絶対属性と相対属性に大別される。

絶対属性とは、各データに対してある属性に対する絶対的な値を付与する考え方であり、例えば、色の属性や形状の属性、年齢や性別と行った属性に対する直接的な値を付与することが考えられる。Liu らは、顔のパーツを属性として用いて、人が写った風景写真から顔の位置を特定しそのパーツを識別する手法を提示している [24]。Lampert らは、訓練例の無い動物の画像を色や形などの属性を基に 50 種類の動物クラスに分類する手法を提案している [25]。

また、属性に対してその属性の強さを、入力データ対の間で相対的にラベル付けを行う考え方を相対属性と呼ぶ。Parikh らは人の画像や風景の画像に対し、笑顔や自然といった属性にて相対的なラベル付けを行い、それを用いて機械学習を行うことで新規の画像に対して各属性の相対的な強さを予測するモデルを作成している [13]。また、Fan らは大型の放置荷物を発見する用途で使用しており、監視カメラの映像内に映りこむ物体や人に対して静止度等の相対的な属性付けを行うことで誤警報を削減するといった手法を提示している [26]。このように絶対的なラベル付けを行うことが困難な属性に対しても、相対的なラベルであれば付与可能な場合が存在する。

2.3 ランキング学習

ランキング学習とは、情報検索において順位付きモデルを構築する際に用いられる機械学習のことを指す。Cao らは listwise なランキング手法を提案しており、医療分野についてまとめられたクエリに対する検索結果に基づき、情報検索においてペア方式よりもリスト方式の方が優れていると論じている [27]。Burgess らは勾配降下法を用いたランキング学習手法を提案しており、おもちゃの検索問題においてその有用性を提示している [28]。また、ランキング学習は、情報検索分野以外にも積極的に用いられている。

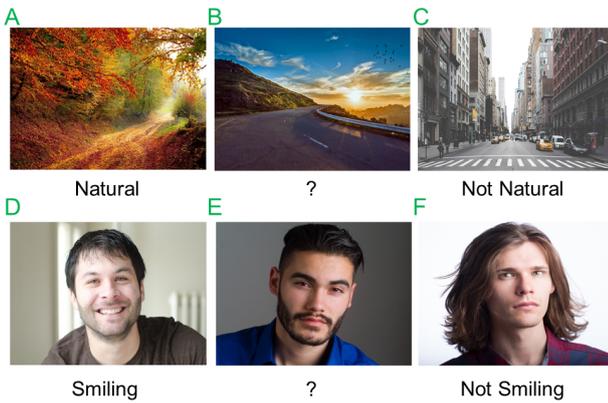


図 1 二種類の属性にて分別された画像

Martin と Xiang は、歩容認証の 1 対 1 認証の問題において、PrimalRank サポートベクターマシンを適用することで、荷物所持状況や服装による影響を低減する試みを行っている [29]. Yang らは、局所性疼痛症候群の患者の歩行解析において、主成分分析 (PCA) とランキング学習を組み合わせて用いている [30]. 他にも医療分野において歩行解析をランキング学習で行っている研究が多数存在しているが、人の主観に基づいた属性に対する学習にはあまり用いられていない。

3. 相対歩容属性のデータ収集

本節ではデータ収集のために実施したアノテーション実験について述べる。

3.1 データ収集の手法

歩容には人によって様々な個性が表れる。それらの歩容に対して、人の主観に基づく「綺麗さ」などを含む様々な属性を設定し、歩行映像からそれらの各属性に対する良さを推定することが、本研究の目的となる。したがって、歩行映像に対して、各属性に対するランキング学習を行う必要があり、そのためには、各歩行データに対する属性のアノテーションが必要となる。本研究では、そのアノテーションにおいて、相対属性の枠組みを用いた。

ここで、相対属性について、概説する。例えば、図 1 のような A-F の 6 枚の画像があるとすると、これらの内、A と C の 2 枚の画像はそれぞれ自然の画像と自然ではない画像としてラベル付けが可能で、D と F の画像は笑顔の画像と笑顔ではない画像にラベル付けが可能である。しかし、B と E の画像を見ると、それぞれ自然の画像であるかそうではないか、画像の中の人物は笑顔であるかそうではないかの 2 値にて、即ち絶対値でのラベル付けを行うことは困難である。仮に実施したとしても、アノテーターによるばらつきにより、安定したアノテーション結果が得られるとは限らない。

そのような場合に有効な考え方が、相対属性である。あ

る属性において、絶対的なラベル付けが困難である画像を相対的にラベル付けすることで、より柔軟な評価を可能とする。図 1 の例で考えると、「B は A より自然ではない」が「B は C より自然である」、「E は D より笑顔ではない」が「E は F よりも笑顔である」といったラベル付けが可能である。

本研究で使用する歩容の属性に関しても、同様に絶対値でのラベル付けが困難であるといえる。仮に、歩容の「綺麗さ」という属性に対して、どのような歩容が綺麗に見えるのかを各歩行映像にラベル付けを行うとする。その際に歩容の綺麗さを数値化して示すのは困難であり、全データに対して一貫した正確なアノテーションは殆ど不可能であるといえる。そのため相対的なラベル付けを行う相対属性の枠組みを採用するに至った。

3.2 提示画像の検討

データ収集において、アノテーターに対し提示する歩行映像としてどのようなものが適しているかを精査することで、精度の高いデータ収集や推定結果の精度向上などが期待できる。ここではアノテーション時に提示する歩行映像についての内容や形式についての検討過程や検討理由述べる。

3.2.1 提示画像内容の検討

人の歩容は、年齢や性別によって大きく異なる。特に成人するまでは体の成長も早く、数歳の違いでも歩容は大きく変わり得る。このような点を踏まえて提示する歩行映像の内容を決定する必要がある。

まず年齢についての検討を行った。機械学習を行う上でデータ数はできるだけ多い方が良いことから、年齢幅を狭めてしまうとアノテーション時に提示できる被験者数が不足する懸念がある。そのため、使用する映像を年代ごとで分割し、使用する年代について検討した。先に述べたように成人するまでは体の成長による変化が大きいことから、成長による要素と属性そのものの要素が混じる恐れがあったため、対象から外した。また、60 代以降は老いによる歩行の差が大きく、やはり属性以外の要素が大きくなることから除外した。最後に、残りの 20 代から 50 代までの年代について、実際に各年代の歩行映像を目視で確認し、最も歩行が安定している 30 代を採択した。

次に、性別についての検討を行った。まず初めに考慮した点が服装である。女性は足元まで隠れるようなロングスカートをはいている可能性があり、また、妊娠による体の厚みや歩容の変化なども鑑みて、やはり、属性そのものの以外の要素が大きく関係することから、男性のデータを用いるものとした。よって、今回の実験には「30 代 男性」の歩行映像を使用する。

3.2.2 提示画像形式の検討

人の主観によるアノテーションを実施するにあたって、

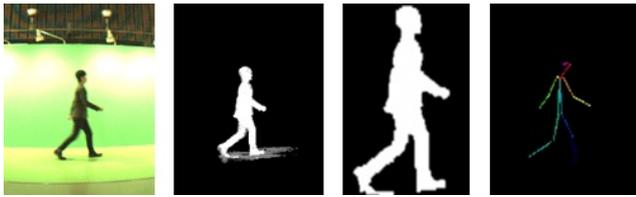


図 2 提示形式の種類. 左から順に (a) 原画像 (b) シルエット画像 (c) 正規化シルエット画像 (d) キネマティクス情報

提示する歩行映像の形式は重要である. 提示する映像形式として図 2 の 4 種類を選考対象とした.

今回の実験では歩容に関する各属性に対する評価を人の主観に基づいて行うため, 歩容評価に関係のない情報を含む映像を使用すると正確なデータが得られない可能性が大きい. そのため, 容姿や服装のセンスなどの良し悪しが各属性の評価に影響を及ぼしてしまう可能性が考えられる. (a) の原画像に関しては, 言わずもがなそれらの情報が含まれているため提示する映像としては不適切であると考えた.

続いて考慮した点が体型である. 容姿や服装のセンスと同様に体の厚みなどによって評価に影響を及ぼすことを懸念し, (d) のキネマティクス情報も候補とした. ここで言うキネマティクス情報とは歩行映像より骨格のみを抽出したものを指す. しかし, 生成した骨格のみの歩行映像を確認したところ, フレーム毎の骨格抽出の精度が必ずしも十分ではなく, 時系列で見た際には不安定であることから, 性格に属性情報を読み乗ることが困難であると考えられた. また, 歩容と体型の要素については, 相互に影響を及ぼすことから, 歩いている際の姿という意味では両者を含めても良いものと考えた. したがって, (d) のキネマティクスを対象外として, (b) のシルエット画像と (c) の正規化シルエット画像の二つを候補として残した. ここで, 正規化シルエット画像は, 原画像サイズのシルエット画像に対して, 人物領域の位置合わせや大きさの正規化を行ったものを指す. アノテーションを行う上で, 映像が正規化されていると比較しやすいといった利点がある一方, 位置合わせと大きさの正規化が行われているため, 歩行者の歩行速度など属性推定に必要な情報が失われてしまっているという欠点もある. 以上より 不必要な情報を削りつつも比較的评价の容易な (b) のシルエット画像をアノテーターに提示するものとした.

3.3 歩容属性の検討

人の主観に基づく歩容属性について, 映像からアノテーターが感じられるかどうかや, 被験者間である程度のバリエーションがあるかどうかといった点を考慮しつつ, 検討した. 結果として, 歩容の全体的な「綺麗さ」を含む, 以下の 8 属性を用いるものとした.

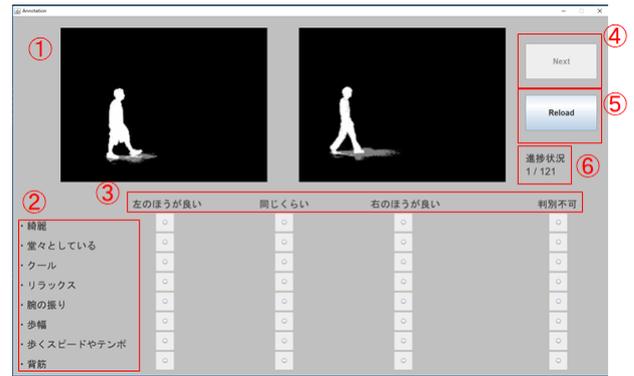


図 3 アノテーションツール

- 綺麗さ
- 堂々とした
- クールな
- リラックスした
- 背筋
- 腕の振り
- 歩幅
- 歩くスピードやテンポ

候補の内, 前半 4 つは歩容全身に関する属性を, 後半 4 つは体の部位や特定項目に関する「綺麗さ」に関する属性を表す. ここで, 各属性値について, 必ずしもその値が大きければ大きいほど良いとは限らない点に注意されたい. 例として「歩幅」を挙げると, 歩行時の歩幅が大きければ良いというわけではなく, アノテーター自身の考える歩容の綺麗さに対して最適な歩幅に近い歩行を行っているかどうかを判断基準とした. 今回の実験では 30 代男性の歩行映像を使用するため, 予備実験でのアノテーターの意見を加味して, 最も適していると考えられた上記の 8 属性を選択した.

3.4 アノテーションツール

歩容による属性のアノテーションのために, 図 3 に示す自作のアノテーションツールを作成した.

本ツールの挙動を概説する. 本ツールを起動すると, 事前に準備された全歩行映像の内, 一対の映像が読み込まれ, ウィンドウの上部 (図 3 の 1) に表示される. アノテーターはこれらの映像を見比べ, 左下に表示される属性 (図 3 の 2) に対して, 「1: 左の方が良い」 「2: 同じくらい」 「3: 右の方が良い」 「4: 判別不可」 (図 3 の 3) いずれかにチェックする (以下総合して「ラベル項目」と呼ぶ). 「2: 同じくらい」では良し悪しにかかわらず同じくらいであればこれを選択する. 「4:判別不可」は上記の 1-3 のいずれとも判断しがたい (判断する自信が無い) と感じた場合にのみを選択する. 以上を属性の個数の分だけ行い, "Next" ボタンを押下することで次の歩行映像へと遷移する (図 3 の 4). "Next" ボタンは全ての属性について回答が終了しない限りグレーアウトしているため, データが不足した状態で次に進むことはできないようになっている. また, 動作環境などにより歩行映像が止まってしまう場合がある.

その場合は "Reload" ボタンにより映像の再読み込みが可能となっている (図 3 の 5). "Reload" ボタンの下には進捗状況が表示されており, 全体の映像ペア数の内どの程度が終了しているかを確認することができる (図 3 の 6).

3.5 データ収集実験

3.5.1 データ収集実験の設定

データセットとして, 30 代男性の歩行映像を 1,200 データを用意し, それぞれを二回ずつ使用し, 同じペアが出来ないよう 1,200 ペアを作成した. アノテーターの人数は 9 人で, 属性は 3.3 節にて最終決定した 8 個を使用した. これにより 1,200 データ \times 8 属性の合計 10,800 ペアのデータセットを構築した. アノテーションは 3.4 節にて説明しているツールを使用した. また, アノテーターは各人で適宜選択した時間帯で教示作業を行い, 概ね 3 週間以内に全データの アノテーションが終了するように依頼した.

4. 相対歩容属性の推定

構築した歩行映像と相対属性のデータセットを学習することで, 歩行映像から属性を推定するランキング学習を行う. 本節では, そのための各手順について説明する.

4.1 特徴抽出

まず, 歩行映像よりシルエット画像列を抽出し, 人物領域の位置合わせやサイズの正規化をすることで, 88 画素 \times 128 画素の正規化シルエット画像列を取得する. ついで, 正規化シルエット画像列の時間方向の正規化自己相関最大にする時間シフト量を歩行周期として算出し, 正規化シルエット画像列を時間方向に足し合わせて歩行周期で平均化した GEI を生成する.

更に, 得られた GEI に対して, 深層学習による特徴抽出を行う. この特徴抽出には 物体認識等の分野で標準的に用いられているネットワーク構造である VGG-16 [31] を使用する. VGG-16 は, 畳み込み層が 13 層と全結合層が 3 層の合計 16 層からなる畳み込みニューラルネットワークである. VGG-16 は, 元々, 物体認識等に用いられるネットワークであるが, 前半の畳み込み層 13 層は ImageNet [32] と呼ばれる大規模な画像データセットにより学習されているため, 画像に特化した特徴抽出器としても使用することができる. 本研究では, GEI を入力として, VGG-16 の畳み込み層を利用して特徴抽出を行い, 出力された特徴ベクトルを, ランキング学習のための入力特徴量として用いる.

4.2 ランキング学習

本実験では, 4.1 節にて説明した VGG-16 によって抽出された特徴と教示された各相対属性ラベルを用いて, ランキング学習モデルの一つであるプライマルランクサポートベクターマシン (Primal Rank Support vector Machine;

Primal Rank SVM) を学習した. これにより, テストデータの歩容特徴を入力することで, 各属性のスコアを出力することが可能となる. 以下に Primal Rank SVM の手法の説明を行う.

画像から特徴を抽出した特徴ベクトルの集合を \mathbf{x}_i とし, 作成された画像ペアに対し, 各属性ごとに画像の順序付きペアの集合 O_m と順序なしのペアの集合 S_m を学習データとして設定する. 即ち, O_m が属性 m における学習画像間の強弱関係集合で, S_m が属性 m が同じ強さをもつ学習画像ペアの集合を示している.

これらの学習データを使用して以下の式を満たすランキングスコア r_m を算出するために, 重み \mathbf{w}_m を学習することが最終的な目標となる.

$$r_m(\mathbf{x}_i) = \mathbf{w}_m^T \mathbf{x}_i \quad (1)$$

その際, 以下の条件を満たすようにする必要がある. 即ち, 強弱関係のある画像間ではランキングスコアに差が生まれ, 同等の強さを持つ画像間ではスコアに差が生まれないようにしなければならない.

$$\forall (i, j) \in O_m : \mathbf{w}_m^T \mathbf{x}_i > \mathbf{w}_m^T \mathbf{x}_j \quad (2)$$

$$\forall (i, j) \in S_m : \mathbf{w}_m^T \mathbf{x}_i = \mathbf{w}_m^T \mathbf{x}_j \quad (3)$$

以上を踏まえて目的関数は以下ようになる. C は誤りをどれほど許容するかというパラメータ定数であり, この数値が小さいほど誤判別を許容する. また, ξ と γ は制約条件の緩和具合を表している.

$$\text{minimize } \left(\frac{1}{2} \|\mathbf{w}_m^T\|_2^2 + C \left(\sum \xi_{ij}^2 + \sum \gamma_{ij}^2 \right) \right) \quad (4)$$

$$\text{s.t. } \mathbf{w}_m^T (\mathbf{x}_i - \mathbf{x}_j) \geq 1 - \xi_{ij}; \forall (i, j) \in O_m \quad (5)$$

$$|\mathbf{w}_m^T (\mathbf{x}_i - \mathbf{x}_j)| \leq \gamma_{ij}; \forall (i, j) \in S_m \quad (6)$$

$$\xi_{ij} \geq 0; \gamma_{ij} \geq 0 \quad (7)$$

これは SVM での分類と類似している. SVM では正例と負例がどの程度離れているか, 即ち, マージンがなるべく大きくなるような境界を求めますが, この手法ではすべての順序ペアに関するランキングスコアの差のうち, 最小のランキングスコアの差を最大化するように境界を求めるという考え方で学習を行う (図 4).

図 4 のように重みベクトル $\mathbf{w}_1, \mathbf{w}_2$ が存在している場合, \mathbf{w}_1 では最小のランキングスコアは \mathbf{x}_1 と \mathbf{x}_3 間で計算される d_1 で, \mathbf{w}_2 では最小のランキングスコアは \mathbf{x}_2 と \mathbf{x}_4 間で計算され d_2 となる. d_1 と d_2 を比較すると d_1 のほうがマージンが大きいので, 重みベクトル \mathbf{w}_1 によるモデルのほうが優れているといえる.

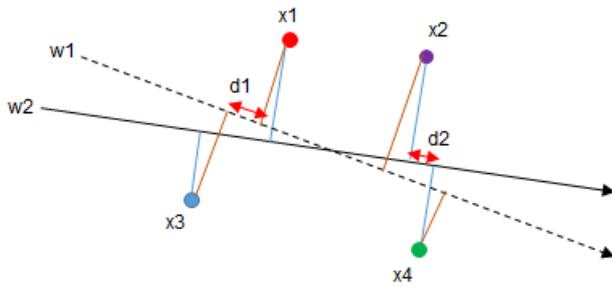


図 4 マージンの例

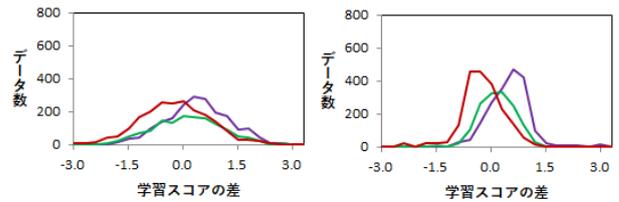


図 6 堂々とした:スコア分布 (左) テストデータ (右) 学習データ

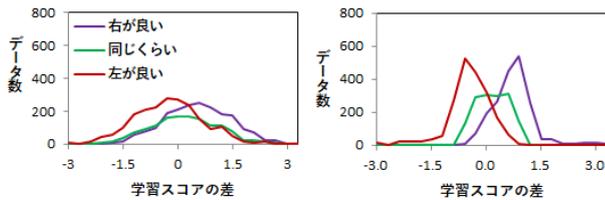


図 5 綺麗さ:スコア分布 (左) テストデータ (右) 学習データ

表 1 各属性の精度

属性	精度 (%)		
	テストデータ	学習データ	閾値
綺麗	60.5	95.8	0.09
堂々とした	59.0	96.0	0.07
クール	46.5	97.8	0.08
リラックス	55.7	99.0	0.07
腕の振り	50.2	95.5	0.24
歩幅	38.2	98.5	0.23
速さとテンポ	42.3	98.0	0.37
背筋	56.8	96.0	0.35

5. 実験

5.1 実験設定

本実験では、データ収集実験にて相対的なラベル付けを行った各歩行映像それぞれに対して属性ごとのスコアを算出する。スコアが大きい歩容程、その属性に対する評価が高いものとなる。

また、データ収集実験にて収集したデータは、学習データ 5,400 データとテストデータの 5,400 データに分割し、学習データのみをランキング学習に使用する。学習データとテストデータには、ID の一致する歩行映像を含めてもよいものとするが、厳密性のため、比較するペアは同じものを含まないものとする。

5.2 定量評価

本実験の結果を定量的に分析した内容を以下に述べる。

5.2.1 スコア分布

まず、精度評価を行う前に、属性に対する推定スコアの分布を、アノテーション結果毎に確認する。ここでは、各属性の内「綺麗さ」「堂々とした」を取り上げ、歩行映像ペアそれぞれから推定されたスコアの差を算出し、そのスコアの差をアノテーション結果毎に可視化した (図 5, 6)。より具体的には、「右が良い」「同じくらい」「左が良い」という三種類のアノテーション毎に、横軸を「学習スコアの差 = 右側の歩行映像の推定スコア - 左側の歩行映像の推定スコア」とした場合のヒストグラムを示す。図より、学習データに対してはアノテーション毎にある程度分離された分布となっている一方で、テストデータに対しては汎化性能が低下していることから、分布が重なっていることが読み取れる。

5.2.2 閾値の決定と精度評価

精度評価を行う上で、歩行映像ペアに対するスコア差の絶対値がある閾値より小さい場合に、システムとしては「同じくらい」として判断することとなる。そこで、そのハイパーパラメタである閾値 Δ を決定するため、学習データに含まれる各歩行映像ペアに対して、複数のアノテーターが付与した相対属性値 (右が良い: +1, 同じくらい: 0, 左が良い: -1) に対する平均値を計算し、その値が閾値 Δ よりも大きければ「右が良い」、その値が閾値 $-\Delta$ よりも小さければ「左が良い」、それ以外であれば「同じくらい」が真値であると見なす。一方で、推定スコアの差の方についても、同様に、閾値 Δ に基づいて「右が良い」「同じくらい」「左が良い」の判定を行い、真値と一致していれば正解、そうでなければ不正解として、精度を算出する。その上で、学習データに対して、その精度が最大となる閾値を採用し、それをテストデータにも適用するものとする。属性毎に得られた閾値とテストデータ・学習データに対する精度を表 1 に示す。結果として、学習データに対しては 95% 以上の精度で識別できている一方、テストデータに対しては汎化誤差が大きいことから、精度低下が見られる。

5.3 定性評価

本実験の結果を定性的に分析した内容を以下に述べる。

図 7 では各属性ごとに結果のスコアが最も高いものと最も低いものを 3 個ずつ載せている。形式は原映像から抽出した正規化シルエットを、1 周期分抽出したものとなる。

5.4 考察

5.4.1 定量評価に対する考察

歩行映像ペアに対する推定スコアの差の分布を見ると、

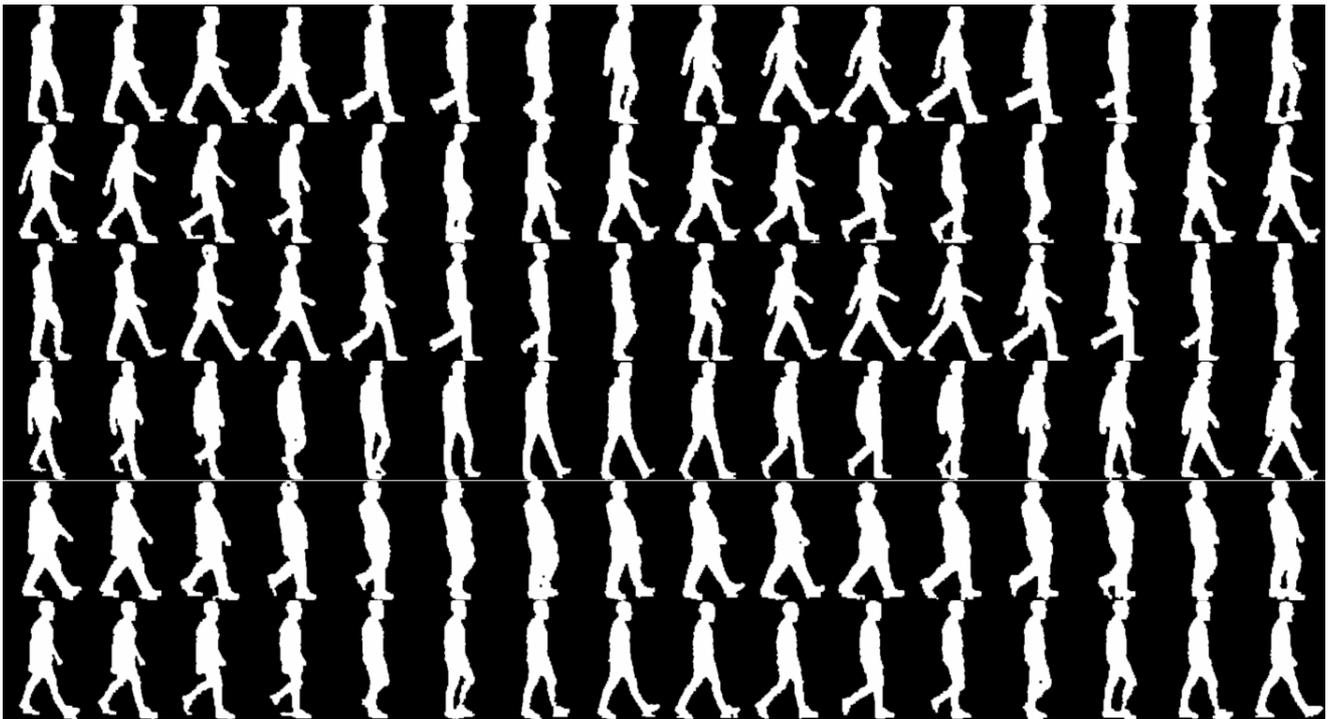


図 7 綺麗さ：スコアの高い歩行(上3つ)/低い歩行(下3つ)

テストデータに対しては、「右が良い」「同じくらい」「左が良い」のアノテーション項目のいずれに対しても分布の広がりが大きく、分布間の重なりも大きい。また、学習データに対しては、テストデータよりは分布が分離されているものの、属性によっては「同じくらい」の項目の分布が「右が良い」「左が良い」の二つ項目の分布と重なっているものも見受けられた。これは、アノテーションの時に比較する歩行映像ペアの間に、当該属性に関して人が明確に知覚することのできる差がない場合に起こる。即ち判別が比較的難しい属性では、アノテーションの結果にばらつきが起きやすいと言いつ換えることもできる。例えば、「クール」や「リラックス」といった、評価が比較的難しい属性では、各ラベル項目のグラフが重なっている部分も多く、アノテーションデータに対してよい学習スコアが出力されているとは言い難い結果となった。「綺麗さ」や「背筋」など、人にとって評価基準が比較的分かりやすいものに関しては他の属性に比べ、比較的良いスコア分布が出力されていることから、属性によって分布の差が大きくなった。

テストデータと学習データの分布形状や精度を比較すると、両者の間に相当の差があることから、汎化誤差は大きいものといえる。よって、今後、学習データ数を増やすことで、汎化誤差を低減することが望まれる。

5.4.2 定性評価に対する考察

「綺麗さ」に対する歩行を観察すると、推定スコアの高い歩行例はスコアの低い歩行例に比べて「腕の振り」が大きく、ある程度「背筋」が伸びていることが分かる。また、30代の男性ということもあり、少し堂々としたよう

な歩行をしているように見える。属性「堂々とした」を見ると、腕を大きく振り、一步一步の歩幅が大きい歩行が高スコアとなっていることが多い。また、「堂々とした」と「歩幅」や、「綺麗」と「背筋」でスコアの低い歩行として同一の歩行映像がそれぞれランクインしている傾向がみられ、以上の各属性の関連性は大きいと考えられる。「クール」では比較的体のパーツが華奢な歩行者の映像のスコアが高くなっていることが多いことから、体型による歩容の差異が強く反映されている可能性がある。「速さとテンポ」と「リラックス」に関しては、良いスコアと悪いスコアの歩行映像を観察したが、明確な差を見つけることが出来なかった。これは、アノテーション自体が困難な属性であることに起因するものと考えられる。

6. おわりに

本論文では、歩行映像に対して人の主観に基づく属性の推定を行った。属性としては、「綺麗さ」「堂々とした」「クール」「リラックス」「腕の振り」「歩幅」「速さとテンポ」「背筋」の8種類を用い、相対属性の枠組みで、歩行映像ペアに対するアノテーションを実施した。ランキング学習モデルの入力には、GEIをVGG-16の畳み込み層に通した特徴を利用し、Primal Rank SVMにより属性スコアの推定値を出力した。実験では、歩容から人の主観に基づく属性を推定できるという可能性を示した。

今後は特徴抽出や学習アルゴリズムを歩容に適したものへと変更し、精度改善を図る予定である。また、人の主観による属性データの収集方法をさらに精査し、評価する属性

の数や質, 提示の仕方などを考えて, 今後の追加データ収集を実施予定である. また, 検証方法としては, より多くの学習データを利用できる, Leave-one-out cross validation (LOOCV) 法を用いて, 精度評価を行う予定である.

謝辞 本研究はJSPS 科研費 JP18H04115 の助成を受けたものです.

参考文献

- [1] Mark S. Nixon, Tieniu N. Tan, and Rama Chellappa. *Human Identification Based on Gait (The Kluwer International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [2] Yasushi Makihara, Darko S Matovski, Mark S Nixon, John N Carter, and Yasushi Yagi. Gait recognition: Databases, representations, and applications. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–15, 1999.
- [3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, Vol. 35, No. 4, pp. 399–458, December 2003.
- [4] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.
- [5] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, Vol. 56, No. 4, pp. 882–889, 2011.
- [6] 黒沢健至. 防犯カメラ映像の解析技術. セイフティ エンジニアリング, Vol. 185, pp. 21–25, Dec. 2016.
- [7] James R Gage. Gait analysis. an essential tool in the treatment of cerebral palsy. *Clinical orthopaedics and related research*, No. 288, pp. 126–134, 1993.
- [8] David H Sutherland and Jon R Davids. Common gait abnormalities of the knee in cerebral palsy. *Clinical orthopaedics and related research*, No. 288, pp. 139–147, 1993.
- [9] Y. Yagi, I. Mitsugami, S. Shioiri, and H. Habe. Behavior understanding based on intention-gait model. *Human-Harmonized Information Technology, Springer*, Vol. 2, pp. 139–172, 2017.
- [10] 満上育久, 周成菊, 丹羽真隆, 大倉史生, 八木康史. 高齢者認知機能推定のためのデュアルタスク解析. 第 60 回システム制御情報学会研究発表講演会, 2016.
- [11] K. Aoki, T. T. Ngo, I. Mitsugami, F. Okura, M. Niwa, Y. Makihara, Y. Yagi, and H. Kazui. Early detection of lower mmse scores in elderly based on dual-task gait. *IEEE Access*, Vol. 7, pp. 40085–40094, 2019.
- [12] ポスチュアウオーキング協会 HP. <http://www.posture.co.jp/>.
- [13] D. Parikh and K. Grauman. Relative attributes. In *2011 International Conference on Computer Vision*, pp. 503–510, Nov 2011.
- [14] A. F. Bobick and A. Y. Johnson. Gait recognition using static, activity-specific parameters. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 1, pp. I-423–I-430 vol.1, 2001.
- [15] N. Spencer and J. Carter. Towards pose invariant gait reconstruction. In *IEEE International Conference on Image Processing 2005*, Vol. 3, pp. III-261–4, Sept 2005.
- [16] S. Lee, Y. Liu, and R. Collins. Shape variation-based frieze pattern for robust gait recognition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.
- [17] Erhu Zhang, Yongwei Zhao, and Wei Xiong. Active energy image plus 2dlpp for gait recognition. *Signal Processing*, Vol. 90, No. 7, pp. 2295–2302, 2010.
- [18] Toby H.W. Lam, K.H. Cheung, and James N.K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, Vol. 44, No. 4, pp. 973 – 987, 2011.
- [19] Ju Man and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322, Feb 2006.
- [20] Niels Lynnerup and Peter Kastmand Larsen. Gait as evidence. *IET Biometrics*, Vol. 3, No. 2, pp. 47–54, 6 2014.
- [21] H. Iwama, D. Muramatsu, Y. Makihara, and Y. Yagi. Gait verification system for criminal investigation. *IPSSJ Trans. on Computer Vision and Applications*, Vol. 5, pp. 163–175, Oct. 2013.
- [22] Jiwen Lu and Yap-Peng Tan. Gait-based human age estimation. *IEEE Transactions on Information Forensics and Security*, Vol. 5, No. 4, pp. 761–770, 2010.
- [23] 万波秀年, 榎原靖, 八木康史. 歩容における性別・年齢の分類と特徴解析. 電子情報通信学会論文誌 D, Vol. 92, No. 8, pp. 1373–1382, 2009.
- [24] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [25] C. H. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 3, pp. 453–465, March 2014.
- [26] Quanfu Fan, Prasad Gabbur, and Sharath Pankanti. Relative attributes for large-scale abandoned object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.
- [27] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning, ICML '07*, pp. 129–136, New York, NY, USA, 2007. ACM.
- [28] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22Nd International Conference on Machine Learning, ICML '05*, pp. 89–96, New York, NY, USA, 2005. ACM.
- [29] Ral Martn-Flez and Tao Xiang. Uncooperative gait recognition by learning to rank. *Pattern Recognition*, Vol. 47, No. 12, pp. 3793 – 3806, 2014.
- [30] M. Yang, H. Zheng, H. Wang, S. Mcclean, and N. Harris. Combining feature ranking with pca: An application to gait analysis. In *2010 International Conference on Machine Learning and Cybernetics*, Vol. 1, pp. 494–499, July 2010.
- [31] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. 2014.
- [32] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, June 2009.