

頭上カメラから観測された身体動揺を用いた性別認識の精度評価

Evaluation of gender classification accuracy using body sway observed from overhead camera

神谷 卓也[†], 山口 優太[†], 中谷 真太郎[†],
西山 正志[†], 岩井 儀雄[†]

Takuya Kamitani[†], Yuta Yamaguchi[†], Shintaro Nakatani[†], Masashi Nishiyama[†] and Yoshio Iwai[†]

あらまし 本論文では、頭上カメラで撮影された映像中の身体動揺を用いることで、立ち止まる人物の性別を識別できるかどうかについて評価する。従来より性別認識において、歩容の映像から抽出された特徴量が用いられている。一方、立ち止まる人物の動きについては、性別認識において、これまで考慮されてこなかった。立ち止まる人物の性別を識別するため、我々が過去に提案した Local Movement (LM) 特徴量を、身体動揺が撮影された映像から抽出する。本論文の目的は、LM 特徴量と識別器との組合せによって、性別を識別できるかを評価することにある。ここでは識別器として線形 SVM を用いた。男女の身体動揺の映像データセットを構築し、性別認識の精度を評価した。実験結果から、LM 特徴量と線形 SVM とを組合せた本評価対象手法は、他の手法と比べて、立ち止まる人物の性別を、身体動揺の映像から高精度に識別できることを確認した。

キーワード：性別認識，身体動揺，頭上カメラ，時空間特徴量

1. ま え が き

カメラ映像を用いて人物の性別を識別する技術¹⁾²⁾が必要とされている。人物の性別を正しく識別することで、セキュリティ強化やマーケティング立案の応用が期待されている。男女を見分ける手掛かりを獲得するため、カメラ映像から観測される人物の動きを利用することが考えられている。

人物の動きは、歩く時と立ち止まる時のそれぞれで発生する。歩く人物のカメラ映像からは、腕振りや足運びによる動きが観測される。この動きは歩容と呼ばれている。一方、立ち止まる人物のカメラ映像からは、身体の微細な動きが観測される。この動きは身体動揺と呼ばれている。以下では、カメラ映像中のそれぞれの動きを用いて性別を識別する場合について議論を進める。

まず、歩容を撮影したカメラ映像から、性別を識別する場合を考える。歩容から男女を見分けるために、既存研究^{3)~5)}では、Gait Energy Image (GEI) 特徴量を抽出している。この GEI 特徴量と識別器とを組合せることで、歩く人物の性別を精度よく識別できることが報告されている。ただし、これらの既存研究は、歩く人物のみを対象としていた。次

に、身体動揺を撮影したカメラ映像から、性別を識別する場合を考える。身体動揺から男女を見分けるための特徴を抽出する研究は、我々が調べた限りでは存在しない。よって本論文では、カメラ映像中の身体動揺から特徴を抽出し、その特徴を性別認識に利用できる可能性があるかについて焦点をあて議論を進める。

身体動揺に男女を見分ける特徴が含まれているかどうかを考える。医学の分析研究^{6)~8)}では、身体動揺を表すパラメータの中には、男女の間で統計的有意差が表れるものが存在することが報告されている。得られたパラメータを、性別認識にそのまま適用することも考えられるが、接触型の圧力センサを床に埋め込む制約があった。

非接触センサの一つであるカメラを用いて、立ち止まる人物から身体動揺を観測することを考える。文献^{9)~13)}では、転倒防止や映像生成や個人識別を目的として、カメラ映像から身体動揺を観測する応用例が報告されている。ただし、これらの文献では、カメラ映像から観測された身体動揺に含まれる特徴が、性別を識別できる可能性を持つかどうかについて評価されていなかった。

そこで本論文では、我々が過去に提案した LM 特徴量¹³⁾と識別器とを組合せた手法を用いることで、身体動揺のカメラ映像から性別を識別できるかどうかを評価する。本評価対象手法では、頭を中心とした上半身を人物領域とし、そのシルエット映像を取得する。身体動揺の男女差を表す特

2020年8月17日受付, 2021年2月7日再受付, 2021年3月18日採録

[†]鳥取大学大学院 工学研究科

(〒 680-8552 鳥取市湖山町南4丁目101番地, TEL 0857-31-5183)

微量を、シルエット映像の時空間変化から抽出する。その特徴量と識別器とを組合せることで、性別認識の精度を獲得する。独自に収集した男女の身体動揺の映像データセットを用いた実験により、本論文の評価対象手法は、身体動揺の微細な動きの特徴を捉え、高い識別精度を達成することを確認した。さらに、本評価対象手法と、医学知見に基づく手法との間で、性別認識の精度を比較した。医学知見に基づく手法では、統計的有意差が見られるパラメータを、カメラ映像から特徴量として抽出し、その特徴量と識別器とを組合せた。実験結果から、本評価対象手法の精度が優れていることを確認した。

2. 関連研究

2.1 性別認識の既存手法

カメラ映像中の人物の性別を識別するために、前節で述べた動きの特徴を用いる場合を考える。文献^{3)~5)}では、歩容の動きから抽出された GEI 特徴量と識別器とを組合せる手法が提案されている。この GEI 特徴量¹⁵⁾とは、歩容の 1 周期から観測される腕や足の動きを、人物シルエットに加算したものである。文献³⁾では、GEI 特徴量をそのまま性別認識に適用している。文献⁴⁾では、歩容の 1 周期を、時間方向の 4 つの区間に分割している。それぞれの区間から GEI 特徴量を抽出し、性別認識に適用している。文献⁵⁾では、腕の動きと足の動きが、性別認識の精度へ寄与する割合が異なると仮定している。腕の動きを表す GEI 特徴量と足の動きを表す GEI 特徴量を算出し、それぞれの特徴量に重みを割り当てている。これらの既存手法^{3)~5)}は、腕や足が歩行で大きく動くことを前提としている。ただし、立ち止まる人物の身体動揺からは、腕や足の大きな動きを観測することは難しい。LM 特徴量¹³⁾と識別器とを組合せた本評価対象手法が、GEI 特徴量と識別器とを組合せた手法に比べ、性別認識の精度が高くなることを確認する。その詳細を 4.2 節の GEI との比較結果で述べる。

人物の動き以外からでも、性別を見分ける特徴を抽出することは可能である。その一つの特徴として、静止画に含まれる人物の見え方そのものが挙げられる。以下では、人物の見え方の特徴を用いて、性別認識を行う手法について述べる。既存手法^{16)~18)}は、静止画中の色や勾配を特徴量として抽出し、それらの特徴量と識別器とを組合せることで性別を識別している。既存手法^{19)~22)}は、深層学習の一つである Convolutional Neural Network (CNN) を用いることで性別を識別している。これらの既存手法は、大量の静止画から訓練できれば高い精度が得られる利点をもつ。見え方と動きの両方を手掛りとする LM 特徴量と識別器とを組合せた本評価対象手法が、見え方のみを手掛かりとする一般的な手法に比べて、性別認識の精度を高めることを確認する。その詳細を 4.2 節の CNN との比較結果で述べる。

2.2 身体動揺の性差に関する分析研究

身体動揺に性差が表れることを明らかにするため、医学

分野において分析研究^{6)~8)}が報告されている。これらの分析研究では、足元に埋め込まれた圧力センサから身体動揺を観測している。圧力センサから出力された重心位置の時系列データにおいて、性差が表れるパラメータを設計し統計的に有意差を検定している。時系列データの周波数特性と軌跡のパラメータ⁶⁾、時系列データの軌跡を楕円近似した時のパラメータ⁷⁾、時系列データの周波数特性の特定帯域パラメータ⁸⁾に、性差が表れることが報告されている。ただし、これらの分析研究は、得られたパラメータの知見を性別認識へ応用することを考慮していなかった。医学知見で得られた周波数特性や軌跡のパラメータをカメラ映像から抽出し、性別認識の特徴量として用いた場合、それらのパラメータ自体の種類が少ないため、高い精度を得られない可能性がある。これを確認するため、医学知見に基づく特徴量と識別器とを組合せた手法の精度について評価した。その詳細を 4.4 節で述べる。

2.3 カメラ映像から観測される身体動揺の応用例

立ち止まる人物のカメラ映像から身体動揺を観測し、様々な目的で応用する例^{9)~13)}が報告されている。文献⁹⁾では、転倒防止のリスク推定を行うため、身体動揺が応用されている。複数のカメラを用いて様々な方向から人物を撮影し、その人物の 3 次元的重心位置の時系列データを観測している。文献¹⁰⁾では、人とインタラクションを行うアバタ映像を生成するため、身体動揺が応用されている。正面に設置されたカメラを用いて人物を撮影し、揺れの中心となる時刻を時系列データから推定している。文献¹¹⁾¹²⁾では、臨床現場で身体バランスを調査するため、身体動揺が応用されている。マイクロソフト社の Kinect から出力された関節位置の時系列データを分析している。文献¹³⁾では、カメラ映像中の人物の動きから個人を識別するため、身体動揺が応用されている。頭上カメラを用いて身体動揺の時系列データを観測し、個人性を捉える特徴を抽出している。身体動揺を用いていないが、頭上カメラを用いた個人識別の有効性は文献¹⁴⁾でも示されている。文献^{9)~13)}では、それぞれの応用例で身体動揺が有効であることが示されているものの、性別認識における有効性は示されていない。身体動揺から得られる時系列データを、他分野の応用例が活用していたように、性別認識の本評価対象手法でも活用できることを確認する。活用するための具体的な手順として、特徴抽出について 3.3 節で述べ、識別処理について 3.4 節で述べる。

3. 評価対象手法

3.1 概要

立ち止まる人物の性別を、身体動揺のカメラ映像から識別する評価対象手法を設計する。本論文では、身体動揺を用いた性別認識のユースケースを次のように想定する。駅のホームに設置された広告を人物が読むケース、商品の購入を人物が検討するケース、建物内のエレベータを人物が

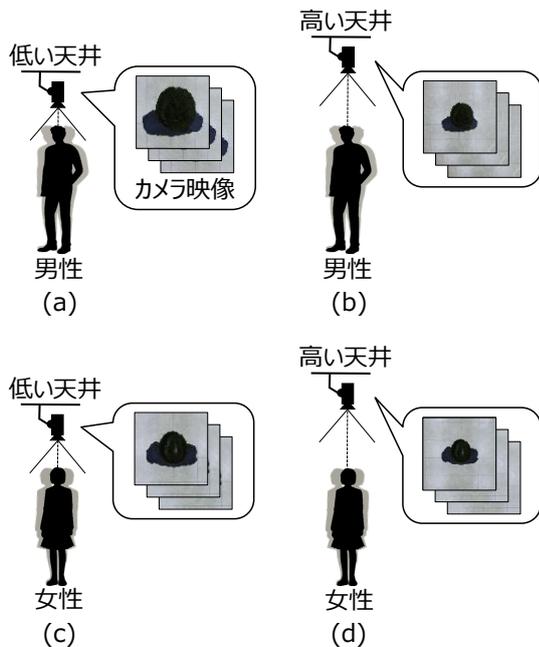


図1 頭上カメラの高さが異なる撮影環境における人物の見かけの大きさの変動例。

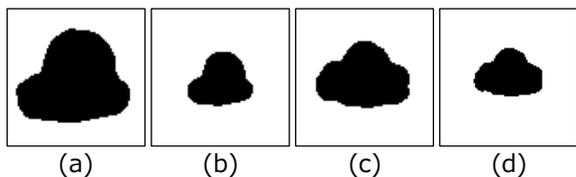


図2 異なる高さの頭上カメラから取得された男女のシルエット画像の例。黒色の画素は頭を中心とした人物領域である。白色の画素は背景領域である。

待つケースを考えている。それぞれのユースケースにおいて、立ち止まる人物の性別を正しく識別することで、マーケティング分野や防犯分野への貢献が期待される。本論文で想定するユースケースにおいて、性別認識の対象となる人物の周りに、複数の人物が存在することがある。天井に取り付けられた一般的な防犯カメラの俯瞰視野では、人物同士の重なりから遮蔽が発生し、対象人物を撮影できない可能性がある。上記ユースケースにおいて、将来的に発生する遮蔽の問題を事前に防ぐため、頭上カメラを用いた撮影を考える。

頭上カメラが天井へ埋め込まれた撮影環境について考える。カメラが埋め込まれる天井の高さは、各撮影環境で同じであるとは限らない。天井の高さが変化すると、人物領域の見かけの大きさが変化し、映像中の身体動揺の見え方に影響を及ぼす。人物領域の見かけの大きさの具体例を図1に示す。図中(a)と(b)では、同じ男性であるにも関わらず、頭上カメラの高さにより見かけの大きさが変化している。図中(c)と(d)の女性でも、同様に見かけの大きさが変化している。本論文では、頭上カメラが設置された天井の高さに観測結果が依存しないように、評価対象手法を設計することを考える。このために評価対象手法を、以下の

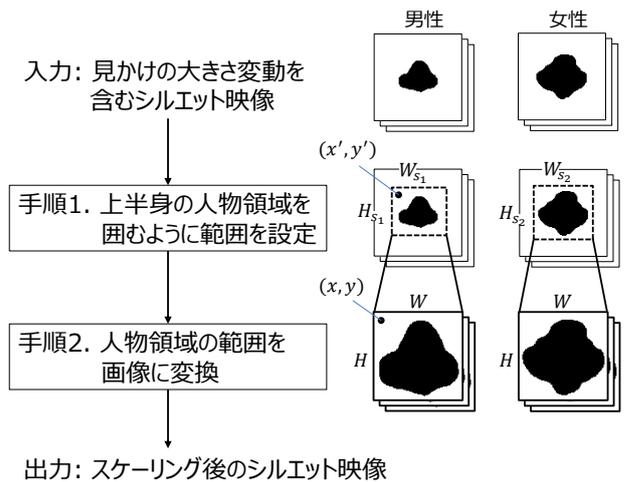


図3 シルエット映像中の人物の見かけの大きさ変動を除去する処理の流れ。

4つのステップで構成する。まず、頭上カメラを用いて、立ち止まる人物のカメラ映像を取得し、頭を中心とした人物領域を表すシルエット映像を推定する。次に、頭上カメラの高さに起因する人物の見かけの大きさ変動をシルエット映像から除去する。その後、見かけの大きさの変動が除去されたシルエット映像から、身体動揺を表す時系列データを観測し、その時系列データから男女を見分ける特徴量を抽出する。最後に、その特徴量と識別器とを組み合わせることで、性別認識の精度を獲得する。シルエット映像中の人物の見かけの大きさの変動について3.2節で述べる。特徴抽出の手順について3.3節で詳細を述べ、識別処理の手順について3.4節で述べる。

3.2 シルエット映像中の人物の見かけの大きさの変動除去

人物の性別を正しく識別するために、男性クラスと女性クラスのそれぞれにおいて、クラス内見え方が類似するように撮影されることが望ましい。ただし、3.1節で述べたように、頭上カメラを埋め込む天井の高さにより、人物の見かけの大きさが変化する。人物領域を表すシルエット画像を、図1の男性と女性から推定した結果を図2に示す。図中(a)と(b)は、ともに男性クラスであるにも関わらず、人物領域の見かけの大きさが異なっている。同様の傾向が女性クラスの(c)と(d)でも見られる。これらのシルエット映像をそのまま用いた場合、クラス内変動が大きくなるため、性別認識の精度低下が引き起こされる。よって本評価対象手法では、シルエット映像中の人物の見かけの大きさの変動を除去することを考える。このために、全てのシルエット映像において、人物の見かけの大きさを近づける処理を加える。この処理の流れを図3に示す。入力されたシルエット映像 s において、頭を中心とした上半身の人物領域を囲むように、縦幅が H_s 、横幅が W_s の範囲を設定する。その際、各時刻において、人物領域が範囲外で存在しないように、マージンを設ける。このマージンは、それ



図 4 身体動揺のシルエット映像から LM 特徴量を抽出する処理の流れ。

それぞれの映像の揺れ幅の最大値を基に決定される。設定された範囲を、縦幅が H 、横幅が W の画像へ式 (1) で変換することで、スケーリングを行う。

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{W}{W_s} & 0 \\ 0 & \frac{H}{H_s} \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (1)$$

なお、 x, y はスケーリング後の画素座標を表し、 x', y' はスケーリング前の画素座標を表す。

3.3 特徴抽出の手順

身体動揺を表す LM 特徴量を抽出するため、文献¹³⁾ のアルゴリズムを用いる。この処理の流れを図 4 に示す。まず、揺れの中心位置が観測される参照時刻 r を決定する。このため、時刻 $t (t = 1, \dots, T)$ のシルエット画像 s_t の中から、互いに最も距離が小さくなる 1 枚を選択する。この選択された画像の時刻を、参照時刻 r とする。参照シルエット画像を s_r とし、その人物領域の重心を求め、揺れの中心位置を設定する。次に、空間方向の特徴を捉えるため、揺れの中心位置を基準として、 s_t を放射線状に I 個の局所ブロックに分割する。その後、時間方向の特徴を捉えるため、それぞれの局所ブロック $i (i = 1, \dots, I)$ において、移動量 d_{it} を式 (2) を用いて算出する。

$$d_{it} = \sum_{\mathbf{x} \in \text{block}(i)} \|s_r(\mathbf{x}) - s_t(\mathbf{x})\|_1 \quad (2)$$

ここで、 $\text{block}(i)$ は i 番目のブロックを表し、 $\|\cdot\|_1$ は L1 ノルムを表す。 \mathbf{x} は画素位置とし、 $s_r(\mathbf{x})$ は参照シルエット画像の位置 \mathbf{x} の画素値を表し、 $s_t(\mathbf{x})$ は時刻 t のシルエット画像の位置 \mathbf{x} の画素値を表す。各時刻において計測された移動量 d_{it} を時間方向に並べることで、時系列データを

表 1 評価データセットに含まれる実験協力者の詳細。

	女性	男性
人数	30 名	30 名
平均年齢	22.4 ± 6.3 歳	21.6 ± 1.3 歳
平均身長	158.7 ± 4.7 cm	170.2 ± 6.4 cm

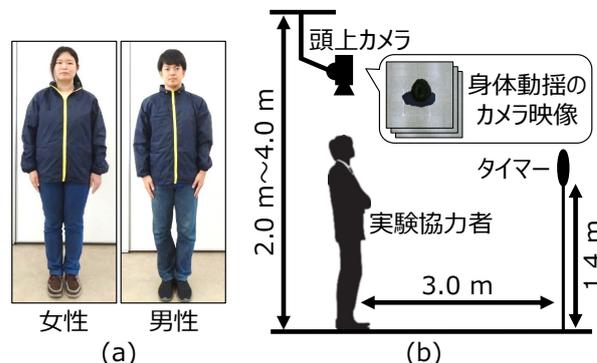


図 5 人物の身体動揺の撮影環境。

獲得する。この時系列データに対して、長さ L の窓関数を畳み込み、Power Spectral Density (PSD)²³⁾ を推定する。なお各局所ブロックの PSD の次元数は $L/2$ となる。それぞれの局所ブロックから求めた PSD を結合し、 $IL/2$ 次元の LM 特徴量を決定する。

3.4 識別処理の手順

立ち止まる人物の性別を識別する際、身体動揺の映像から抽出された LM 特徴量を、識別器の入力として用いる。様々な識別器の学習モデルが提案されているが、本評価対象手法では線形 SVM を用いる。このモデルでは、入力された特徴量の各次元が、識別にどれだけ寄与するかを表す重みを、簡便に確認することができる。この重みを用いることで、LM 特徴量の各次元における男女間の差異を 4.5 節で可視化する。

4. 実験

4.1 データセット

立ち止まる人物の性別を、カメラ映像から観測された身体動揺を用いて識別できるかについて評価した。このために、男性と女性の身体動揺のカメラ映像を独自に収集した。実験協力者は男女それぞれで 30 名ずつとし、合計で 60 名とした。実験協力者の詳細を表 1 に示す。カメラ映像の撮影中は、図 5 (a) に示す直立姿勢 (ロンベルグ肢位) を維持するように実験協力者へ指示した。背景差分による人物領域の推定精度を安定させるため、ナイロン製で紺色の上着を着用するように、全ての実験協力者へ指示した。髪型は自由とし、髪留めやヘアゴムはそのままよいとした。頭が観測されるように、帽子を身に着けないように指示した。撮影中の顔向きの変動を抑制するため、3.0 メートル先に設置されたタイマーを見るように実験協力者へ指示した。なお、タイマーの高さは 1.4 メートルとした。本データセットの撮影環境を図 5 (b) に示す。カメラの高さが床

から 2.0 メートルから 4.0 メートルの間となるようランダムに設置した。カメラの解像度を 1920×1080 画素とし、サンプリング周波数を 30 ヘルツとした。カメラの内部パラメータは一定とした。カメラ光軸が床面と垂直となるようにキャリブレーションを行った。人物 1 名あたりの撮影時間を 60 秒とした。

4.2 性別認識の精度評価

性別認識の精度を評価するため、実験協力者 60 名をランダムにシャッフルし、50 名を訓練サンプルに、10 名をテストサンプルに用いた。訓練サンプルとテストサンプルの間で、同じ人物が含まれないようにした。このランダムシャッフルを 30 回試行し、性別認識の精度の平均と標準偏差を算出した。ここでは、3 つの手法について比較を行った。以下、それぞれの手法について述べる。

評価対象手法 (LM): 立ち止まる人物の身体動揺を捉える特徴量として、3.3 節で述べた LM 特徴量を用いた。LM 特徴量を抽出するため、局所ブロック数を $I = 8$ とし、窓関数の長さを $L = 64$ (2.1 秒) とした。シルエット映像をカメラ映像から背景差分を用いて取得した。その際の縦幅 H を 100 画素、横幅 W を 100 画素とした。識別器には線形 Support Vector Machine (SVM) を用いた。正則化パラメータを $C = 1.0$ とした。線形 SVM の訓練サンプルとして、男性から抽出された 25 個の LM 特徴量と、女性から抽出された 25 個の LM 特徴量とを用いた。

比較手法 1 (GEI): 歩容の動きで性別を識別する代表的な特徴量として、GEI 特徴量を用いた。GEI 特徴量を抽出するため、60 秒間のシルエット映像を時間方向に平均化し、1 枚の平均画像を算出した。シルエット映像は評価対象手法と同じものを用いた。識別器には線形 Support Vector Machine (SVM) を用いた。正則化パラメータを $C = 1.0$ とした。線形 SVM の訓練サンプルとして、男性から抽出された 25 個の GEI 特徴量と、女性から抽出された 25 個の GEI 特徴量とを用いた。

比較手法 2 (CNN): 静止画を用いて性別を識別する代表的な手法として文献²¹⁾ の CNN を用いた。CNN の構造を、畳込み層 4 個とプーリング層 4 個とした。CNN の訓練サンプルとして、男性 45,000 枚と女性 45,000 枚の静止画を用いた。静止画のサイズは 100 × 100 画素とし、各画素は RGB 値とした。識別には二値交差エントロピーを用いた。最適化アルゴリズムとして確率的勾配降下法を用いた。本論文で想定する頭上からの撮影条件において、深層学習を用いた性別認識の一般的な手法が、どの程度の精度が得られるかを検証する。

各手法を用いた場合の性別認識の精度を図 6 に示す。本評価対象手法の精度は $90.3 \pm 1.3\%$ であり、比較手法 1 の精度の $67.7 \pm 0.8\%$ に比べて、22.6 ポイント高かった。比較手法 1 の GEI 特徴量は、歩く人物の腕や足の大きな動きを捉えるように設計されているため、立ち止まる人物の微細な動きを十分に捉えることができなかつたと考えられる。

一方、評価対象手法の LM 特徴量は、身体動揺の微細な動きを捉えており、精度が大幅に向上したと考えられる。また、比較手法 2 の精度の $84.8 \pm 2.7\%$ に比べて、本評価対象手法の精度は 5.5 ポイント高かった。比較手法 2 の CNN は、人物の空間的な見え方から特徴を抽出しているため、時間方向の動きの変化を捉えていなかったと考えられる。一方、評価対象手法は、身体動揺の空間方向と時間方向の特徴を共に考慮しているため、識別精度が改善されたと考えられる。

比較手法 1 の識別精度が低い原因が、線形 SVM の訓練サンプル数の不足によるものであるかを確認する。線形 SVM の訓練サンプル数を変化させた場合の精度を調査した。ここでは、訓練サンプル数を 40 個と 4 個に設定した。訓練サンプル数以外の実験条件を、4.2 節 1 段落で述べた条件と同じにした。その結果、本評価対象手法の精度は、訓練サンプル数が 40 個の場合で $85.7 \pm 2.5\%$ 、4 個の場合で $64.0 \pm 2.5\%$ となった。比較手法 1 の精度は、訓練サンプル数が 40 個の場合で $67.0 \pm 1.6\%$ 、4 個の場合で $58.0 \pm 3.1\%$ となった。本評価対象手法と比較手法 1 のどちらについても、訓練サンプル数の減少に伴い、精度低下が確認された。実験結果から、訓練サンプル数を極端に減少させた場合は、比較手法 1 において、識別精度の低下を引き起こす可能性があると考えられる。

さらに比較手法 1 について、ブートストラップサンプリングを用いて、仮想的に訓練サンプルを生成した場合の精度を調査した。1 個のシルエット映像を構成する 1800 フレームの中から、重複を許してランダムにフレームを指定数だけ選択し、それらのフレームから 1 個の GEI 特徴量を生成した。このランダム生成を 1 人当たり 100 回試行することで、合計 $50 \times 100 = 5000$ 個の訓練サンプルを生成した。その結果、サンプリング指定数が 500 フレームの場合は $68.0 \pm 3.7\%$ 、1000 フレームの場合は $67.0 \pm 1.3\%$ 、1500 フレームの場合は $68.0 \pm 2.7\%$ となった。訓練サンプル数をブートストラップサンプリングで仮想的に増加させても、精度は大きく変化しないことが分かった。大量の訓練サンプル数を、ブートストラップサンプリングではなく、実際に準備した場合の精度を、今後は調査していく必要がある。

4.3 時空間特徴量と識別器とを組合せた手法、および、時空間を捉える深層学習を用いた手法の精度評価

身体動揺を時空間で捉えることで、性別認識の精度が向上することが 4.2 節で確認された。人物の動きを時空間で捉える方法は、歩容認証や行動認識の分野で既に多く提案されている。LM 特徴量と識別器とを組合せた本評価対象手法が、他の時空間特徴量と識別器とを組合せた手法や、時空間を捉える深層学習を用いた手法と比較して、立ち止まる人物の性別を正しく識別できるかどうかを調査した。以下、本実験で用いた手法について述べる。

LM: 4.2 節で述べた LM 特徴量と線形 SVM との組合せ

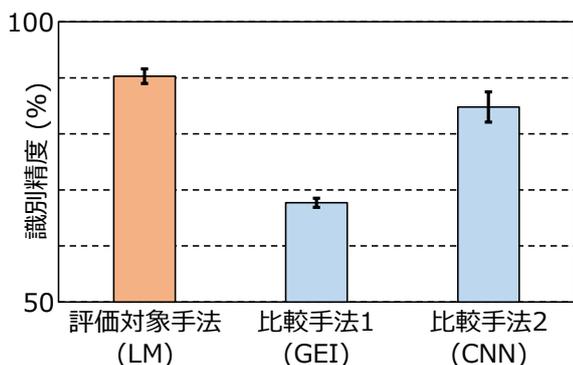


図 6 各手法の性別認識の精度比較. 図中の各棒グラフに付属する縦線は, 精度の標準偏差を表す.

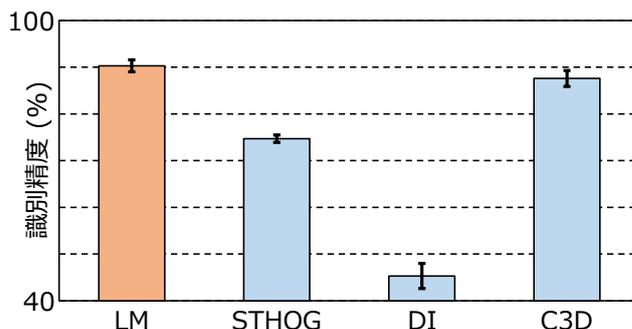


図 7 時空間特徴量と識別器とを組合せた手法の精度, および, 時空間を捉える深層学習を用いた手法の精度. 図中の各棒グラフに付属する縦線は, 精度の標準偏差を表す.

を用いた.

STHOG: 歩容認証で利用されている Spatio-Temporal Histograms of Oriented Gradient (STHOG)²⁴⁾ 特徴量と識別器との組合せを用いた. STHOG 特徴量のパラメータについて, セルサイズを 25(画素)×25(画素)×450(フレーム), ブロックサイズを 2(セル)×2(セル)×2(セル), ウィンドウサイズを 2(ブロック)×2(ブロック)×2(ブロック)(100(画素)×100(画素)×1800(フレーム)) とした. なおパラメータ記述の様子は文献²⁵⁾²⁶⁾ に従った. グレースケール映像のサイズを 100 × 100 画素とした. 識別器には線形 SVM を用いた. 正則化パラメータを $C = 1.0$ とした. 線形 SVM の訓練サンプルとして, 男性から抽出された 25 個の STHOG 特徴量と, 女性から抽出された 25 個の STHOG 特徴量を用いた.

DI: 行動認識で利用されている Dynamic image²⁷⁾ 特徴量と識別器との組合せを用いた. 60 秒間のカラー映像に対して, Rank SVM²⁸⁾ を適用することで DI 特徴量を算出した. カラー映像のサイズを 100 × 100 画素とした. 識別器には線形 SVM を用いた. 正則化パラメータを $C = 1.0$ とした. 線形 SVM の訓練サンプルとして, 男性から抽出された 25 個の DI 特徴量と, 女性から抽出された 25 個の DI 特徴量を用いた.

C3D: 行動認識で利用されている C3D²⁹⁾ を用いた. 深層学習のネットワーク構造を, 三次元畳み込み層 4 個と三次元プーリング層 4 個にした. 60 秒のカラー映像から, 16 枚の連続フレームで構成される短カラー映像をランダムに生成した. C3D は 4.2 節で述べた比較手法 2 の CNN を時間方向に拡張したものである. 訓練サンプルとして, 男性 2800 個と女性 2800 個の短カラー映像を用いた. 短カラー映像のサイズを 100 × 100 画素とした. 学習時の損失項として二値交差エントロピー, 最適化アルゴリズムとして確率的勾配降下法を用いた.

データセットなど, 上記の手法以外の実験条件を 4.2 節と同じとした.

時空間特徴量と識別器とを組合せた手法の精度, および, 時空間を捉える深層学習を用いた手法の精度を図 7 に示

す. STHOG 特徴量と線形 SVM とを組合せた手法の精度の $78.3 \pm 2.1\%$ に比べて, 本評価対象手法の精度は $90.0 \pm 1.3\%$ であり, 12.0 ポイント高かった. STHOG 特徴量は, 4.2 節の GEI 特徴量と同様に, 歩く人物の腕や足の大きな動きを表すように設計されているため, 立ち止る人物の微細な動きを十分に捉えることができなかつたと考えられる. DI 特徴量と線形 SVM とを組合せた手法の精度の $46.0 \pm 2.3\%$ に比べて, 本評価対象手法の精度は 44.3 ポイント高かった. DI 特徴量は, スポーツのように腕や足を含む身体の姿勢が大きく変化する動きを表すように設計されているため, 身体動揺の微細な動きを捉えることができなかつたと考えられる. また, C3D を用いた手法の精度の $87.1 \pm 1.5\%$ に比べて, 本評価対象手法の精度は 3.2 ポイント高かった. 本評価対象手法の精度と, C3D を用いた手法の精度との間で, 対応ありの t 検定を実施した. その検定の結果, 有意差 ($p < 0.05$) が見られた. C3D は, 微細な動きを捉えるように設計されていないため, LM 特徴量と線形 SVM とを組合せた場合と比べて識別精度が低下したと考えられる. 以上の結果から, LM 特徴量と線形 SVM とを組合せた本評価対象手法は, 他の時空間特徴量と線形 SVM とを組合せた手法や, 時空間を捉える深層学習を用いた手法と比較して, 身体動揺に含まれる僅かな動きを捉えることができ, 立ち止まる人物の性別認識に貢献したと考えられる.

STHOG 特徴量と線形 SVM とを組合せた手法の精度は, セルのフレーム数に関係すると考えられる. ここではセルのフレーム数を 60, 450, または, 900 とした時, 性別認識の精度を求めた. フレーム数以外の実験条件を, 4.3 節 1 段落で述べた条件と同じにした. その結果, STHOG 特徴量のセルのフレーム数が 60 の場合は $74.7 \pm 2.5\%$, 450 の場合は $78.3 \pm 2.1\%$, 900 の場合は $78.0 \pm 1.4\%$ であった. セルのフレーム数が 450 の場合, 60 の場合と比べて精度が高くなるものの, 900 の場合と比べるとほぼ同じであった. 身体動揺は微小な動きであるため, フレーム数を大きくしても, それほど精度が変化しないと考えられる.

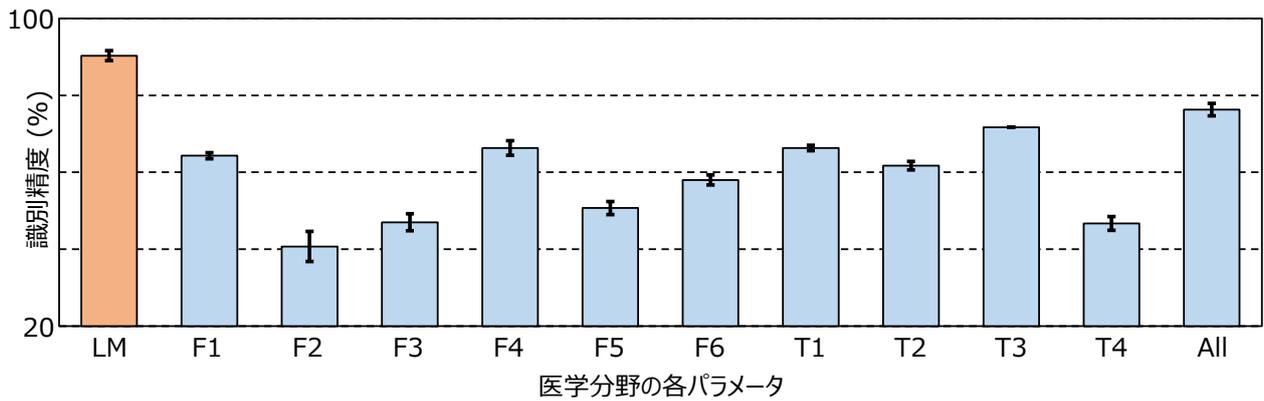


図 8 医学分野の各パラメータと線形 SVM とを組合せた手法の精度.

4.4 医学分野で得られた知見を性別認識に応用した場合との精度比較

医学の分析研究^{6)~8)}では、2.2 節で述べたように、重心位置の時系列データにおける周波数特性や軌跡のパラメータに、性差が表れることが報告されている。これらの分析研究では、各パラメータを獲得するため、圧力センサを使用していたが、カメラを使用することも考えられる。周波数特性や軌跡のパラメータをカメラ映像から特徴量として抽出し、その特徴量と線形 SVM とを組合せた手法を検討した。この実験では、シルエット画像中の人物領域の中心を重心位置の代わりに用いた。揺れの中心となる参照時刻の重心位置を原点に設定し、画像の横方向と縦方向をそれぞれ x 軸と y 軸とした。なお、画像上部が人物の前面、画像下部が人物の背面、画像右部が人物の右手、画像左部が人物の左手に対応している。ただし、圧力センサから観測された身体動揺の信号と、カメラ映像から観測された身体動揺の信号とが、どの程度類似しているかの評価はなされていないため、さらなる評価を今後は進める必要がある。

性別認識に応用する医学の知見として、男女の間で有意差が見られる下記の 10 個のパラメータ (周波数特性の F1 から F6、および、軌跡の T1 から T4) を用いた。

F1: 重心位置から原点までの距離の時系列データに対して Fast Fourier Transform (FFT) を適用し、直流成分の振幅を特徴とした。

F2: 重心位置から原点までの距離の時系列データに対して FFT を適用し、得られた振幅を低周波から加算していき、全体の 95% を含む時の周波数を特徴とした。

F3: 重心の移動速度の大きさの時系列データに対して FFT を適用し、直流成分の振幅を特徴とした。

F4: 重心の移動速度の x 軸成分の時系列データに対して FFT を適用し、0.2 ヘルツよりも小さな周波数帯の振幅の合計値を特徴とした。

F5: 重心の移動速度の y 軸成分の時系列データに対して FFT を適用し、0.2 ヘルツよりも小さな周波数帯の振幅の合計値を特徴とした。

F6: 重心の移動速度の x 軸成分の時系列データに対して

FFT を適用し、2 ヘルツよりも大きな周波数帯の振幅の合計値を特徴とした。

T1: 全時刻の重心位置から描かれる軌跡に対して、全体の 95% が含まれるように楕円近似を適用した。得られた楕円の面積を特徴とした。

T2: 軌跡を楕円近似した時の長軸の長さを特徴とした。

T3: 軌跡を楕円近似した時の短軸の長さを特徴とした。

T4: 軌跡の長さを合計し、その値を特徴とした。

各パラメータで性別認識を行う時は、それぞれのパラメータを 1 次元の特徴量とした。識別精度の向上を狙い、F1 から T4 までのパラメータを並べた 10 次元の特徴ベクトル (All) も比較対象とした。特徴量以外の実験条件は、4.2 節の評価対象手法と同じとした。

医学分野の各パラメータと線形 SVM とを組合せた手法の精度を図 8 に示す。F1 から F6 までの各パラメータと線形 SVM とを組合せた手法の精度は、それぞれ $64.3 \pm 0.8\%$ 、 $40.7 \pm 3.9\%$ 、 $47.0 \pm 2.2\%$ 、 $66.3 \pm 1.9\%$ 、 $50.7 \pm 1.7\%$ 、 $58.0 \pm 1.3\%$ であった。T1 から T4 までの各パラメータと線形 SVM とを組合せた手法の精度は、それぞれ $66.3 \pm 0.7\%$ 、 $61.7 \pm 1.1\%$ 、 $71.7 \pm 0.0\%$ 、 $46.7 \pm 1.8\%$ であった。また、全パラメータを並べた All と線形 SVM とを組合せた手法の精度は $76.3 \pm 1.6\%$ であった。なお本評価対象手法の精度は $90.3 \pm 1.3\%$ であった。身体動揺を観測するため、医学分野の分析研究では圧力センサを用いており、本論文の実験ではカメラ映像を用いた違いがあるものの、同じ目的をもって算出されたパラメータの中で、どれを特徴量として線形 SVM と組合せたとしても、本評価対象手法の精度の方が高いことを確認した。

4.5 評価対象手法の特徴量に与える重みの可視化

評価対象手法で用いた LM 特徴量のどの次元が識別時に重要であるかを確認するため、線形 SVM の重みを可視化した。特徴量を算出する時、局所ブロック数を $I = 8$ 、窓関数の長さを $L = 64$ とした。次元数は、 $I \times L / 2 = 8 \times 32 = 256$ 次元とした。分割された局所ブロック数の通し番号 (P1 から P8) を図 9 (a) に示す。なお、P2 が人物の左手、P4 が人物の背面、P6 が人物の右手、P8 が人物の前面に対応す

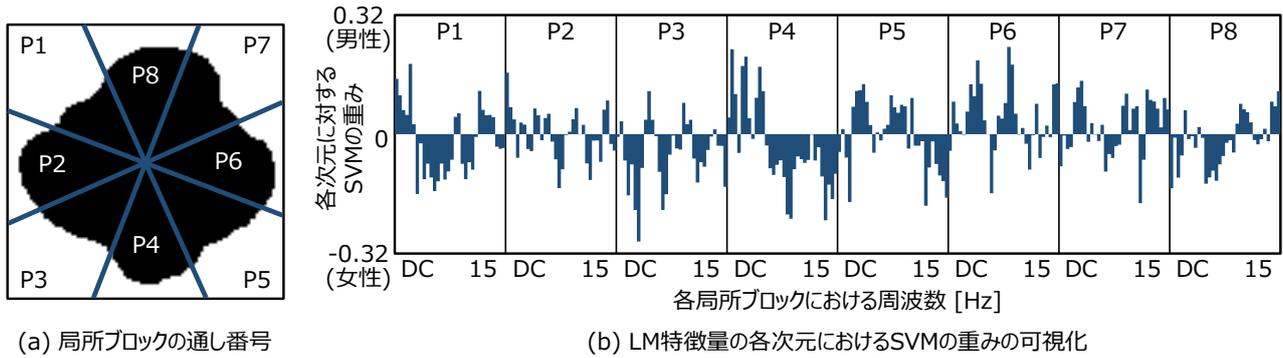


図 9 身体動揺を表す特徴量のどの次元が性別認識に重要であるかを確認するため、線形 SVM の重みを可視化。

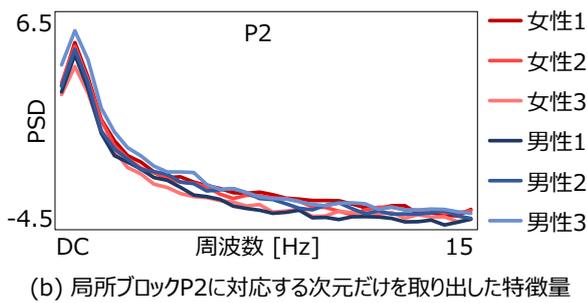
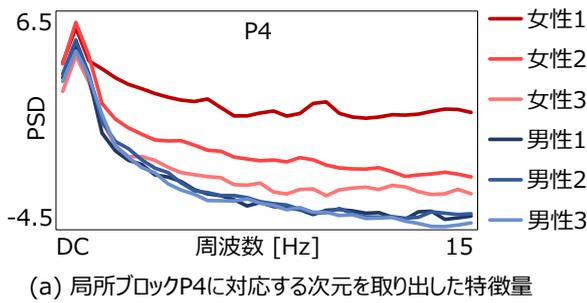


図 10 局所ブロック P2 と P4 のそれぞれに対応する LM 特徴量の例。

る。それぞれの局所ブロックに対応する LM 特徴量の各次元について、SVM の重みを可視化した結果を図 9 (b) に示す。図中の横軸は、各局所ブロックにおける周波数を表す。各局所ブロックにおいて、横軸の最も左の周波数は直流、最も右は 15 ヘルツである。図中の縦軸は、各次元に対する SVM の重みを表す。正の向きに値が大きいほど男性の識別に寄与し、負の向きに値が大きいほど女性の識別に寄与する。

個別に各周波数成分を見る前に、各局所ブロックが性別認識にどの程度寄与するかを評価した。各局所ブロックにて、SVM の重みの絶対値の和を求めた。その結果、P1 から P8 で 2.65, 1.72, 2.51, 3.44, 2.36, 2.29, 2.11, 1.81 であった。絶対値の和が大きい局所ブロックほど、性別認識に寄与すると考えられる。局所ブロック P4 の和が最も大きかったことから、人物の背中側が特に重要であると考えられる。

次に、局所ブロック P4 の中で各周波数に対する SVM の

重みを考察する。図 9 (b) の P4 から、低周波側は男性の重みが強く、高周波側は女性の重みが強いことが分かる。実際に各人物の特徴量を可視化し、男女の違いが周波数の帯域で表れることを確認した。局所ブロック P4 に対応する特徴量の次元を取り出した結果を図 10 (a) に示す。SVM の識別境界面から離れた特徴量をもつ男性 3 名と女性 3 名を用いた。ここでは比較のため、局所ブロック内の重みの絶対値の和が小さかった P2 での特徴量を、図 10 (b) に示す。P4 では、3 ヘルツから 15 ヘルツの周波数帯域に男女の間で特徴量に大きな違いが見られた。一方、P2 では、特に顕著な違いは見られなかった。P4 における男女間の違いについて理由を考察する。P4 は人物の背面に対応する局所ブロックであり、その人物の後頭部を含んでいた。女性の後頭部は、男性の後頭部と比べて、後ろ髪が長い場合が多く見られた。長い髪は動きの自由度が高いため、高周波帯の特徴が発生しやすいと考えられる。このため、女性は男性と比べて、高周波帯の成分に該当する特徴量に違いが表れたと考えられる。誤識別となった女性の事例を確認したところ、髪の短い女性が含まれていた。この女性から得た LM 特徴量は、髪の短い男性の LM 特徴量と近くなり、識別誤りが発生したと考えられる。ただし、男性が長髪である場合も考えられるため、データセットを拡充した上で、さらなる今後の調査が必要である。

4.6 男女間の身長差が本評価対象手法に与える影響の調査

頭上カメラを用いる本評価対象手法は、身体動揺の男女の違いではなく、身長の男女の違いを捉えることで、性別を識別している可能性が考えられる。まず実験で用いたデータセットにおける男女間の身長差を確認した。表 1 に示した男性の平均身長と女性の平均身長との間で、ウィルコクソンの順位と検定を実施したところ、有意差 ($p < 0.05$) が見られた。このため、実験で用いたデータセットにおいて、男女間で身長差があると言える。さらに実験を進めるため、身長差が大きい場合と身長差が小さい場合のサブセットを用意した。

サブセット 1: 男女間の身長差が小さい 20 名 (男性 10 名, 女性 10 名) を対象とした。男性の平均身長は 163.3 ± 3.3 cm,

女性の平均身長は 161.1 ± 2.0 cm であった。男女間で、ウィルコクソンの順位和検定を適用したところ、有意差 ($p < 0.05$) が見られず、サブセット 1 には身長差があるとは言えない。

サブセット 2: 男女間の身長差が大きい 20 名 (男性 10 名, 女性 10 名) を対象とした。男性の平均身長は 176.8 ± 4.2 cm, 女性の平均身長は 153.9 ± 1.5 cm であった。男女間で、ウィルコクソンの順位和検定を適用したところ、有意差 ($p < 0.05$) が見られ、サブセット 2 には身長差があると言える。

なお、サブセット間で同じ人物が含まれないとした。テストサンプルとして、20 名から男性 1 名と女性 1 名とをランダムに選択し、残りの 18 名を線形 SVM の訓練サンプルとした。各サブセットにおいて、このランダムな選択を 50 回試行し、精度の平均と標準偏差を算出した。なお、本評価対象手法のパラメータは 4.2 節と同じにした。

本評価対象手法を用いた性別認識の精度は、身長差が小さいサブセット 1 で $90.0 \pm 3.5\%$, 身長差が大きいサブセット 2 で $88.0 \pm 2.7\%$ となった。サブセット間で平均精度にウィルコクソンの順位和検定を適用したところ、有意差 ($p < 0.05$) が見られず、精度に違いがあるとは言えなかった。このことから、本評価対象手法の精度は男女間の身長差に依存する可能性は小さいと考えられる。

5. 本評価対象手法における制約と課題

本論文の実験では、研究の初期段階であるため、下記の制約を人物に課した。

制約 1 人物が直立姿勢を 60 秒間維持

制約 2 人物が同じ服装を着用

このため 3.1 節で述べた想定ユースケースに本評価対象手法をそのまま適用することはできないと考えられる。よって今後は以下の課題を解決していく必要がある。

課題 1 姿勢変化を許容する特徴量の開発

想定ユースケースにおいて、人物が直立姿勢を 60 秒間維持するケースは非常に少ない。片足に重心をかけ休みをとる姿勢、吊革に挟まって立ち止まる姿勢など、様々な立位姿勢が考えられる。時間経過で変化する立位姿勢の中から、不変な特徴量を抽出するアルゴリズムの開発が必要である。

課題 2 服装変化を許容する特徴量の開発

想定ユースケースにおいて、制服やスーツなどの似た種類の服装を着用する場合と、私服のように人物毎で異なる種類の服装を着用する場合がある。特に後者の場合、上半身のシルエット画像が多様になり、特徴量自体が変動するため、精度が低下する可能性が大きい。服装が変化しても、不変な特徴量を抽出するアルゴリズムの開発が必要である。

6. むすび

本論文では、頭上カメラを用いて撮影された映像から身体動揺の動きを表す特徴量を抽出し、その特徴量と識別器とを組合せることで、立ち止まる人物の性別を識別できるかを評価した。本評価対象手法では、映像中の人物の見え方に関する変動を除去するために、見かけの大きさを揃える処理を施した。空間方向の特徴を捉えるために、人物領域を複数の局所ブロックに分割し、また、時間方向の特徴を捉えるために、移動量の時間データを各局所ブロックから観測した。本評価対象手法の識別精度を確認するため、身体動揺のカメラ映像のデータセットを構築した。実験結果から、本論文の評価対象手法は、LM 特徴量と線形 SVM とを組合せることで、立ち止まる人物の性別を高い精度で識別できることを確認した。さらに評価対象手法の精度について、医学知見に基づく特徴量と線形 SVM とを組合せた手法の精度と比べることで、カメラ映像から観測される身体動揺は、精度向上のための詳細な特徴を捉えられることが分かった。

本論文で述べた評価対象手法は、立ち止まる人物の性別を認識するユースケースを想定しているものの、研究の初期段階であり現状では多くの制約が設けられている。今後の課題として、頭上カメラのみではなく既設の防犯カメラの利用も想定し、本評価対象手法を改良していくことが求められる。また、様々な髪型の男女においても正しく性別を認識できるかの調査が求められる。さらに、人物の年代や服装などの性別以外の属性に対して、身体動揺が識別に有効であるかどうかを調査する必要がある。

(文 献)

- 1) X. Wang, S. Zheng, R. Yang, B. Luo and J. Tang : "Pedestrian Attribute Recognition: A Survey", CoRR, (2019)
- 2) C.B. Ng, Y.H. Tay and B. Goi : "Recognizing human gender in computer vision: a survey", In Proceedings of the Pacific Rim International Conference on Artificial Intelligence, pp.335-346, (2012)
- 3) C. Shan, S. Gong and P.W. McOwan : "Fusing gait and face cues for human gender recognition", Neurocomputing, **71**, 10-12, pp.1931-1938, (2008)
- 4) R. Martín-Félez, R.A. Mollineda and J.S. Sánchez : "Gender classification from pose-based GELs", In Proceedings of the International Conference on Computer Vision and Graphics, pp.501-508, (2012)
- 5) S. Yu, T. Tan, K. Huang, K. Jia and X. Wu : "A study on gait-based gender classification", IEEE Transactions on Image Processing, **18**, 8, pp.1905-1910, (2009)
- 6) J.W. Kim, G.M. Eom, C.S. Kim, D.H. Kim, J.H. Lee, B.K. Park and J. Hong : "Sex differences in the postural sway characteristics of young and elderly subjects during quiet natural standing", Geriatrics & Gerontology International, **10**, 2, pp.191-198, (2010)
- 7) M. Plandowska, M. Lichota and K. Górniak : "Postural stability of 5-year-old girls and boys with different body heights", PLoS ONE, **14**, 12, pp.1-10, (2020)
- 8) T. Kitabayashi, S. Demura, M. Noda and T. Yamada : "Gender differences in body-sway factors of center of foot pressure in a static upright posture and under the influence of alcohol intake", Journal of physiological anthropology and applied human science, **23**, 4, pp.111-118, (2004)
- 9) F. Wang, M. Skubic, C. Abbott and J.M. Keller : "Body sway measurement for fall risk assessment using inexpensive webcams", In

- Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology, pp.2225–2229, (2010)
- 10) M. Nishiyama, T. Miyauchi, H. Yoshimura and Y. Iwai : “Synthesizing realistic image-based avatars by body sway analysis”, In Proceedings of the Fourth International Conference on Human Agent Interaction, pp.155–162, (2016)
 - 11) L.F. Yeung, K.C. Cheng, C.H. Fong, W.C. Lee and K.Y. Tong : “Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway”, *Gait & Posture*, **40**, 4, pp.532–538, (2014)
 - 12) Z. Lv, V. Penades, S. Blasco, J. Chirivella and P. Gagliardo : “Evaluation of Kinect2 based balance measurement”, *Neurocomputing*, **208**, pp.290–298, (2016)
 - 13) T. Kamitani, H. Yoshimura, M. Nishiyama and Y. Iwai : “Temporal and spatial analysis of local body sway movements for the identification of people”, *IEICE Transactions on Information and Systems*, **102**, 1, pp.165–174, (2019)
 - 14) 李鵬, 呉海元 : “SHAL: 俯瞰視画像から個人再識別のための特徴記述法”, *映像情報メディア学会誌*, **74**, 4, pp.719–728, (2020)
 - 15) J. Han and B. Bhanu : “Individual Recognition Using Gait Energy Image”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**, 2, pp.316–322, (2006)
 - 16) L. Cao, M. Dikmen, Y. Fu and T.S. Huang : “Gender recognition from body”, In Proceedings of the 16th ACM International Conference on Multimedia, pp.725–728, (2008)
 - 17) J.E. Tapia and C.A. Perez : “Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of LBP, intensity, and shape”, *IEEE Transactions on Information Forensics and Security*, **8**, 3, pp.488–499, (2013)
 - 18) M. Collins, J. Zhang, P. Miller and H. Wang : “Full body image feature representations for gender profiling”, In Proceedings of the 12th International Conference on Computer Vision Workshops, pp.1235–1242, (2009)
 - 19) G. Levi and T. Hassner : “Age and gender classification using convolutional neural networks”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.34–42, (2015)
 - 20) G. Ozbulak, Y. Aytar and H.K. Ekenel : “How transferable are CNN-based features for age and gender classification?”, In Proceedings of the International Conference of the Biometrics Special Interest Group, pp.1–6, (2016)
 - 21) G. Antipov, S.A. Berrani, N. Ruchaud and J.L. Dugelay : “Learned vs. hand-crafted features for pedestrian gender recognition”, In Proceedings of the 23rd ACM International Conference on Multimedia, pp.1263–1266, (2015)
 - 22) M. Nishiyama, R. Matsumoto, H. Yoshimura and Y. Iwai : “Extracting discriminative features using task-oriented gaze maps measured from observers for personal attribute classification”, *Pattern Recognition Letters*, **112**, pp.241–248, (2018)
 - 23) P. Welch : “The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms”, *IEEE Transactions on Audio and Electroacoustics*, **15**, 2, pp.70–73, (1967)
 - 24) C. Hua, Y. Makihara and Y. Yagi : “Pedestrian detection by using a spatio-temporal histogram of oriented gradients”, *IEICE Transactions on Information and Systems*, **96**, 6, pp.1376–1386, (2013)
 - 25) 川合諒, 榎原靖, 八木康史 : “STHOG 特徴を用いた複数カメラ間での人物照合”, *研究報告コンピュータビジョンとイメージメディア*, pp.1–8, (2011)
 - 26) 柏本雄士朗, 村松大吾, 八木康史 : “カメラ間人物照合における STHOG 特徴の可視化と解析”, *研究報告コンピュータビジョンとイメージメディア*, pp.1–8, (2017)
 - 27) H. Bilen, B. Fernando, E. Gavves and A. Vedaldi : “Action Recognition with Dynamic Image Networks”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, pp.2799–2813, (2017)
 - 28) A.J. Smola and B. Schölkopf : “A tutorial on support vector regression”, *Statistics and Computing*, **14**, 3, pp.199–222, (2004)
 - 29) D. Tran, L. Bourdev, R. Fergus, L. Torresani and M. Paluri : “Learning spatiotemporal features with 3d convolutional networks”, In Proceedings of the IEEE International Conference on Computer Vision, pp.4489–4497, (2015)

かみたに たくや
神谷 卓也 2018 年 鳥取大学大学院博士前期課程修了。2019 年 同大学大学院博士後期課程に在学中。ソフトウェアメトリクスの研究に従事。2017 年 画像の認識・理解シンポジウム学生奨励賞を受賞。

やまくち ゆうた
山口 優太 2019 年 鳥取大学工学部電気情報系科卒業。2019 年 鳥取大学大学院持続性社会創生科学研究科工学専攻博士前期課程に在学中。

なかにしん たるう
中谷真太郎 2016 年兵庫県立大学大学院工学研究科博士後期課程修了。2014 年日本学術振興会特別研究員(DC2)。2016 年から鳥取大学大学院工学研究科助教、2019 年同講師となり、現在に至る。博士(工学)。生体信号の計測とその応用に関する研究に従事。計測自動制御学会、電気学会、IEEE などの会員。

にしやま まさし
西山 正志 2000 年 岡山大学工学部情報工学科卒業。2002 年 同大学院博士前期課程了。同年株式会社東芝入社。同社研究開発センターを経て、現在鳥取大学大学院工学研究科准教授。2011 年 東京大学大学院学際情報学府にて博士(学際情報学)を取得。カメラを用いた人物認識を始めとするパターン認識およびインタラクションの研究に従事。山下記念研究賞や画像センシングシンポジウム優秀学術賞など受賞。電子情報通信学会、情報処理学会各会員。

いわい よしお
岩井 儀雄 1992 年(平成 4 年)大阪大学基礎工学部情報工学科卒業。1997 年(平成 9 年)大阪大学大学院基礎工学研究科博士課程後期修了。同年同大学院助手。2003 年(平成 15 年)同大学院助教授。2004 年(平成 16 年)5 月～2005 年(平成 17 年)3 月英国ケンブリッジ大学客員研究員。2007 年(平成 19 年)同大学院准教授。2011 年(平成 23 年)鳥取大学大学院工学研究科教授。コンピュータビジョン、パターン認識の研究に従事。博士(工学)