

視線位置分布を用いた重み付きランダムフォレストによる性別認識

山口紗也加[†] 西山 正志^{†,††a)} 岩井 儀雄^{†,††}

Gender Classification using Weighted Random Forest with Gaze Distribution

Sayaka YAMAGUCHI[†], Masashi NISHIYAMA^{†,††a)}, and Yoshio IWAI^{†,††}

あらまし 本論文では、人物画像の性別を精度よく識別するために、視線位置分布を組み込んだ重み付きランダムフォレストの学習手法について述べる。訓練サンプルにおいて、人物領域の周囲に存在する背景に偏りが含まれる場合、その背景自体を特徴として誤学習し、性別認識の精度が大きく低下する問題がある。提案手法では、人間が人物画像の性別を判別する際の視線位置分布をランダムフォレストの学習アルゴリズムへ重みとして組み込む。これにより、人間が性別を判別する時に見ている領域から、性別を正しく見分けるための特徴をランダムフォレストの各決定木から抽出できる。視線位置分布を用いた重み付けランダムフォレストの有効性を検証するために、特徴選択を重み付ける考え方、特徴量を重み付ける考え方、および、情報利得を重み付ける考え方を比較した。公開データセットを用いた実験により、情報利得を視線位置分布で重み付けるランダムフォレストにより、性別認識の精度が大きく改善されることを確認した。

キーワード 性別認識, ランダムフォレスト, 重み付け, 視線位置分布

1. ま え が き

公共空間における人物検索や店舗におけるマーケティングに向けて、人物画像から性別を精度良く推定する技術が期待されている。性別認識の既存手法 [1]~[4] では、深層学習や機械学習の技術を用いることで高い精度を実現している。これらの技術を用いて性別認識の精度を向上させるためには、大量の訓練サンプルを収集することが前提条件となっている。訓練サンプルを大量に収集するためには多大なコストを要する。ただし、コストをかけて収集した訓練サンプルであるにも係わらず本来は意図していなかった偏りが含まれる場合がある。特定の場所で撮影された理由により訓練サンプルの背景に偏りが含まれた場合、背景を性別認識の特徴として誤学習する問題が発生する可能性がある。

識別器の学習時において訓練サンプルの背景が引き起こす問題を避けるために、性別認識で必要な特徴を

人物画像から抽出することが重要である。人間は背景に係わらず身体部位を重点的に見ることで性別を正しく判別することができる。このような人間の視覚特性を性別認識の識別器に利用できれば、訓練サンプルとテストサンプルで背景が異なる場合において、識別に必要な特徴のみが重点的に学習され認識精度が向上すると考えられる。近年、人間が画像を見た際の視線位置分布を組み込むことで、識別に有効な特徴量を抽出する手法 [5]~[7] が提案されている。これらの文献では訓練サンプルの背景の偏りについて言及されていないが、本論文で取り組む課題に対しても有効に働くと想定される。特に文献 [7] は、文献 [5], [6] が取り扱っていなかった人物画像を対象としており、実験協力者が人物画像の性別を判別する際の視線位置分布を特徴抽出に用いている。計測された視線位置分布をマップ化し、視線が集まる身体領域に大きな重みを与えている。しかし、この既存手法は識別器の前処理として視線位置分布の重みを人物画像の画素値そのものに加えていたため、その識別器の学習アルゴリズムに適した形で視線位置分布を活用しているとは言えなかった。

本論文では視線位置分布を組み込む学習アルゴリズムとして、複数の決定木を弱識別器として組み合わせるランダムフォレスト [8] に焦点を当てる。ランダム

[†] 鳥取大学大学院持続性社会創生科学研究科
Graduate School of Sustainability Science, Tottori University, Tottori 680-8550, Japan

^{††} 鳥取大学工学部附属クロス情報科学研究センター, 鳥取市
Cross-informatics Research Center, Tottori University

a) E-mail: nishiyama@tottori-u.ac.jp

フォレストは、文献[9],[10]でも述べられているように、単純な構造であるにも係わらず様々な実応用で良好な結果を得ている。ランダムフォレストの学習アルゴリズムは、一様分布に従ったランダム性を多用しているが、本論文では視線位置分布に従ったランダム性を用いる。この考え方を実現する手法として、本論文では重み付けランダムフォレスト[11]~[13]に着目する。その一つである既存手法[11]では、認識精度に寄与する訓練サンプルに大きな重みを与えている。また、既存手法[12]では、認識精度に寄与する決定木の投票結果に大きな重みを与えている。ただしこれらの既存手法は、画像中に存在する特徴量の位置を考慮していないため、視線位置分布を重みとして単純には組み込むことはできなかった。既存手法[13]では、ランダムフォレストの学習時に、特徴量や情報利得にランダム重みを与えることで、訓練サンプルに含まれるノイズへの耐性を高めている。特徴量や情報利得は画像中の位置に依存するため視線位置との関連性が高いと仮定できる。ただし既存手法[13]では一様分布に従うランダム重みを用いるため、視線位置分布を重みとしてどのように組み込むべきかを検討していなかった。

そこで本論文では、ランダムフォレストの学習アルゴリズムに、視線位置分布を用いた重み付けを導入することで、背景に偏りを含む訓練サンプルを用いた場合でも性別認識の精度を向上させる手法を提案する。学習アルゴリズムの改良方針として、特徴選択を重み付ける考え方、特徴量を重み付ける考え方、および、情報利得を重み付ける考え方を検討した。公開データセットを用いた性別認識の実験を行い、それぞれの重み付けの考え方について認識精度を比較した。その結果、訓練サンプルとテストサンプルの背景が大きく異なる場合、視線位置分布に従い情報利得を重み付ける考え方が認識精度を大幅に改善することを確認した。以下、2.で訓練サンプルの背景の偏りについて具体例を紹介し、3.で提案手法の詳細について述べる。次に、4.で性別認識の精度を評価し、最後に5.でまとめる。

2. 訓練サンプルの背景の偏り

性別認識の訓練サンプルを収集する際、人物領域の周囲の背景に含まれる物体に注意する必要がある。図1のように、特定物体が背景へ含まれる場合を考察する。男性の訓練サンプルには画像下部に特定物体が含まれており、女性の訓練サンプルにはその物体は含まれていないとする。このような状況は、例えば商

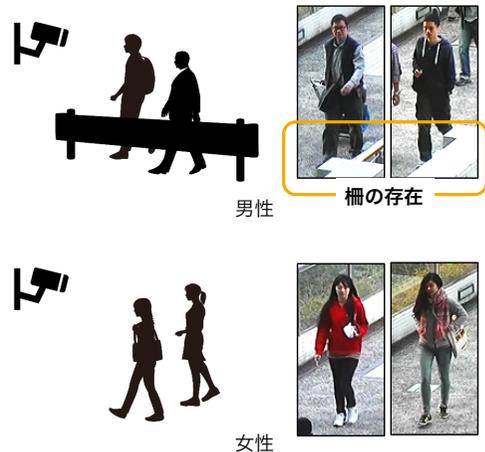


図1 背景に偏りが含まれる訓練サンプルの例。背景の特定物体として男性の手前に柵が存在する状況を想定する。

Fig. 1 Examples of training samples containing the bias in background. We assume a situation that there is a fence in front of men as a specific object in the background.

業施設において、あるカメラ付近には女性が多く出現し、別のカメラ付近には男性が多く出現する場合に発生すると考えられる。偏りが含まれる訓練サンプルを用いて性別認識のためのランダムフォレストを学習した場合、人物領域の違いではなく、特定物体の有無を性別の特徴と誤ってしまうことを予備実験で確認している。学習されたランダムフォレストへ、背景に特定物体が含まれている女性のテストサンプルを入力した場合、本来は女性であるのにも係わらず誤って男性と判定される確率が高いことを予備実験で確認している。この問題を避けるためには、背景差分やセグメンテーションにより人物領域のみを抽出することが考えられる。ただし人物領域のみを正確に抽出することは難しいため、様々な背景を含む訓練サンプルを収集しておくことが望ましい。訓練サンプルの背景が明らかに偏っている場合、画像撮影の収集計画に修正を加えればよいが、事後の学習結果から意図しない訓練サンプルの偏りが発見される場合も存在する。訓練サンプルの収集は非常に手間がかかるため、収集された訓練サンプルを有効に活用することが望ましい。よって本論文では、前節でも述べたように、視線位置分布を用いた重み付けランダムフォレストを適用することで、背景に偏りを含む訓練サンプルでも性別を正しく識別することを狙う。

3. 提案手法

3.1 既存のランダムフォレストの流れ

ランダムフォレスト [8] の学習アルゴリズムについて概要を述べる。まず、訓練サンプルの集合にブートストラップサンプリングを適用することで複数のサブセットを生成する。次に、それぞれのサブセットから決定木を生成する。ここで、人物画像の位置 (x_i, y_i) における色 c_i の画素値を特徴量 $f_i(x_i, y_i, c_i)$ と呼ぶ。なお画像認識の分野では、二つの位置から求めた画素値差分 [14], [15] を特徴量とすることが多いが、本論文ではアルゴリズムを単純にするため画素値そのものを特徴量とする。

訓練サンプルの各サブセットから決定木を学習する際、 $f_i(x_i, y_i, c_i)$ をランダムに M 個選択し、しきい値 t_j をランダムに N 個設定することで情報利得を計算する。親ノードにおいて情報利得が最大値をとるように、左右どちらかの子ノードに訓練サンプルを分岐させる。分岐条件の候補となる情報利得 $I_{i,j}(x_i, y_i, c_i, t_j)$ は、式 (1) で表される。

$$I_{i,j}(x_i, y_i, c_i, t_j) = H(S) - \sum_{k \in \{L, R\}} \frac{|S^k|}{|S|} H(S^k) \quad (1)$$

ここで、 S は分岐前の訓練サンプルの集合、 S^k は分岐後の訓練サンプルの集合、 $H(\cdot)$ はエントロピー、 $||$ は集合の要素数とする。ただし、 $S^L \cup S^R = S$ 、 $S^L \cap S^R = \phi$ とし、 ϕ は空集合を表す。訓練サンプルを $s \in S$ とすると集合 S^L, S^R は次式で表される。

$$S^L = \{s | f(x_i, y_i, c_i) \geq t_j\} \quad (2)$$

$$S^R = \{s | f(x_i, y_i, c_i) < t_j\} \quad (3)$$

なおエントロピーは式 (4) で表される。

$$H(S) = -p(a_f) \log(p(a_f)) - p(a_m) \log(p(a_m)) \quad (4)$$

ここで、 $p(a_f), p(a_m)$ は集合 S に含まれる訓練サンプルの中で女性ラベル a_f または男性ラベル a_m の訓練サンプルが含まれる比率を表す。情報利得が最大となる候補を、親ノードから子ノードへの分岐条件とする。決定木の深さが指定された値になるまで分岐を繰り返す。

3.2 視線マップ

提案手法では、人が性別を認識する際の視線位置分

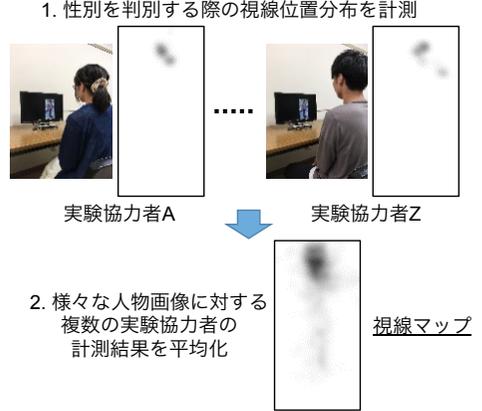


図2 性別認識のタスクを与えた時の視線マップの生成例。
Fig. 2 Generation of a gaze map when a gender classification task is given.

布を、文献 [7] の視線マップを用いて表す。実験協力者に性別認識のタスクを与え、様々な人物画像を提示した際の視線位置を計測する。複数の実験協力者の計測結果を平均化することで視線マップを生成する。ここで視線マップの位置 (x_i, y_i) における値を $g(x_i, y_i)$ で表す。 $g(x_i, y_i)$ が取り得る値の範囲は $[0, 1]$ とし、視線マップの大きさを人物画像と同じとする。図2に性別認識のタスクを与えた時の視線マップの生成例を示す。図中では、視線マップの黒色の領域は、実験協力者からの視線が集まることを表している。逆に白色の領域は視線が集まらないことを表している。文献 [7] では、性別を判別する際の実験協力者の視線は人物画像中の頭部付近へ集まることが報告されている。

3.3 特徴選択を重み付ける考え方

ランダムフォレストの学習アルゴリズムにおいて、特徴選択を重み付ける手法について述べる。以下では、既存のランダムフォレストの学習手法 [8], [13] と対比させながら提案手法を説明する。既存手法 [8], [13] を画像認識に適用する場合、各特徴量に対応する画像中の位置を、一様分布に従ってランダムに選択する。提案手法では、実験協力者の視線が集まる位置の特徴量が選択され易いように改良を加える。即ち、各特徴量に対応する画像中の位置を、一様分布ではなく視線位置分布に従ってランダムに選択する。提案手法の考え方を図3に示す。視線マップの中で視線が集まっていることを表す黒色の領域と同じ位置 (x_i, y_i) にある人物画像の特徴量 $f(x_i, y_i, c_i)$ を選択され易くする。一方、白色の領域ではその逆となり、特徴量を選択され

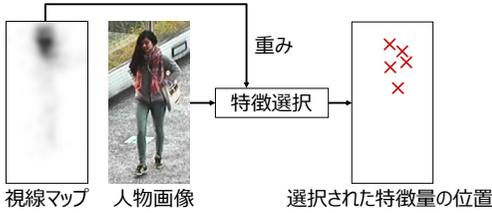


図 3 特徴選択を重み付け.
Fig. 3 Give weights for feature selection.

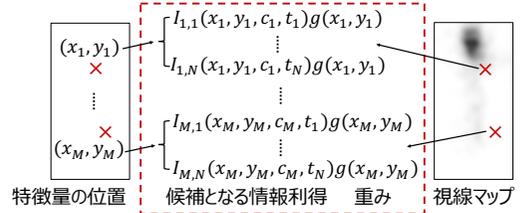


図 5 情報利得を重み付け.
Fig. 5 Give weights for information gains.

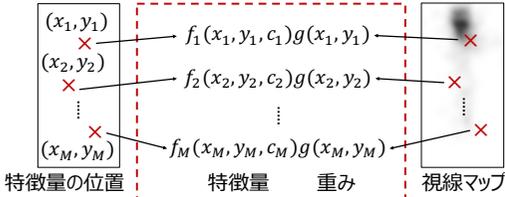


図 4 特徴量を重み付け.
Fig. 4 Give weights for feature values.

難しくする。

3.4 特徴量を重み付ける考え方

ランダムフォレストの学習アルゴリズムにおいて、特徴量そのものを重み付ける手法について述べる。以下では、既存のランダムフォレストの学習手法 [8], [13] と対比させながら提案手法を説明する。既存手法 [8] では、特徴量には何も変更を加えずそのままの値を用いている。既存手法 [13] では、一様分布に従うランダムな重みを特徴量に与えている。一方、提案手法では、位置 (x_i, y_i) に対応する視線マップ $g(x_i, y_i)$ の値で特徴量 $f(x_i, y_i, c_i)$ へ重みを与える。

$$f'(x_i, y_i, c_i) = f(x_i, y_i, c_i)g(x_i, y_i) \quad (5)$$

これにより、視線マップ上で大きな値をもつ位置に対応する特徴量の値は大きくなる。提案手法の考え方を図 4 に示す。なお、上記の手続きにより、文献 [7] で提案されている前処理と等価の結果が得られる。

3.5 情報利得を重み付ける考え方

ランダムフォレストの学習アルゴリズムにおいて、情報利得を重み付ける手法について述べる。以下では、既存のランダムフォレストの学習手法 [8], [13] と対比させながら提案手法を説明する。既存手法 [8] では、分岐条件の候補となるパラメータの組み合わせ $\{x_i, y_i, c_i, t_j\}$ の中から、情報利得が最大となるものを用いている。既存手法 [13] では、一様分布に従うランダム重みを情報利得へ与えた後、分岐条件を決定して

いる。一方、提案手法では、特徴量の位置 (x_i, y_i) に対応した視線マップ $g(x_i, y_i)$ の値で重み付けされた式 (6) の情報利得 $I'_{i,j}(x_i, y_i, c_i, t_j)$ を用いる。

$$I'_{i,j}(x_i, y_i, c_i, t_j) = I_{i,j}(x_i, y_i, c_i, t_j)g(x_i, y_i) \quad (6)$$

ノードの分岐条件のパラメータとして、最大の $I'_{\max}(x_i, y_i, c_i, t_j)$ に対する $\{x_i, y_i, c_i, t_j\}$ を用いる。提案手法の考え方を図 5 に示す。視線マップ $g(x_i, y_i)$ の中で視線が集まっていることを表す黒色領域に対応する位置の $I'_{i,j}(x_i, y_i, c_i, t_j)$ は値が大きくなり、白色領域に対応する位置の $I'_{i,j}(x_i, y_i, c_i, t_j)$ は値が小さくなる。

4. 実験

4.1 評価データセット

提案手法の有効性を確認するために、PETA データセット [16] に含まれる CUHK データセットを用いて性別認識の性能を評価した。背景に偏りを含む訓練サンプルとして、通路に存在する柵の有無を対象とした。訓練サンプルとして以下の条件を用いた。

- T (女性:柵無し, 男性:柵有り)
- また、テストサンプルとして以下の条件を用いた。

- P1 (女性:柵無し, 男性:柵有り)
- P2 (女性:柵有り, 男性:柵無し)

性別認識に用いる訓練サンプルとテストサンプルの例を図 6 に示す。CUHK データセットには同一人物の画像が含まれている。ただし人物 ID が教示されていないため、属性ラベルを用いて同一人物と想定される画像を除外した。全ての属性ラベルが一致する人物画像が複数存在する場合、その中から一枚のみをランダムに選択し実験に用いた。実験に用いたデータセットには、背景に柵無しの女性の画像が 419 枚、柵有りの女性の画像が 355 枚、背景に柵無しの男性の画像が 476 枚、柵有りの男性の画像が 426 枚、が含まれていた。人物画像の大きさは 80×160 画素であり、RGB

の画素値をもつ。

実験に用いた人物画像の位置合わせがどれくらいなされているかを調査した。柵無しの女性、柵有りの女性、柵無しの男性、柵有りの男性の人物画像について、カラーからグレースケールに変換し輝度平均を求めた。その結果を図7に示す。どの条件でも人物シルエットが表れていることから、身体部位という意味での位置合わせはほぼなされていると言える。

視線マップとして、図2下部で示したものを実験で用いた。文献[7]の順に従い視線マップを生成した。訓練サンプルとテストサンプルに含まれていないCUHKデータセットの人物画像8枚を刺激とした。実験協力者14名に対して視線位置分布を計測した。人物画像を表示するために24インチディスプレイ(サイズ53×30cm, 解像度1920×1080画素)を用い、そのディスプレイ下部に視線計測装置GP3 Eye Tracker(サンプリングレート60Hz)を設置した。実験協力者の目からディスプレイまでの距離は水平方向に65cmとした。実験協力者の目の位置は座高の違いにより床からの高さが110cmから120cmの間で変化していた。なお、人物画像の原サイズは80×160画素であるが、ディスプレイ上では480×960画素に拡大して表示した。計測時にディスプレイ中心に視線が集まるセンターバイアスを避けるため、ディスプレイ中心から水平方向に±720画素、垂直方向に±60画素の範囲で人物画像をランダムな位置で表示した。実験前に視線計測の分解能を調査した。カタログに記載されている誤差は約1度であり、ディスプレイ上では42×42画素の領域に相当していた。ディスプレイ上での頭領域の平均解像度は136×145画素であることから、本実験における視線計測の分解能は頭領域の上中下の3段階程度はあると考えられる。

4.2 性別認識の性能

性別認識の性能を、以下の手法で比較した。

- ベースライン(文献[8])
- ランダム重み(文献[13])
- 提案手法1(特徴選択を重み付け)
- 提案手法2(特徴量を重み付け)
- 提案手法3(情報利得を重み付け)

訓練サンプルとテストサンプルを4.1節のデータセットからランダムに選択し5個の評価セットを作成した。各評価セットには、男性323枚と女性323枚の訓練サンプル、男性17枚と女性17枚のテストサンプルが含まれるように設定した。特徴量選択の回数 M を196、



図6 性別認識に用いる訓練サンプル T とテストサンプル P1 と P2 の例。

Fig. 6 Examples of training samples T and test samples P1, P2 for gender classification.

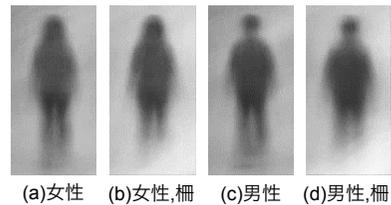


図7 実験で用いた人物画像の輝度平均。

Fig. 7 Examples of training samples T and test samples P1, P2 for gender classification.

ある特徴量に対するしきい値の選択回数 N を50、決定木の数を200、決定木の深さを5とした。

性別認識の精度を図8に示す。図中(a)と(b)の実験結果から、ベースラインとランダム重みの認識精度はほぼ同じであることが分かった。図中(b)の実験結果から、訓練サンプルとテストサンプルで背景が異なる場合、視線マップを組み込むことで性能が向上することが確認された。一方、図中(a)の実験結果から、視線マップを組み込むと精度が低下する場面があることが分かった。ただし、既存手法のベースラインとランダム重みは、次節で明らかにするように柵の有無で

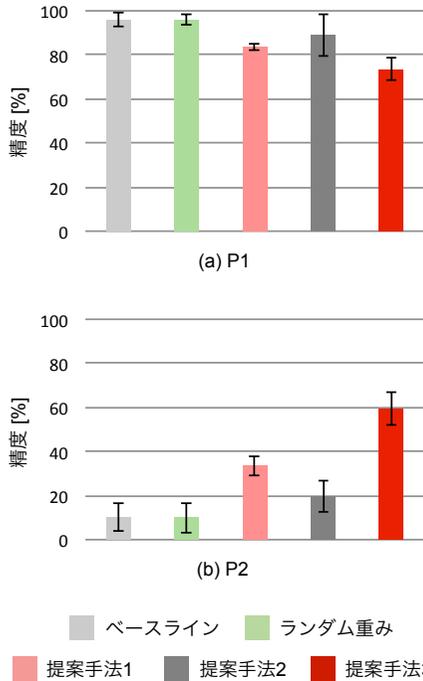


図8 テストサンプル P1 と P2 における性別認識の精度。
Fig. 8 Accuracy of gender classification for training samples P1 and P2.

性別をほぼ見分けているため、これらの既存手法が優れているとは言えない。実験結果を検証するために生成した変数重要度マップについて次節で述べる。

4.3 変数重要度マップを用いた検証

変数重要度マップを以下の手順で生成した。

- (1) 訓練サンプルと同じ大きさの変数重要度マップを準備し各位置 (x_i, y_i) の値を 0 で初期化した。
- (2) 各決定木を学習する際、ブートストラップサンプリングで選択されなかった訓練サンプル (Out-Of-Bag) をバリデーションサンプルとした。
- (3) 学習された各決定木にバリデーションサンプルを入力した。
- (4) 各決定木においてバリデーションサンプルが通過したノードの情報利得 $I_{i,j}(x_i, y_i, c_i, t_j)$ または $I'_{i,j}(x_i, y_i, c_i, t_j)$ を計算した。
- (5) 変数重要度マップ中の位置 (x_i, y_i) の値に、ステップ (4) で求めた情報利得の値を加算した。加算する際は (x_i, y_i) の 8 近傍の位置にも同じ値を加算した。
- (6) 最後に、各位置において加算値の平方根を求めることで、極端に大きい値を抑え小さい値を強調

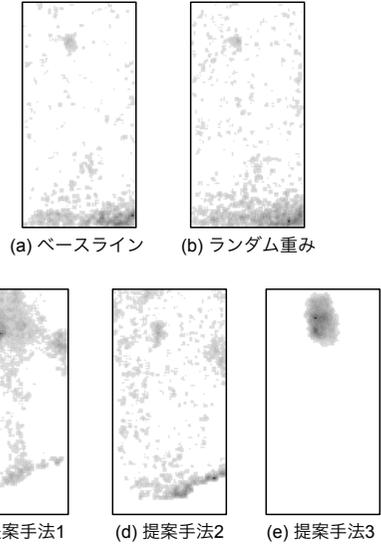


図9 各手法の変数重要度マップ。
Fig. 9 Maps of variable importance for each method.

した。

なお、訓練サンプルとして 4.1 節の条件 T を用いた。

前節の実験で用いたベースライン、ランダム重み、提案手法 1、提案手法 2、提案手法 3 の変数重要度マップを図 9 に示す。図中の黒い領域はランダムフォレストの各決定木で識別時に重視されたことを表し、白い領域は逆に重視されなかったことを表す。図中 (a), (b) の既存手法で、背景の柵が存在する領域に対する重要度が高くなっていることが確認された。この結果より、図 8(a) におけるベースラインとランダム重みの既存手法では、柵の有無で性別を見分けている可能性が非常に高いと言える。図 9(d) の提案手法 2 において、柵の一部の領域に対する重要度が高くなっていた。この原因として、視線マップの重みが 0 より僅かにでも大きい場合、重み付けされた特徴量の値は小さくなったとしても、その範囲の中で情報利得が高くなる分岐条件のパラメータが採用されたためであると考えられる。

図 9(c) の提案手法 1 と (e) の提案手法 3 の変数重要度マップは、人物領域の中でも実験協力者が性別を認識する際に見る頭部近辺に存在する領域を重視していることが確認された。ただし、図中 (c) では、人物の右肩付近の背景領域もランダムフォレストの各決定木において重視されていた。実験で用いた訓練サンプルには、背景の分かりやすい違いである柵の有無だけ

でなく、それ以外の背景にも当初想定していなかった偏りがあることが分かった。以上の結果より、視線位置分布に従うランダム重み付けを導入した提案手法の有効性を確認した。

図 9(e) の変数重要度マップは、実験協力者から計測した視線マップと見え方が近いと言える。実際に視線マップと変数重要度マップとの間で正規化相関を計算したところ、ベースラインとは -0.04 、ランダム重みとは -0.06 、提案手法 1 とは 0.40 、提案手法 2 とは -0.02 、提案手法 3 とは 0.76 であった。情報利得の重みを変更することで、視線マップと変数重要度マップの相関値が高まることが分かった。

4.4 身体部位を含む領域との比較

ランダムフォレスト学習時に視線位置分布を重みとして用いることの有効性を確認するために、身体部位を含む領域を人手で指定した場合と比較実験を行った。文献 [17] で報告されているように、人物画像全体を用いるのではなく、身体部位を含む矩形領域を人手で指定した方が性別認識の精度が高くなる場合がある。ここでの実験では以下の領域を設定することで精度を比較した。

全身: 評価に用いた CUHK データセットの全人物画像から求めた平均人物画像の画素値を二値化することで、全身が含まれる領域を設定した。その例を図 10(a) に示す。

上半身: 文献 [18] で述べられている head-shoulder と torso に従い頭から腰までの身体部位が含まれる領域を全身領域から設定した。その領域の例を図 10(b) に示す。

頭-肩: 同じく文献 [18] で述べられている head-shoulder に従い頭と肩の身体部位が含まれる領域を全身領域から設定した。その領域の例を図 10(c) に示す。

なお図中の黒色領域では画素値を 0 とした。訓練サンプルとテストサンプルの両方に同じ領域を設定し、文献 [8] のベースライン手法でランダムフォレストの学習を行った。領域設定を除いた実験条件は、4.1 節と同じとした。比較のため、4.2 節の実験で最も精度が高かった提案手法 3 を用いた。

身体部位を含む領域を用いた比較手法と提案手法 3 の性別認識の精度を図 11 に示す。図中 (a) のテストサンプルの背景が訓練サンプルの背景と同じである P1 において、提案手法 3 の精度は、全身より低く、上半身と頭-肩より高かった。図中 (b) のテストサンプルの

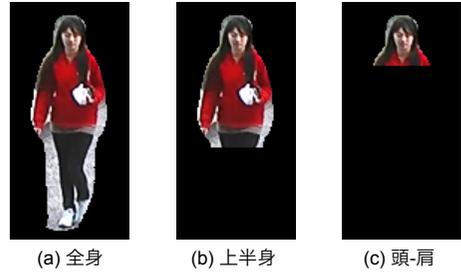


図 10 身体部位を含む領域を設定した人物画像の例。
Fig. 10 Pedestrian images when using regions containing body parts.

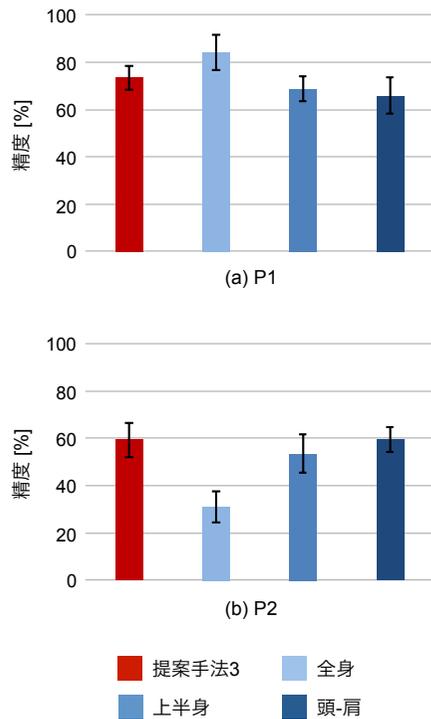


図 11 テストサンプル P1 と P2 において、身体部位を含む領域を用いた手法と認識精度を比較。
Fig. 11 Accuracy of gender classification when using regions containing body parts.

背景が訓練サンプルの背景と異なる P2 において、提案手法 3 の精度は、全身や上半身より高く、頭-肩とほぼ同等であった。頭-肩の領域をマスキングする比較手法により、P2 において認識精度が大きく改善される効果が確認された。ただし、頭-肩を用いる比較手法は P1 において精度の低下がみられたため、視線マップを用いる提案手法にも有効性があると考えられる。

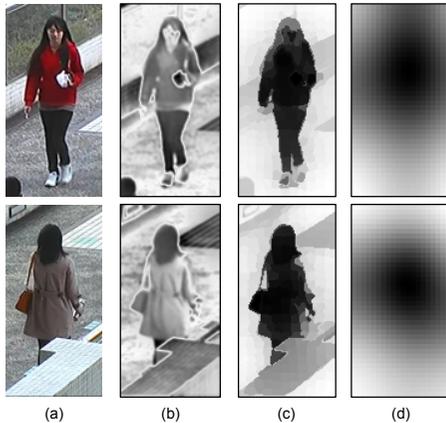


図 12 訓練サンプルから生成した顕著性マップの例。図中 (a) の人物画像に対し、(b) 既存手法 1, (c) 既存手法 2, (d) 既存手法 3 で顕著性マップを生成した。

Fig. 12 Examples of saliency maps acquired from training samples.

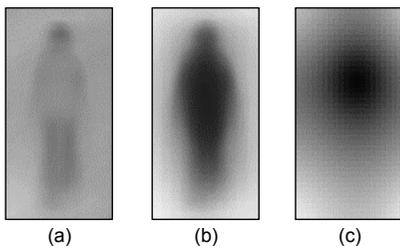


図 13 平均顕著性マップを (a) 既存手法 1, (b) 既存手法 2, (c) 既存手法 3 で生成した結果。

Fig. 13 Average saliency maps.

今後、人物領域セグメンテーションを提案手法へ組み込むことで、有効性のさらなる検証が必要であると言える。

4.5 顕著性マップを用いた場合との比較

画像中で注目が集まり易い領域を推定する手法として顕著性マップが提案されている。ここでは性別認識の精度について、視線マップの代わりに顕著性マップを用いた場合と比較する実験を行った。顕著性マップを生成するために文献[19]の既存手法 1, 文献[20]の既存手法 2, 文献[21]の既存手法 3 を用いた。図 12 に生成された顕著性マップの例を示す。図中では、黒色の領域は顕著性が高く、白色の領域は顕著性が低いことを表す。各訓練サンプルから顕著性マップを生成し、それらの平均から平均顕著性マップを算出した。図 13 に平均顕著性マップを示す。本実験では、視線

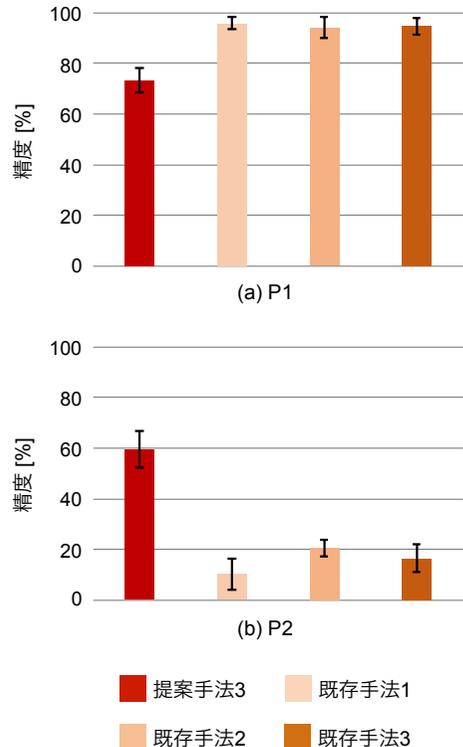


図 14 平均顕著性マップを用いた場合の性別認識の精度比較。

Fig. 14 Accuracy of gender classification when using the average saliency maps.

マップの代わりに平均顕著性マップを用いて情報利得へ重み付けを行った。認識精度の比較対象として、4.2 節の実験で最も精度が高かった提案手法 3 を用いた。訓練サンプル、テストサンプル、および、ランダムフォレストのパラメータは 4.1 節と同じとした。

平均顕著性マップを用いた場合と視線マップを用いた場合の性別認識の精度を図 14 に示す。図中 (a) のテストサンプルの背景が訓練サンプルの背景と同じである P1 において、平均顕著性マップを用いた既存手法 1 から 3 の精度が高かった。一方、図中 (b) のテストサンプルの背景が訓練サンプルの背景と異なる P2 では、視線マップを用いた提案手法の精度が高かった。識別時にどこを重視するか可視化するために、既存手法 1 から 3 の変数重要度マップを生成した。その結果を図 15 に示す。図中 (a) から (c) において、背景の柵が存在している足元領域を重視していたことが確認された。また図中 (c) から、身体領域に加えて右側

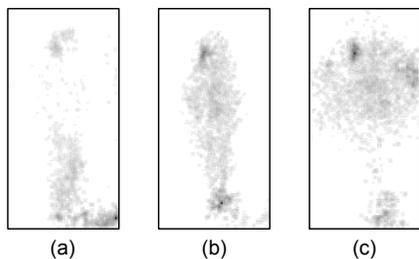


図 15 平均顕著性マップを用いた場合の変数重要度マップ。
Fig. 15 Maps of variable importance when using the average saliency maps.

に存在する背景領域も識別時に重視していたことがなかった。よって図 14(a)の既存手法 1,2,3 では、背景の偏りを性別認識の特徴と誤って学習したと考えられる。これらの結果より、実際に人間から計測された視線マップを用いた場合は、平均顕著性マップを用いた場合と比較して、性別認識に有効な特徴を学習できると言える。

4.6 畳み込みニューラルネットワークを用いた手法との比較

畳み込みニューラルネットワーク (CNN) に視線プーリングを組み込む文献 [5] の手法と認識精度を比較した。訓練サンプル数が非常に少ないため浅いネットワーク構造である文献 [22] の *Mini-CNN* を用いた。前段に畳み込み層を 2 層設け、それぞれの間に最大プーリング層を配置し、後段に全結合層を 2 層設けた。全結合層の前に配置された最大プーリング層から出力される特徴マップに対して、視線マップを用いて重みを与えることで視線プーリングを適用した。なお、文献 [5] で用いられているグローバル平均プーリングは精度低下がみられたため本実験では用いないこととした。比較対象として、4.2 節の実験で最も精度が高かった提案手法 3 の結果を用いた。

視線プーリングを CNN に適用した場合の性別認識の精度を表 1 に示す。訓練サンプル、テストサンプルとして、4.1 節の T, P1, P2 を用いた。まず、視線プーリングを用いた場合と用いない場合の CNN の精度を比較する。テストサンプルの背景が訓練サンプルの背景と同じである P1 では、視線プーリングを用いることで精度低下が見られた。一方、テストサンプルの背景が訓練サンプルの背景と異なる P2 では、視線プーリングを用いることで精度が大きく向上していた。さらに P2 において提案手法 3 では、視線プーリングを

表 1 畳み込みニューラルネットワークに視線プーリングを組み込んだ手法との認識精度 (%) の比較。

Table 1 Accuracy (%) of convolutional neural network using gaze pooling.

手法	P1	P2
視線プーリングなし CNN	89.0 ± 9.2	31.4 ± 9.2
視線プーリングあり CNN	79.3 ± 4.5	51.7 ± 9.4
提案手法 3	73.5 ± 4.9	59.4 ± 7.3

用いた CNN と比べて、認識精度が改善されていた。本実験で用いたデータセットのようにサンプル数が非常に少ない場合は、視線マップを用いた重み付けランダムフォレストは CNN と比較して有効であるケースが存在することが確認された。

4.7 異なる特徴量を用いた場合との比較

前節までは画素値をランダムフォレストの特徴量としていたが、本節では異なる特徴量を評価した。性別認識の特徴量として文献 [23] で用いられているエッジマップと histogram of oriented gradients (HOG) を比較対象とした。各特徴量について以下で説明する。
画素値: 各画素の RGB 値を特徴量とした。4.2 節の実験で最も精度が高かった提案手法 3 で用いた特徴量である。

エッジマップ: Canny 法により抽出された各画素の勾配強度を特徴量とした。勾配強度を用いて算出された情報利得に対して、提案手法 3 と同様に視線マップの重みを適用した。

HOG: 勾配ヒストグラムを計算する際、1 個のセルの大きさを 8×16 画素とし、1 個のブロックに含まれるセル数を 2×2 個とした。勾配ヒストグラムの各ビンの頻度を特徴量とした。ビンの頻度から算出された情報利得に対して、視線マップの重みを適用した。重みを決定する際、勾配ヒストグラムに含まれるセルに対応する視線マップの領域において重みの平均を求めた。

異なる特徴量を用いた場合の性別認識の精度を表 2 に示す。訓練サンプル、テストサンプルとして、4.1 節の T, P1, P2 を用いた。テストサンプルの背景が訓練サンプルの背景と同じである P1 では、視線マップによる重み付けを行わない方が、どの特徴量でも認識精度が高かった。これは、背景の偏りを性別認識の特徴と誤って学習したためと考えられる。一方、テストサンプルの背景が訓練サンプルの背景と異なる P2 では、視線マップにより重み付けを行うことで、どの特徴量でも認識精度が向上していた。これは、背景よりは頭部付近の特徴量から算出された情報利得を重視し

表 2 異なる特徴量を用いた場合の性別認識の精度 (%).
Table 2 Accuracy (%) when using different features.

特徴量	視線マップ	P1	P2
画素値	あり	73.5 ± 4.9	59.4 ± 7.3
	なし	95.9 ± 3.0	10.0 ± 6.3
エッジ	あり	56.5 ± 8.2	58.3 ± 5.7
	なし	70.0 ± 2.2	44.1 ± 4.2
HOG	あり	75.9 ± 11.4	45.3 ± 9.9
	なし	97.1 ± 1.9	14.1 ± 5.7

たためと考えられる。ランダムフォレストに入力する特徴量を変更したとしても、背景に偏りがある場合は視線マップが有効であることを確認した。

5. む す び

本論文では、人物画像の性別を精度よく認識するために、ランダムフォレストの学習アルゴリズムに視線位置分布を重みとして組み込む手法について述べた。訓練サンプルとテストサンプルの画像の背景が異なる場合、視線位置分布を重み付けランダムフォレストの情報利得に組み込むことで精度が大幅に向上することを実験で確認した。

今後の課題として、様々なデータセットでの精度検証、性別以外の属性での評価実験、ランダムフォレスト以外の学習手法への組み込みなどが挙げられる。

謝辞 本研究の一部は、総務省SCOPE(No. 172308003)およびJSPS科研費JP17K00238の助成を受けたものである。

文 献

- [1] P. Sudowe, H. Spitzer, and B. Leibe. Person attribute recognition with a jointly-trained holistic cnn model. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 87–95, 2015.
- [2] T. Matsukawa and E. Suzuki. Person re-identification using cnn features learned from combination of attributes. In *Pattern Recognition 23rd International Conference on*, pp. 2428–2433, 2016.
- [3] A. Schumann and R. Stiefelhagen. Person re-identification by deep learning attribute-complementary information. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1435–1443, 2017.
- [4] Y. Lu, A. Kumar, S. Zhai, Y. Cheng, T. Javidi, and R. Feris. Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1131–1140, 2017.
- [5] H. Sattar, A. Bulling, and M. Fritz. Predicting the category and attributes of visual search targets using deep gaze pooling. In *Proceedings of IEEE International Conference on Computer Vision Workshops*, pp. 2740–2748, 2017.
- [6] N. Murrugarra-Llerena and A. Kovashka. Learning attributes from human gaze. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pp. 510–519, 2017.
- [7] 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*, Vol. 112, pp. 241 – 248, 2018.
- [8] L. Breiman. Random forests. *Machine learning*, Vol. 45, No. 1, pp. 5–32, 2001.
- [9] A. Verikas, A. Gelzinis, and M. Bacauskiene. Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, Vol. 44, No. 2, pp. 330 – 349, 2011.
- [10] L. Rokach. Decision forest: Twenty years of research. *Information Fusion*, Vol. 27, pp. 111 – 125, 2016.
- [11] D. Amaratunga, J. Cabrera, and Y. S. Lee. Enriched random forests. *Bioinformatics*, Vol. 24, No. 18, pp. 2010 – 2014, 2008.
- [12] S. J. Winham, R. R. Freimuth, and J. M. Biernacka. A weighted random forests approach to improve predictive performance. *Statistical Analysis and Data Mining*, Vol. 6, No. 6, pp. 496–505, 2013.
- [13] J. Maudes, J. J. Rodríguez, C. García-Osorio, and N. García-Pedrajas. Random feature weights for decision tree ensemble construction. *Information Fusion*, Vol. 13, No. 1, pp. 20 – 30, 2012.
- [14] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky. Hough forests for object detection, tracking, and action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 11, pp. 2188–2202, 2011.
- [15] V. Lepetit, P. Laguerre, and P. Fua. Randomized trees for real-time keypoint recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 775–781, 2005.
- [16] Y. Deng, P. Luo, C.C. Loy, and X. Tang. Pedestrian attribute recognition at far distance. In *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 789–792, 2014.
- [17] M. Li, S. Bao, W. Dong, Y. Wang, and Z. Su. Head-shoulder based gender recognition. In *Proceedings of IEEE International Conference on Image Processing*, pp. 2753–2756, 2013.
- [18] Bo Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *Proceedings of Tenth IEEE International Conference on Computer Vision*, Vol. 1, pp. 90–97, 2005.

- [19] S. Montabone and A. Soto. Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image and Vision Computing*, Vol. 28, No. 3, pp. 391 – 402, 2010.
- [20] W. Zhu, S. Liang, Y. Wei, and J. sun. Saliency optimization from robust background detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2814–2821, 2014.
- [21] M. Kummerer, T. S. A. Wallis, L. A. Gatys, and M. Bethge. Understanding low- and high-level contributions to fixation prediction. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4799–4808, 2017.
- [22] 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.
- [23] 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.

(平成 xx 年 xx 月 xx 日受付)

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)



山口紗也加

2018年鳥取大学工学部知能情報工学科卒業。2018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。



西山 正志 (正員：シニア会員)

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。山下記念研究賞など受賞。電子情報通信学会、情報処理学会各会員。



岩井 儀雄 (正員)

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