

ランダムフォレストを利用した
パーツベースオンライン学習による人物追跡
Human Tracking By Parts-Based Online Learning Using Random Forests

中島 健太郎† 梅澤 猛† 大澤 範高†
Kentaro Nakajima Takeshi Umezawa Noritaka Osawa

1. はじめに

ランダムフォレストは統計的学習法による人物検出の機械学習手法として、大量のサンプルに対しても効率的に学習が可能な手法として知られている。三井らの手法[1]はあらかじめ用意したデータセットを人体パーツに分割して勾配特徴量を学習し、マルチ識別器を作成することによって、オクルージョンに頑健な手法である。どこにオクルージョンが発生するかが未知でも検出精度を保つことができるため、実環境での人物検出に有効であると考えた。しかし実環境では人物の服装、歩行による人物像の変化も考えられ、特定人物を追跡する場合、オフライン学習のみでは十分でない場合が多い。そこで勾配特徴に加えて、時間方向の特徴量である歩容特徴量[2]も用いたオンライン学習手法を提案する。ランダムフォレストは学習する特徴量をランダム選択するという特性から、ノイズがある学習サンプルでも効率的に学習できることで知られているため、歩容特徴量をランダムフォレストで学習することで、これまでノイズの多さから無視されていた高周波数成分を用いて識別することができると考えている。より高精度な個人識別が可能となり、服装や人物像の変化にほとんど依存しない識別が期待できる。また、逐次的に学習をすることで、あらかじめ対象人物のデータセットを用意しなくとも、個人識別が可能となる。

2. 提案手法

提案手法はオフライン学習とオンライン学習の2つのステップからなる。まず画像中の人物を検出するため、あらかじめオフライン学習を行うことで識別モデルを1つ構築する。検出された画像群の位置情報を用いることで、フレーム間の同一人物判定をし、オフライン学習で構築されたモデルをもとにオンライン学習を行うことで人物ごとにモデルを更新し、出現する人数分新しいモデルを作成する。

2.1. オフライン学習

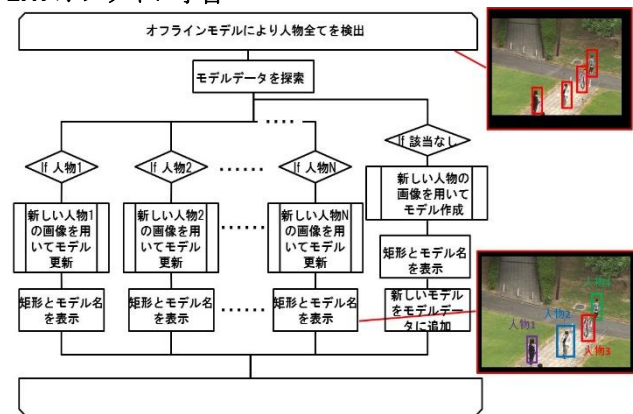


図1: 提案手法概要

2.1.1. HOG 特徴量

まずフレーム内の人物全てを検出するため、人物が共通して持つような特徴量を学習する必要がある。HOG 特徴量[3]は、画像の局所領域から輝度勾配・輝度強度を取り出す特徴量であり、物体の大きな形状を表現可能である。歩行者や人工物など、物体を検出するために広く使われており、人物検出の分野では高い精度が認められている特徴量である。まず、画像領域をあらかじめ設定したピクセルごとに分割し、それをセルと呼ぶ。次に、そのセルをいくつか組み合わせるものをブロックとし、このブロック内のセル内の勾配方向 $\theta(x, y)$ ごとの勾配強度 $m(x, y)$ を式(1), (2)により算出する。ここで (x, y) は勾配強度と勾配方向を求めるピクセルの座標を表し、 $f_x(x, y)$, $f_y(x, y)$ は (x, y) の x 軸方向、 y 軸方向の傾きで、輝度値を $L(x, y)$ としたとき式(3)で求められる。ここで求められた輝度勾配を0から180°まで20°間隔で9つに分割してヒストグラムを作成する。このヒストグラム $v(n)$ を正規化する式は、1ブロックに含まれる縦、横のセルの数を s 、正規化後のヒストグラムを $v'(n)$ としたとき式(4)のようになる。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (1)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (2)$$

$$\begin{cases} f_x(x, y) = L(x+1, y) - L(x-1, y) \\ f_y(x, y) = L(x, y+1) - L(x, y-1) \end{cases} \quad (3)$$

$$v'(n) = \frac{v(n)}{\sqrt{(\sum_{k=1}^{s \times s \times N} v(k)^2) + 1}} \quad (4)$$

2.1.2. 学習方法

オフライン学習ではパーツベース手法を基にモデルを構築する。学習サンプルをパーツ分割し、それぞれのパーツにおいてHOG特徴量を計算し、ランダムフォレストで学習することによってパーツごとに判別可能なマルチ識別器を1つ作成する。パーツごとに出力された値を平均することで最終的な出力値とする。

2.2. オンライン学習

オフライン学習により作成された1つの識別器を用いて入力画像から人物領域の画像を得る。得られた人物画像を利用し、図1のように各人物ごとに学習モデルの更新を行う。作成された各人物ごとのモデルは保持し、新しい人物が出現するたびに増えていく。つまりオンライン学習によりフレーム内に出現した人物の数だけ、モデルを作成することになる。

2.2.1. 新規データの作成

1フレームごとに更新を行うため、ただ新規データを入力するのみだと1枚の画像しか新規データがないことになり、モデルへの貢献度が少ないと考えられる。ここで、新規データは所与の時間間隔(フレーム間隔)で発生する離散的な事象であるので、 i 番目のフレームの画像から得られた特徴量を v_i 、その画像が人か人ではないかを表す教師データを t_i とし、ポアソン分布に従って新規データ(v_i, t_i)が発生していると仮定することでフレームごとに新規データを k 枚だけポジティブサンプルとネガティブサンプルを生成する。ポジティブサンプルは検出された人物領域に加え、その領域を中心に少しだけずらした画像を用い、ネガティブサンプルは入力フレーム中の人物領域以外の場所からランダムに選択する。ここで、 k の確率変数 $P(k)$ は式(5)で表され、パラメータ $\lambda=1$ のポアソン分布に従う。

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!} (\lambda=1) \quad (5)$$

2.2.2. 歩容特徴量

オンライン学習における v_i にはHOG特徴量のみではなく、歩容特徴量も含む。人それぞれで特徴のある歩容特徴量を追加学習することによって個人識別が可能なモデルを構築することができる。なおこの時、後述する汎化誤差推定フェーズのために識別用のデータを k 枚の新規データ(v_{ij}, t_{ij}) ($0 \leq j \leq k$) 除いておく。このデータをOOB(Out of Bag)データと呼ぶ。歩容特徴量の算出は大きく以下の3ステップからなる。

GSV(Gait Silhouette Volume)の作成

まず、各ピクセルの値が $[0,1]$ となる歩行シルエット画像群(GSV)を作成する。入力画像から人物領域矩形を取り出し、あらかじめ設定した大きさの形中心に人物領域が収まるように調整をする。最後にシルエット画像群に時間方向動平均フィルタをかける。図2はGSVの例である。

歩行周期の検出

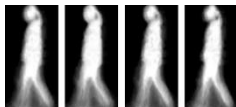


図2: GSV

次に、作成したGSVから歩行周期を計算する。歩行周期 N_{gait} は式(6)~(7)を用い、GSVの時間軸方向の正規化自己相関を計算することで算出する。ここで $C(N)$ は N フレーム分シフトしたときの正規化自己相関で、 $T(N)$ は正規化自己相関を計算する最終フレーム番号、 $g(x, y, n)$ は n フレーム目の空間位置 (x, y) のGSVの値である。また N_{total} はGSVの総フレーム数で、 N_{min}, N_{max} は走査するシフト量の範囲で、あらかじめ設定する。

$$T(N) = N_{total} - N - 1 \quad (6)$$

$$C(N) = \frac{\sum_{x,y} \sum_{n=0}^{T(N)} g(x,y,n)g(x,y,n+N)}{\sqrt{\sum_{x,y} \sum_{n=0}^{T(N)} g(x,y,n)^2} \sqrt{\sum_{x,y} \sum_{n=0}^{T(N)} g(x,y,n+N)^2}} \quad (7)$$

求めた $C(N)$ ($N_{min} < N < N_{max}$) から最大値 $M(n)$ ($n \in N$) とし、式(8)により歩行周期 N_{gait} を計算する。ここで N_m は $M(n)$ のデータ数である。

$$N_{gait} = \frac{\sum_k M(k)}{N_m} \quad (8)$$

周波数特徴の抽出

周波数特徴抽出のためには、まず全体シーケンス S から歩行周期 N_{gait} フレームごとにサブシーケンス $\{S_i\}$ ($i = 1, 2, \dots, N_{sub}$)を取り出す。 i 番目のサブシーケンスのフレーム範囲は $[iN_{gait}, (i+1)N_{gait} - 1]$ となる。次に各サブシーケンスの各ピクセルに対して時間軸方向の離散フーリエ変換(DFT)を行い、さらに $G_i(x, y, k)$ の振幅 $|G_i(x, y, k)|$ を歩行周期 N_{gait} で正規化した振幅スペクトル $A_i(x, y, k)$ を式(9), (10), (11)を用いて計算する。ここで ω_0 は歩行周期 N_{gait} に対応する基本角周波数であり、 $g(x, y, k), G_i(x, y, k)$ はDFTの k 倍周波数に対応する成分である

$$G_i(x, y, k) = \sum_{n=iN_{gait}}^{(i+1)N_{gait}-1} g(x, y, k) e^{-j\omega_0 kn} \quad (9)$$

$$\omega_0 = \frac{2\pi}{N_{gait}} \quad (10)$$

$$A_i(x, y, k) = \frac{1}{N_{gait}} |G_i(x, y, k)| \quad (11)$$

2.2.3. 新しい決定木の作成

新しく得られた新規データから得られた特徴量を元にランダムフォレストのアルゴリズムを用いて新しい決定木を1つ作成する。

2.2.4. 不要決定木の削除

新たに作成された決定木を加えて決定木の数は $T+1$ になる。決定木をフレームごとに増やしていくと、識別の処理時間が長くなっていく。そこで、決定木の数を変えないように、2.2.2. であらかじめ識別用に取っておいたOOBデータ I を用いて、決定木ごとに汎化誤差推定を行い、一番精度が低い決定木を削除してランダムフォレストのモデルを更新し、更新を終了する。

3. まとめ

歩容特徴量をランダムフォレストで学習することの利点として、ランダムフォレストの特徴量をランダム選択して学習するという特性を利用して既存の歩容認証では使われていなかった高周波成分を用いた人物追跡ができるのではないかと考える。しかし、高周波成分が識別精度にどれほど寄与するのかが明らかになっていないため、今後提案手法と既存の歩容認証において高周波成分を用いた場合との比較実験を行い、ランダムフォレストを用いたことの高周波成分の識別精度への寄与について評価をする予定である。

参考文献

- [1] 三井 相和, 藤吉 弘亘, "Randomized Trees を用いたパーツベースによる人検出法", 動的画像処理実用可ワークショップ 2011(DIA), O3-1, 2011.
- [2] 梶原靖, 佐川立昌, 向川康博, 越後富夫, 八木康史 "周波数領域における方向変換モデルを用いた歩容認証" 情報処理学会研究報告.コンピュータビジョンとイメージメディア(CVIM), 2006(5), 117-124, 2006-01-19
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", IEEE Computer Vision and Pattern Recognition, vol. 1, pp. 886-893, 2005.