

人体シルエットの生成型追加学習による人検出の高精度化

瀬瀬 直也[†] 山内 悠嗣[†] 藤吉 弘亘[†]

[†] 中部大学 〒487-8501 愛知県春日井市松本町 1200

E-mail: [†]tuna@vision.cs.chubu.ac.jp, yuu@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

あらまし 本稿では、3次元人体モデルから生成した学習サンプルを用いた生成型追加学習による物体検出法を提案する。一般的な人検出では、学習サンプル収集時の環境と検出対象の環境が大きく異なる場合において、検出精度が低下するという問題がある。本稿では、一般的なデータベースを用いて汎用性のある識別器を事前学習で構築し、3次元人体の生成モデルに検出対象の環境のカメラ位置や背景などの情報を入力し、その環境に特化した学習サンプルを生成して追加学習を行う。これにより、学習サンプルと検出対象の環境の違いをできるだけ低減させることで、特定の環境に特化した識別器を構築する。評価実験の結果、提案手法は INRIA Person Dataset を用いて学習した手法と比較して FPPW が 0.01 において検出率を約 13.8% 向上させることができた。

キーワード 人検出, 生成型追加学習, 人体シルエット

Improvement in Human Detection by Generation Addition Training of Human Silhouette

Naoya KOUKETSU[†], Yuji YAMAUCHI[†], and Hironobu FUJIYOSHI[†]

[†] Chubu University, 1200 Matsumototyo, kasugai-shi, Aichi, 487-8501

E-mail: [†]tuna@vision.cs.chubu.ac.jp, yuu@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract An object detection method that involves additional learning with a training sample generated by a 3-D human model is proposed. In general human detection, the environment collection of the training sample may differ greatly from the detection environment, resulting in low detection accuracy. In the proposed method, a general classifier is learned in advance, and then camera position, a background image, and other such information on the detection environment is input to generate training samples that are specialized for that environment for use in additional training. That makes it possible to learn a final classifier that is specific to a particular environment. The results of evaluation experiments show that the proposed method increased the detection rate by 13.8% relative to a training method that used the INRIA Person Dataset, with a FPPW of 0.01.

Key words Human detection, Generation addition training, Human silhouette

1. はじめに

画像中から自動的に人を検出する技術は、セキュリティやマーケティングなどの多種多様な分野で実現が期待され、人検出の高精度化に関する研究が数多く提案されている [1]~[7]。近年提案された人検出法は、人の形状に着目した手法 [6], [8], [9] や、形状の関連性に着目した手法 [7], [10], パーツベースの人検出法 [4], [5], [11], 人の動きに着目した手法 [1]~[3], [12], 色情報を利用した手法 [13], [14] など、人の識別に対して有効な特徴量をどのように捉えるかを検討した手法が多い。これらの手法は、できるだけ多種多様の学習サンプルを用意し統計的学習手法を用いて識別器を構築し、人検出を困難とする人の姿勢や

体型、衣服の違いなどの“人の個体差に関する要因”を吸収しつつ、人らしさを捉えることで検出性能の向上に貢献した。しかし、学習データセットと人検出システムが稼働するシーンが異なる場合、人検出性能が低下するという問題がある。この問題を解決するためには、システムが稼働する環境からデータを採取して、識別器を再学習する必要があるが、人検出器を学習するためのデータセットをシーン毎に作成することは大きな労力を必要とし、実際には難しい問題である。

そこで、本稿では特定シーンを考慮した生成型追加学習による人検出法を提案する。提案手法は、カメラの位置や俯角を考慮し、人体モデルから特定シーンに特化した人体シルエット画像を自動生成する。そして、これらの学習データを用いて識別

器を学習し、事前に汎用性のあるデータセットで学習した識別器と組み合わせる。これにより、追加学習に必要とするデータの取得等の手間や学習時間のコストを削減しながらも、特定シーンに頑健な識別器を実現することができる。

2. 関連研究

本章では、提案手法に関連する研究について、2つの観点からまとめる。

シーンを考慮した人検出

シーンを考慮することで、高精度な人検出を実現した手法が報告されている。Hoiem らは、画像から検出した人や推定したジオメトリ情報、推定した画像中の水平線より、人の位置と大きさの関係を取得し、人検出に利用することで検出性能を向上させた[15]。また、Li らは、設置したカメラの俯角を考慮することで、検出性能を向上させた[16]。カメラの俯角が大きいほど、画像中での人の見え方が大きく変化するため、このようなシーンにおいては俯角の影響を考慮するアプローチは有効だと考えられる。他にも、オンライン学習を利用した人検出法[17]が提案されており、オンラインでサンプルを収集するアプローチもシーンを考慮した一手法と言える。

これらの手法は、シーンを考慮することで、人検出の高精度化を実現しているが、カメラパラメータもしくはシーンに含まれる背景情報のどちらかのみを考慮している。提案手法は、人体モデルとシーンの背景情報を直接利用することで、両方について考慮することが可能であるため、より高精度な人検出が期待できる。

生成型学習

学習に利用するサンプルを自動的に生成する手法[18]～[20]が提案されている。これらの手法は、実環境下において起こりえる変化を想定してサンプルを生成する。これにより、容易に大量の学習サンプルを得ることが可能である他、実画像で観測される可能性があるノイズも考慮しているため、これらのサンプルを用いて学習した識別器は、高い汎用性が得られると考えられる。しかし、従来の手法で生成するサンプルの対象は、標識等の比較的単純な2次元パターンのみである。提案手法は、複雑な形状を持つ人体のシルエット画像を生成するために、3次元の人体モデルを利用する。3次元の人体モデルの歩行動作中のシルエット画像を抽出することで、自然で自由な姿勢を得ることが可能となる。

3. 人体シルエット画像の生成モデル

提案手法は、人体モデルから特定シーンに特化した学習データを自動生成するために、3次元の人体モデルを用いる。図1に示すような人体モデルを用いることで、様々な視点からの自然で自由な姿勢の人体シルエット画像を得ることができる。

3.1 人体モデル

提案手法で使用する人体モデルには、形状モデルやモデルの各パーツの階層構造、動作データなどが含まれている。人体の形状モデルは、19のパーツが存在し、これらのパーツは階層的な構造で表現される。そのため、例えば右肩を動かした場合、

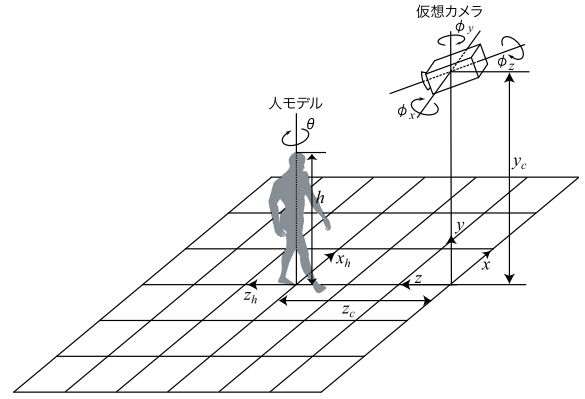


図1 人体モデル

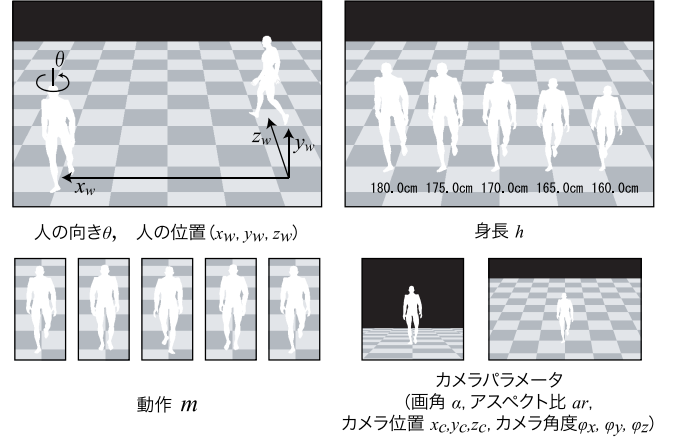


図2 生成モデルのパラメータとこれに対応する人体モデル

右腕や右手が右肩の影響を受けて動く。本研究では、19のパーツに歩行動作のパラメータを与えることで、歩行姿勢としてモデルを表現する。また、下記のパラメータを与えることで図2のように、任意の視点から撮影したモデルの姿勢を得ることができる。

- カメラパラメータ

画角 α , アスペクト比 $aspect$, カメラ位置 x_c, y_c, z_c , カメラ角度 ϕ_x, ϕ_y, ϕ_z

- 人体形状パラメータ

身長 h , 人の向き θ , 人の位置 x_h, y_h, z_h

- テクスチャ

背景のテクスチャ T_{bg} , 人体のテクスチャ T_{in}

上記のパラメータの中で、人の向きと位置は事前に決定できないパラメータであるため、ランダムに設定する。人体のテクスチャは、衣服などを考慮することも考えられるが、様々な種類の衣服を事前に用意する事が難しい。人体のテクスチャを設定しないことも考えられるが、学習に使用する人のサンプルにテクスチャがない場合、テクスチャがないことが人らしさとして学習される可能性がある。そこで、本研究では事前に用意した自然画像のテクスチャをランダムで設定する。また、人の身長に関しては、平均身長 171.9cm とした文部科学省の統計調査に基づいたパラメータを与える。



図 3 実環境を考慮したシルエット画像の生成の例

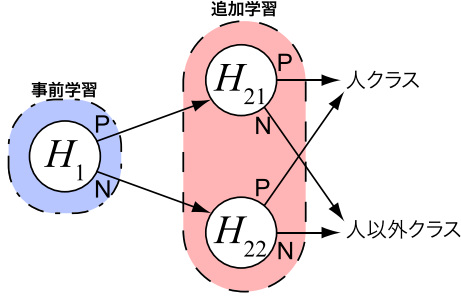


図 4 提案手法により学習した識別器の構成

3.2 特定シーンに特化した人体シルエット画像

特定シーンに特化した人体シルエット画像を得るために、検出対象のカメラに関するパラメータを設定する。これにより、特定シーンに特化した人体モデルを生成することができる。

本研究では、固定カメラを想定し、事前に得たカメラパラメータをモデルに与える。また、設置したカメラから得られる画像を、背景のテクスチャとして用いる。これにより、複雑なテクスチャを持つ背景に対して頑健な検出が可能となる。

図 3 はカメラの高さ $y_c = 5.5m$ 、カメラ角度 $\phi_x = 21^\circ$ でカメラを設置した特定シーンにおける人体シルエットを生成した例である。

4. 生成型追加学習

3.1 にて述べたように、自動的に特定シーンに特化した人体シルエット画像を生成することができる。そこで、容易にサンプルを生成できるメリットを維持しつつ、必要最小限の学習コストにするために、我々は生成サンプルを利用した追加学習法を提案する。提案する生成型追加学習は、2 段階の識別器を用い、その構成を図 4 に示す。1 段階目は、図 5 の Step1 に示すように汎用性のあるデータベースを用いて識別器 H_1 を学習する事前学習である。2 段階目は、図 5 の Step2 のように識別器 H_1 にて誤識別した生成シルエット画像を加えて、図 5 の Step3 に示すように識別器 H_2 を学習する追加学習である。最後に図 5 の Step4 に示すように構築した識別器 H_2 のしきい値を変えることで識別器 H_{21} 、識別器 H_{22} とする。このように、2 段階に識別器を構成することにより、1 段階目の識別器で誤識別をした場合においても、2 段階目の識別器で正しく識別できれば、最終的に正しい出力が得られる。

4.1 事前学習

1 段階目の学習パートである事前学習は、汎用性のあるデータベースを用いて識別器 H_1 を学習する。1 段階目の識別結果は、識別器の出力をしきい値 th_1 により判定した結果となる。

$$H_1(x) = \begin{cases} 1 & H_1(x) > th_1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

1 段階目の学習は、特定シーンを考慮しない事前学習するため、シーンに特化するようなサンプルやパラメータを使用しない。そのため、予め 1 段階目の識別器を学習することにより、識別器を複数のシーンで使うことが可能であり、学習時間を削減することができる。

4.2 生成サンプルを用いた追加学習

2 段階目の学習では、図 4 に示すように、1 段階目の識別器の結果に応じて入力先の識別器が変わるため、2 つの識別器が必要となる。しかし、2 つの識別器の学習するには大きな学習コストが必要となる。そこで、提案手法では次式のように、2 つの識別器は同一の識別器 H_2 を使用し、識別器のしきい値 th_{21} 、 th_{22} を変えることで 2 つの異なる識別器 H_{21} 、 H_{22} とする。

$$H_{21}(x) = \begin{cases} 1 & H_2(x) > th_{21} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$H_{22}(x) = \begin{cases} 1 & H_2(x) > th_{22} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

2 段階目の追加学習は、汎用性のあるデータベースに特定シーンに特化した人体モデルによる生成サンプルを加えたデータを使用する。生成サンプルは、3.1 にて述べたように、複数あるパラメータを変えることでシルエット画像を無制限に生成することが可能である。そのため、本研究では無制限に生成できるポジティブサンプルとネガティブサンプルから、学習に寄与するサンプルを選択するために、事前に学習した識別器 H_1 により識別し、誤識別したサンプルのみを使用する。

4.3 しきい値の自動決定法

ここまで述べてきた学習法により、図 4 に示すような構成の識別器を学習することができる。ここで重要となる点は、識別器 H_1 のしきい値 th_1 である。このしきい値によって、2 段階目の識別器の学習に用いるサンプルが変化するため、提案手法の検出性能に大きな影響を与えるためである。ここでは、適切なしきい値に決定するために、無制限に生成できるサンプルを生かしてしきい値を自動的に決定する。

まず、特定シーンに特化したポジティブサンプルとネガティブサンプルを大量に生成する。次に、事前に学習した識別器 H_1 のしきい値を網羅的に変化させ、全ての生成したサンプルを識別する。そして、最も識別誤差が小さくなるしきい値を、識別器 H_1 のしきい値 th_1 として設定する。これにより、特定シーンに対して頑健なしきい値を自動的に決定することが可能となる。

5. 評価実験

本章では、提案手法の有効性を示すために従来法との比較実

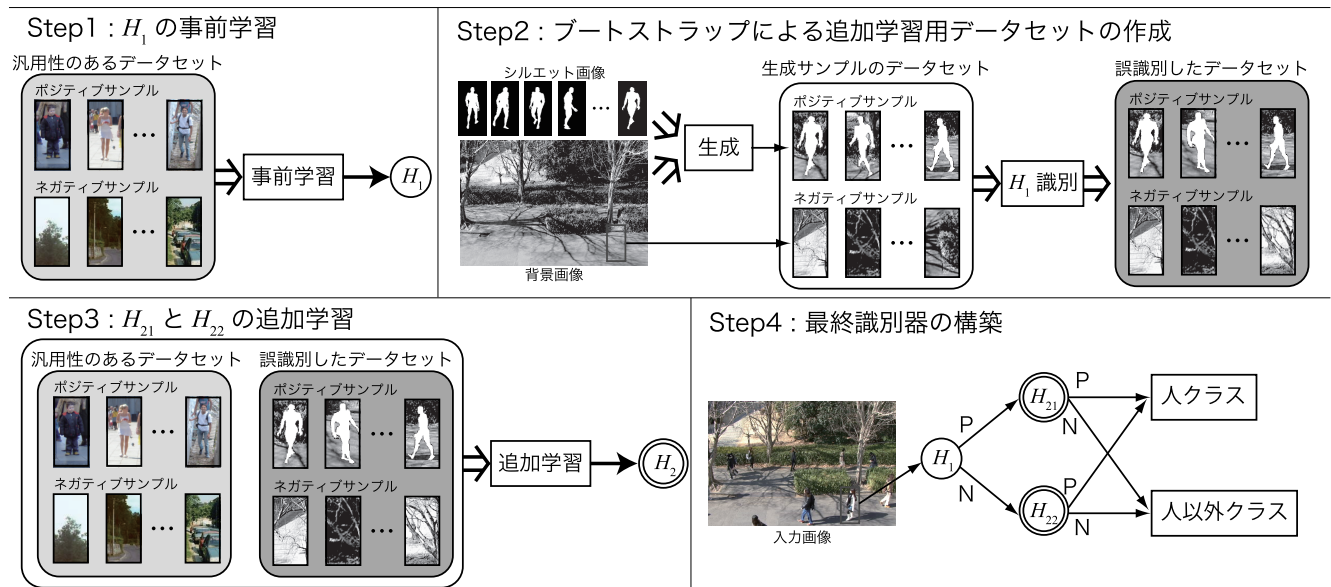


図 5 提案手法による学習の流れ

験を行う。実験は、特定シーンに特化した際の検出性能の評価実験と、提案手法の汎用化性能を確かめる実験、False Positive Per Image による比較の 3 つについて実験する。

5.1 実験概要

評価実験では、下記に示す手法を比較する。

- 事前学習の識別器 H_1 のみ (従来法)
- 追加学習の識別器 H_2 のみ
- 提案手法 H_1, H_{21}, H_{22}

評価実験に用いる特徴量数は 300 個に統一し、比較手法である事前学習の識別器 H_1 のみと、追加学習の識別器 H_2 のみでは弱識別器を 300 個とする。提案手法では事前学習の識別器 H_1 200 個、追加学習の識別器 H_2 100 個とし 300 個の特徴量を用いて実験を行う。一つ目と二つ目の実験の評価は、miss rate と False Positive Per Window を DET カーブにより比較する。これは、1 つの検出ウィンドウ当たりの未検出率を比較することができる。三つ目の実験の評価は、miss rate と False Positive Per Image を DET カーブにより比較する [21]。これは、カメラから得られる 1 枚の画像当たりの未検出率を比較することができる。人検出システムを動作させる場合、1 枚の画像に対して検出ウィンドウをラスタスキャンさせ、人として検出されたウィンドウを統合することで、最終的な結果とする。False Positive Per Window による評価では、検出ウィンドウの統合が含まれない評価であるため、実際のシステム動作時の検出性能とは異なる問題がある。False Positive Per Image による評価は、検出ウィンドウの統合処理を含んでいるため、近年、物体検出法の評価に多く利用されている。今回、検出ウィンドウの統合処理には、Mean Shift クラスタリング [22] を使用した。

5.2 データベース

実験に使用する人画像データベースは、汎用性のある画像データベースとしては、Web 上で一般公開されている INRIA Person Dataset [9] を使用し、シーンに特化した画像データベー

表 1 データセットの内訳 [枚]

データベースの種類	学習用		評価用	
	人	人以外	人	人以外
INRIA	2,415	12,180	1,126	4,530
シーンに特化したデータセット	2,415	12,180	450 フレーム	

スとしては、人体モデルと屋外で撮影した映像から得られる人体シルエット画像を使用する。

INRIA Person Dataset は、人検出法のベンチマークデータベースとして利用され、人の姿勢や向き、視点、照明の変動や背景テクスチャなど、様々なバリエーションを含んだ大規模データセットである。

シーンに特化した画像データベースを作成するために、屋外にカメラを設置し、歩行者を撮影した。比較的人の通行量が多い屋外の通りに、高さ 5.5m、俯角 21° となるようにカメラを設置し、数万フレームの画像を撮影した。生成するサンプルは、人体モデルにカメラの高さと俯角、背景テクスチャを与えることで生成した。表 1 に、学習と評価に使用する画像データベースの枚数を示す。シーンに特化したデータセットの評価画像は、撮影した数万フレームの中からランダムで 450 枚を選択したデータセットである。この 450 枚に対して、予め画像中の人の位置と大きさの真値を与え、評価する際には検出ウィンドウをラスタスキャンさせて、識別結果と真値を照合する。図 6 に、学習に使用したデータセットの一部を示す。

5.3 提案手法の実装

提案手法は、識別器を学習するために画像特徴量と統計的学習手法が必要となる。特に制約がないため、どのような手法を利用してもよいが、本研究では画像特徴量に Histograms of Oriented Gradients(HOG) [9]、統計的学習手法には Real AdaBoost [23] を用いる。生成する人体シルエット画像の人体のテクスチャには、自然画像のテクスチャを用いる。



(a) INRIA ポジティブサンプル



(b) INRIA ネガティブサンプル



(c) 生成 ポジティブサンプル



(d) 生成 ネガティブサンプル

図 6 学習用データセットの例

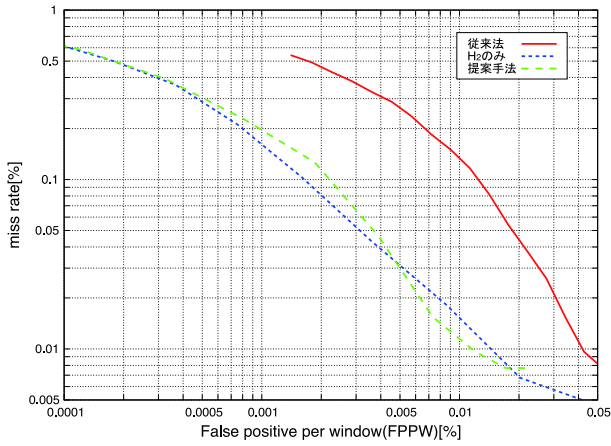


図 7 識別器の性能評価実験結果

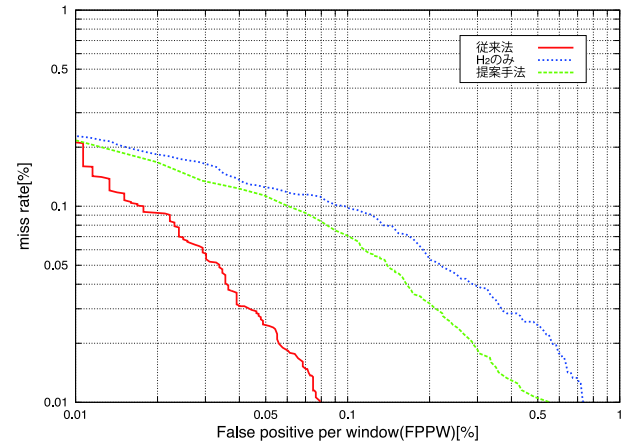


図 8 汎用性の評価実験結果

5.4 実験 1：特定シーンに特化した際の検出性能

図 7 に実験結果を示す。DET カーブでは、原点に近いほど検出性能が良いことを示す。そのため、提案手法が従来法に比べ全体的に検出性能が高いことがわかる。誤検出率 1% の場合、従来法事前学習の識別器のみでは未検出率 15.0% に対し、提案手法では未検出率 1.2% となり、13.8% 検出精度を向上させることができた。

5.5 実験 2：提案手法の汎用化性能

提案手法では特定シーンに特化した識別器を構築することで、特化したシーンに対して検出性能を大きく向上することが確認できた。しかし、特定シーンに特化した識別器を構築するため汎用化性能に影響が出ると考えられる。このため、特定シーンに特化した提案手法の汎用化性能について評価実験を行う。

図 8 に実験結果を示す。結果から従来法より汎用性が低下しているが、提案手法と H_2 のみを比較すると、提案手法が検出精度が高い結果となった。以上より、追加学習した識別器を 2 段階で構築することで汎用性をやや向上させることができた。

5.6 False Positive Per Image による評価

実験 1 で用いたシーケンス 100 フレームを用いて、False Positive Per Image により提案手法を評価する。表 2 に実験結果を示す。提案手法は従来法と比べ、FPPI を約 4.0 倍、miss rate を約 1.6 倍に低減することができた。図 9 に人検出例を示す。結果から従来法に比べ誤検出を抑制しつつ、検出性能が向

上していることがわかる。

表 2 実環境における人検出結果

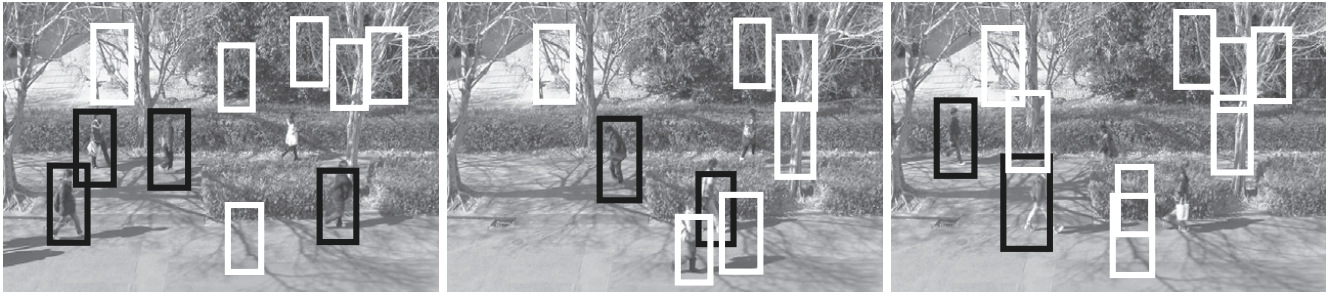
	実画像 100 フレーム	
	FPPI[個]	miss rate[%]
従来法	7.5	56.0
提案手法	1.9	35.1

6. ま と め

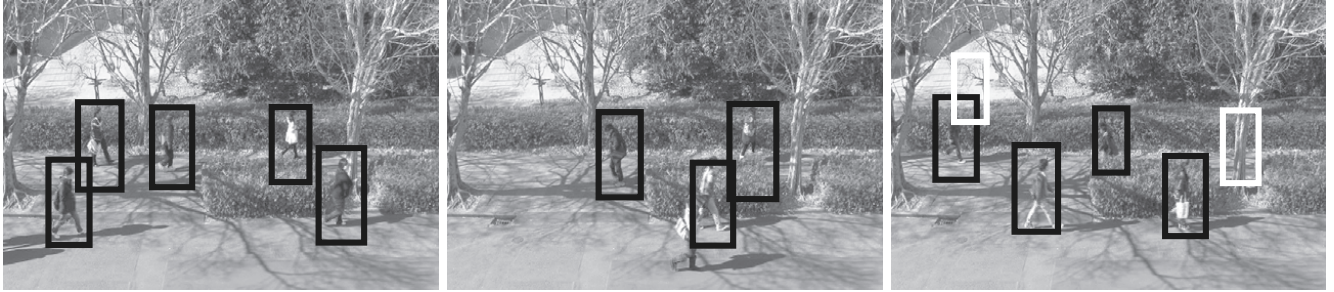
本稿では、特定のシーンを考慮した生成型追加学習による人検出法を提案した。特定シーンに特化した人体シルエット画像を自動生成したデータを用いて、追加学習を行い識別器を組み合わせるにより、特定シーンでの検出性能の向上を確認した。また、生成型追加学習を用いて 2 段階の識別器を構築することで、識別器の汎用性低下を抑制することが可能となり、提案手法の有効性を示した。今後はオンライン学習への展開や、特定シーンにおける人物存在事前確率 [24] を用いて検出性能の高精度化を行う予定である。

文 献

- [1] P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” *International Journal of Computer Vision*, pp.153–161, 2005.
- [2] A. Ess, B. Leibe, K. Schindler, and L. van Gool, “Moving



(a) 従来法



(b) 提案手法

図9 人検出例 (黒枠は正解した検出ウィンドウ, 白枠は誤検出した検出ウィンドウを表わす)

- obstacle detection in highly dynamic scenes,” IEEE International Conference on Robotics and Automation, 2009.
- [3] C. Wojek, S. Walk, and B. Schiele, “Multi-cue onboard pedestrian detection,” IEEE Computer Vision and Pattern Recognition, 2009.
 - [4] B. Babenko, M.H. Yang, and S. Belongie, “Visual tracking with online multiple instance learning,” IEEE Computer Vision and Pattern Recognition, 2009.
 - [5] Z. Lin, G. Hua, and L.S. Davis, “Multiple instance feature for robust part-based object detection,” IEEE Computer Vision and Pattern Recognition, 2009.
 - [6] T. Watanabe, S. Ito, and K. Yokoi, “Co-occurrence histograms of oriented gradients for pedestrian detection,” Proc. Of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology, pp.37–47, 2009.
 - [7] 河合富美, H.C. Keng, “オブジェクトの同一性を考慮したエッジ特徴による人物検出,” 画像の認識・理解シンポジウム (MIRU), pp.2084–2091, 2010.
 - [8] K. Levi, and Y. Weiss, “Learning object detection from a small number of examples: the importance of good features,” IEEE Computer Vision and Pattern Recognition, vol.2, pp.53–60, 2004.
 - [9] N. Dalal, and B. Triggs, “Histograms of oriented gradients for human detection,” IEEE Computer Vision and Pattern Recognition, vol.1, pp.886–893, 2005.
 - [10] B. Wu, and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detections,” IEEE Computer Vision and Pattern Recognition, vol.1, pp.90–97, 2005.
 - [11] P. Felzenszwalb, R. Girshick, and D. McAllester, “Cascade object detection with deformable part models,” IEEE Computer Vision and Pattern Recognition, 2010.
 - [12] N. Dalal, B. Triggs, and C. Schmid, “Human detection using oriented histograms of flow and appearance,” European Conference on Computer Vision, 2006.
 - [13] P. Ott, and M. Everingham, “Implicit color segmentation features for pedestrian and object detection,” IEEE International Conference on Computer Vision, 2009.
 - [14] T. Deselaers, and V. Ferrari, “Global and efficient self-similarity for object classification and detection,” IEEE Computer Vision and Pattern Recognition, 2010.
 - [15] D. Hoiem, A.A. Efros, and M. Hebert, “Putting objects in perspective,” International Journal of Computer Vision, 2008.
 - [16] Y. Li, B. Wu, and R. Nevatia, “Human detection by searching in 3d space using camera and scene knowledge,” International Conference on Pattern Recognition, pp.1–5, 2008.
 - [17] P.M. Roth, H. Grabner, D. Skocaj, and H. Bischof, “On-line conservative learning for person detection,” IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005.
 - [18] D.M. Gavrila, and J. Giebel, “Virtual sample generation for template-based shape matching,” IEEE Computer Vision and Pattern Recognition, pp.676–681, 2001.
 - [19] 野田雅文, 高橋友和, 井手一郎, 目加田慶人, 村瀬洋, “生成型学習を用いた車載カメラ映像中の路面標示認識の検討,” 電子情報通信学会総合大会, p.220, 2008.
 - [20] 道満恵介, 出口大輔, 高橋友和, 目加田慶人, 井手一郎, 村瀬洋, “色変動を考慮した生成型学習による道路標識検出器の構築,” 画像の認識・理解シンポジウム (MIRU), 2009.
 - [21] P. Dollá, C. Wojek, S. Bernt, and P. Perona, “Pedestrian detection: A benchmark,” Computer Vision and Pattern Recognition, pp.1-8, 2009.
 - [22] D. Comaniciu, and P. Meer, “Mean shift : A robust approach toward feature space analysis,” Pattern Analysis and Machine Intelligence, pp.603 – 619, 2002.
 - [23] R.E. Schapire, and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” Machine Learning, pp.297 – 336, 1999.
 - [24] 中河秀仁, 渡部済, 木戸出正繼, “人物の移動軌跡と画像の色情報をを用いた人物存在事前確率の効率的な獲得,” 画像の認識・理解シンポジウム (MIRU), 2009.