

人検出のための生成型学習と Negative-Bag MILBoost による 学習の効率化

土屋 成光[†] 山内 悠嗣[†] 藤吉 弘亘[†]

[†] 中部大学 春日井市

Graduate School of Engineering, Chubu University, 1200 Matsumoto, Kasugai, Aichi, 487-8501 Japan.

E-mail: [†]{tsuchiya,yuu}@vision.cs.chubu.ac.jp, ^{††}hf@cs.chubu.ac.jp

あらまし 人検出に用いられる統計的学習法は大量の学習サンプルを必要とするため、サンプルの収集コストが高い。また、学習サンプルを収集した環境と人検出システムが稼働する環境が異なる場合には、人の見え方が変化するため検出性能が低下することがある。そこで、本稿では 3 次元人体モデルを用いた学習サンプルの自動生成と Negative-Bag MILBoost による生成型学習法を提案する。本研究では、3 次元人体モデルを用いて特定シーンに特化した学習用ポジティブサンプルを自動的に生成する。学習用ネガティブサンプルは、映像からランダムに自動的に切り出して収集するが、人画像を収集することがある。このような誤って付与されたラベルを持つ学習サンプルを用いて学習する場合、識別能力が低下するという問題がある。そこで、本研究では誤ラベルのサンプルが混在しても悪影響を受けない Negative-Bag MILBoost を用いて実現する。評価実験の結果、3 次元人体モデルから生成した学習サンプルを用いて学習した識別器は、人手で切り出した学習サンプルを用いて学習した識別器よりも高い検出性能であることを確認した。また、提案手法は学習用ネガティブサンプル中に人画像が混在した状態においても、識別性能の低下を抑制することができた。

キーワード 3次元人体モデル, Negative-Bag MILBoost, 人検出, 生成型学習

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]~[10]。近年提案された人検出法は、人の形状に着目した特徴量 [4], [5], [6] や、人の動きに着目した特徴量 [1]~[3], [7], 色情報を利用した特徴量 [8], [9], 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。これらの特徴量は、人の姿勢や向き、体格、衣服等による異なる人の見え方となる“人の個体差に関する要因”を吸収しつつ、人らしい形状を捉えることで検出性能の向上に貢献した。しかし、学習用データベースを収集した環境と人検出システムが稼働するシーンが異なる場合、それぞれの環境で観測される人の見え方が異なるために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人画像を収集し、識別器を再学習する必要がある。しかしながら、シーン毎に人検出器を学習するためのデータセットを作成することは大きな労働力と多大な時間を必要とするため実際には難しい問題である。

これらの問題を解決するアプローチとして、少数の学習サンプルからスケール変化や回転、ノイズの付加などの実環境で測定されうる変動を含むように変形させた学習サンプルを生成し、生成したサンプルを用いて識別器を学習する生成型学習法 [14] が提案されている。文

献 [11] では、認識対象である車載カメラから見える路面上の標識に対して、光学ぼけや動きぼけ等を考慮したサンプルを生成して学習に利用した。また、文献 [12] では道路標識に対して、位置ずれや回転などの形状の変化、背景などのテクスチャの変化、反射や影などの色の变化を考慮した生成モデルを用いて学習サンプルを生成した。しかし、これらの手法が認識対象とするのは比較的簡単な 2 次元パターンであり、人のような非剛体で複雑な形状や姿勢を持つ物体に対しては、同様のアプローチで学習サンプルを生成するのは困難である。

そこで、本稿では 3 次元人体モデルを用いた学習サンプルの自動生成と Negative-Bag MILBoost による生成型学習法を提案する。本研究では、文献 [13] と同様に 3 次元人体モデルを用いて特定シーンに特化した学習用ポジティブサンプルを自動的に生成する。学習用ネガティブサンプルは、映像からランダムに自動的に切り出して収集するが、人画像を収集することがある。このような誤って付与されたラベルを持つ学習サンプルを用いて学習する場合、識別能力が低下するという問題がある。そこで、本研究では誤ラベルのサンプルが混在しても悪影響を受けない Negative-Bag MILBoost を用いて実現する。

本稿では、2 章で人体モデルを用いて特定シーンに特化した学習サンプルを生成する方法について述べる。3 章では、誤ラベルが付与された学習サンプルの混在を考慮した Negative-Bag MILBoost による学習について述

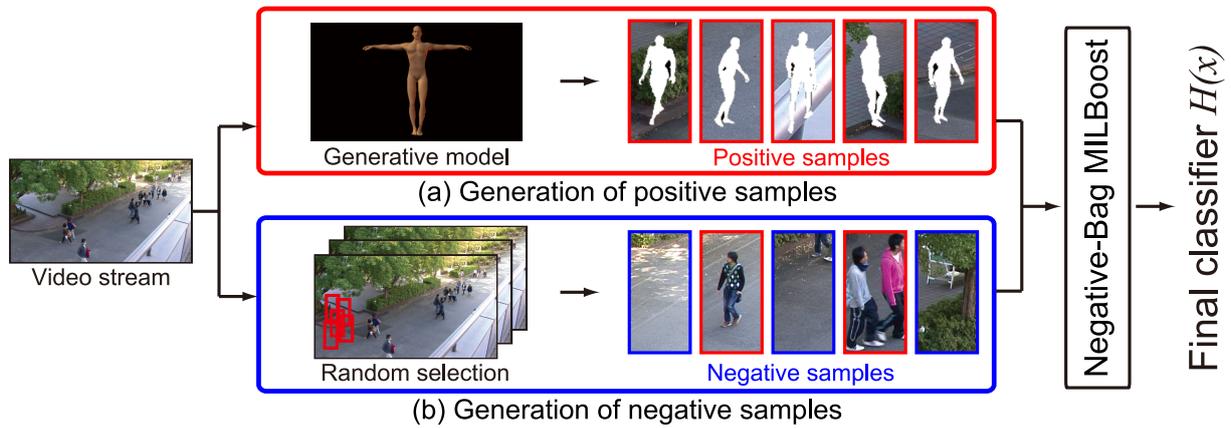


図1 提案手法による生成型学習の流れ.

Fig. 1 Generative learning procedure in the proposed method.

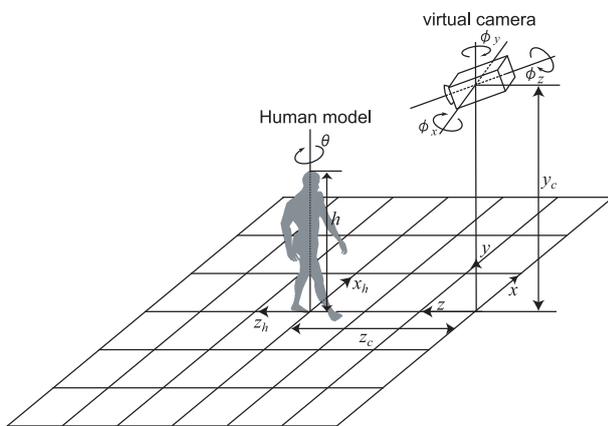


図2 3次元人体モデル.

Fig. 2 3D human model.

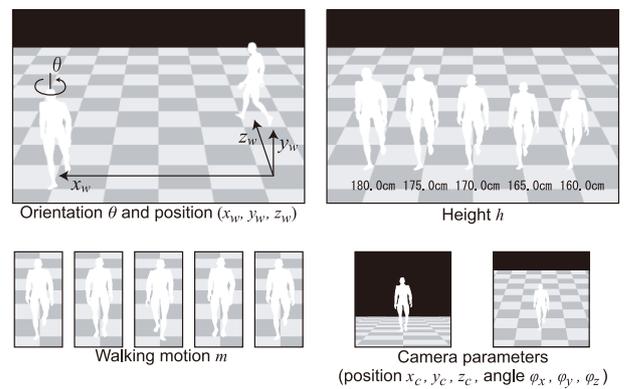


図3 パラメータに対応した人体モデル.

Fig. 3 Adapting the 3D human model to various parameters.

べる。4章で提案手法の有効性を示すために3つの評価実験を行う。

2. 生成モデルを用いた学習用人画像の生成

提案手法による識別器を学習するまでの流れを図1に示す。提案手法は、特定シーンに特化した学習サンプルを自動生成するために、ポジティブサンプルは3次元の人体モデルを用いて人体シルエット画像を生成する(図1(a))。ネガティブサンプルは映像中から切り出す(図1(b))。そして、生成したサンプルをNegative-Bag MILBoostに入力することで識別器を学習する。

学習用ポジティブサンプルの生成には、図2に示すような人体モデルと仮想カメラを用いることで、様々な視点からの自由な姿勢の人体シルエット画像を生成することが可能となる。

2.1 3次元人体モデル

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、右腕や右手が連動して動く。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、歩行姿勢として人体モデルを表現する。また、下記のパラメータを与えることで図3に示すように任意の視点から撮影した人体モデルの姿勢を得ることができる。

- カメラパラメータ
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}

2.2 ポジティブサンプルの生成

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、実環境に設置したカメラのパラメータを3次元人体モデルに入力する。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。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、一様ランダムとする。人の身長は、文部科学省の統計調査より平均身長171.9cmとす

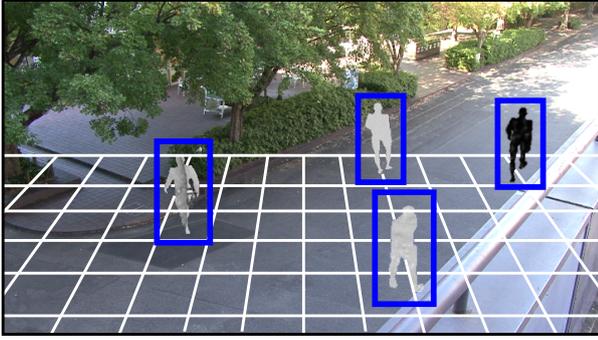


図 4 実環境を考慮したシルエット画像の生成の例 (カメラ位置 $(x_c, y_c, z_c) = (0m, 6.2m, 0m)$, カメラ角度 $(\phi_x, \phi_y, \phi_z) = (21^\circ, 0^\circ, 0^\circ)$).

Fig. 4 Examples of silhouette images generated specifically for the real environment (camera position $(x_c, y_c, z_c) = (0m, 6.2m, 0m)$, camera angle $(\phi_x, \phi_y, \phi_z) = (21^\circ, 0^\circ, 0^\circ)$).

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、多種多様な種類の衣服を用意する事が難しい。また、人体のテクスチャを張り付けないことも考えられるが、このようなサンプルを用いて学習した場合、人の内部はテクスチャがないものとして学習される。しかし、本研究で用いる HOG 特徴量で重要なのは人の形状を捉えることであるため、ネガティブサンプルを十分に用意すれば問題はない。そこで、本研究では人体内部のテクスチャを用意しない。もし色や人体テクスチャに依存する特徴量を使用する場合はそれに合わせたテクスチャの付与が必要である。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。

図 4 はカメラの高さ $y_c = 6.2m$, カメラ角度 $\phi_x = 21^\circ$ とした際の人体シルエット画像を生成した例である。このように生成した人体シルエットを中心に切り出した画像を学習用ポジティブサンプルとして用いる。

2.3 ネガティブサンプルの生成

ネガティブサンプルは、撮影した映像中からランダムで収集する。しかし、ランダムにサンプルを収集した場合、ネガティブサンプルとして人画像が収集される問題がある。この問題を解決するために、本研究では、誤って付与されたラベルを持つサンプルの混在を考慮した Negative-Bag MILBoost により識別器を学習する。

3. 学習サンプルの混在を考慮した MILBoost による学習

本章では誤ラベルが付与されたサンプルが混在してしまう問題を解決するために、Negative-Bag MILBoost による識別器の学習法について述べる。

3.1 MILBoost [16]

物体検出に利用される統計的学習手法は、各サンプルに対して与えられたラベルに基づき学習する。これに対

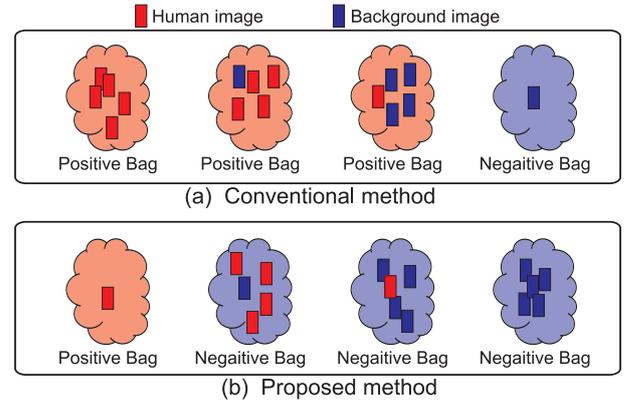


図 5 提案手法による Bag の構成

Fig. 5 Construction of Bag by the proposal method.

して Multiple Instance Learning (MIL) [15] では、複数のサンプルから構成される Bag に対してラベルを付与する。Bag 中のサンプルに一つでもポジティブが含まれていればポジティブ Bag となり、それ以外はネガティブ Bag とラベル付けされる。MIL では、Bag に与えられたラベルに基づき識別器を学習する。そのため、ラベルが付与されていない未知のサンプルを含むデータに対しても学習が可能なアプローチとなる。本研究では、MIL を Boosting に導入した MILBoost [16] をベースとした Negative-Bag MILBoost により学習する。

MILBoost は MIL の学習モデルを Boosting に導入した学習アルゴリズムである。Viola らは MILBoost により効率的に顔検出器を学習する方法を提案した。この手法は、顔周辺を適当にサンプリングすることでポジティブ Bag を作成し、ポジティブサンプルの収集を簡略化した。MILBoost は、Boosting により学習する際、各 Bag と各サンプルに対するクラス尤度を求め、サンプルに対する重みの更新時に、クラス尤度を用いて誤って付与されたラベルのサンプルの重みを小さくする。これにより、誤って付与されたラベルの影響を抑制することが可能となる。

3.2 MILBoost の改良

MILBoost [16] では、ポジティブサンプルのラベルが付与できない問題としているが、提案手法ではネガティブサンプルのラベルが付与できない問題である。そのため、提案手法の問題設定に適用するために MILBoost を改良する。ここでは、まず Bag の作成方法について述べ、次に本研究の問題設定に適用した Negative-Bag MILBoost の学習アルゴリズムについて述べる。

3.2.1 Bag の作成方法

本研究の問題設定では、ネガティブサンプルに正しいラベルを必ずしも与えることができない。そのため、図 5 に示すように文献 [16] の Bag の構成を変更する。提案手法のポジティブ Bag は、人体モデルから生成した人画

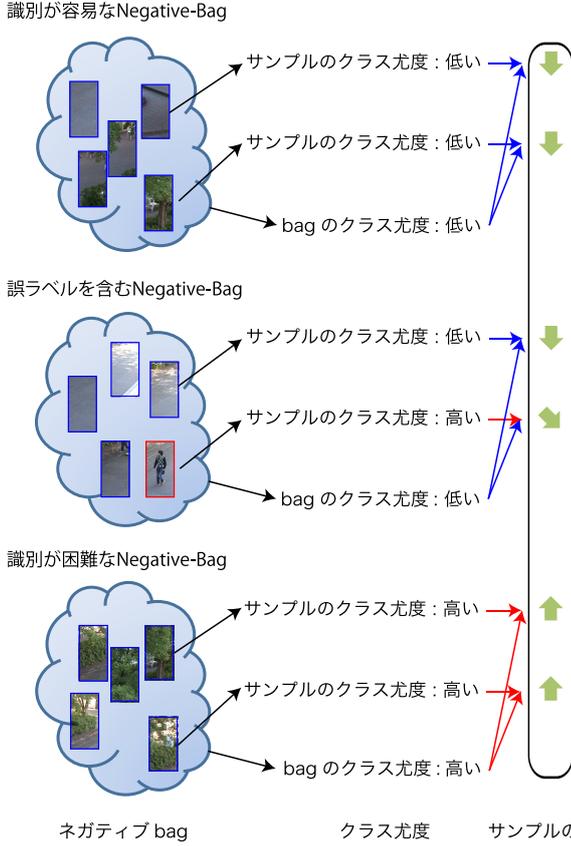


図 6 Negative-Bag MILBoost による学習サンプルの重み付け.

Fig. 6 Update weight of training samples by advanced MILBoost.

像を一つの Bag として扱う。ネガティブ Bag は、映像中から切り出した複数のサンプルを一つの Bag として扱い、この中には人画像が含まれる可能性もある。

3.2.2 学習

MILBoost の学習は、学習サンプルに対する重みの更新以外は Boosting と共通のアルゴリズムである。本研究では、学習サンプルに対する重みの更新処理を本研究の問題設定に適用するように変更する。Real AdaBoost [17] をベースとした Negative-Bag MILBoost の学習アルゴリズムを図 7 に示す。

まず、学習の事前準備として J 個の学習サンプルが所属する I 個の Bag を用意する。そして、これらの Bag に対してラベルを与える。次に、 i 番目の Bag に所属する j 番目のサンプルの重み w_{ij} を式 (1) により初期化する。

次に、Negative-Bag MILBoost により強識別器を学習する。まず、特徴量 x を入力とした確率密度関数 W_+ , W_- を式 (2), (3) により作成する。本研究では、特徴量として Histogram of Oriented Gradients(HOG) 特徴量 [6] を用いる。確率密度関数 W_{\pm} は、1 ブロックの HOG 特徴量に対して得られる 1 次元のヒストグラムにより表現され、学習サンプルの重み $w_t(i, j)$ をそれぞれラベル y 毎に累積することで作成する。ここで t は学習回数、 k は 1 次元ヒストグラムの BIN の番号を表す、学習サン

1. Input

- Assign a correct label $y \in \{1, 0\}$ to I bags containing J learning samples.

2. Initialization

- Initialize the weight $w_t(i, j)$ for the learning samples

$$w_1(i, j) = \frac{\text{Bag of class}}{\text{Bag of all}} \quad (1)$$

3. Training

for $t = 1, 2$ to T [T learning iterations] do

for $l = 1, 2$ to L [L weak classifier candidates] do

- Create the probability density function W_{\pm} of weak classifier candidate $h_t(x)$

$$W_+^k = \sum_{i, j: k \in K \wedge y_i = 1} w_t(i, j) \quad (2)$$

$$W_-^k = \sum_{i, j: k \in K \wedge y_i = 0} w_t(i, j) \quad (3)$$

- Calculate evaluation value Z_l

$$Z_l = 2 \sum_{k=1}^K \sqrt{W_+^k W_-^k} \quad (4)$$

end for

- Select weak classifier $h_t(x)$ with smallest Z_l

$$h_t(x) = \arg \min_{l \in \mathcal{L}} Z_l \quad (5)$$

- Weak classifier $h(x)$

$$h(x) = \frac{1}{2} \ln \frac{W_+^k + \varepsilon}{W_-^k + \varepsilon} (\varepsilon = 1/J) \quad (6)$$

- Update weights $w_t(i, j)$ of learning samples

$$w_{ij} = \begin{cases} -p_{ij} & \text{if } y_i = 1 \\ \frac{p_{ij} \times (p_i)}{1 - p_i} & \text{if } y_i = 0 \end{cases} \quad (7)$$

$$p_i = \prod_{j \in \text{Bag}_i} p_{ij} \quad (8)$$

$$p_{ij} = \frac{1}{1 + \exp(-H_t(x))} \quad (9)$$

$$w_{ij} = \begin{cases} w_{ij} + \min w_+ & \text{if } y_i = 1 \\ w_{ij} + \min w_- & \text{if } y_i = 0 \end{cases} \quad (10)$$

end for

4. Output

- Final classifier $H(x)$

$$H(x) = \text{sign} \left(\sum_{t=1}^T h_t(x) \right) \quad (11)$$

図 7 学習アルゴリズム.

Fig. 7 Training algorithm.

ルの重み $w_t(i, j)$ は、サンプルの最大値を K 分割した際に対応する BIN k の w_t^k に累積される。確率密度関数 W_{\pm} を作成後、式 (4) より弱識別器 $h(x)$ の評価値 Z を求め

表 1 学習用データセットの内訳.
Table 1 Examples of learning data sets.

	Positive	Negative
DB1	INRIA(2,416)	生成 (12,180)
DB2	実環境 (2,416)	生成 (12,180)
DB3	生成 (2,416)	INRIA(12,180)
DB4	生成 (2,416)	生成 (12,180)

る. 評価値 Z は弱識別器の識別性能を表し, 小さいほどポジティブサンプルとネガティブサンプルを分離する能力があることを表している. これを全ての弱識別器候補について評価値を求め, 式 (5) に示すように最も識別性能が高い弱識別器候補を t 個目の弱識別器 $h_t(x)$ (式 (6)) とする.

弱識別器を選択後, 式 (8), (9) から得られる Bag のクラス尤度 p_i とサンプルのクラス尤度 p_{ij} を用いて学習サンプルの重みを更新する. 以上の処理を T 回繰り返すことで, 式 (11) 最終識別器 $H(x)$ を得る.

3.2.3 Negative-Bag を用いた誤ラベルへの対応

図 6 に Negative-Bag MILBoost による学習サンプルの重みの更新の概念図を示す. MILBoost では, サンプルに対するクラスラベルが付与されていないため, 学習サンプルの重み w_{ij} は式 (7) に従い Bag のラベルに基づき更新される. ポジティブ Bag に含まれるサンプルに対しては, サンプルのクラス尤度 p_{ij} により重みを更新する. クラス尤度は, 高い値であるほど人画像, 低い値であるほど背景画像の可能性が高いことを表している.

ネガティブ Bag に含まれているサンプルは, サンプルのクラス尤度 p_{ij} と Bag のクラス尤度 p_i により重みを更新する. サンプルのクラス尤度 p_{ij} の値が総じて低く, 結果, Bag のクラス尤度 p_i の値が低い場合, その各サンプルは識別が容易なネガティブサンプルであるとし Bag 中のサンプルの重み w_{ij} は低下する. また, その中に誤ラベルされたポジティブサンプルが含まれていた場合, Bag のクラス尤度 p_i が十分に低ければ, そのサンプルはノイズであると捉え, サンプルの重み w_{ij} は低下する. 一方, サンプルのクラス尤度 p_{ij} の値が総じて高く, 結果, Bag のクラス尤度 p_i の値が高い場合, その各サンプルは識別が困難なネガティブサンプルであるとし Bag 中のサンプルの重み w_{ij} を向上させ, 重点的に学習を行う. このように, Bag 単位での尤度を用いることでノイズの影響を低減することができる.

3.2.4 識別

識別時は, 従来法である Real AdaBoost と同様に計算する. 式 (11) に示すように複数の弱識別器の線形和をしきい値処理し, しきい値より大きい場合は人, そうでない場合は背景として識別結果を出力する.

4. 評価実験

提案手法の有効性を示すために 2 つの評価実験を行う.

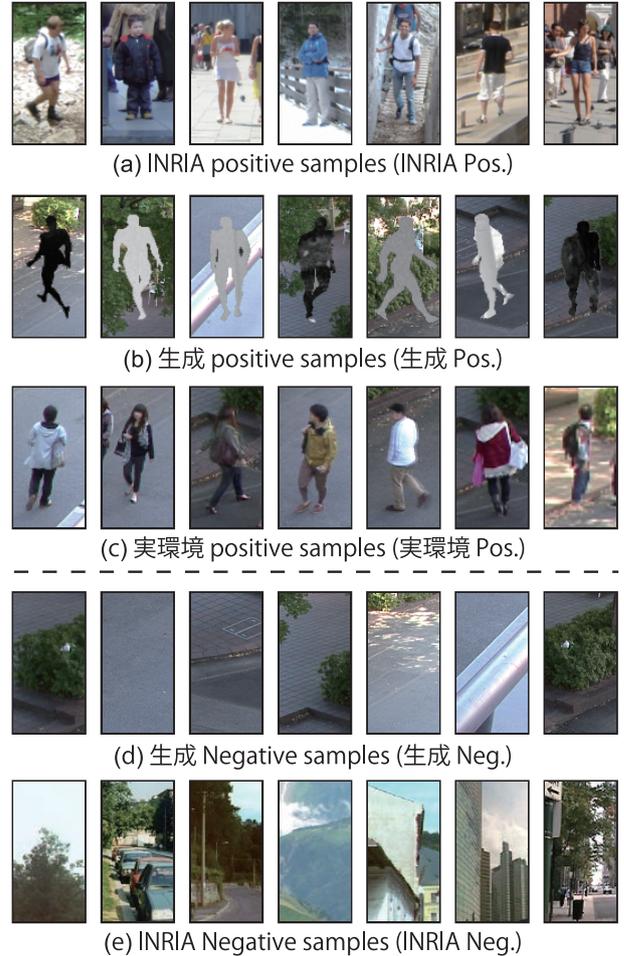


図 8 学習用データセットの例.
Fig. 8 Examples of learning data sets.

1 つ目の実験では, 特定シーンに特化した学習サンプルを生成する生成学習と Negative-Bag MILBoost の組み合わせにより実環境へ対応することの有効性を示す. 次に, 2 つ目の実験では, Negative-Bag MILBoost による検出器の学習法の有効性を, 対ノイズ性に着目した評価実験により示す.

4.1 実験 1: 自動生成の評価

4.1.1 実験概要

特定シーンに特化した学習サンプルの自動生成による有効性を評価する. 下記のデータベース毎に比較する.

- Database1: INRIA Pos. + 生成 Neg.
- Database2: 実環境 Pos. + 生成 Neg.
- Database3: 生成 Pos. + INRIA Neg.
- Database4: 生成 Pos. + 生成 Neg.

INRIA Pos. と INRIA Neg. は, Web 上で一般公開されている INRIA Person Dataset [6] に含まれている人画像と背景画像である. このデータベースは, 多様な人の姿勢, 向き, 視点, 照明の変動や背景テクスチャを含んでおり, 非常に汎用性の高いデータベースである. 実環境 Pos. は, 実環境下で撮影した映像から人手により切り出した人画像である. 生成 Pos. は, 2 章にて述べた

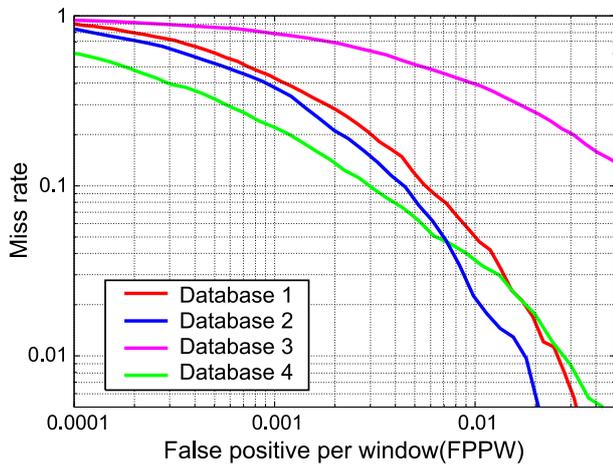


図9 各学習データベースの実験結果。

Fig.9 Experimental results obtained with each learning database.

3次元人体モデルを用いて生成した人画像である。生成Neg. は、実環境画像からランダムに生成した背景画像である。本実験に使用する実環境下の映像は、人の通行量が多い屋外の通路を撮影したものである。カメラの高さ6.2m、俯角21°としてカメラを設置し、約1時間の映像を撮影した。提案手法により生成するサンプルは、人体モデルに上記の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に学習に使用する画像データベースの枚数と種類を示す。図8に学習に使用した各データセットのサンプル例を示す。評価用データベースには、実環境下で撮影した動画画像からランダムで選択した450フレームを評価用データベースとして用いる。

実験結果の比較には Detection Error Tradeoff(DET)カーブを用いる。DETカーブは横軸に False Positive Per Window(FPPW)、縦軸に miss rate を表わし、左下の原点に近いほど検出性能が高いことを示す。

4.1.2 実験結果

DETカーブを図9に示す。まず、ネガティブサンプルが同一の Database 1, Database 2, Database 4 を比較すると、検出性能が最も高いのは人体モデルから生成したサンプルを用いた Database 4 であった。これは、実環境下で撮影した映像に対応した人の見えを生成できたからといえる。実環境下の映像から人手で切り出したサンプルを用いた Database 2 は、自動生成よりも低い結果となった。これは、人画像を人手で大量に切り出す際には、切り出し基準が曖昧になることがあり、これが識別器に悪影響を及ぼしたと考えられる。汎用性のあるデータベースを用いた Database 1 の結果が最も低い検出率となった。これは、学習用データベースの INRIA Person Dataset は実験環境とカメラ位置が異なるため、サンプル中の人の見えも大きく異なるからといえる。

次に、Database 3 と Database 4 を比較すると、実環境下で撮影した映像の背景を用いた Database 4 の方が良い結果が得られた。これは、Database 4 では実環境か

ら生成した学習用ネガティブサンプルを用いているため、実環境のシーンに特化した識別器となり検出性能が大きく向上したといえる。

以上により、人体モデルから生成されたサンプルを用いることで、実環境に特化した識別器を学習することができた。これにより、汎用的なデータセットを用いた場合と比べて検出性能を向上させることができた。提案手法と従来法それぞれの人検出結果例を図10に示す。

4.2 実験2：誤サンプルの影響の評価

4.2.1 実験概要

ネガティブの誤サンプルに対応した MILBoost による学習法の有効性を評価する。提案手法と Real AdaBoost を比較する。

- 従来法：Real AdaBoost
- 提案手法：Negative-Bag MILBoost

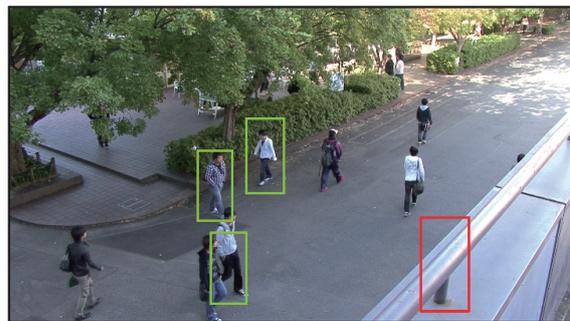
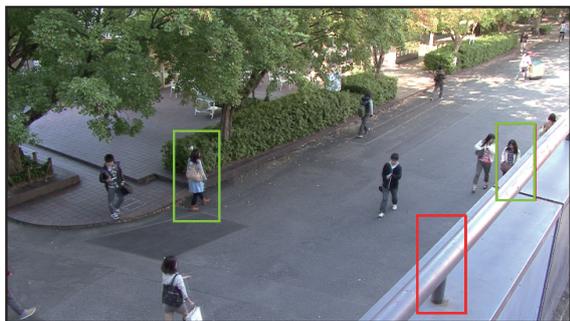
Negative-Bag MILBoost の有効性を確認するために、学習用のネガティブサンプルへ故意に人画像を混在させて識別器を学習する。その際の人画像の割合を0%~30%まで変動させ、その際の識別結果を比較する。実験に使用するデータベースは、ポジティブサンプルには INRIA Person Dataset のポジティブサンプル1,200枚を用いる。ネガティブサンプルには INRIA Person Dataset のネガティブサンプル4,000枚を用いる。混在させる画像は、INRIA Person Dataset のポジティブサンプルに使用していない1,200枚を0%~30%の割合で混在させる。評価に使用するデータベースは、INRIA データベースにおける評価データセットと同じものを使用する。

実験結果の比較には Equal Error Rate(EER)を用いる。EERは、Miss rate と FPPW が等しい時の値である。

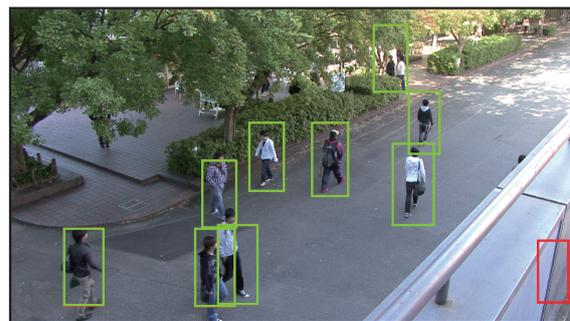
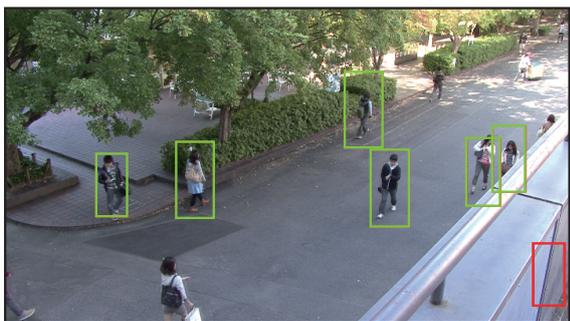
4.2.2 実験結果

実験結果を図11に示す。実験結果より、ネガティブサンプル中に人画像の含有率が高くなるに従って従来法では EER が高くなるが、提案手法 (NB MILBoost) では EER の増大を抑制していることがわかる。人画像の含有率が15%の場合を比較すると、提案手法は従来法よりも EER が6.1%低い。以上より、提案手法はネガティブサンプル中に人画像が含まれていても、識別器の学習に及ぼす悪影響を低減することができた。また、混入率0%時点においても提案手法の EER がわずかに低いのは、ノイズ低減効果による差であると考えられる。

同じネガティブ Bag に含まれる人画像と背景画像の強識別器の出力と、サンプルのクラス尤度の遷移を図12に示す。横軸は学習回数、左縦軸はサンプルの重み、右縦軸は強識別器の値を表す。図12より、学習回数を重ねる毎にネガティブ Bag の人画像は人として強く識別され、サンプルの重みが急速に低下していることがわかる。これは、Bag 単位での尤度によって本サンプルをノイズとして学習に利用しないことを意味する。その結果、ネガティブサンプルに混入されているにもかかわらず、識



(a) 従来法



(b) 提案手法

図 10 人検出例 (青枠は正解した検出ウィンドウ, 赤枠は誤検出したウィンドウを表す).

Fig. 10 Examples of human detection (Green window is true detection, red window is false detection).

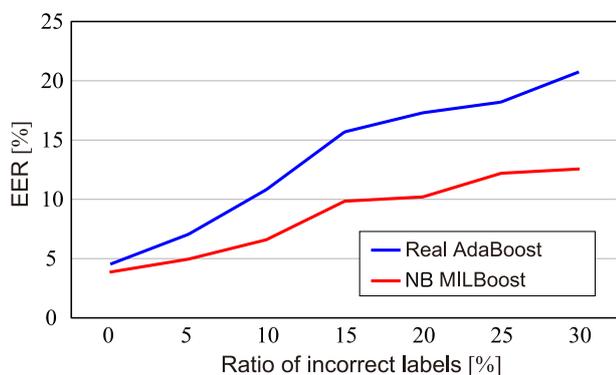
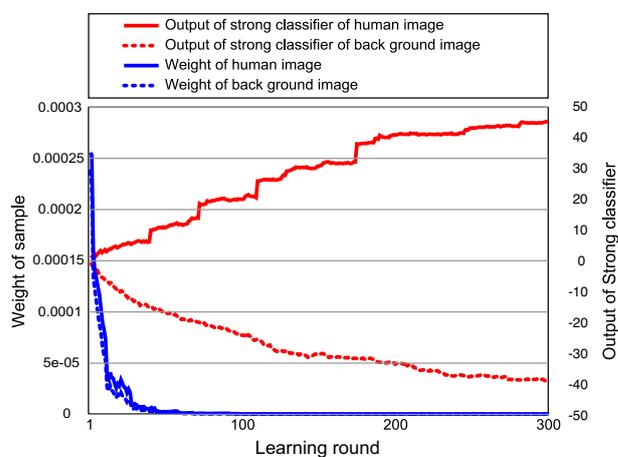


図 11 誤ラベルの割合を変化させた際の性能比較.

Fig. 11 Performance in the ratio of incorrect labels.



Human image



Back ground image

図 12 識別器の出力と重みの遷移.

Fig. 12 Changes in classifier output and weighting.

別の際は人らしい出力値を保つことができた. 一方, 背景画像は背景と識別され, 出力, 重み共に低下している. ネガティブ Bag においては, サンプルのクラス尤度が低いほど, 学習サンプルの重みが低下する. そのため, 提案する Negative-Bag MILBoost の学習アルゴリズムは, 誤って付与されたラベルを持つ人画像の悪影響を低減できていることがわかる.

4.3 考察

各実験結果より, 実シーンにおいて様々な変化を伴う人画像を擬似的に生成し, それを学習することで効率的な学習を実現可能であるといえる. しかし, ネガティブ

サンプルについての省力化は自動収集する他ない. それに伴う誤ラベル問題に付いて, 提案手法では MILBoost を Negative-Bag に適用することで対処した.

その狙いとするところは誤ラベルされたサンプルをノ

イズとして取り扱うことであり、実際にノイズ低減効果によって通常の2クラス Boosting に比べ性能を改善している。これは、ネガティブサンプルとして収集する画像における人画像領域の存在率が低いことを見込んでのことであり、提案手法においても誤ラベルされたサンプルの混入率に比例して性能低下が起こることは避けられない。

想定しているケースにおいては、多量の人画像が無作為に自動抽出したネガティブサンプルに混入することは考えがたい。しかし、交通量の多い交差点など、群衆の存在が予想される設置箇所ではその限りではない。もしポジティブサンプルのみで構成された Negative-Bag が存在すれば、それは式(7)に従い重要なネガティブサンプルとして学習される。この点に関しては更に事前学習、コンテキストなど何らかの方法で人画像の事前確率を推定するなど、半教師付き学習的アプローチを取り入れて対処する必要があると考えられる。

5. おわりに

本稿では、3次元人体モデルから学習サンプルの自動生成と Negative-Bag MILBoost を用いた生成型学習法を提案した。特定シーンにおいて、3次元人体モデルから生成した学習サンプルを用いることにより、実環境に特化した識別器を学習することができた。さらに、Negative-Bag MILBoost を用いることにより、誤ラベルを付与されたサンプルに対して悪影響を受けない学習を実現した。今後は、転移学習によるさらなる学習の省力化や、オンライン学習への展開、特定シーンにおける人物存在事前確率[18]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 Conference on Computer Vision, pp.153-161, 2005.
- [2] A. Ess and B. Leibe and K. Schindler and L. van Gool. "Moving Obstacle Detection in Highly Dynamic Scenes", IEEE International Conference on Robotics and Automation, 2009.
- [3] C. Wojek, S. Walk, and B. Schiele, "Multi-cue on-board pedestrian detection", IEEE Computer Vision and Pattern Recognition, 2009.
- [4] T. Watannabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection", Image and Video Technology, pp.37-47, 2009.
- [5] 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.
- [6] N. Dalal and B. Triggs: "Histograms of oriented gradients for human detection", IEEE Computer Vision and Pattern Recognition, pp. 886-893, 2005.
- [7] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance", European Conference on Computer Vision,

- 2006.
- [8] P. Ott, and M. Everingham, "Implicit color segmentation features for pedestrian and object detection", IEEE International Conference on Computer Vision 2009.
- [9] T. Deselaers, and V. Ferrari, "Global and efficient self similarity for object classification and detection", IEEE Computer Vision and Pattern Recognition, 2010.
- [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, 1, pp. 90-97, (2005).
- [11] M. Noda, T. Takahashi, D. Deguchi, I. Ide, H. Murase, Y. Kojima, T. Naito, "Recognition of Road Markings from In-Vehicle Camera Images by a Generative Learning Method", IAPR Conference on Machine Vision Applications, pp.514-517, 2009.
- [12] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide and H. Murase, "Construction of cascaded traffic sign detector using generative learning", International Conference on Innovative Computing Information and Control, pp. 889-892, 2009.
- [13] F. M. Tur, D. Vazquez, D. Geronimo and A. M. Lopez, "Learning Appearance in Virtual Scenarios for Pedestrian Detection", IEEE Computer Vision and Pattern Recognition, 2010.
- [14] 村瀬洋, "画像認識のための生成型学習", 情報処理学会論文誌, pp.35-42, 2005.
- [15] T. G. Dietterich, R. H. Lathrop, T. Lozano-Perez and A. Pharmaceutical, "Solving the Multiple-Instance Problem with Axis-Parallel Rectangles", Artificial Intelligence, Vol. 89, pp. 31-107, 1997.
- [16] P. Viola, John C. Platt and Cha Zhang, "Multiple instance boosting for object detection", Neural Information Processing Systems 18, pp.1419-1426, 2006.
- [17] R. E. Schapire, and Y. Singer, "Improved Boosting algorithms using confidence-rated predictions", Machine Learning, pp.297-336, 1999.
- [18] 中河秀仁, 渡部済, 木戸出正継, "人物の移動軌跡と画像の色情報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.