

第5章

生成型学習に基づく学習の効率化

4章では、特徴量の貢献度を用いて、再学習することなく簡易に識別器を調整するアプローチを提案した。しかし、カメラの設置角度が大きく変化する場合、画像上の見えは、物体のパーツの位置やサイズも大きく変化するため、調整での対応が困難となる。このような場合、再学習を必要とするが、再学習には大量の学習サンプル収集等の大きな人的コストが必要となるという問題がある。

本章では、特定シーンに対する再学習を効率化、高精度化する手法を提案する。5.1では、人体モデルのCG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。5.2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べる。これにより、特定シーンから自動的に大量の学習サンプルを収集し、人的コストを大きく削減可能となることが期待できる。

5.1 学習サンプルの自動生成による効率化

カメラの設置視点の変化等の大きな変動があった場合、4章で述べた調整での対応は困難となり、再学習を必要とする。再学習の際に問題となるのが、画像中の人の位置やサイズなどが標準化された学習サンプルを大量に収集するために必要な膨大な人的、時間的コストである。

そこで、本章では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。学習サンプルの自動生成法として、Marら [32] はCGを用いた人体画像生成法を提案している。本手法は、低コストで大量に学習サンプルを得ることを目的としており、ゲームで用いられる街のシーンのCGから、人画像の切出しを行う手法である。そのため、設置環境の特定シーンを反映した学習サンプルの収集は不可能である。そこで、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し、ネガティブサンプルも同様に特定シーンから収集する。

このように、特定シーンに特化した学習サンプルを学習に用いることで、特定シーンにおいては、一般的なデータベースで学習した識別器と比べて精度を向上することができる。また、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ため、人画像が含まれることがある。このように誤ったラベルを付与された学習サンプルを用いて学習すると、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 を用いる。

提案手法の流れを図5.1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を、3次元の人体モデルのCGを用いて生成する(図5.1(a))。ネガティブサンプルは映像中から切り出した画像を用いる(図5.1(b))。このように生成した学習サンプルを Negative-Bag MILBoost に入力することで識別器を学習する。

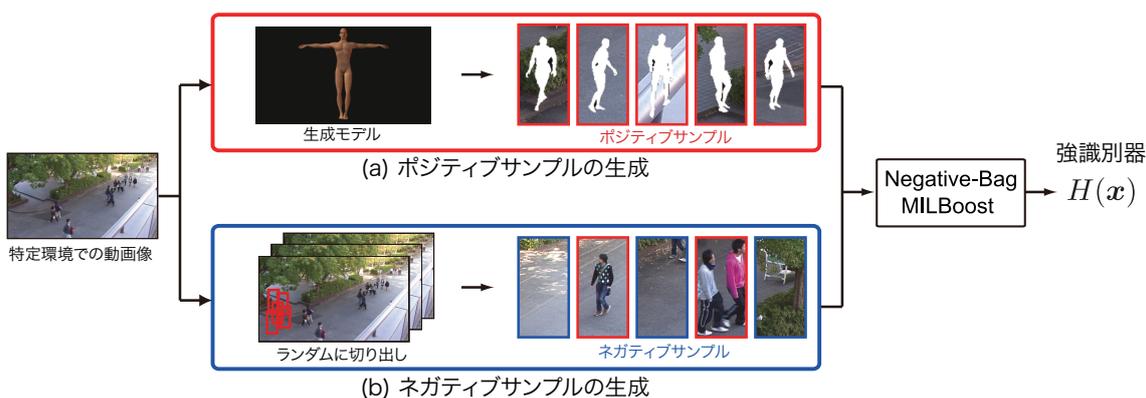


図 5.1: 生成型学習の流れ。

学習用ポジティブサンプルの生成には、図5.2に示すような人体モデルと仮想カメラを用いる。これにより、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。

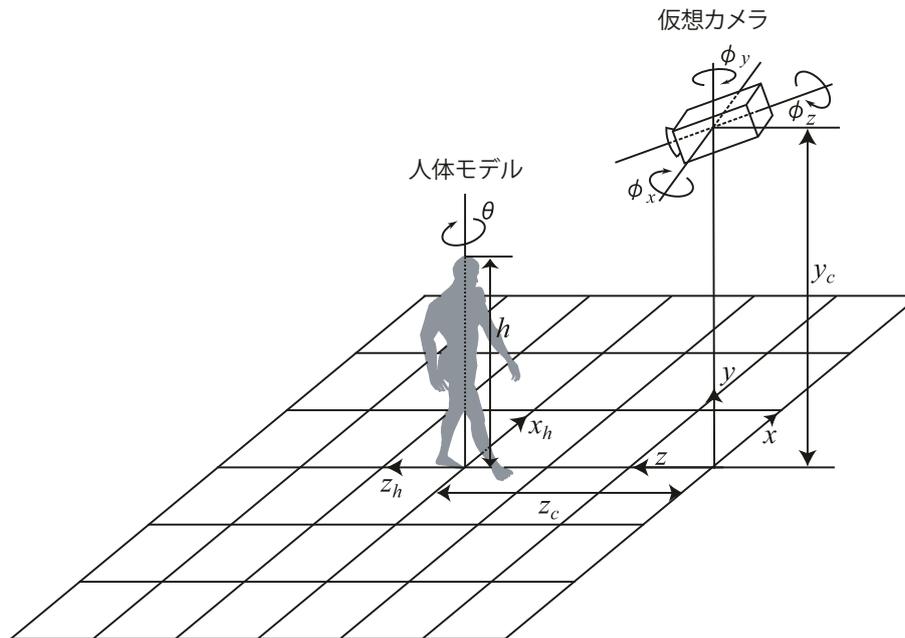


図 5.2: 3次元人体モデル.

5.1.1 3次元人体モデルによる人画像生成

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、歩行姿勢の人体モデルを表現する。また、下記の各パラメータを与えることで、図5.3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生成することができる。

～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}

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ており、事前に得たカメラパラメータを用いる。人形状パラメータにおける人の向きと位置は事前に決定できないため、一様にランダム

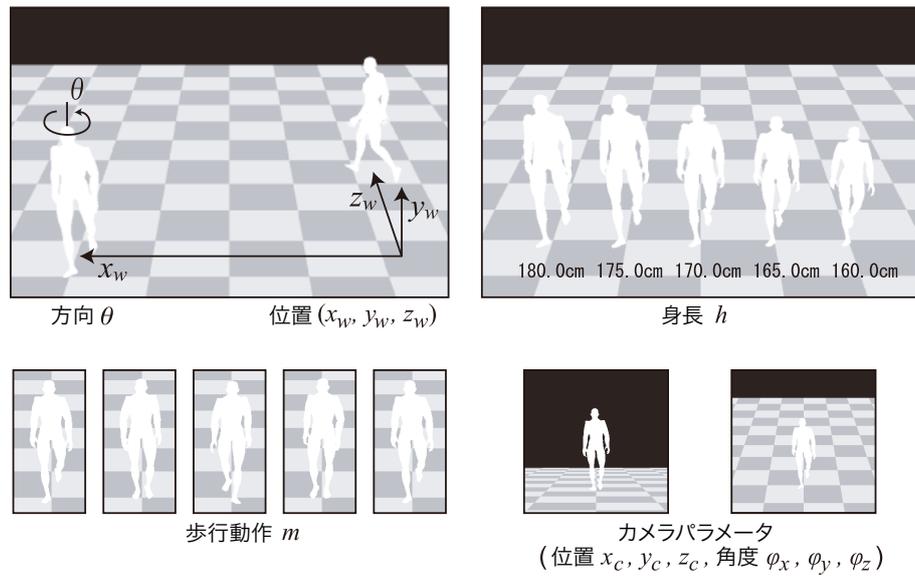


図 5.3: 3次元人体モデル生成時のパラメータ。

設定とする。人の身長は、文部科学省の統計調査 [33] より平均身長 171.9cm とする。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、多種多様な衣服を用意することが難しい。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。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貼り付ける。これにより、衣服によるテクスチャ変化の影響を受けずに、人の輪郭を学習する。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 5.4 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。このように特定シーン画像において人体シルエットを生成し、人体領域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。

5.1.2 ランダムサンプリングによる背景サンプルの自動収集

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。これにより、特定シーンにおいては、一般的なデータベースで学習した識別器と比べ、特定シーンに特化したネガティブサンプルを収集できるため、識別器の高精度化が可能となる。しかし、ランダムにサンプルを収集した場合、図 5.1(b) に示すように、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識別器の学習法として、Negative-Bag MILBoost を提案する。

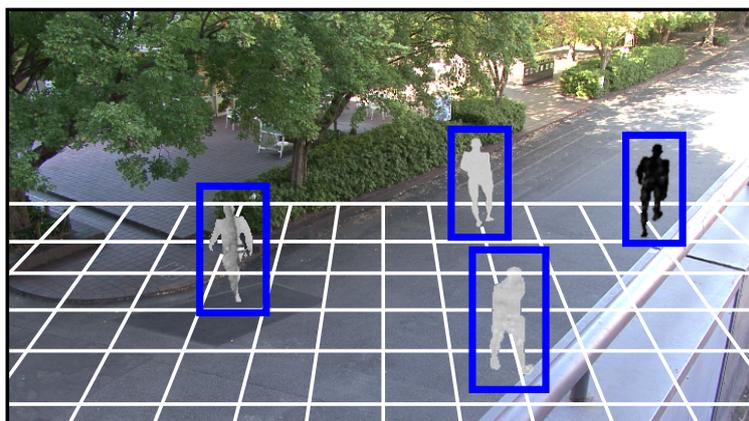


図 5.4: 実環境に対する 3 次元人体モデルの生成例.

(camera position $(x_c, y_c, z_c) = (0m, 6.2m, 0m)$, camera angle $(\phi_x, \phi_y, \phi_z) = (21^\circ, 0^\circ, 0^\circ)$)

5.2 Multiple Instance Learning の導入によるノイズ低減

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。

5.2.1 MILBoost

Multiple Instance Learning (MIL) [34] 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 は画像座標上の位置が近傍関係にある複数のサンプルにより構成され、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 とラベル付けされ、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 [35] を拡張した Negative-Bag MILBoost を提案する。

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り、Viola らは顔検出器の学習に適用している。顔検出器の学習サンプルは顔画像を大量に必要とするため、顔画像の収集を対象領域とその周辺領域を切り出し集めた Bag として扱い、ポジティブサンプルの収集を容易にしている。このポジティブサンプルには、画像中に背景の比率が高い顔画像等が含まれるが、MIL を導入することによって、それらの影響を低減し、高精度な顔検出を実現している。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。

5.2.2 Negative-bag MILBoost

MILBoost[35]では、ポジティブサンプルのラベルに誤ラベルが混在する問題を対象としている。しかし、提案手法ではネガティブサンプルのラベルに誤ラベルが混在する問題である。そこで、本研究の問題設定にあわせて、ネガティブサンプルにMILBoostのBagを適用する。ここでは、まずBagの作成方法について述べ、次にNegative-Bag MILBoostの学習アルゴリズムについて述べる。

■ 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誤ラベルを含むため、文献[35]のBagの構成を図5.5に示すように変更する。図5.5(a)に示す従来法では、ポジティブサンプルが誤ラベルを含むため、複数のポジティブサンプルをBagとし、ラベル付けが確実なネガティブサンプルは単体の画像を1つのBagとして学習に利用する。一方、提案手法においては図5.5(b)で示すように、人体モデルから生成した人画像を一つの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一つのBagとして扱う。このネガティブBagの中に人画像が含まれる可能性がある。

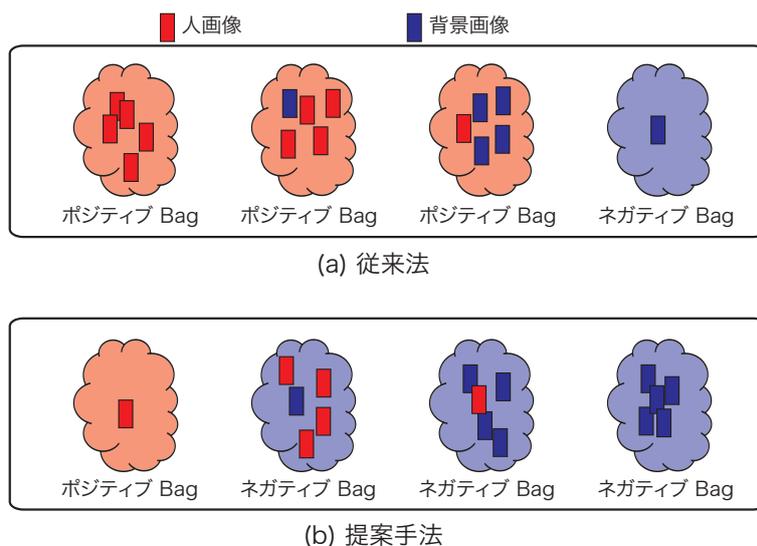


図 5.5: 提案手法における Bag の構成。

■ Negative-Bag MILBoost の学習

MILBoost の学習は、通常の Boosting と共通のアルゴリズムであるが、学習サンプルの重みの更新法が異なる。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應する

ように拡張する。Real AdaBoost [6] をベースとした Negative-Bag MILBoost の学習アルゴリズムを**アルゴリズム 5.1** 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。これらの Bag は、ネガティブサンプルを 1 つでも含んでいればネガティブ Bag、それ以外はポジティブ Bag としてラベルが与えられている。次に、 i 番目の Bag に所属する j 番目のサンプルの重み $w_t(i, j)$ を式 (5.1) により初期化する。

次に、学習サンプルから抽出した特徴ベクトル \mathbf{x} を入力とした確率密度関数 W_+ , W_- を式 (5.2), (5.3) により作成する。特徴量には、Histogram of Oriented Gradients(HOG) 特徴量 [3] を用いる。確率密度関数 W_{\pm} は、HOG 特徴量で得られる特徴次元毎にヒストグラムとして表現される。学習サンプルの重み $w_t(i, j)$ は、そのサンプルの特徴量の値を K 段階で量子化した値 k とラベル y に基づき、確率密度関数 W_+^k または W_-^k に累積される。ここで t は学習回数を表す。確率密度関数 W_{\pm} を作成した後、式 (5.4) より弱識別器 $h(\mathbf{x})$ の評価値 Z を求める。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す。 L 個の弱識別器候補について評価値を求め、最も識別性能が高い弱識別器候補を t 個目の弱識別器 $h_t(\mathbf{x})$ (式 (5.5)) とする。1 つの弱識別器で 1 つの特徴量を観測し、その全てについて評価を行うため、 L は特徴量の次元数と同数である。

弱識別器を選択後、式 (5.7), (5.8) 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 (5.9) より最終識別器 $H(\mathbf{x})$ を得る。

アルゴリズム 5.1: Negative-Bag MILBoost の学習アルゴリズム.

1. 入力: J 個の学習サンプルを含む I 個の Bag とラベル $y \in \{1, 0\}$ を用意

2. 初期化: 学習サンプルの重み $w_t(i, j)$ を初期化

$$w_1(i, j) = \frac{\text{Bag of class}}{\text{Bag of all}} \quad (5.1)$$

3. 学習:

For $t = 1, \dots, T$ // 学習ラウンド

For $l = 1, \dots, L$ // 弱識別器の候補数

• 弱識別器の確率密度関数 W_{\pm} の算出

$$W_+^k = \sum_{i,j:k \in K \wedge y_i=1} w_t(i, j) \quad (5.2)$$

$$W_-^k = \sum_{i,j:k \in K \wedge y_i=0} w_t(i, j) \quad (5.3)$$

• Z_l の評価

$$Z_l = 2 \sum_{k=1}^K \sqrt{W_+^k W_-^k} \quad (5.4)$$

End for

• Z_l が最小の $h_t(\mathbf{x})$ を弱識別器として選択

• 弱識別器 $h(\mathbf{x})$

$$h(\mathbf{x}) = \frac{1}{2} \ln \frac{W_+^k + \varepsilon}{W_-^k + \varepsilon} \quad (\varepsilon = 1/J) \quad (5.5)$$

• 重み $w_t(i, j)$ の更新

$$w_t(i, j) = \begin{cases} p_{ij} & \text{if } y_i = 1 \\ \frac{p_{ij} \times (p_i)}{1 - p_i} & \text{if } y_i = 0 \end{cases} \quad (5.6)$$

$$p_i = \prod_{j \in \text{Bag}_i} p_{ij} \quad (5.7)$$

$$p_{ij} = \frac{1}{1 + \exp(-H_t(\mathbf{x}))} \quad (5.8)$$

End for

4. 出力

• 強識別器 $H(\mathbf{x})$

$$H(\mathbf{x}) = \text{sign} \left(\sum_{t=1}^T h_t(\mathbf{x}) \right) \quad (5.9)$$

■ Negative-Bag を用いた誤ラベルへの対応

図 5.6 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。

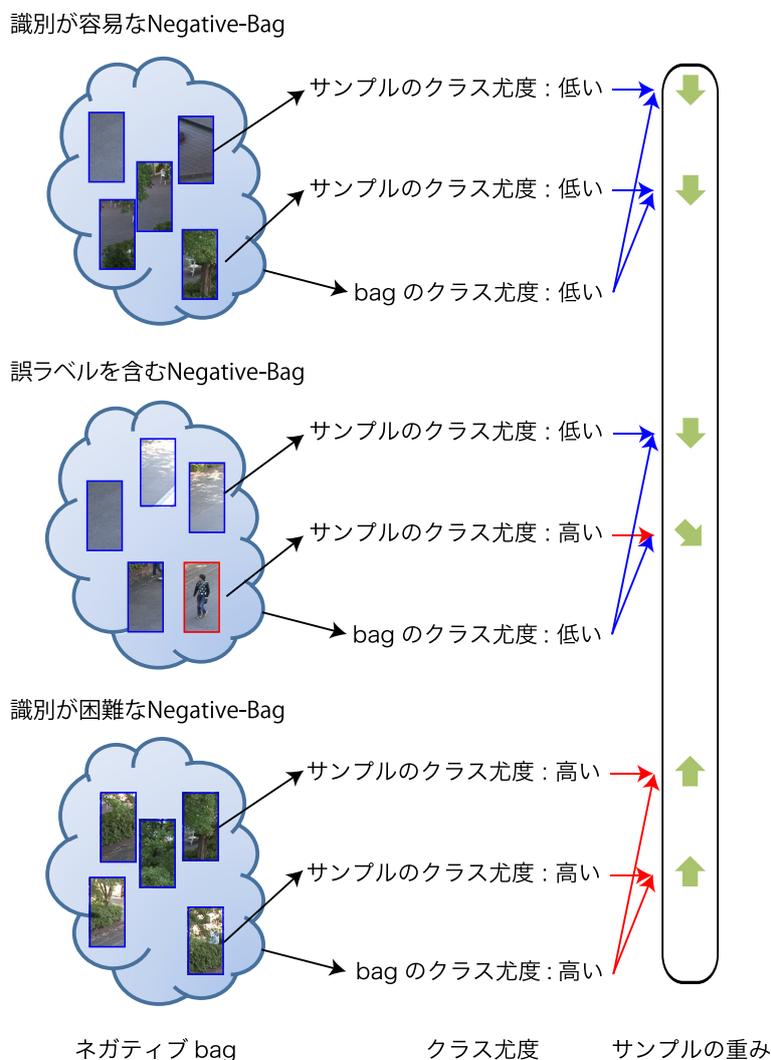


図 5.6: Negative-Bag MILBoost による学習サンプルの重みの更新.

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、 i 番目の Bag に含まれる j 番目の学習サンプルの重み $w_i(i, j)$ は式 (5.6) 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} の値が総じて低いと Bag のクラス尤度 p_i の値も

低くなる。クラス尤度の低いサンプルは識別が容易なネガティブサンプルであるため、Bag 中のサンプルの重み $w_t(i, j)$ は低下する。もし、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ても、Bag のクラス尤度 p_i が十分に低ければ、そのサンプルをノイズであると捉え、サンプルの重み $w_t(i, j)$ は低下する。一方、サンプルのクラス尤度 p_{ij} の値が総じて高いと Bag のクラス尤度 p_i の値が高くなる。こ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ため、Bag 中のサンプルの重み $w_t(i, j)$ 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。

■ 識別

識別時は、従来法である Real AdaBoost と同様に計算する。式 (5.9) に示すように複数の弱識別器の線形和を求め、しきい値により判定する。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。

5.3 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。1つ目の実験では、特定シーンに特化した学習サンプルの自動生成の有効性を示す。2つ目の実験では、誤ラベルに対する評価実験により、Negative-Bag MILBoost による学習法の有効性を示す。

5.3.1 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。4種類のデータベースを用いて識別器を学習し、識別性能を比較する。Database 1はINRIA Person Dataset[3]に含まれている人画像をポジティブサンプルとして使用し、提案手法により自動収集した特定シーンの背景画像をネガティブサンプルとして使用する。Database 2は実環境で収集した人画像をポジティブサンプル、提案手法により実環境から自動収集した背景画像をネガティブサンプルとして使用する。Database 3は提案手法により生成した人画像をポジティブサンプル、INRIA person datasetの背景画像をネガティブサンプルとして使用する。Database 4は提案手法により生成した人画像をポジティブサンプル、提案手法により実環境から自動収集した背景画像をネガティブサンプルとして使用する。一般的なデータベースの人画像を用いた Database1、実環境の人画像を用いた Database2 と提案手法である Database 4 との比較により、提案手法で生成する人画像を用いて、実環境でサンプルを収集したものと同等に特定シーンの学習が可能であることを示す。また、提案するネガティブサンプルの自動収集についても評価するため、一般的なデータベースである INRIA Person Dataset のネガティブサンプルと生成人画像の組み合わせである Database3 と提案手法である Database 4 を比較する。データベース毎に使用するポジティブサンプル、ネガティブサンプルをまとめると以下のようになる。

～ Database 1 : INRIA Pos. + 生成 (Generated) Neg.

～ Database 2 : 実環境 (Real environment) Pos. + 生成 Neg.

～ Database 3 : 生成 Pos. + INRIA Neg.

～ Database 4 : 生成 Pos. + 生成 Neg.

実験結果の比較には DET カーブを用いる。

■ データベース

実験に用いるデータベースの詳細について述べる。INRIA Pos. と INRIA Neg. は、INRIA Person Dataset 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、5.1にて述べた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 5.1 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 5.7 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。

表 5.1: 学習用データベースの内訳

	ポジティブ	ネガティブ
Database 1	INRIA Pos.(2,416)	生成 Neg.(12,180)
Database 2	実環境 Pos.(2,416)	生成 Neg.(12,180)
Database 3	生成 Pos.(2,416)	INRIA Neg.(12,180)
Database 4	生成 Pos.(2,416)	生成 Neg.(12,180)



(a) INRIA positive samples (INRIA Pos.)



(b) 生成 positive samples (生成 Pos.)



(c) 実環境 positive samples (実環境 Pos.)



(d) 生成 Negative samples (生成 Neg.)



(e) INRIA Negative samples (INRIA Neg.)

図 5.7: 学習データベース例.

5.3.2 生成画像に基づく特定シーン学習の有効性

各データベースを用いて学習した際の DET カーブを図 5.8 に示す。

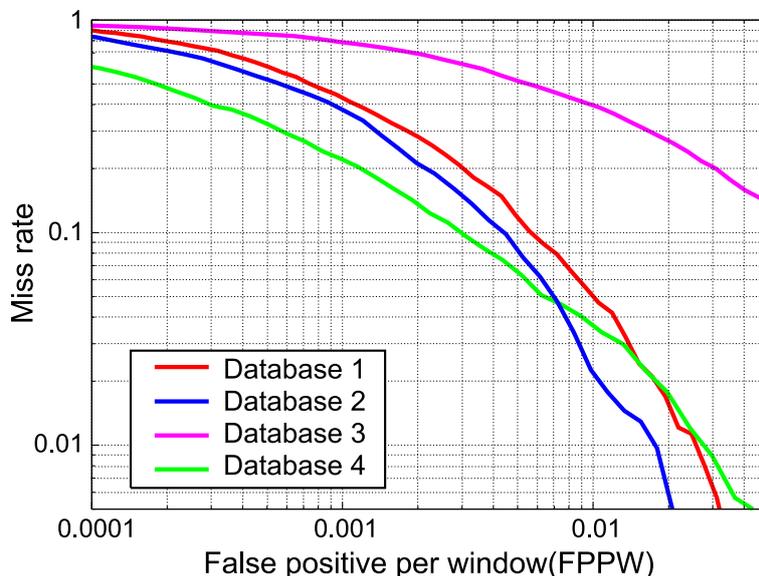


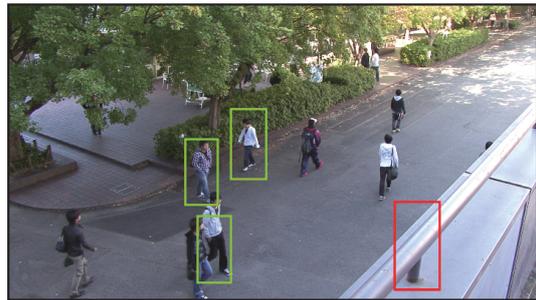
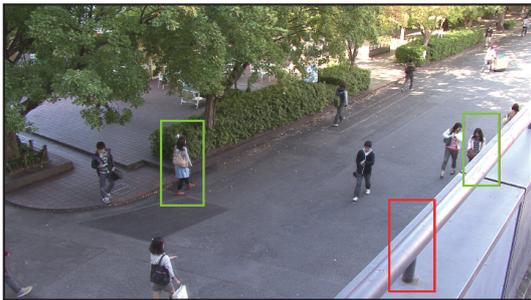
図 5.8: 各データベースで学習を行った際の DET カーブ。

まず、ネガティブサンプルが同一の Database 1, Database 2, 提案手法である 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提案手法であった。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、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。

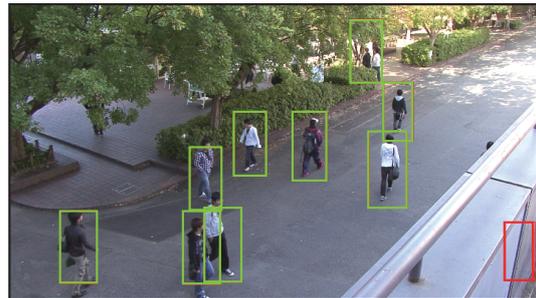
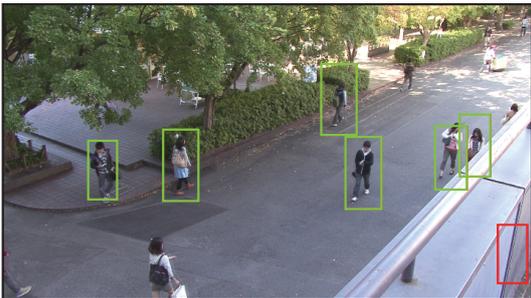
次に、Database 3 と提案手法を比較すると、実環境下で撮影した映像の背景を用いた提案手法の方が良い結果が得られた。これは、提案手法では実環境の画像から切り出した背景画像を学習用ネガティブサンプルとして用いているため、実環境のシーンに特化した識別器となり検出性能が大きく向上したといえる。

提案手法と従来法それぞれの人検出結果例を図 5.9 に示す。

図 5.9 より、同一画像において FPPW を等しくした時、提案手法はより多くの人を正しく検出できていることがわかる。



(a) 従来法



(b) 提案手法

図 5.9: 人検出結果例 (緑：正検出ウインドウ, 赤：誤検出ウインドウ).

5.3.3 Negative-bag MILBoost によるノイズ低減の有効性

ネガティブの誤サンプルに対応した Negative-Bag 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人画像を混在させて識別器を学習する。ネガティブサンプルにおける人画像の割合を 0% ∈ 30% まで変動させ、その際の識別性能を比較する。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 の人画像 1,200 枚を用いる。ネガティブサンプルには INRIA Person Dataset の背景画像 4,000 枚を用いる。ネガティブサンプルのノイズとして用いる人画像は、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。本実験の比較には Equal Error Rate(EER) を用いる。EER は、DET において miss rate と FPPW が等しい時の値であり、この値が低いほど性能が高いことを表す。DET との違いは、1つの手法の評価が1つの値で表現できることであり、本実験のように、複数のバリエーションをもつ学習サンプルを用いて手法を比較する際に用いられる。

■ 実験結果

実験結果を図 5.10 に示す。実験結果より、従来法は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 EER が高くなる。一方、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。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といえる。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 5.11 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出力値を表す。図 5.11 より、ネガティブ Bag の人画像の識別結果は人らしい高い出力値を持つことがわかる。通常の Boosting の場合、ラベルと異なる識別結果を出力すると、次の弱識別器の選択では識別困難なサンプルとして重みが上昇する。しかし、図 5.11 より、提案手法では逆に重みが低下していることがわかる。これは、Bag 単位での尤度を用いて本サンプルをノイズと判断して、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することができた。

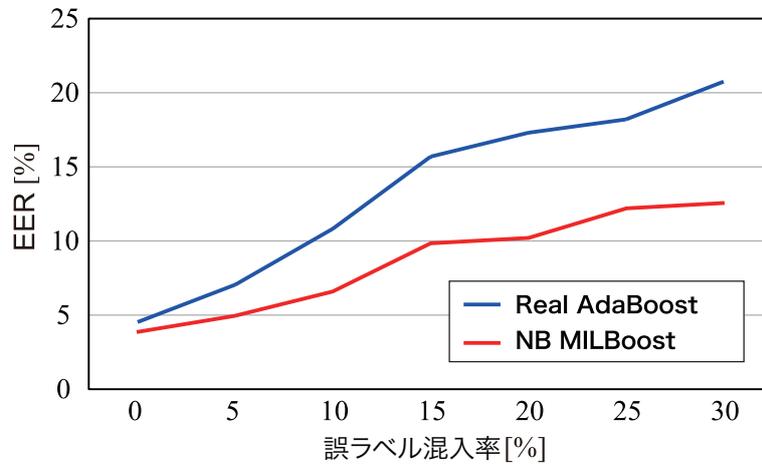


図 5.10: ノイズを混入した際の識別性能.

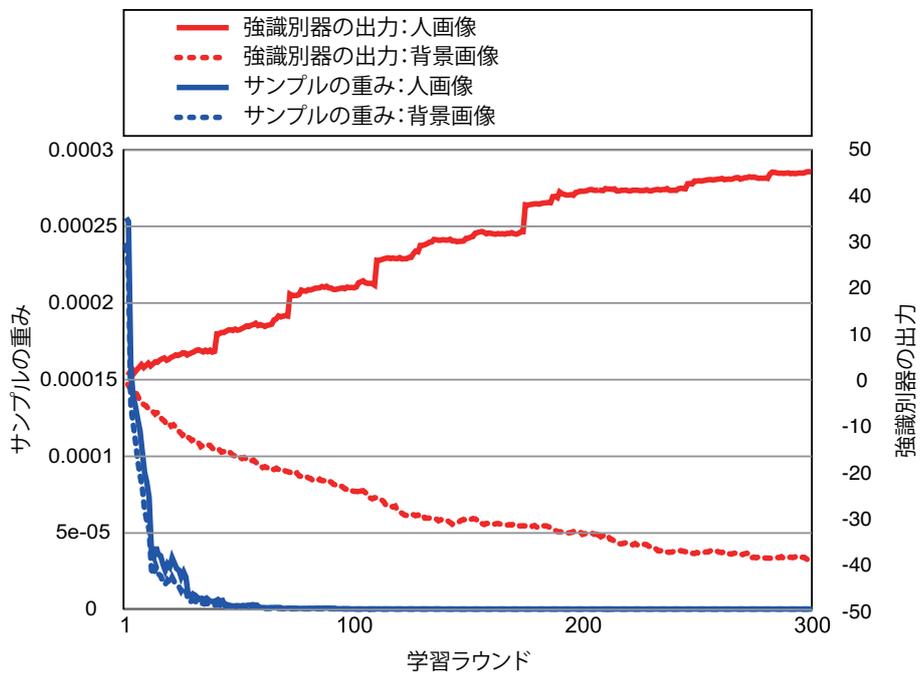


図 5.11: 識別器の出力と重みの遷移.

5.4 まとめ

本章では、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について述べた。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可能である。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可能であることを述べた。しかしながら、現場環境への適応の際に再度識別器を学習する計算コストは学習用画像の枚数と特徴量の次元数に比例して大きくなり、この計算コストについても実用化の障害であるといえる。そこで、6章では転移学習による高速な再学習を導入して計算コストを効率化する。