

距離画像セグメンテーションに基づくリアルタイム人物検出

柴田 雅聡^{*1} 生形 徹^{*1} 寺林 賢司^{*2}
モロ アレサンドロ^{*3} 梅田 和昇^{*3}

Real-Time Human Detection Based on Range Image Segmentation

Masatoshi Shibata^{*1}, Toru Ubukata^{*1}, Kenji Terabayashi^{*2},
Alessandro Moro^{*3} and Kazunori Umeda^{*3}

In this paper, a human detection method that combines information of segmented range image and human detectors based on image local feature is proposed. The method uses a stereo vision system called Subtraction Stereo, which extracts foreground information and range image. Range image is segmented for every object by Mean Shift Clustering. The proposed method reduces processing time and false detection by limiting the search range to the object. Joint HOG feature is used for human detection and reduces undetected human by integration of detection windows in consideration of occlusion. The proposed method is evaluated by experiments comparing with the method using the conventional Joint HOG feature.

Key Words: Human Detection, Stereo Vision, Joint HOG Feature, Occlusion, Segmentation

1. 序 論

近年、監視カメラの映像から不審者の検出や人の行動解析などを行う技術が期待されており、その基盤技術として画像中から人物をリアルタイムに検出する技術が求められている。人物検出精度の高い手法の一つとして、DalalとTriggsによって提案されたHOG特徴量[1]に基づく人物検出がある。HOG特徴量はエッジ情報を用いるため照明変化に頑強な局所特徴量である。また、オブジェクト検出のための様々な局所特徴量として、Haar-like特徴量[2]、EOH特徴量[3]、Edgelet特徴量[4]なども提案されている。このような局所特徴量をBoosting[5]やSVM[6]などの統計的学習法を用いて学習・識別する手法が盛んに研究されている[7][8]。しかし、このような局所特徴量に基づく人物検出は画像全体に対して検出ウィンドウを走査して特徴量を抽出する必要がある。同時に、画像中の人物の大きさが未知なため、検出ウィンドウのスケールを変化させながら画像全体を複数回スキャンする必要がある。そのため、特徴量算出に多くの処理時間を必要とする。また複雑な背景や人物の

全身の形状を取得することが困難な遮蔽（オクルージョン）を生じていた場合、検出精度が低下する。

高速な人物検出手法として、Benensonら[9]は、ステレオカメラを用いた高速な距離計測と距離画像から算出した物体の幾何学的なつながりを用いて、135 [fps]での人物検出を行っている。しかし、オクルージョンに対しての対応がなく、計測シーンも限定されてしまっている。特徴量算出の処理時間を削減する手法として、鈴木ら[10]はステレオ処理により得られた距離情報を活かし、限られた領域と検出ウィンドウのサイズのみを走査する手法を提案している。しかし、カメラの地面からの高さや仰角によって画像内での人物の見え方が異なるため、検出ウィンドウの高さと幅を補正する透視投影モデルを仮定する必要がある。また、池村ら[11]はTOFカメラより得られた距離情報を用いた効率のよい検出ウィンドウの走査手法を提案している。しかし、人物検出とは関係のない背景領域を検出ウィンドウが走査するため、処理速度の低下や誤検出の増加を引き起こす可能性がある。Vineetら[12]はグラフィカルモデルの一種である条件付き確率場（Conditional Random Fields）を用いて画像内セグメンテーションを行っている。提案されている手法はビデオ映像から高精度なセグメンテーションを可能にしている。しかし、検証されている画像は人物の上半身のみなど人物が大きく映るシーンのみを用いており、監視カメラを想定した際に人物の映る大きさによっては正しくセグメンテーションが行えない可能性がある。また、局所特徴量を用いた人物検出を行う手法として、Wangら[13]は、人物検出を行う前にオクルー

原稿受付 2013年10月2日

^{*1}中央大学大学院

^{*2}静岡大学

^{*3}中央大学

^{*1}School of Science and Engineering, Chuo University

^{*2}Faculty of Technology, Shizuoka University

^{*3}Faculty of Science and Engineering, Chuo University

■ 本論文は有用性で評価されました。

ジョンを検出することで、HOG 特徴と LBP (Local Binary Pattern) [14] を用いた人物検出結果を向上させている。学習には線形 SVM を用いており、INRIA [15] データセットの画像に対して高い検出精度を実現している。しかし、処理速度の面ではオンライン処理することができず、またデータセットに対する評価にとどまっている。

我々は、先行研究において、背景差分法により前景領域（移動物体領域）を検出し、前景領域に限定した距離画像を取得する差分ステレオ [16] を提案している。本稿では、差分ステレオより得られる前景領域に限定した距離画像をオブジェクトごとに領域分割（セグメンテーション）[17] した情報と、局所特徴量ベースの人物検出を組み合わせることで、高精度な人物検出とリアルタイム処理を実現する手法を提案する。

提案手法では、あらかじめ Mean Shift Clustering [18] により距離画像をオブジェクトごとにセグメンテーションし、局所特徴量を算出する検出ウィンドウの走査範囲を限定して処理時間の削減と誤検出の低減を行う。その際、距離情報を用いることで擬似透視投影モデルを仮定し、検出ウィンドウの高さと幅を補正する。また、分割されたオブジェクト間の距離を比較することで人物間で発生するオクルージョンを検出し、オクルージョンが発生している割合に応じて遮蔽部における識別器の寄与を抑制して検出精度を向上する。さらに、人物として識別された検出ウィンドウを遮蔽を考慮して統合することにより未検出を低減する。本論文では、人物検出に用いる局所特徴量は Joint HOG 特徴量を用いる。また、Joint HOG 特徴量を用いた人物検出 [19] を従来手法と位置付け、比較実験を行うことで提案手法の有用性を示す。

本論文の構成は以下に示す通りである。まず、2 章で差分ステレオを用いた前景検出について説明する。次に、3 章でステレオカメラから取得した距離画像をオブジェクトごとにセグメンテーションする手法を説明する。そして、4 章で距離画像セグメンテーションの結果に基づくオクルージョンを考慮した人物検出手法を提案する。5 章では提案手法の有効性を確認するために、単純な背景と複雑な背景の屋内での 2 シーンと照明環境が変化しやすい屋外シーンで評価し、従来手法と検出精度を比較する。最後に 6 章で結論と今後の展望について述べる。

2. 前景検出

2.1 差分ステレオ

本研究では、画像中の前景領域（移動体領域）を検出する手法として、差分ステレオ [16] を用いている。差分ステレオの基本アルゴリズムを Fig. 1 に示す。差分ステレオでは、まず左右カメラそれぞれで背景差分によって前景領域を抽出し、その後抽出された前景領域をマッチングすることで距離情報を得る。マッチングする領域を背景差分で得られた前景領域に限定することで、マッチングの誤対応と処理時間を削減することができる。差分ステレオによって検出された距離情報を持った前景領域を Fig. 2 (b) に示す。Fig. 2 (b) において、カメラからオブジェクトまでの距離が灰色は近く、白になるにつれ遠くなるよう色分けされている。また、黒は前景領域が抽出されないために、距離計測を行っていない領域である。

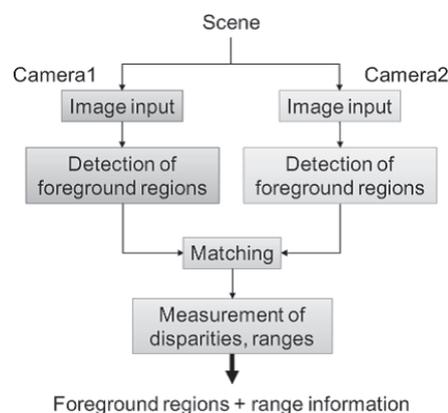


Fig. 1 Basic algorithm of the subtraction stereo [16]

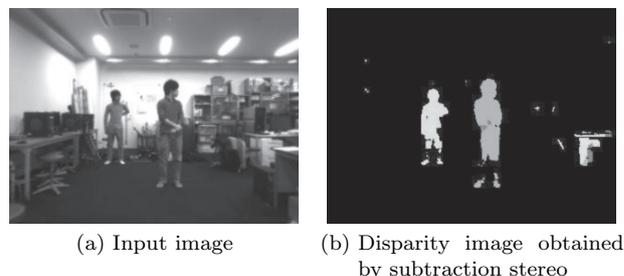


Fig. 2 Subtraction stereo

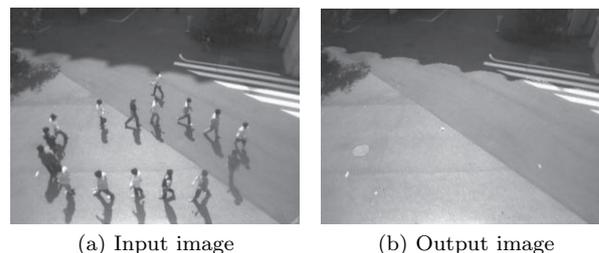


Fig. 3 Background updating

2.2 背景画像の自動更新

背景差分は、適切な前景領域を抽出するには照明変化や背景自体の変化に対応する必要がある。そこで、輝度値のヒストグラムを直接扱うノンパラメトリックな背景更新 [20] を行い、この問題に対応する。

入力画像の色情報を画素ごとにヒストグラムとして保持し、背景情報は高い頻度で入力画像に含まれるという仮定のもとに、背景画像を決定する。この仮定は、真の背景画像と入力画像の差が生じる頻度は低いと換言することができ、この頻度が高いときに背景画像をヒストグラム情報に基づき更新する。このとき、GPU を用いて処理を行うことで計算量の問題を解決している。今回用いた背景更新の詳細は文献 [20] にて述べられている。Fig. 3 (a) に対して背景更新を行った結果を Fig. 3 (b) に示す。

2.3 影検出

背景差分により検出された前景領域はオブジェクトの影を含むため、影検出により前景領域から除去する。前景領域から影を除去することにより、次章で説明する距離画像セグメンテーションの際の処理速度と精度の向上を図る。

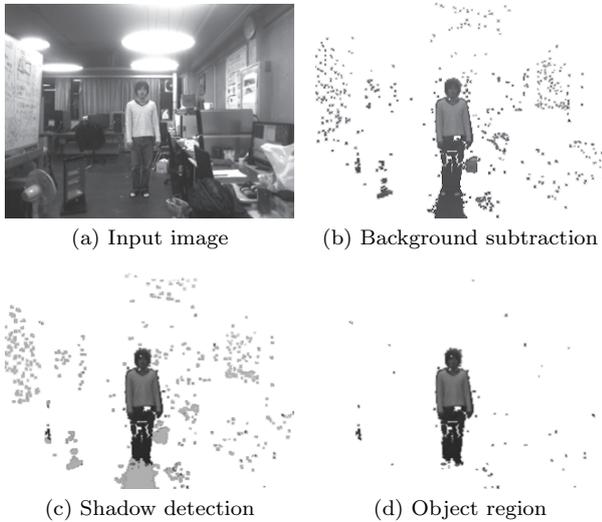


Fig. 4 Shadow detection

画像座標 (x, y) における輝度値を $I_{(x,y)}$ 、背景画像における同位置の輝度値を $I'_{(x,y)}$ とし、次式に示す評価関数で影を判定する。

$$\theta_{(t+1,x,y)} = \begin{cases} \alpha\Psi_{(x,y)} + \beta\Lambda_{(x,y)} + (1 - \alpha - \beta)\theta_{(t,x,y)}, & \text{if } \frac{I_{(x,y)}}{\eta} < I'_{(x,y)} \\ \infty, & \text{otherwise.} \end{cases} \quad (1)$$

ここで、 θ は影と判断するためのスコアを表し、 θ が閾値以下となる画素を影と判定して前景領域から除去する。 Ψ は前景の近傍画素と背景の近傍画素値の相違度を表し、 Λ は前景の色相と背景の色相の相違度を表す。また、 α, β, η はそれぞれ、テクスチャ、カラー、明度に重みを与える定数である。それぞれの定数は、屋内、屋外で同様に正しく影検出が行える値を実験的に求めた。今回用いた影検出手法の詳細は文献 [21] にて述べられている。

Fig. 4 (a) に対して、背景差分を行った結果を Fig. 4 (b) に、影検出を行った結果を Fig. 4 (c) に示す。図中において、灰色は影と判定された領域を表わす。この影が除去された前景領域 (Fig. 4 (d)) に対してラベリングを行い、連結領域ごとに一つのオブジェクトとして扱う。

3. 距離画像セグメンテーション

前景検出で抽出された領域から重なりが生じたオブジェクトを個々に検出するため、前景領域の距離画像をセグメンテーションする手法を説明する。

3.1 距離情報の投影とヒストグラムの構築

2.3 節で説明したように、前景領域内の連結領域ごとに一つのオブジェクトとして認識される。しかし、このオブジェクトには Fig. 5 (a) のように複数人物が含まれる可能性がある。そこで本手法では、差分ステレオにより得られる前景領域の距離画像をセグメンテーションすることで、個々のオブジェクトを検出する。

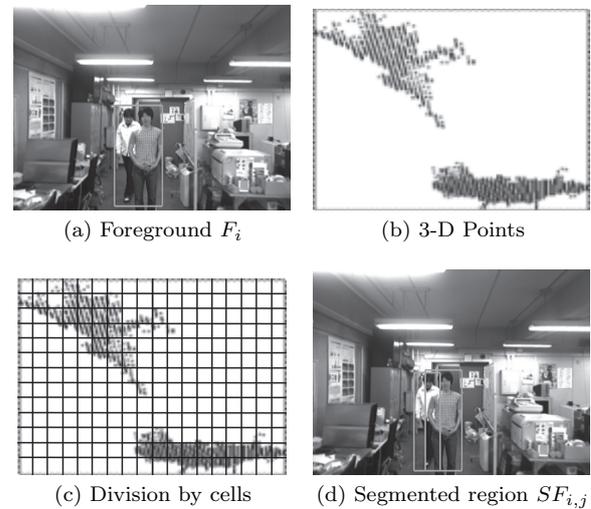


Fig. 5 Range image segmentation

検出されたオブジェクトの総数を n とし、これらのオブジェクトを $F_i (i = 1, \dots, n)$ と定義する。以降の処理はこの F_i ごとに行う。 F_i を構成している各画素はステレオカメラによって計測された距離情報を備えている。これらの画素を世界座標系の $X-Y$ 平面 (地面) に投影する。Fig. 5 (a) の矩形内に含まれる前景領域の距離情報を $X-Y$ 平面に投影した画像を Fig. 5 (b) に示す。さらに、Fig. 5 (c) のように $X-Y$ 平面を一定サイズ (実験では $5[\text{cm}] \times 5[\text{cm}]$) のセルで分割し、各セルに含まれる画素数からヒストグラムを構成する。また、作成されたヒストグラムに対してしきい値処理し、オブジェクトの持つ距離情報の外れ値を除去することで、次節で説明する Mean Shift Clustering の処理時間短縮を図る。

3.2 Mean Shift Clustering によるセグメンテーション

人物が直立していると仮定すると、ヒストグラムのピーク周辺に人物が存在する可能性が高い。そこで、ヒストグラムの頻度に対して Mean Shift Clustering [18] を行い、ピーク位置とピーク数を自動推定する。投影平面上の任意のセル c の位置ベクトルを \mathbf{P}_c とすると、重心位置 \mathbf{v} における Mean Shift ベクトル $\mathbf{m}(\mathbf{v})$ は次式で表される。

$$\mathbf{m}(\mathbf{v}) = \frac{\sum_c \mathbf{P}_c H_c g\left(\left\|\frac{\mathbf{v} - \mathbf{P}_c}{\sigma}\right\|^2\right)}{\sum_c H_c g\left(\left\|\frac{\mathbf{v} - \mathbf{P}_c}{\sigma}\right\|^2\right)} - \mathbf{v} \quad (2)$$

ここで、 H_c はセル c におけるヒストグラムの頻度、 g はガウシアンカーネル、 σ はガウシアンカーネルで用いる標準偏差を表す。この Mean Shift ベクトル $\mathbf{m}(\mathbf{v})$ を用い、次のステップで距離画像をクラスタリングする。

- (1) 投影面の大きさに応じてカーネルの初期位置、配置数を推定投影面の行方向のセル数を $Width$ 、列方向のセル数を $Depth$ としたとき、カーネルの配置数 $Number$ は

$$Number = \frac{Width}{Kernelwidth} \times \frac{Depth}{Kerneldepth}$$

とする。このとき、 $Kernelwidth$ 、 $Kerneldepth$ はそれ

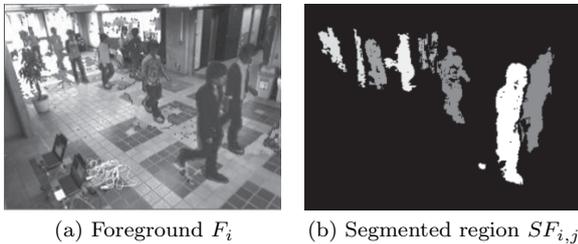


Fig. 6 Range image segmentation in a crowded scene

ぞれ、カーネルの行方向のセル数、列方向のセル数である。またカーネルの初期位置は、投影面の左上を基準にカーネルが重ならないように敷き詰めて配置する。

- (2) 各カーネルを式 (2) の反復計算により移動させ、ヒストグラムのピーク位置を推定
- (3) 近傍のカーネルを統合し、重心位置より一定範囲内のセルを同じクラスタとする

分類されたクラスタごとに、セルに含まれる投影点を画像上に逆投影することで、前景領域をセグメンテーションする。Fig. 5 (a) に示す画像に対するセグメンテーション結果を Fig. 5 (d) に示す。この手法によって F_i が m_i 個に分割された領域を $SF_{i,j} (j = 1, \dots, m_i)$ と定義する。画像中に多数の人物が映っている状態において、セグメンテーション手法を用いた例を Fig. 6 に示す。

4. オクルージョンを考慮した人物検出

距離画像セグメンテーションの結果を利用し、特徴量の算出時間を削減してリアルタイムでの人物検出手法を構築する。また、セグメンテーション結果からオクルージョンを検出し、遮蔽部における識別器の寄与を抑制することで検出率の向上を図る。

4.1 Joint HOG 特徴量を用いた人物検出

HOG 特徴量 [1] は画像中の検出ウィンドウ内をセルに分割し、各セルにおける輝度勾配を勾配方向ごとにヒストグラム化することで特徴量を得る。単一の HOG 特徴量では人の対称的な形状や連続的な形状を表現することが困難なため、本研究では、位置の異なる二つのセル内の HOG 特徴量の共起性 [22] を同時に捉える Joint HOG 特徴量 [19] を用いる。

共起を表現した特徴量を全セルの組み合わせに対し求め、1 段階めの Real AdaBoost [23] により識別に有効な特徴の組み合わせを選択し、Joint HOG 特徴量を生成する。その後、生成された Joint HOG 特徴量から 2 段階めの Real AdaBoost により識別に有効な特徴量のみを選択し、強識別器 $H(\mathbf{X})$ を構築する。

$$H(\mathbf{X}) = \sum_{t=1}^T h_t(\mathbf{X}) \quad (3)$$

ここで、 \mathbf{X} は選択された Joint HOG 特徴、 T は 2 段階めの学習回数、 $h_t(\mathbf{X})$ は 1 段階めの学習から得られる弱識別器を表す。

4.2 検出ウィンドウの走査・統合

本手法では、検出ウィンドウの走査を Fig. 6 (b) のようなオブジェクトごとに分割された領域 $SF_{i,j} (j = 1, \dots, m_i)$ ごとに

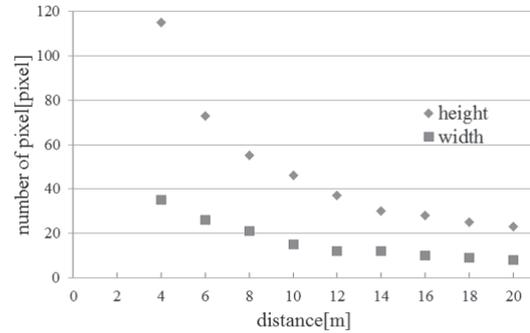


Fig. 7 Transition of the size of human

行う。ここで、 m_i は前景領域 F_i が分割された数である。また、領域ごとの距離情報をもとに、ウィンドウサイズを動的に変化させることで走査回数を削減する。

検出ウィンドウのサイズは、画像中の人物の大きさがカメラからの距離と反比例の関係にあることを利用して推定する。また、カメラの仰角による人物の見えの変化が擬似透視投影に従うと仮定し、下記の式でウィンドウサイズ R_h (高さ)、 R_w (幅) を算出する。

$$R_h = \frac{k_h}{W_Y(i, j)} (\cos \theta - y \sin \theta) \quad (4)$$

$$R_w = \frac{k_w}{C_Z(i, j)} \quad (5)$$

ここで、 $W_Y(i, j)$ は世界座標系でのカメラから領域 $SF_{i,j}$ の重心位置までの距離、 $C_Z(i, j)$ はカメラ座標系でのカメラから領域 $SF_{i,j}$ の重心位置までの距離、 θ はカメラの仰角、 y は画像の縦幅を正規化したときの画像座標を表す。また、Fig. 7 において、それぞれのピクセル数の推移が距離に反比例していることから比例定数 k_h 、 k_w を算出した。

Fig. 6 (b) での領域 $SF_{i,j}$ ごとに提案手法を用いて走査した結果を Fig. 8 (a) に示す。特徴量算出の時間を削減するとともに、背景からの誤検出を低減している。

特徴量算出後、人物と識別された検出ウィンドウは、近傍のウィンドウと統合されて検出結果となる。Fig. 8 (a) での検出ウィンドウの統合結果を 8 (b) に示す。検出ウィンドウの統合を画像全体で行ってしまうと、オクルージョンが発生している場合、一つのウィンドウに統合されてしまう可能性がある。そこで、ウィンドウの統合を分割された領域 $SF_{i,j}$ ごとに行う。すなわち、Fig. 8 (a) で異なる色の検出ウィンドウ間では統合が行われない。これにより、画像全体でウィンドウの結合を行った際に生じるウィンドウの過統合をなくし、オブジェクトの未検出を減らしている。

4.3 オクルージョンを考慮した識別器の補正

画像中で人物同士が重なる場合、遮蔽により後方の人物が隠れてしまい、局所特徴量が正しく取得できない。そこで、検出ウィンドウ内に複数の領域 $SF_{i,j}$ が含まれるとき、 $SF_{i,j}$ をもとにオクルージョンが発生しているセルを検出し、識別器の出力を補正することで、検出精度を向上させる [11]。

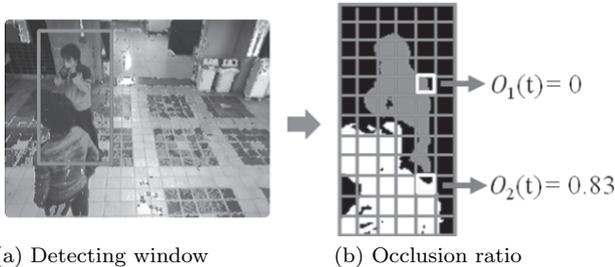
Fig. 9 (b) に示すように、走査対象となる領域 (灰色) より手前にある領域 (白色) をオクルージョンが発生している領域と



(a) Scanning region



(b) Integration of windows

Fig. 8 Detecting window

(a) Detecting window

(b) Occlusion ratio

Fig. 9 Occlusion detection

して検出する. このとき, 各セルでオクルージョンを含む割合 (Fig. 9 (b) で白い領域を含む割合) を算出する. Joint HOG 特徴量は組み合わせた二つのセル内の特徴量から弱識別器 $h_t(\mathbf{X})$ の出力をもとめるが, その際に各セルで算出されたオクルージョンの割合を用いて, 最終識別器 $H'(\mathbf{X})$ を下記の式で補正する.

$$H'(\mathbf{X}) = \sum_{t=1}^T \{h_t(\mathbf{X}) \cdot (1 - O_1(t)) \cdot (1 - O_2(t))\} \quad (6)$$

ここで, $O_1(t)$, $O_2(t)$ は弱識別器 $h_t(\mathbf{X})$ で使用される二つのセルにおけるオクルージョンの割合を表す. オクルージョンの割合が大きいく程, 弱識別器の出力が小さくなることから, 遮蔽部の識別器の出力を抑制することができる. この最終識別器 $H'(\mathbf{X})$ に対し閾値を設け, 人物か否かを識別する.

5. 評価実験

5.1 実験条件

学習には NICTA Pedestrian Dataset [24] を使用し, ポジティブサンプル 7,892 枚, ネガティブサンプル 30,000 枚を用

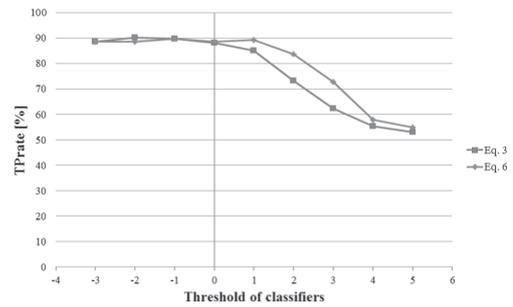


(a) Positive

(b) Negative

Fig. 10 Samples used for training

(a) Experimental scene



(b) Comparison of classifiers

Fig. 11 Evaluation of classifiers

いた. 使用したサンプル例を **Fig. 10** に示す. 学習は 1 段階めで 10 回, 2 段階めで 300 回行い, 識別器を構築した. 実験時のステレオカメラは Bumblebee2 (Point Grey Research) を使用し, 処理には Intel Core 2 Duo CPU (3.06 [GHz]) を搭載した PC を用いた. また, 評価結果の TPrate (True Positive rate) は正しい検出率, FPrate (False Positive rate) は誤検出率を表す.

5.2 オクルージョンを考慮した識別器の評価

4.3 節で述べたオクルージョンを考慮した識別器の有用性を検証するため, 式 (3) と式 (6) の識別器を用いて検出精度を比較する. **Fig. 11** (a) に示す屋内で, オクルージョンが発生するシーン 400 フレームで評価を行った. 識別器に用いる閾値の値を変化させていったときの評価結果を Fig. 11 (b) に示す. また, 誤検出がでない限界の閾値での検出精度を **Table 1** に示す. このときの後方の人物に対するそれぞれの識別器の出力を **Table 2** に示す.

Fig. 11 (b) より, オクルージョンを考慮することで検出精度が向上できていることが見てとれる. これは, **Table 2** の識別器の出力から分かるように, 遮蔽部の識別器出力を抑制したこ

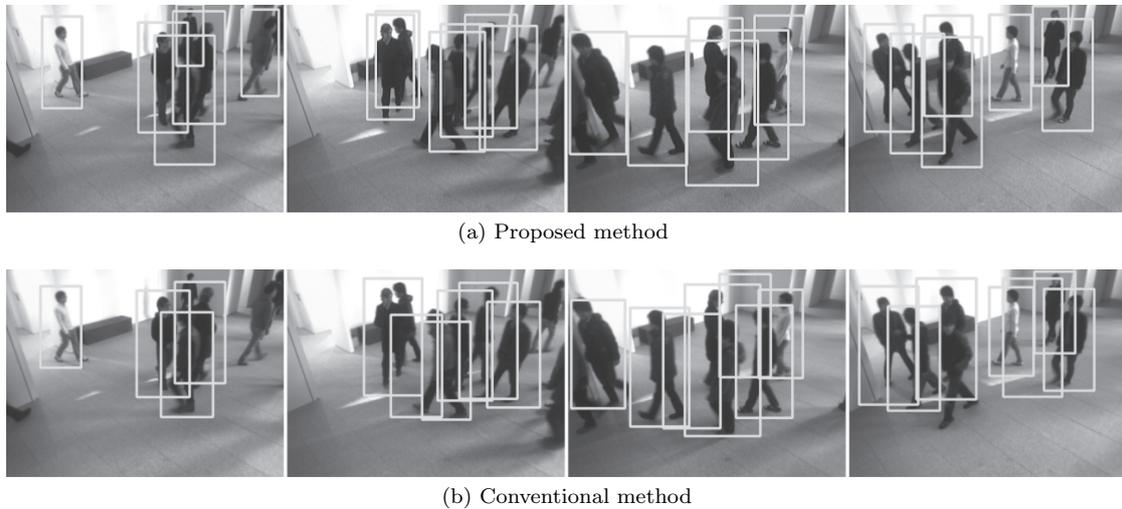


Fig. 12 Examples of detection in simple background

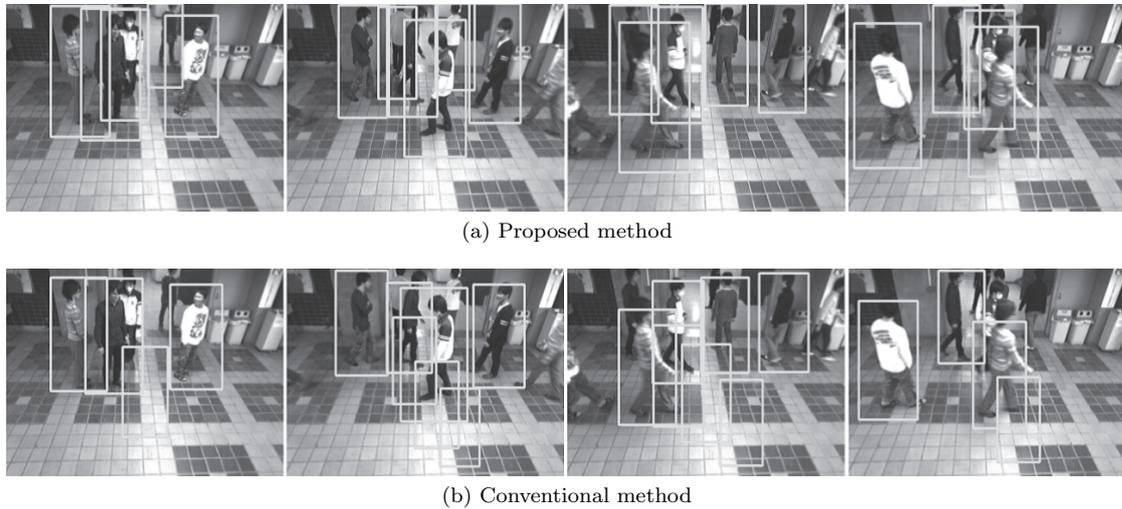


Fig. 13 Examples of detection in complicated background

Table 1 Comparison with and without occlusion handling

Classifier	TPrate [%]	FPrate [%]
$H(x)$ (Eq.3)	71.3	0.0
$H'(x)$ (Eq.6)	89.1	0.0

Table 2 Results of the output by classifier

Classifier	$H(x)$ (Eq.3)	$H'(x)$ (Eq.6)
Average value	-4.70	-3.71
Maximum value	18.8	17.3
Minimum value	-16.6	-15.3
Standard deviation	5.61	5.36

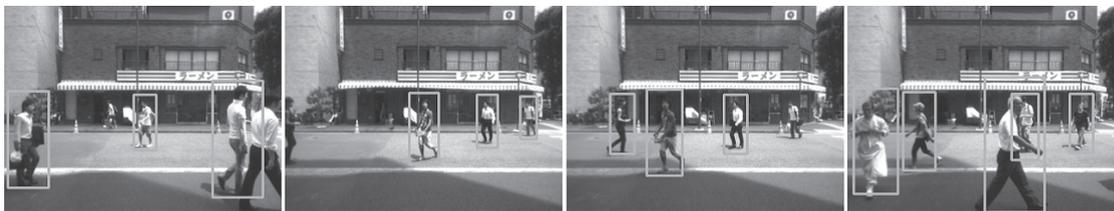
とが要因として考えられる。

5.3 提案手法の評価

提案手法の有用性を検証するために、画像全体で検出ウィンドウをスキャンする手法（従来手法）との精度を比較する。評

価には学習データセットと異なるシーンを用い、屋内環境下の単純な背景と複雑な背景のシーン、屋外環境下での照明変化の大きいシーンにおいてそれぞれ 2,000 フレーム用いた。閾値は単純な背景の動画において、検出精度が最も良くなるしきい値をそれぞれの手法で実験的に求め、すべてのシーンでの実験に用いた。屋内環境下での単純な背景での検出結果の例と評価結果を Fig. 12, Table 3 に、複雑な背景での検出結果の例と評価結果を Fig. 13, Table 4 に示す。また、屋外環境下での照明変化の大きいシーンでの検出結果の例と評価結果を Fig. 14, Table 5 に示す。

Table 3, Table 4, Table 5 の結果より従来手法では背景の複雑化により検出精度が低下しているのが分かる。これは Fig. 13 (b) に示すように、背景に人物の形状に似た形状（図の中心にある十字模様など）が映ると誤検出を誘発してしまうことが原因である。それに対し、提案手法では処理領域を限定することで誤検出を低減している。また、オクルージョンを考慮しているため、Fig. 12 (b), Fig. 13 (b), Fig. 14 (b) においてオクルージョンにより未検出になっている人物も、Fig. 12 (a),



(a) Proposed method



(b) Conventional method

Fig. 14 Examples of detection in outdoor

Table 3 Evaluation result in simple background

Method	TPrate [%]	FPrate [%]
Proposed	80.0	1.3
Conventional	73.0	9.8

Table 4 Evaluation result in complicated background

Method	TPrate [%]	FPrate [%]
Proposed	83.1	3.1
Conventional	63.5	65.9

Table 5 Evaluation result in outdoor

Method	TPrate [%]	FPrate [%]
Proposed	68.8	1.1
Conventional	48.8	68.5

Fig. 13 (a), Fig. 14 (a) では正しく検出していることが見てとれる。Fig. 14 (a) においては、シーン内での背景、照明の変化(建物の影など)が起きているが、それによる誤検出は見られなかった。しかし、人物がカメラに近づきすぎた場合、人物の全身の形状が検出ウィンドウ内に収まらないために、未検出が多く発生した。

5.4 処理速度

Table 6 に各プロセスにおける計算時間を示す。画面全体をスキャンする従来手法と比較し、処理領域を限定することで処理速度が大幅に向上していることが分かる。また、提案手法では約 11[fps] で動作することから、オンラインでの人物検出が可能である。

6. 結論と今後の展望

本論文では、様々な環境下において人物を検出することを目標とし、距離画像セグメンテーションに基づいた人物検出手法

Table 6 Computation time

Process	Proposed method [ms]	Conventional method [ms]
Capture	16.3	16.3
Background subtraction	0.4	-
Stereo matching	18.8	-
Shadow detection	6.7	-
Segmentation	11.5	-
Joint HOG	30.4	502.2
Others	4.0	-
Total	88.1	518.5

を提案した。差分ステレオより検出される前景領域をオブジェクトごとにセグメンテーションし、検出ウィンドウの走査範囲を限定することで人物検出精度と処理速度の性能を向上した。また、前景領域の距離情報とカメラの外部パラメータより前景領域に合った検出ウィンドウのサイズを求めることで、処理速度の性能が向上した。さらに、オクルージョンを考慮した検出ウィンドウの統合により未検出を減らし、オクルージョン割合に応じた識別器出力の抑制により検出精度を向上させた。

今後は、人物のパーツごとの検出と体全体の検出結果を組み合わせて、人物の見えの変化により対応できる識別器を構築していく予定である。

参 考 文 献

- [1] N. Dalal and B. Triggs: "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.886-893, 2005.
- [2] P. Viola and J. Jones: "Rapid object detection using a boosted cascade of simple features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.511-518, 2001.
- [3] K. Levi and Y. Weiss: "Learning object detection from a small number of example: The importance of good feature," IEEE Computer Society Conference on Computer Vision and Pattern

- Recognition, vol.2, pp.53–60, 2004.
- [4] B. Wu and R. Nevatia: “Detection of multiple, partially occluded human in a single image by bayesian combination of edgelet part detectors,” IEEE International Conference on Computer Vision, vol.1, pp.90–97, 2005.
- [5] Y. Freund and M. Jones: “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” EuroCOLT’95, pp.23–37, 1995.
- [6] B.E. Boser, I. M.Guyon and V.N. Vapnik: “A Training Algorithm for Optimal Margin Classifiers,” Proc. 5th Annual Workshop on Computational Learning Theory, pp.144–152, 1992.
- [7] C. Papageorgiou, et al.: “A Trainable System for Object Detection,” IEEE International Conference on Computer Vision, vol.38, no.1, pp.15–33, 2000.
- [8] P. Viola, et al.: “Detecting Pedestrians Using Patterns of Motion and Appearance,” IEEE International Conference on Computer Vision, pp.734–741, 2003.
- [9] R. Benenson, et al.: “Pedestrian detection at 100 frames per second,” IEEE International Conference on Computer Vision and Pattern Recognition, pp.2093–2910, 2012.
- [10] 鈴木一生, 呉海元, 和田俊和: “ステレオカメラを用いた顔領域の高速化”, 情報処理学会研究報告 CVIM, pp.107–112, 2008.
- [11] 池村翔, 藤吉弘亘: “距離情報に基づく局所特徴量によるリアルタイム人物検出”, 電子情報通信学会論文誌, vol.J93-D, no.3, pp.355–364, 2010.
- [12] V. Vineet, J. Warrell, L. Ladicky and P. Torr: “Human Instance Segmentation from Video using Detector-based Conditional Random Fields,” Proc. of the British Machine Vision Conference, pp.80.1–80.11, 2011.
- [13] X. Wang, T. X. Han and S. Yan: “An HOG-LBP Human Detector with PartialOcclusion Handling,” Proc. of IEEE International Conference on Computer Vision, pp.32–39, 2009.
- [14] T. Ahonen, A. Hadid and M. Pietikinen: “Face description with local binary patterns: Application to face recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.28, no.12, pp.2037–2041, 2006.
- [15] INRIA Person Dataset, <http://pascal.inrialpes.fr/data/human/>
- [16] K. Umeda, et al.: “Subtraction Stereo—A Stereo Camera System That Focuses On Moving Regions—,” Proc. of SPIE-IS & T Electronic Imaging, vol.7239 Three-Dimensional Imaging Metrology, 723908, 2009.
- [17] T. Ubukata, et al.: “Multi-Object Segmentation in a Projection Plane Using Subtraction Stereo,” Proc. International Conference on Pattern Recognition, pp.3296–3299, 2010.
- [18] Y. Cheng: “Mean Shift, Mode Seeking, and Clustering,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.17, no.8, pp.790–799, 1995.
- [19] 三井相和, 山内悠嗣, 藤吉弘亘: “Joint 特徴量を用いた 2 段階 Boosting による物体検出”, 電子情報通信学会論文誌, vol.J92-D, no.9, pp.1591–1601, 2009.
- [20] 寺林賢司, モロアレサンドロ, 梅田和昇: “GPU を用いたノンパラメトリック背景更新”, 平成 22 年電気学会電子・情報・システム部門大会予稿集, TC10-7, pp.398–399, 2010.
- [21] A. Moro, et al.: “Auto-adaptive threshold and shadow detection approaches for pedestrians detection,” Proc. AWSVCI, pp.9–12, 2009.
- [22] T. Mita, T. Kaneko, B. Stenger and O. Hori: “Discriminative Feature Co-occurrence Selection for Object Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.30, no.7, pp.1257–1269, 2008.
- [23] R.E. Schapire and Y. Singer: “Improved Boosting Algorithm Using Confidence-rated Predictions,” Machine Learning, no.37, pp.297–336, 1999.
- [24] G. Overett, L. Petersson, N. Brewer, L. Andersson and N. Pettersson: “A new pedestrian dataset for supervised learning,” Proc. IEEE Intelligent Vehicle Symposium, pp.373–378, 2008.



柴田雅聡 (Masatoshi Shibata)

2012 年中央大学理工学部精密機械工学科卒業, 2014 年同大学院理工学研究科精密工学専攻博士前期課程修了。ロボットビジョン等の研究に従事。



生形 徹 (Toru Ubukata)

2010 年中央大学理工学部精密機械工学科卒業, 2012 年同大学院理工学研究科精密工学専攻博士前期課程修了。ロボットビジョン等の研究に従事。



寺林賢司 (Kenji Terabayashi)

2004 年北海道大学大学院システム情報工学専攻修士課程修了。2008 年東京大学大学院工学系研究科精密機械工学専攻博士課程修了, 博士 (工学)。同年中央大学理工学部精密機械工学科助教, 2012 年より静岡大学工学部機械工学科助教。ヒューマンインタフェース, ロボットビジョン等の研究に従事。

日本機械学会, 精密工学会, 日本バーチャルリアリティ学会, IEEE 等の会員。 (日本ロボット学会正会員)



モロ アレサンドロ (Alessandro Moro)

He received his M.S. degree from the University of Udine, Italy, in 2006, and Ph.D. degree from University of Trieste, in 2010. He is currently a member of Ritecs. His research interests include robotics and image processing.



梅田和昇 (Kazunori Umeda)

1989 年東京大学工学部精密機械工学科卒業, 1994 年同博士課程修了。同年中央大学理工学部精密機械工学科専任講師, 2006 年より同教授。2003–2004 年カナダ NRC Visiting Worker。ロボットビジョンの研究に従事。2004 年 MIRU 長尾賞受賞, 博士 (工学)。日本機械学会 (フェロー), 精密工学会, 電子情報通信学会, IEEE 等の会員。 (日本ロボット学会正会員)

電子情報通信学会, IEEE 等の会員。 (日本ロボット学会正会員)