

距離情報に基づく局所特徴量によるリアルタイム人検出

Real-Time Human Detection Using Local Features Based on Depth Information

池村翔† 藤吉弘亘†

Sho Ikemura†, Hironobu Fujiyoshi†

† 中部大学 大学院 工学研究科

†Department of Computer Science, Chubu University

E-mail: si@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract

近年, 統計的学習法と局所特徴量を用いた人検出に関する研究が多く取り組まれている. 従来の人検出法では, HOG(Histograms of Oriented Gradients) 特徴量のような勾配情報に基づく特徴量を用いる手法が多く, 人の重なりや複雑な背景に対して人検出が困難となる場合がある. さらに, 検出時には検出ウィンドウのスケールを変化させながら画像上を複数回にわたりラスタスキャンするため処理コストが高く, リアルタイム処理が困難である. そこで, 本稿ではTOF カメラから得られる距離情報を用いて人の重なりや複雑なシーンに頑健なリアルタイム人検出手法を提案する. 本研究では, 距離画像から2つの局所領域の距離関係を捉えることができる距離ヒストグラム特徴量を抽出する. 抽出された特徴量を用いて Real AdaBoost 識別器を構築し, 人の識別を行う. 検出時には3次元実空間における検出ウィンドウのラスタスキャンをすることで, 人の大きさに合わない検出ウィンドウを削除して, 高速化を実現する. 3次元実空間における Mean-Shift クラスタリングにより人と識別された検出ウィンドウを統合することで, 人がどこにいるかを自動的に検出することができる. 評価実験の結果, 誤検出率 5.0%において検出率 98.9%となり, HOG 特徴量を用いた従来法と比較して 4.9%検出率を向上させることができた. また, 提案手法は約 10fps でリアルタイムに人検出が可能であることを確認した.

1 はじめに

近年, オフィスや家, 公共施設などの空間において, その空間内の人の意図を理解し行動を支援する技術の実現が期待されている. このような人を観る技術 [1] を実現するためには, 空間内のどこに人がいるかを知る必要があり, 高精度かつ高速な人検出の実現が課題となっている. 人検出は従来から盛んに研究されており, さま

ざまな手法が提案されている [2]-[5]. 従来の可視光カメラを用いた人検出手法では, 局所特徴量と Boosting による統計的学習からなる手法が多い. 局所特徴量には勾配ベースの特徴量である EOH[2], HOG[3], Edgelet[4] 等が多く用いられており, 高精度な人検出を実現できることが報告されている. しかし, 可視光カメラ画像による勾配ベースの特徴量は, 複雑な背景や人の重なりが多い場合, 人の形状を捉えることが困難となり, 検出精度が低下する場合がある. この問題に対して, Essらはステレオカメラにより得られる距離情報を用いて混雑したシーンにおける高精度な人検出法を提案した [6]. しかし, ステレオカメラによる距離情報の取得は, カメラ較正やステレオマッチングによる画像間の対応計算が必要であるため, 処理コストが高く, リアルタイム検出が不可能である. また, 従来の人検出手法においても, 画像中における人のサイズが未知であるため, 検出ウィンドウのスケールを変化させながら繰り返しラスタスキャンを行うことにより, 計算コストが増大し, リアルタイム処理が困難となる.

そこで, 本稿ではTOF カメラから得られる距離情報を用いて人の重なりや複雑なシーンに頑健なリアルタイム人検出手法を提案する. 提案手法はTOF カメラ(Time of Flight) から得られる距離情報を用いて, 局所領域間の距離関係を捉える距離ヒストグラム特徴量を抽出し, Real AdaBoost により識別器を構築する. 構築された識別器を用いて人検出を行い, 3次元実空間における検出ウィンドウのラスタスキャンによる高速化と, オクルージョン領域の考慮による検出精度の向上を実現する.

2 距離情報に基づく局所特徴量

TOF カメラは, カメラの周囲に付いた LED より照射される赤外光が対象物に反射し, カメラで観測されるまでの時間を計測することにより, 物体までの距離を計測するカメラである. 本研究では, TOF カメラとし

て MESA 社の SR-3000 を用いる。SR-3000 は、0.3m ～7.5m(分解能:3m において 22mm) までの距離情報をリアルタイムで取得することができる。人を検出する際、TOF カメラより得られる距離情報を用いて人体と背景の距離関係を捉えることが有効であると考えられる。そこで提案手法では、2つの局所領域間の距離分布の類似度から得られる距離ヒストグラム特徴量を提案する。

2.1 距離ヒストグラム特徴量

距離ヒストグラム特徴量は、2つの局所領域から得られる距離ヒストグラムの類似度を特徴量としたものである。図1に示すように、距離画像を 8×8 [pixel] のセルに分割し、セルで構成される2つの矩形領域を選択する。選択された2つの矩形領域の距離情報から各距離ヒストグラムを算出し、それぞれの距離ヒストグラムの合計が1となるように正規化する。算出された m ビンから成る2つの正規化距離ヒストグラム p, q の各ビンを p_n, q_n とし、Bhattacharyya 距離 [7] による類似度 S を算出し、これを距離ヒストグラム特徴量とする。

$$S = \sum_{n=1}^m \sqrt{p_n q_n} \quad (1)$$

これを全ての矩形領域の組み合わせに対して行うことにより、距離ヒストグラム特徴量ベクトルを算出する。距離ヒストグラム特徴量は、異なる領域の類似度から得られる特徴量であるため、カメラからの距離に依存しない2つの領域の相対的な距離関係を表現する特徴量となる。

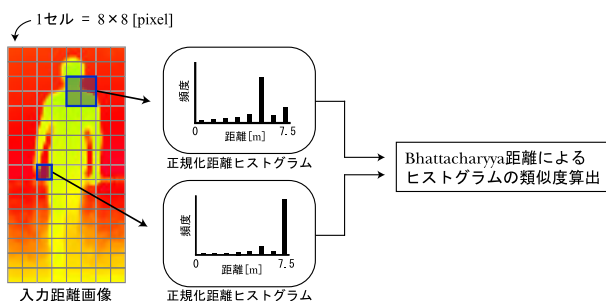


図1 距離ヒストグラム特徴量の抽出

2.2 矩形領域サイズの可変

距離ヒストグラム特徴量では、図2に示すように距離ヒストグラムを算出する際、式(2)を用いてヒストグラムの正規化を行うため、大きさの異なる矩形領域同士における類似度を算出することが可能となる。ここで、 p_n を距離ヒストグラムの n 番目のビンとすると、正規化距離ヒストグラムの n 番目のビン p'_n は、次式に

より求めることができる。

$$p'_n = \frac{p_n}{\sum_{i=1}^m p_i} \quad (2)$$

局所特徴量の矩形領域のサイズを変化させることで、人の肩や頭部のような1つのセルでは表現できない領域における距離関係を捉えることが可能となる。本手法では検出ウィンドウサイズを 64×128 [pixel] とするため、 8×16 のセルに分割できる。距離ヒストグラムの算出矩形領域を 1×1 から 8×8 セルまでセル単位で変化させることにより、492個の矩形領域が存在する。492個の矩形領域の組み合わせから距離ヒストグラム特徴量を算出するため、1つの検出ウィンドウ内から ${}_{492}C_2 = 120,786$ 個の特徴量候補が抽出される。

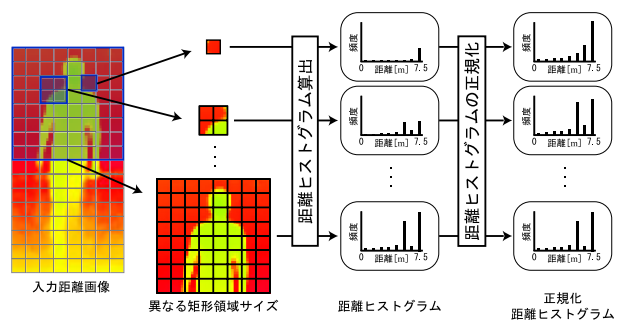


図2 異なる矩形領域サイズにおける距離ヒストグラムの正規化

2.3 Integral Histogram による距離ヒストグラム計算の高速化

特徴量算出時の計算コストを抑えるために、本手法では Integral Histogram [8] を用いて高速に距離ヒストグラムを算出する。まず、距離画像の各画素における距離を 0.3m 間隔で量子化する。本研究では 0m ～7.5m を 0.3m 間隔とするため、25個のビンからなる距離ヒストグラムを算出することになる。次に、図3に示すように、ビン n に対応した25枚の量子化画像 I^n を作成し、各量子化画像 $I^n(u, v)$ から積分画像 $ii^n(u, v)$ を式(3)(4)により算出する。

$$s^n(u, v) = s^n(u, v-1) + i^n(u, v) \quad (3)$$

$$ii^n(u, v) = ii^n(u-1, v) + s^n(u, v) \quad (4)$$

ここで、 $s^n(u, v)$ はビン n における行の画素の総和、 $ii^n(u, v)$ は列の s^n の総和を表わす。ただし、 $s^n(u, -1) = 0$, $ii^n(-1, v) = 0$ とする。図3のDの領域から距離ヒストグラムの n 番目のビン D^n を算出するには、 n 番目の積分画像 ii^n から式(5)により4点から和を求めれば良いことになる。

$$D^n = (ii^n(u, v) + ii^n(u-W, v-H)) - (ii^n(u-W, v) + ii^n(u, v-H)) \quad (5)$$

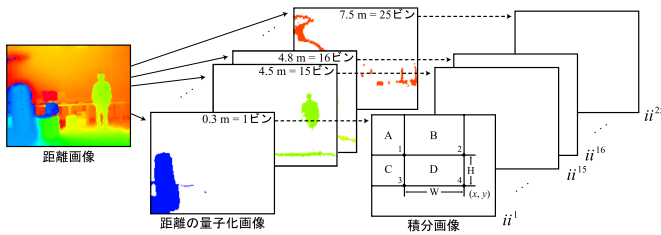


図3 Integral Histogramの算出

これにより、領域の大きさに関係なく距離ヒストグラムの n 番目のビンの値を高速に求めることが可能となる。

3 距離情報を用いた人検出

提案手法における人検出の流れを図4に示す。提案手法は、まず距離画像上において検出ウィンドウをラスタスキャンし、各検出ウィンドウから距離ヒストグラム特徴量を算出する。次に算出された特徴量のオクルージョン判定を行い、Real AdaBoostを用いて各ウィンドウが人か人以外かの識別を行う。ラスタスキャン後、3次元実空間におけるMean-Shiftクラスタリングにより人と識別された検出ウィンドウを統合し、人領域を決定する。

3.1 Real AdaBoostによる識別器の構築

提案手法では、人の識別にReal AdaBoost[10]を用いる。図5にReal AdaBoostによる識別器の構築アルゴリズムを示す。Real AdaBoostは、ポジティブクラスの特徴量とネガティブクラスの特徴量の各次元の確率密度分布から分離度を求め、ポジティブクラスとネガティブクラスを最も分離できる特徴量を弱識別器として選択する。このとき分離度を評価値とするため、実数による識別結果の出力が可能である。学習により選択された弱識別器を $h_t(x)$ とすると、構築される最終識別器 $H(x)$ は式(6)となる。

$$H(x) = \text{sign}\left(\sum_{t=1}^T h_t(x)\right) \quad (6)$$

3.2 3次元実空間におけるラスタスキャン

提案手法では、図6に示すように $y_w = 0$ とした地面上において、 $60 \times 180[\text{cm}]$ の検出ウィンドウを z_w を変化させながら x_w 方向へのラスタスキャンを繰り返すことにより、3次元実空間の地面上をラスタスキャンする。3次元実空間中のラスタスキャンにより得られる検出ウィンドウの3次元座標を式(7)を用いて画像座標 $[u, v]^T$ に投影し、投影された座標位置に対する領域内

Real AdaBoost のアルゴリズム

1. 前処理
 N 個の学習サンプル $x_1, x_2, \dots, x_N$
 クラスに対応したラベル $y \in \{+1, -1\}$ の用意
2. 学習サンプルの重み D_t を初期化
 $D_{1n} = 1/N$
3. For $t = 1, 2, \dots, T$ T 回の学習回数
 For $m = 1, 2, \dots, M$ M 個の弱識別器
 3.1 弱識別器 h_t の確率密度分布 W_t の作成
 $W_+^j = \sum_{i: j \in J_i^+ = +1} D_t(i)$
 $W_-^j = \sum_{i: j \in J_i^- = -1} D_t(i)$
 3.2 弱識別器 $h(x)$
 $h(x) = \frac{1}{2} \ln \frac{W_+^j + \epsilon}{W_-^j + \epsilon}$
 3.3 弱識別器選択のための評価値 Z_m の算出
 $Z_m = 2 \sum_j \sqrt{W_+^j W_-^j}$
 4. 最も Z_m が小さい弱識別器を選択
 $h_t = \arg \min Z_{t,m}$
 5. 学習サンプルの重み $D_t(i)$ の更新
 $D_{t+1}(i) = D_t(i) \exp[-y_i h_t(x_i)]$
 6. 学習サンプルの重み $D_t(i)$ の正規化
7. 最終識別器
 $H(x) = \text{sign}\left(\sum_{t=1}^T h_t(x)\right)$

図5 Real AdaBoostのアルゴリズム

の距離情報から特徴量を計算し、識別を行う。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \quad (7)$$

$$P = A[R|T] \quad (8)$$

行列 P は透視投影行列であり、カメラキャリブレーションにより得られた内部パラメータ A と、外部パラメータである回転行列 R と平行移動ベクトル T から算出される。TOFカメラは画像座標 (u, v) と、対応する $5 \times 4 \times 7.5[\text{m}]$ の空間内における世界座標 (x_w, y_w, z_w) を取得することができるため、容易にカメラキャリブレーションを行うことが可能である。

3.3 オクルージョンに対応した識別

複数の人が重なるような混雑したシーンにおいては、対象とする人領域にオクルージョンが発生する場合がある。オクルージョン領域から抽出される距離情報は、弱識別器の誤った応答を出力する原因となる。従って、このようなオクルージョン領域を捉える弱識別器の出力は、そのまま最終識別器に統合しないようにする。提案手法では検出ウィンドウを3次元実空間においてラ

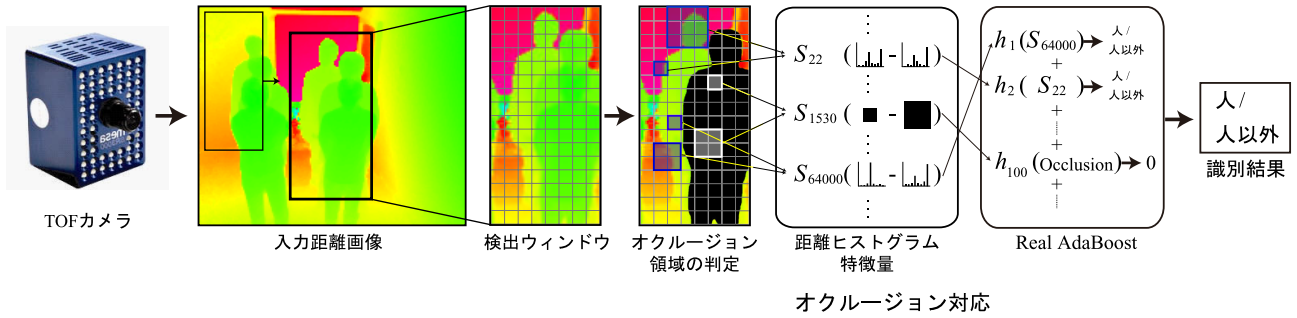


図 4 距離情報を用いた人検出の流れ

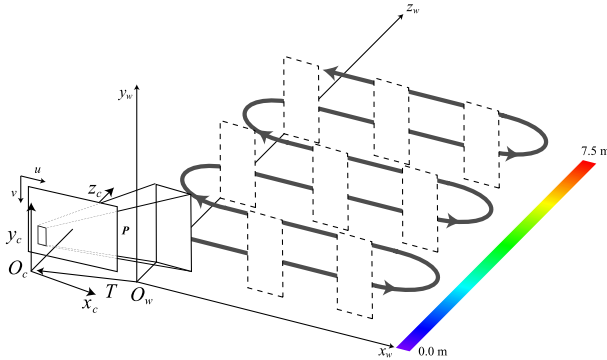


図 6 3次元実空間におけるラスタスキャン

スタスキャンするため，ウィンドウの世界座標が既知である．そこで，TOF カメラから得られる距離情報と比較することで，カメラに対して検出ウィンドウより手前に存在する物体領域をオクルージョンとして判別し，識別に利用する．

3.3.1 オクルージョン領域の抽出

3次元実空間におけるラスタスキャン時の検出ウィンドウの距離 z_w を用いて，オクルージョン判定の閾値を決定する．検出ウィンドウ内の各画素を (u, v) ，そのデプスマップを $d(u, v)$ とすると，各座標におけるオクルージョンラベル $O(u, v)$ は式 (9) で表される．

$$O(u, v) = \begin{cases} 1 & \text{if } d(u, v) < z_w - 0.3 \text{ m} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

抽出されたオクルージョン領域 (黒色の領域) を図 7 に示す．

3.3.2 オクルージョン領域を考慮した識別

t 番目の弱識別器 $h_t(x)$ の対象とする矩形領域 B_t 内に存在するオクルージョン領域の割合をオクルージョン率 OR_t とし，式 (10) より求める．

$$OR_t = \frac{1}{B_t} \sum_{(u,v) \in B_t} O(u, v) \quad (10)$$



図 7 抽出したオクルージョン領域例 (黒色の領域)

算出されたオクルージョン率 OR_t を用いて，オクルージョン領域を考慮した最終識別器 $H'(x)$ は式 (11) で表される．

$$H'(x) = \text{sign}\left(\sum_{t=1}^T h_t(x) \cdot (1 - OR_t)\right) \quad (11)$$

図 8(a) に示すように，オクルージョン領域を考慮せず最終識別器により識別を行うと，多くの弱識別器の出力がマイナスとなり，その結果人以外と誤識別される．一方 (b) は，オクルージョン率を考慮して，最終識別器の出力を求めるため，人と正しい識別が可能となる例である．

3.4 3次元実空間における Mean-Shift クラスタリング

従来の可視光カメラによる人検出 [5] では，人と識別された検出ウィンドウを Mean-Shift クラスタリング [9] により統合し検出結果とする．しかし，画像空間における Mean-Shift クラスタリングのみでは，図 9(b)(d) に示すように人の重なりのある場合に検出ウィンドウが誤って統合されることがある．そこで本手法では，図 9(c)(e) に示すように 3次元実空間における Mean-Shift クラスタリングを行う．これにより，人の重なりがある場合でも距離情報によりクラスタを分離することが可能であるため，検出ウィンドウの誤った統合を抑制することができる．

3次元 Mean-Shift クラスタリングは式 (12) により Mean-Shift Vector $m(\mathbf{x})$ を算出する．ここで， $\mathbf{x}$ は移

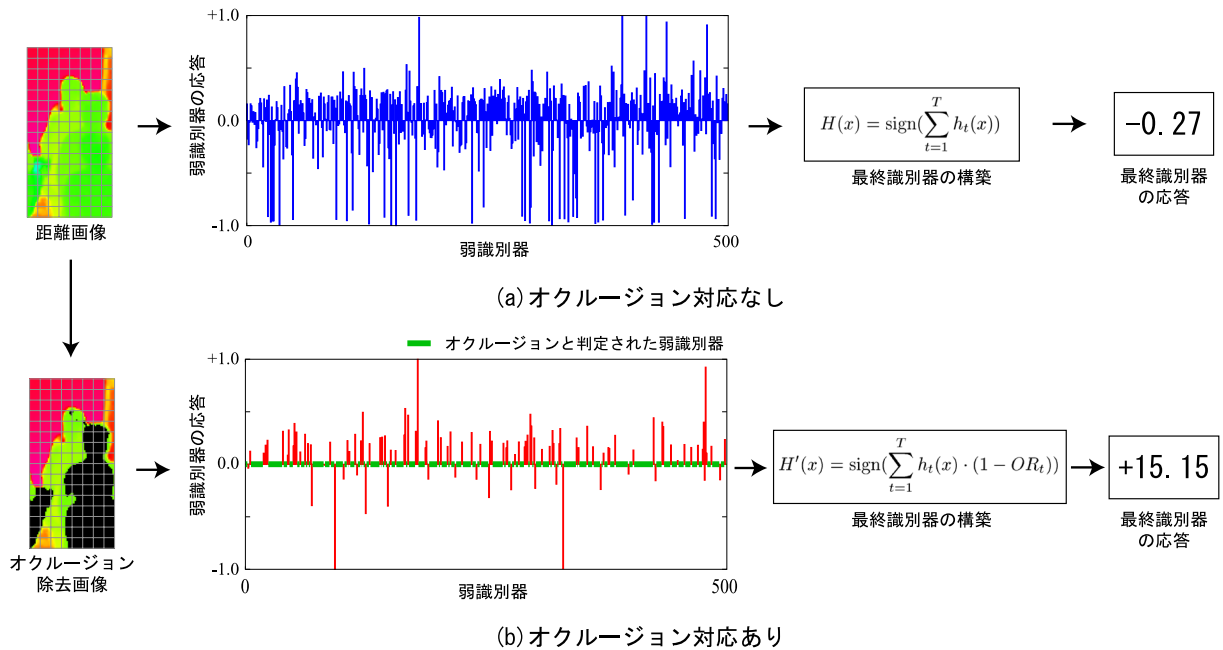


図 8 オクルージョン領域を考慮した識別例

動している検出ウィンドウの中心座標， $\mathbf{x}_i$ は各データの 3 次元座標を表す． k はカーネル関数， h はバンド幅であり，本研究では $h = 0.3\text{m}$ とした．

$$m(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{x}_i k\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^n k\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x} \quad (12)$$

4 識別実験による提案手法の評価

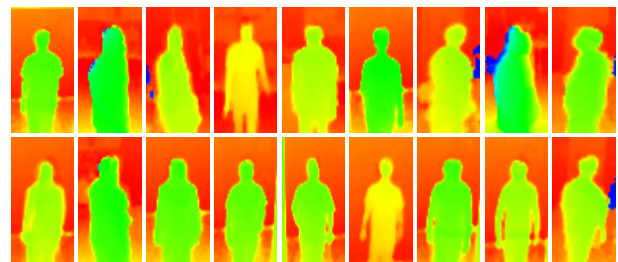
提案手法の有効性を確認するため，評価実験を行う．

4.1 データベース

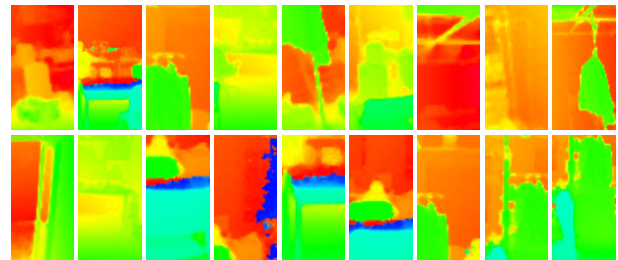
データベースには TOF カメラで撮影したシーケンスを用いる．TOF カメラは屋外での撮影が不可能であるため，屋内でカメラを約 2.5m の高さに設置し，人の歩行シーンと複数の人が重なり合うシーンを対象とした．撮影した屋内のシーケンスから切り出した学習用ポジティブサンプル 1346 枚，学習用ネガティブサンプル 10000 枚を用いる．また，評価には学習用とは別に作成した評価用ポジティブサンプル 2206 枚，評価用ネガティブサンプル 8100 枚を用いる．TOF カメラは屋内において最長 7.5m までの撮影となるため，複数の人の全身を撮影することが困難である．そのため本実験では人の上半身（全身の上部 60%）を検出対象とした．図 10 に学習用，図 11 に評価用のデータベースの一部を示す．

4.2 特徴量の評価実験

評価用データベースを用いて，人の識別実験を行い，提案手法（距離ヒストグラム特徴量）と従来法（距離画



(a) ポジティブサンプル (1346枚)



(b) ネガティブサンプル (10000枚)

図 10 学習用データベースの一部

像の HOG 特徴量) の識別精度による比較を行う．実験結果の比較には，Receiver Operating Characteristic (ROC) カーブを用いる．ROC カーブとは，横軸に誤検出率，縦軸に検出率を表したものである．識別器の閾値を変化させることによって，誤検出率に対する検出率の比較を行うことが可能である．グラフ左上に近いほど検出性能が良いことを表す．

図 12 に特徴量の評価実験結果を示す．距離ヒストグラム特徴量（矩形領域サイズの可変あり）は，誤検出率 5.0%において検出率 98.9%であり，距離画像の HOG 特徴量より 4.9%識別率を向上した．これは，矩形領域サ

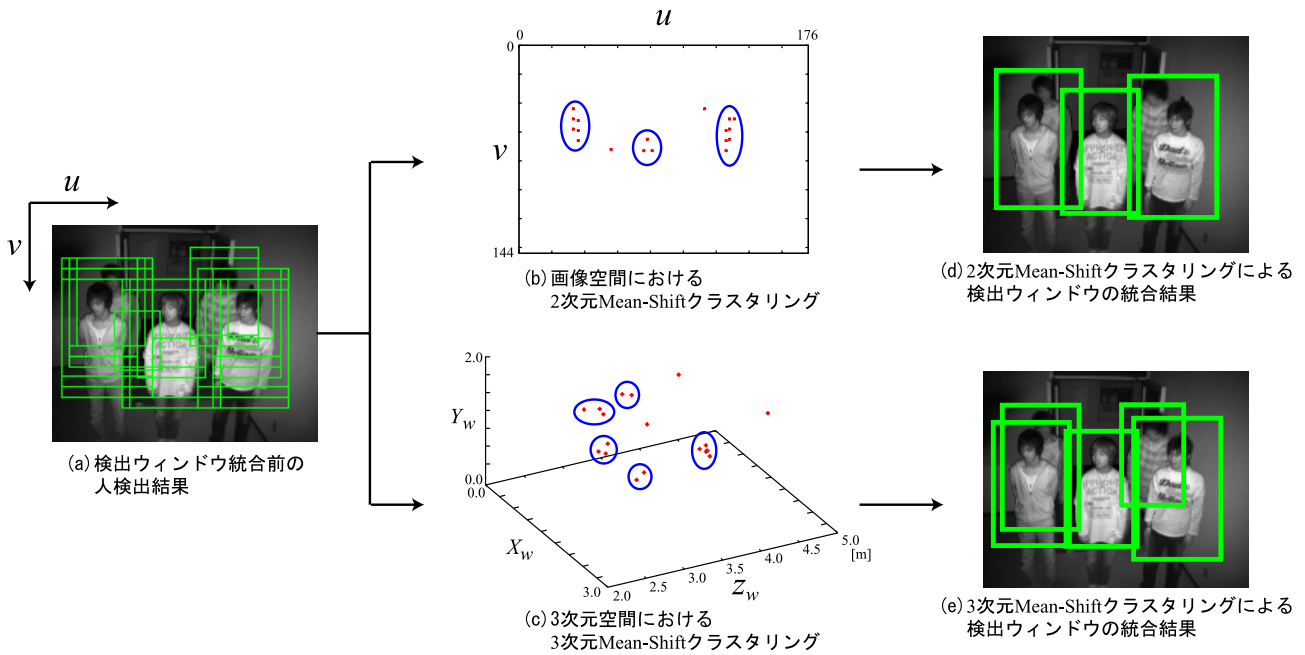


図 9 Mean-Shift クラスタリングによる検出ウィンドウの統合

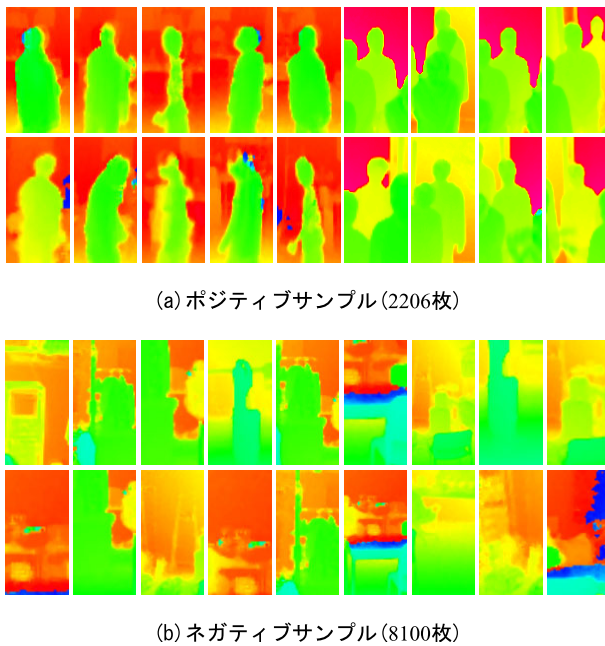


図 11 評価用データベースの一部

イズを可変にすることにより，領域の距離関係を有効に捉えることが可能となるためである．図 13 に誤識別した例を示す．誤識別の傾向として，形状変化の大きい場合や，オクルージョンが発生していることがわかる．

4.3 オクルージョン対応の評価実験

提案手法におけるオクルージョン対応の有効性を示すため人の識別実験により評価を行う．

図 14 にオクルージョン対応ありとなしの評価実験結果を示す．距離ヒストグラム特徴量 (オクルージョン対応あり) は，誤検出率 0.5% において検出率 96.9% であ

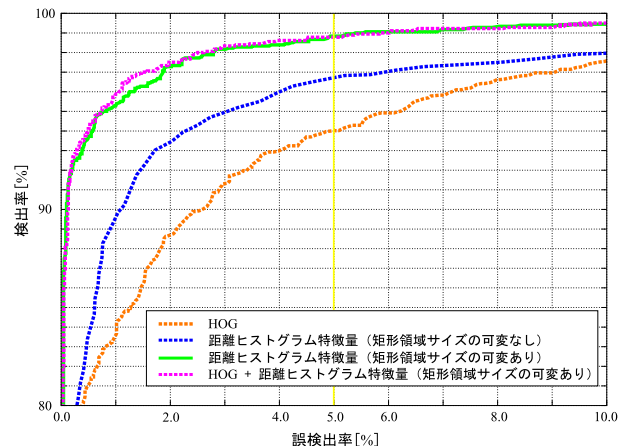


図 12 特徴量の評価実験結果

り，オクルージョンが発生している場合でも高精度に識別が可能であることがわかる．さらに，HOG+距離ヒストグラム特徴量 (オクルージョン対応あり) においても検出率が向上している．これは，オクルージョン率を用いて識別に有効な弱識別器に重みづけし，最終識別器の出力を求めることにより，オクルージョン領域の影響を抑えることができたといえる．

4.4 学習により選択された特徴量

図 15 に提案手法により構築した Real AdaBoost の初期の学習において選択された特徴量を示す．HOG 特徴量 (a) では，人の頭部や肩のエッジが選択されていることから，人の輪郭線を捉えるように特徴量が選択されている．次に，距離ヒストグラム特徴量 (矩形領域サイズの可変なし) (b) では，人の領域と背景領域の組み合わせが多く選択されており，人と背景の境界線を

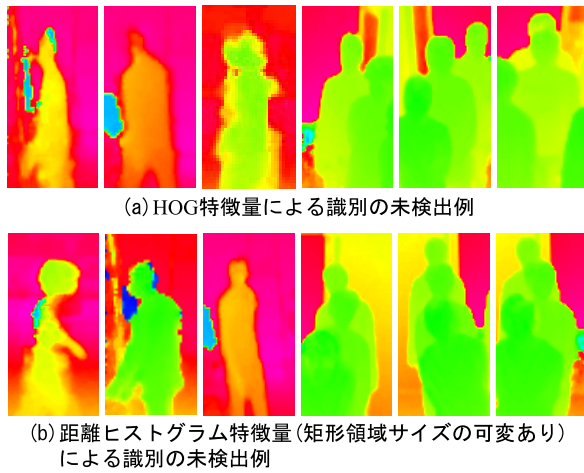


図 13 識別実験における未検出例

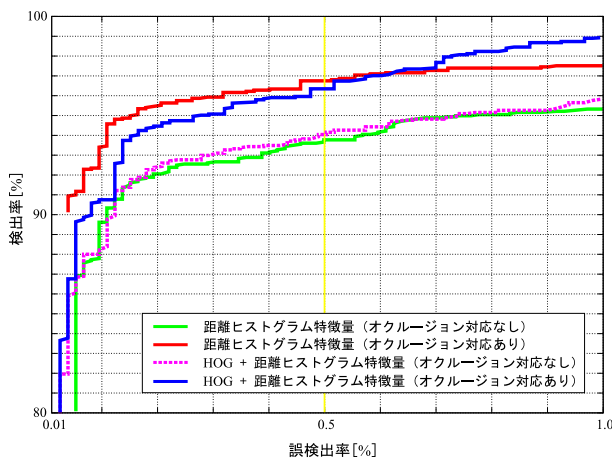


図 14 オクルージョン対応の評価実験結果

捉えていることがわかる．さらに，距離ヒストグラム特徴量 (矩形領域サイズの変更可あり)(c) では，人と背景の境界線を任意のサイズで捉えるように領域が選択されている．これは，人の輪郭線領域の距離関係を捉えることで大まかな人の形状を捉えていると考えられる．また (c) においては，2 回目や5 回目の学習で人の上下，左右の対称的な距離関係を捉えるように領域が選択されている．

4.5 リアルタイム人検出

3次元実空間における検出ウィンドウのラスタスキャンを用いた人検出結果を図16に示す．(a) では，人と同様の高さの物体を誤検出しないで，人のみを検出していることがわかる．さらに (b)(c) では，向きの異なる人の重なりが存在しても，それぞれの人とその3次元位置を正確に検出できていることがわかる．また，表1にIntel Xeon CPU 3.00GHzを用いた際の1フレーム (検出ウィンドウ 361 個) の処理時間を示す．提案手法は約 100ms で処理を行うため，約 10fps でのリアルタイム検出が可能である．

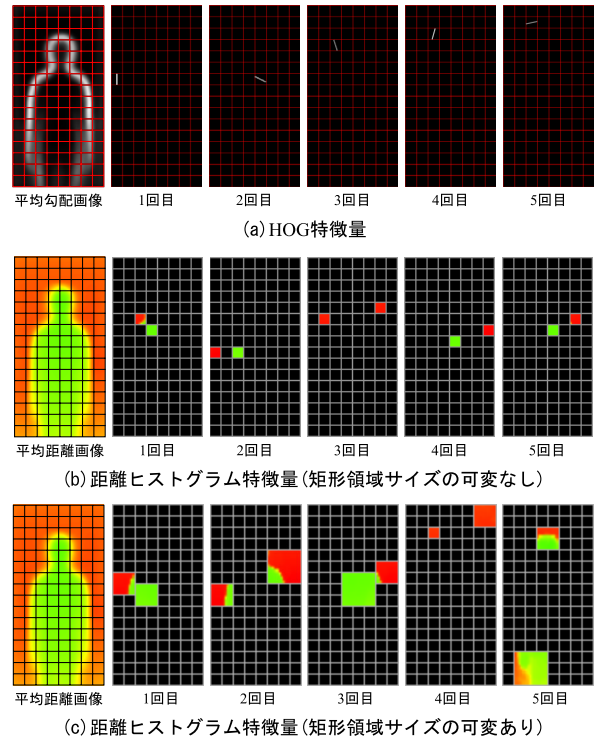


図 15 学習により選択された特徴量

表 1 1 フレームにおける処理時間 [ms]

	検出ウィンドウ1個あたりの処理時間	合計 (361個)
特徴量算出	0.067	24.31
識別	0.125	45.34
検出ウィンドウの統合	—	31.97
合計	—	101.62

5 おわりに

本稿では TOF カメラから得られる距離情報を用いて人の重なりや複雑なシーンに頑健なリアルタイム人検出手法を提案した．評価実験の結果，従来法である HOG 特徴量より 4.9% 検出率を向上させることができた．さらに，提案手法は，特徴量算出，識別，検出ウィンドウの統合における全ての処理時間が約 100ms であるため，約 10fps でのリアルタイム検出が可能であることを確認した．今後は，距離情報とその時間変化を用いた動作解析を行う予定である．

参考文献

- [1] 藤吉弘亘, “人を観る技術: People Image Analysis - インビジブルロボットの視覚機能の実現 -”, 精密工学会サマーセミナー 2008, pp. 113-120, (2008).
- [2] K. Levi and Y. Weiss, “Learning Object Detection from a Small Number of Examples: the Importance of Good Features.”, IEEE Computer Vi-

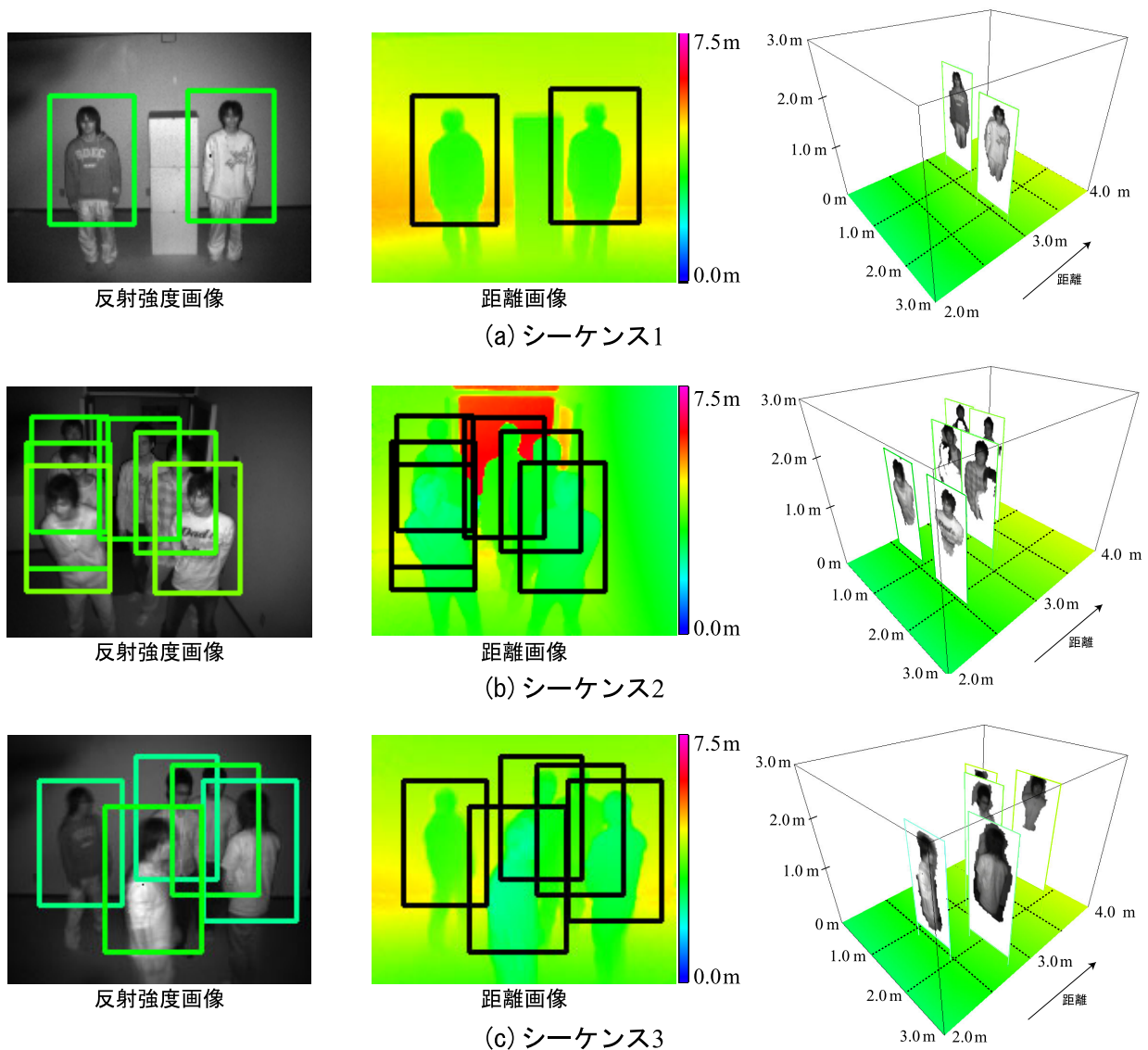


図 16 人検出例

- sion and Pattern Recognition, vol. 2, pp. 53-60, (2004).
- [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).
- [4] B. Wu and R. Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors", IEEE International Conference on Computer Vision, vol. 1, pp. 90-97, (2005).
- [5] 山内悠嗣, 藤吉弘亘, Hwang Bon-Woo, 金出武雄, "アピランスと時空間特徴の共起に基づく人検出", 画像の認識・理解シンポジウム (MIRU2007), pp. 1492-1497, (2007).
- [6] A. Ess, B. Leibe and L. V. Gool, "Depth and Appearance for Mobile Scene Analysis", IEEE International Conference on Computer Vision, pp. 1-8, (2007).
- [7] A. Bhattacharyya, "On a Measure of Divergence between Two Statistical Populations Defined by Probability Distributions", Bull. Calcutta Math. Soc., vol.35, pp. 99-109, (1943).
- [8] F. Porikli, "IntegralHistogram: A Fast Way to Extract Histograms in Cartesian Spaces", IEEE Computer Vision and Pattern Recognition, pp. 829-836, (2005).
- [9] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications", IEEE International Conference on Computer Vision, pp. 1197-1203, (1999).
- [10] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions", Machine Learning, No.37, pp.297-336, (1999).