距離情報を用いたHaar-likeフィルタリングによる人検出

Human Detection by Haar-like Filtering using Depth Information

池村翔　川合俊輔　藤吉弘亘
Sho Ikemura, Shunsuke Kawai, Hironobu Fujiyoshi
中部大学 工学部 情報工学科
Department of Computer Science, Chubu University
E-mail: si@vision.cs.chubu.ac.jp

Abstract

本稿では，TOF カメラにより人を上部から撮影して得られる距離情報を用いて，人の形状を表現する Haar-like フィルタによる高精度な人検出法を提案する．提案手法では，人を上部から撮影するため距離画像の背景差分により大まかな人領域を抽出する．抽出された人領域の距離情報に対し，人の肩，頭，肩という凸形状を表現した Haar-like フィルタの出力を計算する．フィルタリング結果を Mean-Shift クラスタリングにより統合することで人検出を実現する．背景差分により得られた距離情報を Mean-Shift クラスタリングにより統合する手法との比較実験より，提案手法は検出率を 6.6%向上した．このとき提案手法は約 17 fps 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

従来の人検出手法は，HOG 特徴量 [1] に代表される勾配ベースの特徴量と，AdaBoost や SVM 等の統計的学習手法の組み合わせで実現されている [2][3][4]．これらの手法は，人の局所的な形状を捉える人識別器を学習により構築するため，汎化能力が高く，高精度な人検出が可能であることが報告されている．しかしながら，可視光カメラ画像から抽出される勾配ベースの特徴量は，複雑な背景や人同士の重なりが存在する場合に物体の形状を捉えることが困難となる問題がある．背景テクスチャや人同士の重なりに対応するため，距離情報を用いた人検出手法 [5] も提案されているが，完全に人同士が重なる場合には対応できない．また統計的学習手法による人検出手法は，事前に大量の学習用データを必要とするためデータの収集に時間がかかるという問題がある．さらに，作成した学習データと異なるシーンにおける検出は困難となるため，実用化にはこれらの問題を解決する必要がある．

このような問題に対して，人を上部から撮影し，統計的学習手法を用いない人検出法が提案されている．上部から人を撮影することで，人同士の重なりに影響されず，また背景の変化が少ないことから，背景差分処理による物体領域の抽出も容易となる利点がある．技研トラステムより製品化されている人検出システム [6] は，天井に設置されたカメラにより人を上部から撮影し，ベクトル焦点法により人の形状を表現することで，高精度なリアルタイム人検出を実現している．この手法は，事前に人モデルを用意する必要があるため，モデルと大きく異なる人が存在する場合に検出精度が低下すると考えられる．また，Point Grey Research の Censys3D People Tracking System[7] は，ステレオカメラを用いて人を上部から撮影し，得られる距離情報を用いて人の頭部の形状を検出することで人検出を実現している．この手法は，頭部を検出し，その高さから人を検出しているため，同じような高さの物体が存在する場合に誤検出することが考えられる．

そこで，本研究ではリアルタイムで距離情報を取得することができる TOF(Time of Flight) カメラを用いて，歩行者を上部から撮影し，頭や肩の凹凸を表現することで，人の重なりに影響されない高速，高精度な人検出法を提案する．提案手法では，TOF カメラ (Time of Flight) を用いて人を上部から撮影し，得られた距離情報に対し背景差分を行うことで物体領域の抽出を行う．抽出された物体領域に対し人の肩，頭，肩という凸形状を検出するための Haar-like フィルタリングを行う．フィルタリング結果を Mean-Shift クラスタリングにより統合することで，学習を必要としない高精度なリアルタイム人検出を行う．

2　関連研究

人を上部から撮影した画像より人を検出する手法の多くは，人の頭や上半身の形状を捉えるモデルベースの手法が多い．技研トラステムの人検出システム [6] は，ベクトル焦点法を用いて人検出を行う．まず，事前に撮

影した背景画像から一定領域内の輝度値を用いて基本波フーリエ変換により法線ベクトルの算出を行う．次に背景画像に簡易な人モデルを配置した画像においても同様に法線ベクトルの算出を行う．そして，得られた背景画像と人モデルを配置した背景画像の法線ベクトルを比較し，同じ位置と角度の法線ベクトルを削除する．これにより，人モデルの中心点から輪郭上の法線ベクトルへの位置と角度の抽出が可能となる．この人モデルの輪郭上の法線ベクトルを用いて，未知の画像に対してマッチングを行い，モデルとの法線ベクトルのマッチング率を閾値処理することで人の検出を行う．また，見えの変化に対応するために頭部と上半身を分割したモデルを用いてマッチングを行い，パーツごとの検出を行うことで高精度化を実現している．しかし，パロッシーの人検出システムは事前に用意したモデルの輪郭形状を表現しているため，見えの変化が大きくなるにつれ識別精度が低下すると考えられる．

Point Grey Research の Censys3D People Tracking System[7] は，ステレオカメラシステムにより得られる距離情報を用いて人の頭部を検出し，その高さを閾値判定することにより人検出を行う．そのため，人と同じ高さの物体が存在する場合には誤検出する問題がある．また，混雑したシーンにおいて検出精度が著しく低下することが報告されている．

Beleznai らは，背景差分により得られた物体領域内の輝度分布を用いて，Mean-Shift クラスタリングを行うことにより，人の重なりが存在する場合にも高精度に人検出する手法を提案している [8]．しかし，近接した人同士が同じような輝度分布である場合にクラスタを分離できないため誤検出となると考えられる．

3 提案手法

提案手法では，TOF カメラを用いて人を上部から撮影し，得られた距離情報に対し背景差分を行うことで物体領域を抽出する．抽出された物体領域に対し Haar-like フィルタリングにより人の肩，頭，肩という凸形状を検出する．検出された凸形状の頂点の距離情報を Mean-Shift によりクラスタリングすることで人検出を行う．図 1 に提案手法の流れを示す．

3.1 TOF カメラから得られる距離情報

本研究における距離情報の取得には TOF(Time of Flight) カメラを用いる．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-4000 を用いる．SR-4000 は，視野角 $43.6 \times 34.6[\ ^\circ]$ において 0.8〜5.0[m](絶対距離精度：±1%) までの距離情報をリアルタイム（約 30 fps）で取得することができる．

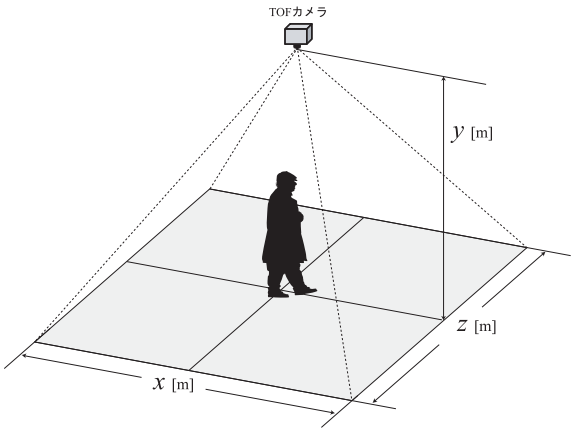


図 2　実験環境

提案手法では図 2 に示すように TOF カメラを設置し，人を上部から撮影する．TOF カメラを設置する高さ y を変化させることにより x，z が変化する．TOF カメラの視野角 (43.6 °× 34.6 °) を $(\theta_h \times \theta_v)$ とすると，ある高さ y における x，z はそれぞれ式 (1)，(2) より算出することができる．

$$x = 2y \tan\theta_h \quad (1)$$

$$z = 2y \tan\theta_v \quad (2)$$

3.2 距離情報を用いた背景差分

人を上部から撮影するシーンでは，カメラから地面までの距離が変化しないため容易に背景差分処理を行うことができる．提案手法は，式 (3) のように入力距離画像と背景距離画像の差分を求めることにより，物体領域を抽出する．物体領域を抽出することで背景領域の処理を削減できるため高速化を実現する．

$$M(u,v) = \begin{cases} I(u,v) & \text{if } |I(u,v) - B(u,v)| > \text{th} \\ \text{Background} & \text{otherwise} \end{cases} \quad (3)$$

ここで，I は現在の距離画像，B は背景距離画像，th は閾値であり，本実験では 0.3m としている．図 3 に背景差分による物体領域の抽出例を示す．

3.3 Haar-like フィルタリングによる凸形状の判別

上部から人を撮影する場合，人は肩，頭，肩という凸型の形状で表現することができる．そこで提案手法では，人の肩，頭，肩の凸形状を捉えるために Haar-like フィルタリングを行う．Haar-like フィルタを用いることにより人の凸形状を捉えることができる．

3.3.1 Haar-like フィルタ

Haar-like フィルタは，Papageorgiou らにより提案されたフィルタである [9]．Viola らは，Haar-like フィル

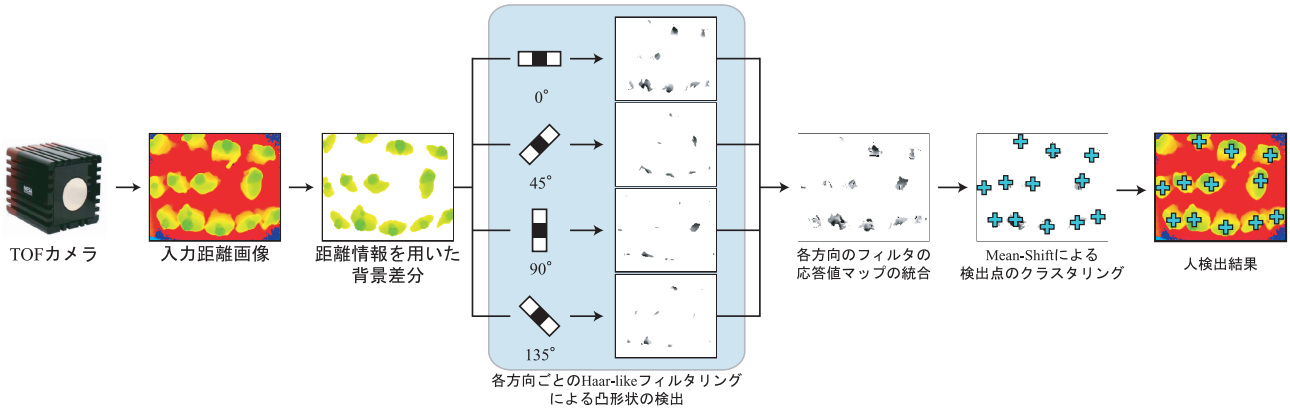


図 1　提案手法による人検出の流れ

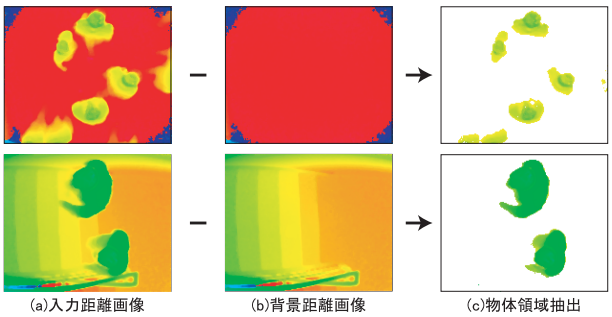


図 3　背景差分による物体域の抽出例

図 4　提案手法で用いる **Haar-like** フィルタ

タの応答値を用いて高速，高精度な顔検出を実現している [10]．顔検出では，Haar-like フィルタにより顔の明暗差を捉えている．

提案手法では，距離画像に対し Haar-like フィルタによりフィルタリングを行い，人の凸形状を捉える．図 4 に示すように，黒の領域 $r1$ と白の領域 $r2$ の 2 つの領域の距離差を式 (4) より算出し Haar-like フィルタの応答値 $H(r1, r2)$ を算出する．

$$H(r1, r2) \quad = \quad S(r1) - S(r2)/2 \tag{4}$$

ここで，$S(r)$ は，領域 r の距離和を算出する関数である．提案手法では，図 4 に示すように，人の向きに対応するため，0 °，45 °，90 °，135 °の 4 方向の Haar-like フィルタを用いる．算出された Haar-like フィルタの応答値 $H(r1, r2)$ を用いて式 (5) により閾値処理することで，凸形状であるかを判別する．

$$F(u, v, d) \quad = \quad \begin{cases} 1 & \text{if} \quad H(r1, r2, d) > \text{th} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

ここで，$F(u, v, d)$ は d 方向の Haar-like フィルタリング結果より得られるフィルタマップであり，座標 (u, v) における凸形状のラベル $y \in (0, 1)$ を与えたものである．また，th は凸形状を判別するための閾値である．

3.3.2　Integral Image を用いた Haar-like フィルタ算出の高速化

Haar-like フィルタリングは，距離画像中の物体領域に対してフィルタリングを行うため，処理コストが膨大となる問題がある．そこで，提案手法では Haar-like フィルタの応答値を高速に算出するために Integral Image[10] を用いる．

縦横方向の Integral Image

距離画像 $I(x, y)$ に対して，式 (6) により計算を行うことで，Integral Image $II(x', y')$ を作成する．

$$II(u', v') = \sum_{u' \leq u,\ v' \leq v} I(u, v) \tag{6}$$

ここで，Integral Image $II(x', y')$ は，原点から注目点 (x', y') までの距離の総和となる．図 5 の D の領域から距離の総和を算出するためには，Integral Image$II(x', y')$ 内の 4 点 P1，P2，P3，P4 を用いて式 (7) より求めることができる．

$$\mathrm{D} = \mathrm{P4} + \mathrm{P1} - \mathrm{P2} - \mathrm{P3} \tag{7}$$

斜め方向の Integral Image

斜め方向の領域から距離の総和を求める場合，図 6 に示すように斜め方向の Integral Image を作成する必要がある．斜め方向の Integral Image は Lienhart ら [11] により提案されている．距離画像 $I(x, y)$ に対して，式 (8) により計算することで，Integral Image $II(x', y')$ を作成する．

$$II'(u', v') = \sum_{u' \leq u,\ u' \leq u - |v - v'|} I(u, v) \tag{8}$$

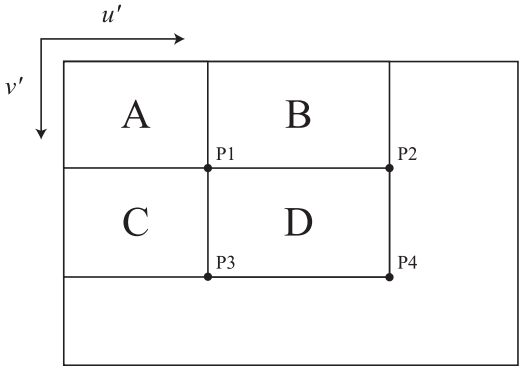


図 5　Integral Image

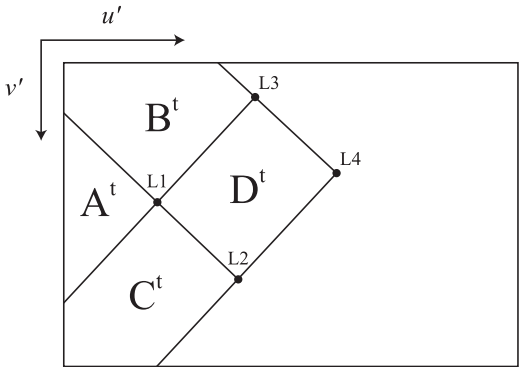


図 6　斜め方向の Integral Image

図 6 の D^{t} の領域から距離の総和を算出するためには，Integral Image $II'(x', y')$ 内の 4 点 L1，L2，L3，L4 を用いて式 (9) より求めることができる．

$$\mathrm{D}^{\mathrm{t}} = \mathrm{L4} + \mathrm{L1} - \mathrm{L2} - \mathrm{L3} \tag{9}$$

3.4　各方向のフィルタマップの統合

Haar-like フィルタリングにより得られた 4 方向のフィルタマップ $F(u, v, d)$ を 1 つに統合する．式 10 に示すように，4 方向のフィルタマップ $F(u, v, d)$ の座標 (u, v) における凸形状ラベルの少なくとも 1 つが 1 であるとき統合後のフィルタマップ $F_{all}(u, v)$ に 1 を与える．

$$F_{all}(u, v) = \delta(\sum_{d} F(u, v, d) > 0) \tag{10}$$

3.5　3 次元実空間における Mean-Shift による距離情報のクラスタリング

統合されたフィルタマップに基づき距離情報をクラスタリングすることで人検出を行う．提案手法では 3 次元の距離情報を Mean-Shift[12] によりクラスタリングする．3 次元 Mean-Shift クラスタリングは式 (11) により Mean-Shift Vector $m(\mathbf{x})$ を算出する．統合されたフィルタマップ $F_{all}(u, v)$ のラベルが 1 である 3 次元座標を $\mathbf{x}_i$ とし，移動している注目点の三次元座標を $\mathbf{x}$ とする．k はカーネル関数，h はバンド幅であり，本研究では $h = 0.15\mathrm{m}$ とする．

$$m(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i k\left(\left|\left|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right|\right|^2\right)}{\sum_{i=1}^{n} k\left(\left|\left|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right|\right|^2\right)} - \mathbf{x} \tag{11}$$

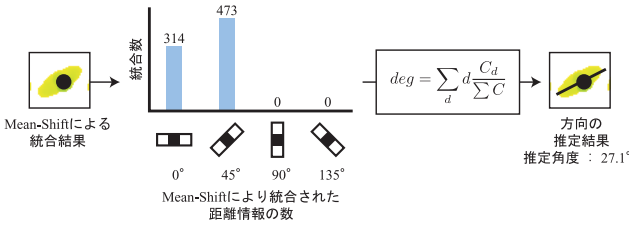


図 7　人の向き推定の流れ

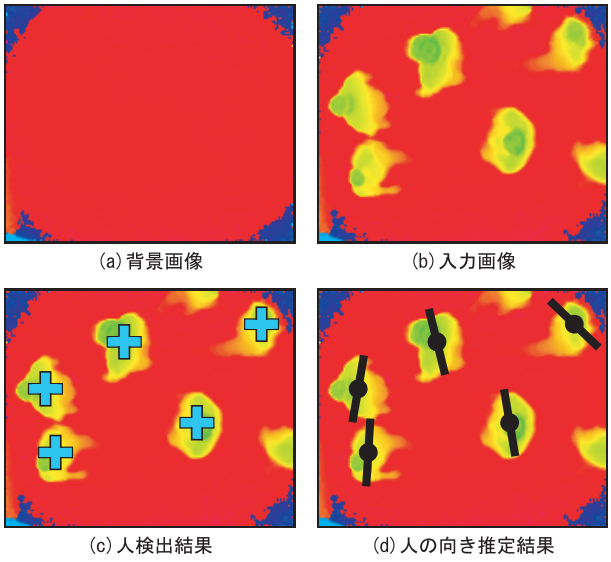


図 8　人の向き推定例

3.6　Harr-like フィルタの応答値による人の向き推定

Mean-Shift によりクラスタリングされた各方向の距離情報の数を利用し，人の向きの推定を行う．図 7 に示すように，Mean-Shift クラスタリングにより人と検出された点にクラスタリングされた距離情報の数を用いて，0 °を基準とした人の向きを算出する．人の向き deg は式 (12) より求めることができる．

$$deg = \sum_{d} d \frac{C_d}{\sum C} \tag{12}$$

ここで，$\{d \mid 0, 45, 90, 135\}$ であり，C_d は検出点に対しクラスタリングされた d 方向の距離情報の数を表す．

3.7　検出点と背景情報を用いた身長推定

提案手法では人検出結果と背景の距離情報を用いることで検出した人の身長を推定する．入力距離画像 I における人と検出された座標の距離を $I(u, v)$，背景距

表 1　データベース

	TOFカメラの高さ[m]	身長[cm]
シーケンス1	4.5	大人 ： 165 ～ 185
シーケンス2	3.2	大人 ： 165 ～ 175 子供 ： 100 ～ 120
シーケンス3 （階段）	3.5	大人 ： 165 ～ 175 子供 ： 100 ～ 120

(a) シーケンス1　(b) シーケンス2　(c) シーケンス3

図 9　各シーケンスの背景画像

離画像 B の同じ座標の距離を $B(u,v)$ とすると，検出された人の身長 h は式 (13) より求めることができる．

$$h = |I(u,v) - B(u,v)| \tag{13}$$

4　人検出実験による提案手法の評価

提案手法の有効性を評価するため，評価実験を行う．

4.1　データベース

本実験では TOF カメラを用いて人を上部から撮影したシーケンスを用いて行う．評価用のデータベースには異なる環境で撮影された 3 種類のデータベースを用いる．表 1 に 3 種類のデータベースの詳細を示す．シーケンス 1 は，大人を対象としたシーケンスとなっており，大勢の人を撮影した混雑しているシーンとなっている．シーケンス 2 は，大人と子供が同時に存在するシーンである．また背景には人以外の物体も存在する．シーケンス 3 は，階段において人が上り下りするシーンであり，階段の高さが変化することにより人の高さも変化する．

4.2　検出精度の比較実験

評価用データベースを用いて人検出実験を行い，検出精度の比較を行う．従来法として，Beleznai らの人検出手法 [8] に基づき，背景差分により得られた物体領域の距離情報を Mean-Shift クラスタリングで統合し，人を検出する手法を用いる．提案手法は背景差分を用いて背景領域を削除しているため，背景を人と誤って検出する誤検出は存在しない．そのため，DET カーブや ROC カーブによる評価ができない．そこで本実験では検出率により 2 つの手法の比較を行う．検出率は式 (14) より算出する．

$$\text{検出率 [\%]} = \frac{\text{検出数 [人]}}{\text{真値 [人]}} \tag{14}$$

表 2　シーケンス 1 の検出精度

	真値 [人]	検出数 [人]	検出率 [%]
Mean-Shift	477	428	89.7
Haar-like + Mean-Shift		471	98.7

表 3　シーケンス 2 の検出精度

	真値 [人]	検出数 [人]	検出率 [%]
Mean-Shift	291	278	95.5
Haar-like + Mean-Shift		286	98.3

表 4　シーケンス 3 の検出精度

	真値 [人]	検出数 [人]	検出率 [%]
Mean-Shift	283	249	88.0
Haar-like + Mean-Shift		271	95.8

表 5　処理時間 [ms]

	背景差分	Haar-like フィルタリング	Mean-Shiftによるクラスタリング	1フレームの処理時間
Mean-Shift	0.356	——	728.438	728.786
Haar-like + Mean-Shift		1.547	56.870	58.773

表 2，3，4 に 3 種類のシーケンスに対する検出実験結果を示す．実験結果より，提案手法である Haar-like + Mean-Shift による 3 種類のシーケンスに対する検出率の平均は，Mean-Shift のみの検出と比較し 6.6%検出率を向上させることができた．これは図 10 に示すように，人同士が近接した場合に Mean-Shift クラスタリングのみを用いて距離情報を統合すると，近接している人同士の距離情報を統合してしまうため，2 人の場合でも 1 人と検出し，未検出となる．一方，提案手法は Haar-like フィルタリングを用いることにより人の凸形状を検出することにより，人の中心部分の距離情報を用いて Mean-Shift クラスタリングを行うため，近接し合う人の距離情報を分離できるため未検出を抑制することができる．

4.3　速度比較

表 5 に 1 フレームあたりの検出処理時間を示す．処理時間の計測には 1 枚当たり平均 9 人のサンプル 500 枚を用いて，1 枚の平均の処理時間を算出した．表 5 より，人検出処理における大部分の計算時間は Mean-Shift クラスタリングであることがわかる．提案手法は，Haar-like フィルタリングにより Mean-Shift によりクラスタリングする距離情報を削減するため，Mean-Shift クラスタリングのみと比較し約 12 倍高速な人検出が可能である．

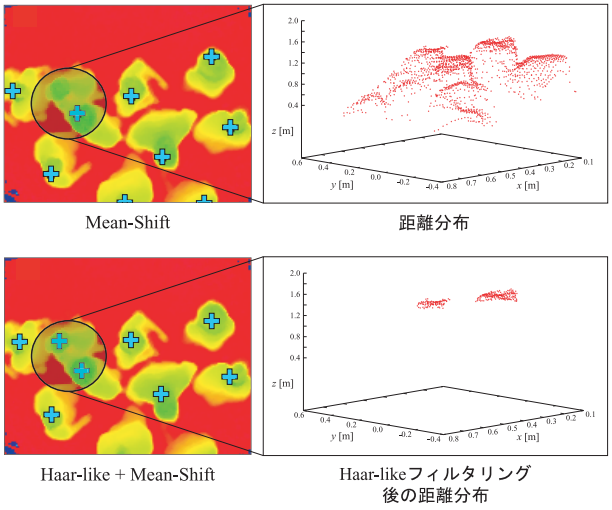


図 **10**　**Haar-like** フィルタリングによる誤統合の改善例

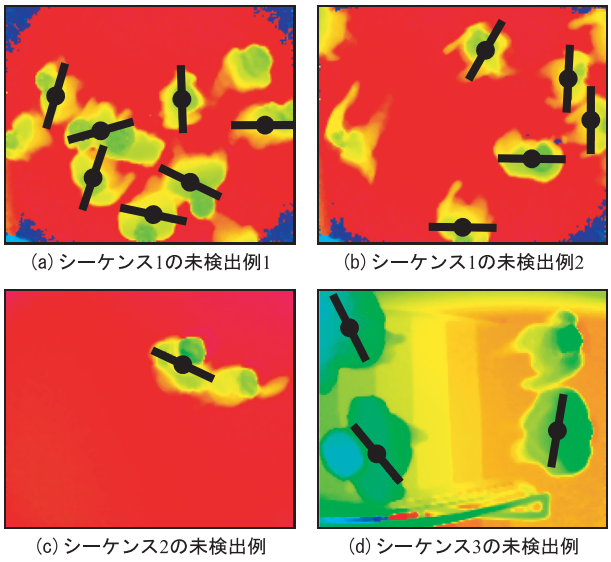


図 **14**　未検出例

4.4　考察

図 11，12，13 に提案手法による人検出例を示す．黒い点が人検出点であり，軸が人の向きを推定した結果である．

提案手法は，図 11(c) のような混雑したシーンにおいて高精度な人検出が可能であることがわかる．これは，Haar-like フィルタリングにより近接する人を分離した後に Mean-Shift クラスタリングで統合するためである．

図 12 のように子供と大人が同時に存在する場合にも Haar-like フィルタのサイズを変化させながら検出することで対応することができる．図 12(b) に示すように，提案手法は事前に背景差分を行うため，背景として人以外の物体が存在しても誤検出しないことがわかる．また，図 12(c) に示すように背景に存在しない台車を押している場合でも，提案手法は Haar-like フィルタリングにより人の凸形状を検出するため，誤検出することなく人のみを検出できている．

図 13 に示すように，階段等の高低差のあるシーンにおいても高精度な人検出が可能となる．これは提案手法が事前に背景差分を用いて物体領域のみを抽出するため，背景の高低差に影響されずに人検出が可能となるためである．

図 14 に提案手法による人検出の未検出例を示す．全ての未検出例に共通して画像の端になるにつれて未検出が増加することがわかる．人が TOF カメラの直下に存在する場合，人の凸形状がよく捉えられるため高精度に検出が可能となる．しかしながら，画像の端になるにつれて側面からの見えに変化することで凸形状を捉えることが困難となるため，未検出が増加する．また，(b)，(c) のように姿勢変化が大きい場合においても凸形状を捉えることができず，未検出の原因となる．

5　おわりに

本研究では，TOF カメラにより人を上部から撮影して得られる距離情報を用いて，人の形状を表現する Haar-like フィルタによる高精度な人検出法を提案した．背景差分により得られた距離情報を Mean-Shift クラスタリングにより統合する手法との比較実験より，提案手法は検出率を 6.6%向上した．このとき提案手法は約 17 fps でリアルタイムに人検出が可能であることを確認した．今後は，本手法の検出結果を用いた人のトラッキングについて検討する予定である．

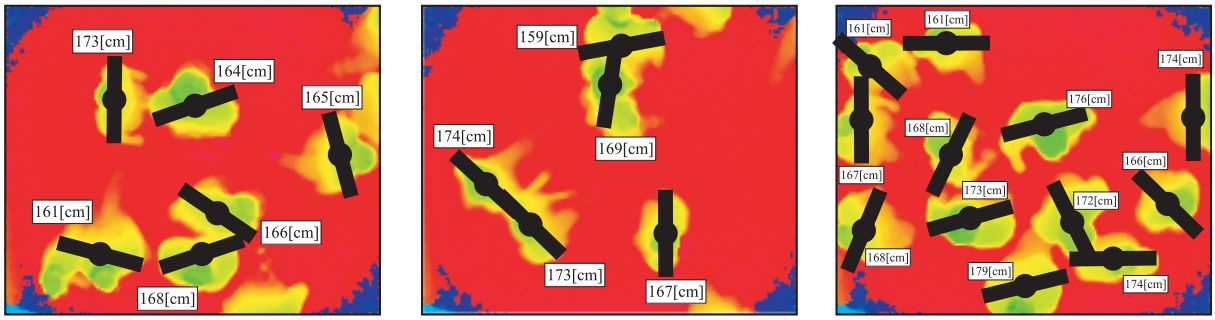
参考文献

[1] N. Dalal and B. Triggs,“ Histograms of Oriented Gradients for Human Detection ”, IEEE Computer Vision and Pattern Recognition, vol.1, pp. 886-893，2005.

[2] 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.

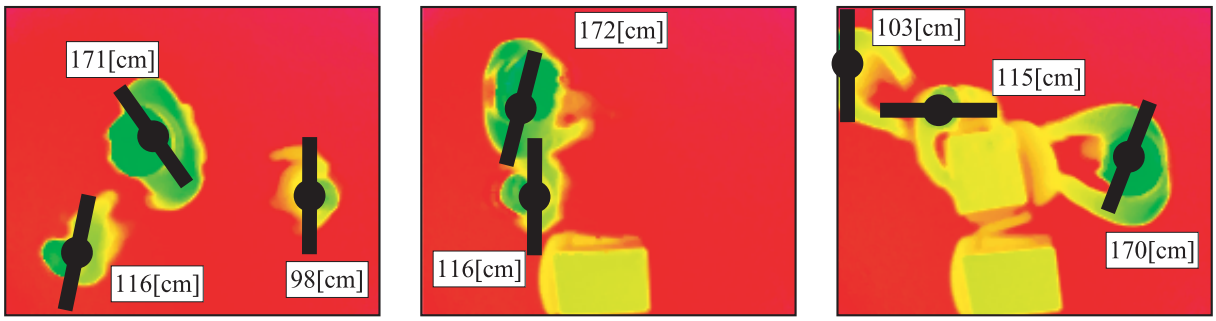
[3] B. Wu and R. Nevatia,“ Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors ”, IEEE International Conference on Computer Vision, vol. 1, pp. 90-97, 2005.

[4] C. Hou, H. Ai, and S. Lao,“ Multiview Pedestrian Detection Based on Vector Boosting ”, Asian Conference on Computer Vision, pp. 210-219, 2007.



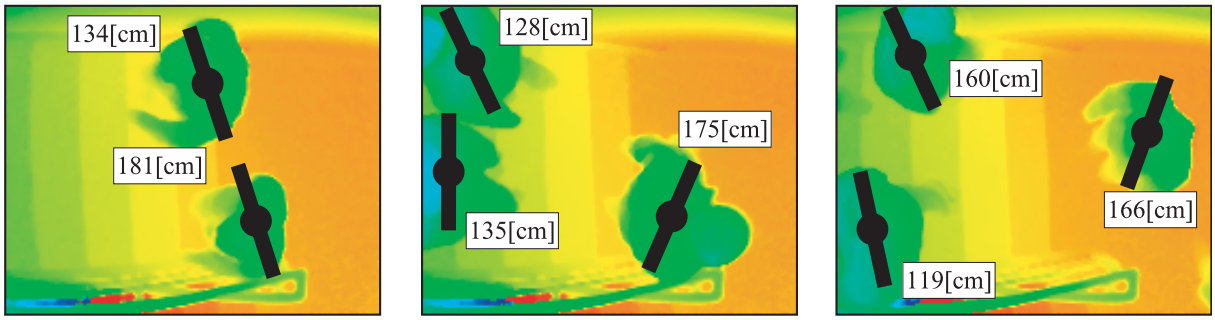
(a) シーン1　(b) シーン2　(c) シーン3

図 **11**　シーケンス **1** の人検出例



(a) シーン1　(b) シーン2　(c) シーン3

図 **12**　シーケンス **2** の人検出例



(a) シーン1　(b) シーン2　(c) シーン3

図 **13**　シーケンス **3** の人検出例

[5] 池村翔 , 藤吉弘亘, “ 距離情報に基づく局所特量によるリアルタイム人検出 ”, 画像センシングシンポジウム (SSII09), IS4-05, 2009.

[6] 技研トラステム , パロッシー , http://www.trastem.co.jp/, 1966.

[7] Point Grey Research Inc, People Tracking System, http://www.ptgrey.com/, 1997.

[8] C. Beleznai, B. Fruhstuck and H. Bischof,“ Human Detection in Groups using a Fast Mean Shift Procedure ”, IEEE International Conference on Image Processing, Vol. 1, pp. 349-352, 2004.

[9] C. P. Papageorgiou, M. Oren and T. Poggio, “ A general framework for object detection ”, IEEE International Conference on Computer Vision, pp. 555-562, 1998.

[10] P. Viola, and M. Jones,“ Robust real-time face detection ”, International Journal of Computer Vision, Vol. 57, no. 2, pp. 137-154, 2002.

[11] R. Lienhart and J. Maydt,“ An Extended Set of Haar-like Features forRapid Object Detection ”, IEEE International Conference on Image Processing, Vol. 1, pp. 900-903, 2002.

[12] D. Comaniciu and P. Meer,“ Mean Shift Analysis and Applications ”, IEEE International Conference on Computer Vision, pp. 1197-1203, 1999.