

1. はじめに

近年、TOF カメラや Kinect といったリアルタイムに距離を計測できる距離画像センサが注目されている。これに伴い、距離センサを利用した人センシング技術への応用が期待されている。従来の距離情報に基づく局所特微量と統計的学習による人検出手法 [1] では、人同士の重なりに対応した人検出を実現しているが、事前に大量の学習データが必要であるという問題がある。また、作成した学習データの撮影環境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、本研究では TOF カメラにより人を上部から撮影した距離画像を用いて、人の肩、頭、肩の凸形状を抽出する Haar-like フィルタリングによる、統計的学習を用いない人検出手法を提案する。

2. 距離情報を用いた Haar-like フィルタリングによる人検出

提案手法では、まず距離画像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が人かどうかを判断するために、人の肩、頭、肩の凸形状を抽出する Haar-like によるフィルタリング処理を行う。Haar-like フィルタリングにより、凸形状と抽出された点を Mean-Shift クラスタリングにより統合することで人検出を行う。図 1 に人モデルに基づく Haar-like フィルタリングによる人検出の流れを示す。

2.1 距離情報を用いた背景差分による物体領域の抽出

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、距離情報を用いることで高精度な背景差分が可能となる。提案手法では、式 (1) のように入力距離画像 I と背景距離画像 B の差分を求めることにより、物体領域抽出画像 M を作成する。

$$M(u, v) = \begin{cases} I(u, v) & \text{if } |I(u, v) - B(u, v)| > th \\ \text{Background} & \text{otherwise} \end{cases} \quad (1)$$

ここで、 th は入力距離画像と背景距離画像の差に対する閾値であり、本研究では 0.3m としている。図 2 に背景差分による物体領域の抽出例を示す。

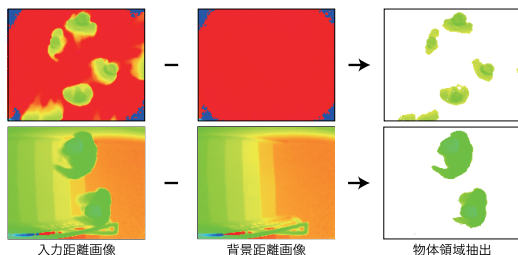


図 2：背景差分による物体領域の抽出例

2.2 Haar-like フィルタリングによる凸形状の抽出

提案手法では、抽出された物体領域が人かどうかを判断するために、人の肩、頭、肩の凸形状を抽出する Haar-like [2] によるフィルタリング処理を行う。Haar-like フィルタとは、黒と白の領域の平均距離の差を応答値とするフィルタであり、黒の領域が高い場合には正の応答値を出力し、白の領域が高い場合には負の応答値を出力する。提案手法では、人の肩、頭、肩の高さの違いを捉えるために白、黒、白と領域を配置した Haar-like フィルタを用いる (図 3)。この Haar-like フィルタの応答値 $H(r1, r2)$ は、黒い領域 $r1$ の平均距離 $S(r1)$ と白い領域 $r2$ の平均距離 $S(r2)$ を用いて次式により算出される。

$$H(r1, r2) = S(r1) - S(r2) \quad (2)$$

算出された応答値が正の値の場合、黒い領域が白い領域に比べ高いことがわかるため、凸形状を抽出するフィルタとなる。提案手法では、人の向きの違いに対応するために、

$0^\circ, 45^\circ, 90^\circ, 135^\circ$ の 4 方向のフィルタリング処理を行うことで凸形状の判別を行う。フィルタリング処理では、算出された Haar-like フィルタの応答値 $H(r1, r2)$ を用いて式 (3) の閾値処理により、凸形状のフィルタマップ $F(u, v, d)$ を作成する。ここで、 d はフィルタの方向を表す。

$$F(u, v, d) = \begin{cases} 1 & \text{if } H(r1, r2, d) > th \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

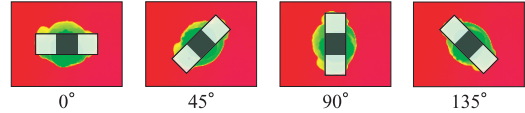


図 3：提案手法で用いる Haar-like フィルタ

2.3 人モデルに基づく Haar-like フィルタの設計

2.2 の Haar-like フィルタでは、全ての画像座標に対して一定のフィルタを用いてフィルタリングを行う。その際に、人が画像のどの位置に立つかにより肩と頭の見え方が変化するため、一定のフィルタリングでは対応することが困難な場合がある。そこで、人の立つ位置や向き、身長の違いに対応するために、事前に人モデルを用いて Haar-like フィルタの黒と白の領域の配置を決定しておく。図 4 に示す人モデルを用いて、画像上の各画素を中心として 0° から 360° まで 45° 毎に向きを変化させた 8 方向の人モデルを配置する。配置した各人モデルの肩と頭の位置を抽出し、その情報を LUT として記録する。LUT には、人モデルの中心座標と人モデルの角度、中心座標と角度に対応した左右の肩と頭の座標が記録されている。また、人モデルの高さを 0.5m から 2.0m まで、0.1m ずつ変化させて同様の LUT を作成することで、人の身長の違いに対応する。

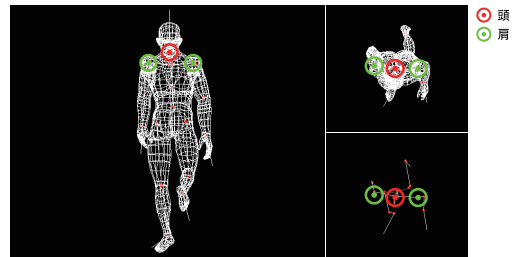


図 4：人モデルからの肩と頭の位置抽出

事前に作成した LUT に基づき、人の位置や向き身長の高さを考慮した Haar-like によるフィルタリング処理を行う。図 5 に示すように、抽出された物体領域の各座標を注目座標とした場合、その注目座標を LUT と照合することで肩と頭の位置を決定する。決定した肩と頭の位置に対して、Haar-like フィルタの白と黒の領域を配置することで、人の見えの違いに対応したフィルタリングが可能となる。

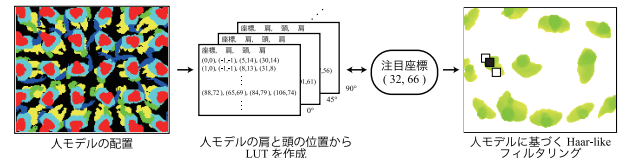


図 5：LUT に基づく Haar-like フィルタリング

2.4 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、繰り返し矩形領域内の総和演算を行うため、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 を用いる。距離画像 $I(x, y)$ に対して、式 (4) により計算を行うことで、Integral Image $II(x', y')$ を作成する。

$$II(u', v') = \sum_{u' \leq u, v' \leq v} I(u, v) \quad (4)$$

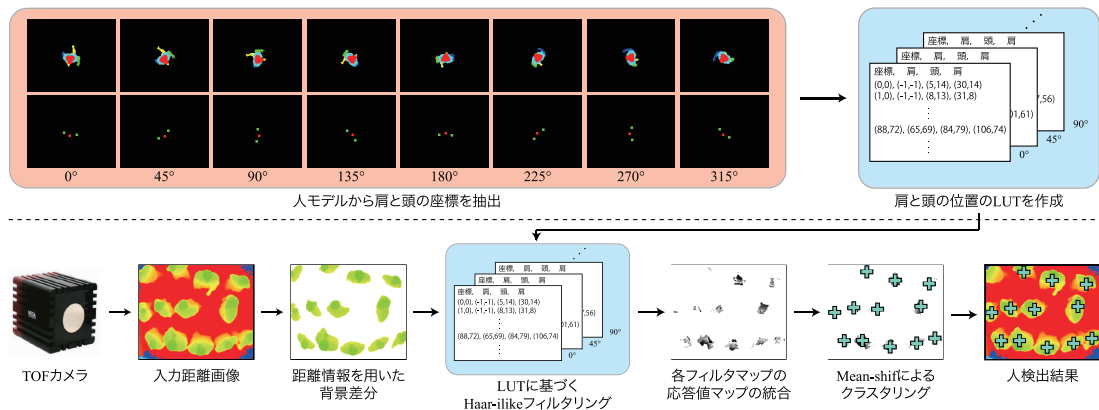


図 1: 人モデルに基づく Haar-like フィルタリングによる人検出の流れ

ここで, Integral Image $I(x', y')$ は, 原点から注目点 (x', y') までの距離の総和となる. ある矩形領域から距離の総和を算出するためには, 矩形領域の左上の角を P1, 右上を P2, 左下を P3, 右下を P4 とすると, 式 (5) より求めることができる.

$$D = P4 + P1 - P2 - P3 \quad (5)$$

2.5 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 d 方向のフィルタマップ $F(u, v, d)$ を 1 つに統合する. 式 (6) に示すように, d 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に少なくとも 1 が 1 つ以上ある場合に統合後のフィルタマップ $F_{all}(u, v)$ に 1 を与える.

$$F_{all}(u, v) = \delta\left(\sum_d F(u, v, d) > 0\right) \quad (6)$$

2.6 Mean-Shift クラスタリングによる抽出点の統合

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 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 (7) 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 3 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 $h = 0.15\text{m}$ とする.

$$m(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{x}_i k\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^n k\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x} \quad (7)$$

3. 評価実験

提案手法の有効性を示すために評価実験を行う.

3.1 実験概要

比較手法には, 従来法として, 背景差分により抽出された物体領域の距離情報を Mean-Shift クラスタリングにより統合する手法 [3], 提案手法 1 として, Haar-like フィルタリングによる人検出手法, 提案手法 2 として, 人モデルに基づく Haar-like フィルタリングによる人検出法を比較する. データベースには, 異なるシーンで撮影した 3 種類のシーケンスを用いる. シーケンス 1 は, 大人を対象とした混雑な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り, 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となっている.

3.2 実験結果

表 1 に各シーケンスに対する従来法, 提案手法 1, 提案手法 2 の人検出精度を示す. 評価実験の結果, 提案手法 1 は従来法と比較して, 検出精度が 3 シーケンスの平均で 2.5% 低下した. これは, 直線的に白と黒を配置した Haar-like フィルタを用いているため, 人の位置の違いによる見えの変化に対応できなかったためである. これに対し, 提案手法 2

は従来法と比較して平均 5.7% 検出精度を向上することができた. これは, 人モデルを用いて, 人の位置に対する見えの変化を考慮した Haar-like フィルタリングを行っているためである.

表 1: 人検出精度 [%]

	シーケンス1	シーケンス2	シーケンス3	平均
従来法	89.7	92.5	88.0	90.1
提案手法1	87.0	90.2	85.6	87.6
提案手法2	98.6	91.3	97.4	95.8

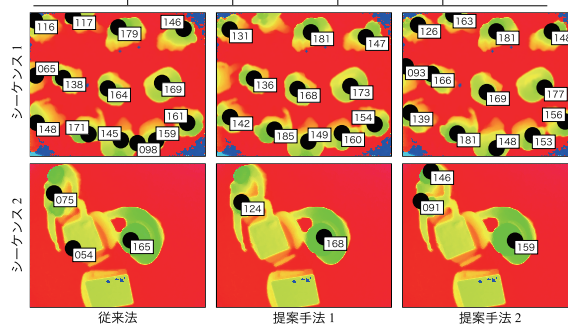


図 6: 各手法による人検出例

表 2 に, 各手法の 1 フレームあたりの処理時間を示す. 提案手法 1, 2 は, Haar-like フィルタリングによりクラスタリングする距離情報を削減するため, 従来法と比べ約 13 倍高速な人検出が可能である.

表 2: 処理時間の比較 [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
従来法	—	—	728.4	728.8
提案手法1	0.4	1.5	56.9	58.8
提案手法2	—	5.5	45.3	51.2

4. おわりに

TOF カメラにより人を上部から撮影することで得られる距離画像を用いた, Haar-like フィルタリングによる人検出手法を提案した. 評価実験の結果, 提案手法は, 従来法と比較し高速, 高精度な人検出が可能であることを確認した.

参考文献

- [1] 池村翔, 藤吉弘巨, “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”, 信学論 D, Vol. J93-D, No. 3, pp.355-364, 2010.
- [2] C. P. Papageorgiou *et al.*, “A general framework for object detection”, IEEE International Conference on Computer Vision, pp. 555-562, 1998.
- [3] C. Belezni *et al.*, “Human Detection in Groups using a Fast Mean Shift Procedure”, IEEE International Conference on Image Processing, Vol. 1, pp. 349-352, 2004.

研究業績

- [1] 池村翔, 藤吉弘巨, “時空間情報と距離情報を用いた Joint Boosting による動作識別”, 電学論 C, Vol.130, No.9, pp.1554-1560, 2010. (他 学術論文 1 件, 国際会議発表 1 件, 国内会議発表 5 件)

受賞

- [1] SSII2009 オーディエンス賞
- [2] 平成 21 年度 電気関係学会東海支部連合大会 IEEE 学生奨励賞
- [3] 平成 22 年度 電子情報通信学会東海支部 学生研究奨励賞