

1. はじめに

従来の人検出法の多くは、可視光カメラにより人を正面から撮影し、形状を表現することにより検出する手法が多く提案されている [1]。しかし、このような手法は人の重なりが存在する場合に検出が困難となる問題がある。そこで、本研究ではリアルタイムで距離情報を取得できる TOF(Time of Flight) カメラを用いて人を上部から撮影することにより、人の重なりに影響されない高速な人検出法を提案する。

2. 距離情報を用いた人検出

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人領域の抽出を行う。抽出された人領域に対し人の肩、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。検出された距離情報を Mean-Shift クラスタリングにより統合することにより人検出を行う。図 1 に提案手法の流れを示す。

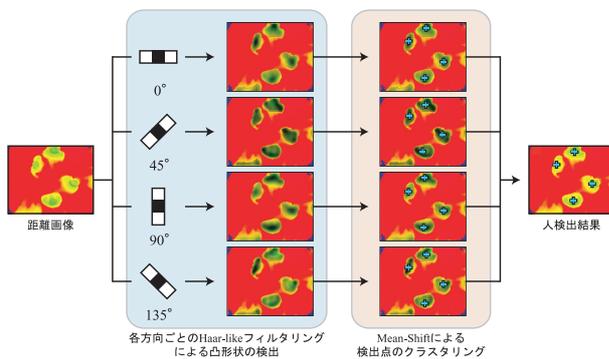


図 1：提案手法による人検出の流れ

2.1 Haar-like フィルタリングによる凸形状検出

提案手法では、人の肩、頭、肩の凸形状を捉えるために Haar-like フィルタリングを用いる。Haar-like フィルタリングは、図 2 に示すように、頭の黒い領域 r_1 と肩の白い領域 r_2 の 2 つの領域の距離差 $H(r_1, r_2)$ を式 (1) より算出し、その応答値を式 (2) により閾値処理することで各画素 (u, v) における人の凸形状ラベル $F(u, v)$ を算出する。このとき、 $S(r)$ は領域 r の輝度を算出する関数であり、 th は凸形状を判定するための閾値である。また提案手法では、人の向きを考慮するために 4 方向の Haar-like フィルタを用いる。

$$H(r_1, r_2) = S(r_1) - S(r_2)/2 \quad (1)$$

$$F(u, v) = \begin{cases} 1 & \text{if } H(r_1, r_2) > th \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Haar-like フィルタリングを行う際、背景差分により得られた人領域の各画素に対しフィルタリングを行う。そのため、非常に多くの Haar-like フィルタの応答値を算出することから、計算コストが増大する問題がある。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。Integral Image は高速に矩形領域内の総和を求める手法であり、積分画像における矩形領域から得られる 4 点を用いた計算により総和を算出する。

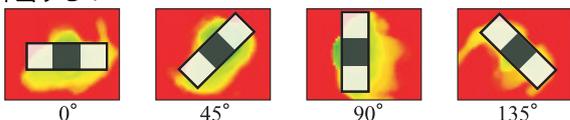


図 2：提案手法で用いる Haar-like フィルタ

2.2 3次元実空間における Mean-Shift クラスタリングによる距離情報の統合

Haar-like フィルタリングにより凸形状と検出された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 (3) に

より Mean-Shift Vector $m(x)$ を算出する。ここで、 x は移動している注目点の 3次元座標、 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 $h = 0.15m$ とする。

$$m(x) = \frac{\sum_{i=1}^n x_i k\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n k\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad (3)$$

3. 評価実験

提案手法の評価実験を行う。比較には、提案手法と人領域の距離情報に対し Mean-Shift クラスタリングを行うことで検出を行う手法の比較を行う。本実験では TOF カメラを 4.5m の高さを設置し、人を上部から撮影したシーケンスを用いる。評価は 130 枚のシーケンスを用いて行う。

表 1：検出精度比較

	真値 [人]	検出数 [人]	検出率 [%]	1フレームの処理時間 [s]
Mean-Shift	477	428	89.7	1.31
Haar-like + Mean-Shift		471	98.7	0.17

表 1 に検出率を示す。提案手法である Haar-like + Mean-Shift による検出は、Mean-Shift のみの検出と比較し 9.0% 検出率を向上させることができた。これは図 3 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、近接した人の距離情報が統合されてしまうため誤検出となる。それに対し提案手法は、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し、人の中心部分の距離情報のみを用いて Mean-Shift クラスタリングを行うことが可能となるため誤検出を抑制できる。

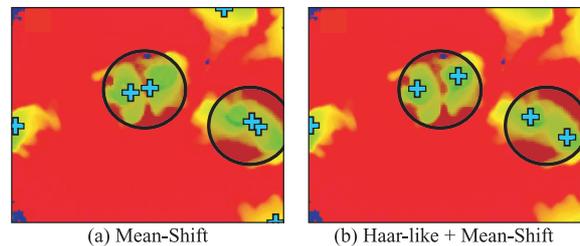


図 3：Haar-like フィルタリングによる誤統合の改善例

図 4 に提案手法による人検出例を示す。人が複数人存在する場合でも高精度に検出が可能であることがわかる。また (b) のように人同士が近接している場合にも誤検出せずに検出できる。このとき提案手法では約 8fps で人検出が可能であることを確認した。

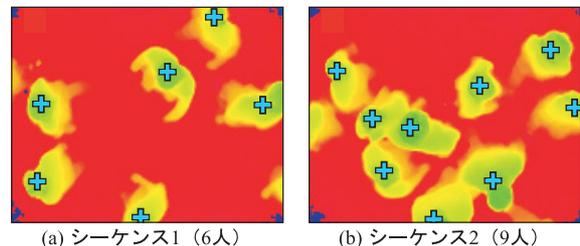


図 4：人検出例

4. おわりに

本研究では、TOF カメラを用いて人を上部から撮影することにより得られる距離情報を用いた人検出法を提案し、高精度かつ高速な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。

参考文献

[1] N. Dalal and B. Triggs: "Histograms of Oriented Gradients for Human Detection", CVPR, pp. 886-893 (2005).