

# 統計的形状モデリングによる 頭上から撮影された距離画像中の歩行者検出

Pedestrian Detection in Top-View Depth Images by Using Statistical Shape Modelings

竹村和将<sup>†</sup>

Kazumasa Takemura

宮城茂幸<sup>†</sup>

Shigeyuki Miyagi

## 1 はじめに

近年、セキュリティやITS (Intelligent Transport Systems, 高度道路交通システム) において、物体の検出や追跡などの動画処理技術が期待されている。特に、映像からの人数計測や人流測定には人検出が不可欠な技術となる。

従来における通常のカメラ画像(濃淡画像)を用いた人検出手法では、照明変動による影響を受けるため、撮影された場所が同じであっても特徴量に変動してしまう。例えば、肌色領域を検出する場合でも、照明変動により色相や明度が増減するため、誤検出や検出漏れの原因となり、正確に肌色を検出するのが困難となる。

このような照明変動に対し、TOF(Time of Flight)カメラから得られる距離画像を用いた手法では、輝度値を用いないため、照明変動による影響を受けず、高精度な人検出を実現できることが報告されている[1],[2]。しかし、濃淡画像、距離画像にかかわらず、正面からの顔や全体像を検出対象とした場合、画像上での重なりによる隠れ(オクルージョン)が生じ、正確に対象物全体の特徴を捉えられないことがある。

そこで、本稿ではTOFカメラから得られる距離画像に基づいた頭上方向からの人検出手法を提案する。頭上方向から撮影した画像を用いることで、人の重なりを減らし、オクルージョンの発生を抑える。

著者らは、HOG(Histograms of Oriented Gradients)特徴量[3]やスパース特徴量[4]のような局所特徴量を用いた検出手法を提案した[5]。HOG特徴量は、勾配情報を利用することで照明変動に頑健となり、物体の大きな形状を得ることができる。しかし、分割した領域ごとに勾配強度、勾配方向を算出し、勾配方向ヒストグラムを作成するため、計算コストがかかるという問題点がある。これに対して、スパース特徴量では、複数の局所領域の画素値の加減算を用いるため、演算回数が少なく、形状変化の少ない物体に対して高速かつ高精度な検出を実現した。

HOG特徴量もスパース特徴量も濃淡画像の特徴を捉えるために開発された特徴量である。距離画像では絶対位置が計測できるため、画像中の物体形状に関する情報が直接含まれている。したがって、汎用的なHOG特徴量やスパース特徴量に頼ることなく、形状を直接表現できる特徴量があるのではないかと考えた。

また、HOG特徴量やスパース特徴量から判別器を構

成するために機械学習がよく用いられる。このとき大量の教師データが必要となる。しかし形状を活かした特徴量を用いることにより、比較的少数の教師データから、短時間にモデル作成ができる手法を考えた。

そこで、本稿では取得した距離画像に対して、頭上方向から見たときの歩行者の立体的形状を活かした検出手法を提案する。頭頂部から肩位置にかけての点群から3軸成分それぞれにおける頭頂部との距離差を取得し、得られた形状データに主成分分析を適用し次元圧縮を行うことにより、特徴空間の部分空間としてモデル化を行う。

最終的には任意の方向へ移動する歩行者の検出を目指しているが、本稿では、最初段階として画像内を上下方向に歩いている歩行者のみを対象とし、提案手法の有効性を検証する。

## 2 歩行者検出の流れ

提案手法における歩行者検出の流れを図1に示す。学習では、あらかじめ正解のわかっている学習用画像を入力したときに、識別能力の高い特徴量を使った識別器を選択することで、検出に用いる強識別器を構築する。検出では、まず入力画像に対して距離値を用いた閾値処理により検出候補領域を絞り込む。次に、これらの候補領域に対して検出窓をラスタスキャンし、各検出窓から特徴量を算出する。算出された特徴量に基づいて識別器により判別することにより人かどうか判定を行う。

距離画像は三次元距離測定カメラであるTOF(Time of Flight)カメラから取得したものをを用いる。本研究ではMESA社のSR-4000[6]を使用した。図2にTOFカメラから取得した距離画像例を示す。

## 3 HOG(Histograms of Oriented Gradients)特徴量

HOG(Histograms of Oriented Gradients)特徴量は、画像の局所領域から輝度勾配強度・輝度勾配方向を取り出す特徴量である。本稿では距離画像から特徴量を算出するため、得られる数値は距離の勾配情報となる。距離の勾配強度は、人の胴体と床のようにカメラからの距離が離れた位置にあるほど大きな値となる。したがって、距離勾配を算出することで物体の輪郭がわかるため、物体の形状を特徴量として得ることができる。

<sup>†</sup>滋賀県立大学大学院工学研究科

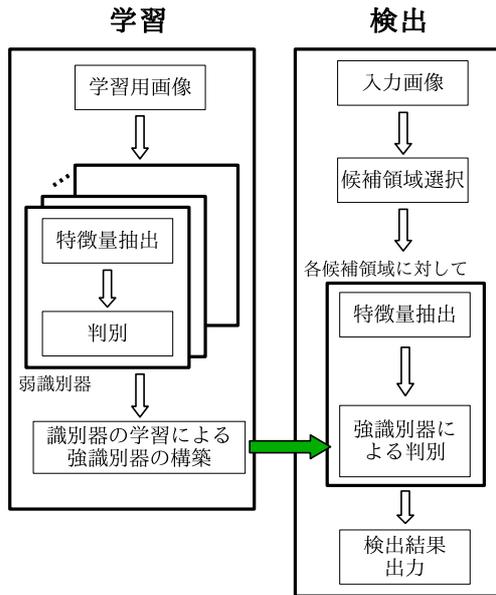


図1 歩行者検出の流れ

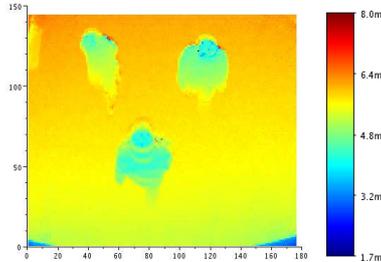


図2 TOFカメラから取得した距離画像例

### 3.1 距離勾配の強度・方向の算出

画像中の位置  $(u, v)$  における距離値を  $L(u, v)$  とすると,  $u, v$  方向の距離勾配  $f_u, f_v$  は次式により算出される.

$$\begin{cases} f_u(u, v) = L(u+1, v) - L(u-1, v) \\ f_v(u, v) = L(u, v+1) - L(u, v-1) \end{cases} \quad (1)$$

このとき, 位置  $(u, v)$  の勾配強度  $m$  と勾配方向  $\theta$  は次式より算出される.

$$m(u, v) = \sqrt{f_u(u, v)^2 + f_v(u, v)^2} \quad (2)$$

$$\theta(u, v) = \tan^{-1} \frac{f_v(u, v)}{f_u(u, v)} \quad (3)$$

ただし, ここで算出される勾配方向は  $0^\circ \sim 360^\circ$  となるが, 勾配方向の向きを考慮する必要が無いいため,  $0^\circ \sim 180^\circ$  に変換して用いる.

### 3.2 セルによるヒストグラム化

算出された勾配方向  $\theta$  を方向ごとに分割し, 各方向に勾配強度  $m$  を加算していく. そして, 近傍の領域 (セル,  $H_{\text{cell}} \times W_{\text{cell}}$ ) ごとにおいて, ヒストグラム化することにより, 距離勾配方向ヒストグラムを作成する. 本稿では,  $8 \times 8$  画素を1セルとし,  $0^\circ \sim 180^\circ$  の勾配方向を

$20^\circ$  ずつに分割するため, 9方向の勾配方向ヒストグラムが得られる.

### 3.3 ブロックによる正規化

各セルで作成した距離の勾配方向ヒストグラムを近傍のセル領域 (ブロック,  $H_{\text{block}} \times W_{\text{block}}$ ) ごとに正規化する. 本稿では, 画像内の物体と床との境界をより捉えやすくするため,  $4 \times 3$  セル,  $3 \times 2$  セル,  $2 \times 1$  セルのように, 水平方向に長い3種類のセル領域を1ブロックとして用いる. ブロック内の  $i$  行  $j$  列に位置するセルの勾配方向ヒストグラムを  $h_{ij}$  とすると,  $h_{ij}$  はブロック内で次式により正規化される.

$$h'_{ij} = \frac{h_{ij}}{\sqrt{\sum_{i=1}^{H_{\text{block}}} \sum_{j=1}^{W_{\text{block}}} h_{ij}^2 + \epsilon}} \quad (4)$$

ここで,  $\epsilon$  は分母が0の場合に計算不能になることを防ぐ定数であり,  $\epsilon = 1$  とする. このように, ブロック単位で正規化した  $h'_{ij}$  を特徴量として抽出する. そのため, 最終的に得られる HOG 特徴量の配列は, ヒストグラムのビン数を  $N_{\text{bin}}$  とすると,  $N_{\text{bin}} \times W_{\text{block}} \times H_{\text{block}}$  の次元を有することになる. ヒストグラムのビン数を9つ,  $8 \times 8$  ピクセルを1セル,  $3 \times 2$  セルを1ブロックと定義した場合を例に考える. この場合, 1セルあたり9次元であり, 水平方向に3セル, 垂直方向に2セル, すなわち, 合計6セルに対して正規化を行うため,  $9 \text{次元} \times 6 \text{セル} = 54 \text{次元}$  を有する.

### 4 スパース特徴量

スパース特徴量は Haar-like 特徴量の一般化であると考えられ, 次式より算出される.

$$F(\boldsymbol{\pi}) = \sum_i \alpha_i p_i(\boldsymbol{\pi}; x, y, s), \quad \alpha_i \in \{-1, +1\} \quad (5)$$

ここで  $p(\boldsymbol{\pi}; x, y, s)$  は, 入力画像  $\boldsymbol{\pi}$  に対して,  $(x, y)$  を左上端の座標とした  $2^s \times 2^s$  画素の局所領域における画素値の総和を意味する. また,  $\alpha$  は加減算の符号を表す結合係数である. したがって, スパース特徴量  $F(\boldsymbol{\pi})$  は,  $i$  個の局所領域における画素値の加減算により算出される. Haar-like 特徴量は図3のような組み合わせを用いているのに対し, スパース特徴量は局所領域の組み合わせに制約がない. したがって, Haar-like 特徴量に比べ, 局所領域の組み合わせの自由度は高い反面, 組み合わせ数が膨大となる.



図3 Haar-like 特徴量の例

### 5 学習手法

HOG 特徴量, スパース特徴量を用いた検出手法では, 統計的学習手法である Real AdaBoost[7] により識別器

を学習する。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関数から分離度を求め、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となる。学習により選択された弱識別器の出力を  $h_t(x)$  とすると、 $T$  個の弱識別器により構築される強識別器の出力  $H(x)$  は次式により求まる。

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

HOG 特徴量を用いた検出手法では、得られた 50 次元のベクトル特徴量に対し、直接 Real AdaBoost を適用した。

一方、スパース特徴量では、局所領域の組み合わせが膨大であるため、直接 Real AdaBoost を用いることは計算量の増大を招く。Huang らは適切な局所領域の組み合わせを探索するアルゴリズムにおいて、初期の局所領域の組み合わせ候補として図 3 に示す Haar-like 特徴量を用いた。それら各候補に対し、add, remove, refine の 3 種類の操作からなる expand と呼ばれる処理を行い、ヒューリスティック探索により、局所領域の組み合わせを決定するという方法を使用している。

本稿では、特徴量選択を効率よく行うために Real AdaBoost を用いるとともに以下の制限を設けた。

1. 初期の局所領域の組み合わせ候補を追加し、図 4 の通りとした。図 4 において E~H が本研究で追加した候補である。
2. 式 (5) により弱識別器を構成する際に、一旦  $s$  を固定し、同一サイズの局所領域を組み合わせた。ただし、Real AdaBoost により弱識別器を選択する際には、複数の  $s$  に対応する弱識別器を準備した。 $s$  の値を 0-3 とした。
3. Real AdaBoost により選択された弱識別器から強識別器を構成する際の弱識別器数に上限を設けた。その選択された弱識別器に対応するスパース特徴量に対し expand 処理を行った。その後、それらの特徴量に対し再度 Real AdaBoost を適用し、弱識別器の選択を行うという処理を繰り返すことにより、識別効果の高い特徴量を絞り込むことにした。本稿では弱識別器の上限を 75 個とした。

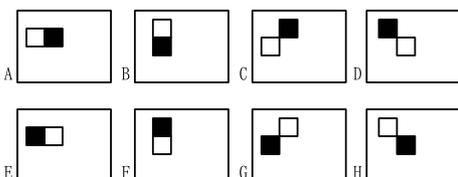


図 4 初期の局所領域の組み合わせ (A,B: Huang らの提案手法, C~H: 本研究で追加した組み合わせ)

## 6 統計的形状モデリング

TOF カメラは、カメラを原点とした 3 軸成分の距離値を取得することができる。そこで、頭頂部を基準とし

たときの頭頂部から肩位置にかけての立体形状に着目し、この形状を直接モデル化することで、見え方変化に頑強な検出手法を提案する。

概略は次の通りである。

まず形状データを取得するために必要な学習データを作成する。学習データより頭頂部から肩にかけて複数の位置での座標を取得し、それらをベクトル特徴量とする。ベクトル特徴量の次元を圧縮するために PCA を適用し、累積寄与率 80% となる固有ベクトルにより張られる部分空間を形状モデルとする。

### 6.1 学習データの作成

まず、撮影した画像から頭頂部があらかじめ設定した検出窓サイズの中央に位置するように歩行者部分を切り取る。

次に、異常値の影響を低減するために  $5 \times 5$  のメディアンフィルタを適用する。頭部の大きさは極端な個人差はないと考え、頭頂部深度方向に一定の距離を閾値とし胴体と頭部のみ含む距離画像を抽出する。図 5(a) に対し、上述の処理を行った結果を図 5(b) に示す。

複数の学習用画像から上述の方法によりあらかじめ切り出しを行い、それらを用いて肩の平均位置を求める。例えば図 5(b) では赤い点で示される位置が肩位置であり、この位置の平均を求めることに相当する。

### 6.2 形状データの取得

節 6.1 で求めた肩の位置と窓の中央、すなわち頭頂部を結ぶ直線上にある点群の絶対座標値を取得する。例えば、図 6 の画像の場合、赤い線で示される部分の座標値を用いる。

このとき、頭頂部の座標を基準とし、各点の 3 次元座標は頭頂部の座標からの相対座標値に換算した。また、点群数は 30 点とした。これらの座標値を直接ベクトル特徴量としたので、その次元は 90 次元とである。

### 6.3 形状データのモデル化

得られた 90 次元の形状特徴ベクトルに主成分分析を適用し次元圧縮を行う。これにより、形状モデルを低次元で表現することができる。本稿では、558 枚の学習用の距離画像から得た形状データに対して主成分分析を行い、累積寄与率が 80% となる固有ベクトルを 8 個選択した。

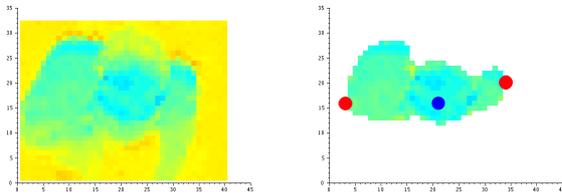
この固有ベクトルと学習データの内積を計算することにより、学習データを 8 次元の部分空間の領域としてモデル化したことになる。

### 6.4 形状モデルによる検出

テストデータが入力されたときの判定方法は次の通りである。

まず、検出窓により切り取られた画像に対し、 $5 \times 5$  のメディアンフィルタを適用する。次に節 6.1 で求めた

肩の位置と画像中心を結ぶ直線上から，節 6.2 と同じ手順により形状特徴ベクトルを抽出する．節 6.3 で求めた固有ベクトルとの内積を計算し，節 6.3 で求めた領域内に存在するかしらないかにより，人か人でないかの判定を行う．



(a) 入力データ (b) 推定結果

図 5 頭頂部および肩位置の推定結果

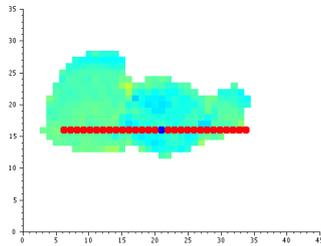


図 6 形状データの取得

## 7 評価実験

従来研究で用いた HOG 特徴量，Haar-like 特徴量，およびスパース特徴量と比較し，提案手法の有効性を評価する．学習では，撮影画像から切り取った  $40 \times 32$  画素のポジティブデータ 558 枚，ネガティブデータ 1652 枚を用い，評価実験には同様にして用意したポジティブデータ 718 枚，ネガティブデータ 1788 枚を用いた．評価実験の結果を表 1 に示す．ここで，処理時間とは Intel Core i5 CPU 3.20GHz を用いた際の検出窓 1 枚に対する処理時間を示す．従来手法の HOG 特徴量に比べ提案手法では検出精度が少し低下しているものの，スパース特徴量とほぼ同等の検出性能が得られた．図 7 に人検出を行った結果を示す．

表 1 評価実験結果

特徴量	正検出率 [%] (True Positive)	誤検出率 [%] (False Positive)	処理時間 [ms/window]
形状モデル	97.5	17.5	1.220
スパース	99.6	19.2	1.163
Haar-like	97.6	28.1	0.460
HOG	99.0	8.7	8.696

## 8 おわりに

本稿では距離情報を利用した統計的形状モデルによる歩行者検出手法を提案した．頭頂部や肩位置の距離情報

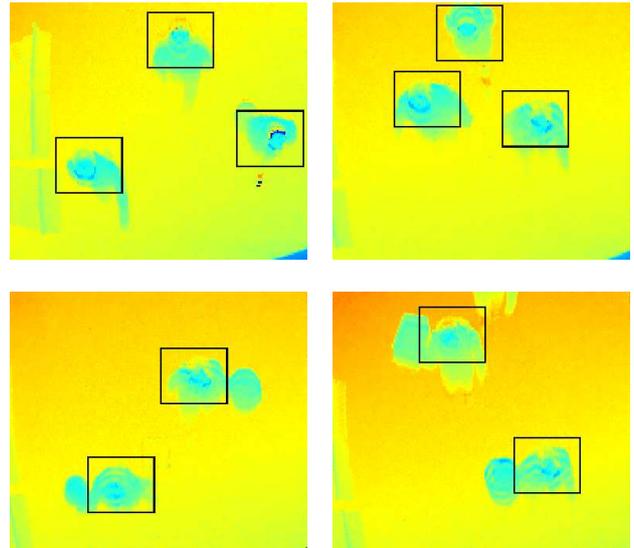


図 7 人検出例

を用いた低次元モデルを用いることで，計算コストが少なく，歩行者の見え方変化に頑強な検出を実現した．今後は，歩行者の向きに依存しない検出手法について検討していく．

## 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，電子情報通信学科論文誌 D，Vol.J93-D，No.3，pp.355–364，2010.
- [2] 稲浦 雄哉，鈴木 昌人，高橋 智一，青柳 誠司：距離画像を用いたオフィス機器の探索 -Bag-of-keypoints と SIFT の利用-，画像の認識・理解シンポジウム (MIRU2011)，pp.298–302，2011.
- [3] Dalal,N. and Triggs,B.：“Histograms of Oriented Gradients for Human Detection”，IEEE Computer Society Conference on Computer Vision and Pattern Recognition，Vol.1，pp.886–892，2005.
- [4] C. Huang, H. Ai, Y. Li, and S. Lao：“Learning Sparse Features in Granular Space for Multi-View Face Detection”，Proc. of the 7th International Conference on Automatic Face and Gesture Recognition, pp.401–406, 2006.
- [5] 竹村 和将，宮城 茂幸：頭上からの距離画像に基づいたスパース特徴量による歩行者検出，電子情報通信学会大会講演論文集，2013，p.188，2013.
- [6] MESA IMAGING AG - SwissRanger SR4000 - miniature 3D time-of-flight range camera. <http://www.mesa-imaging.ch/>
- [7] R.E. Schapire and Y. Singer：“Improved Boosting Algorithms Using Confidence-rated Predictions”，Machine Learning，No.37，pp.297–336，1999.