O5-5

距離画像セグメンテーションに基づくリアルタイム人物検出

○ 生形徹†，有江誠†，アレサンドロ・モロ†，寺林賢司†，梅田和昇†

○ Toru UBUKATA†, Makoto ARIE†, Alessandro MORO†,
Kenji TERABAYASHI† and Kazunori UMEDA†

†：中央大学 / JST CREST，crest@sensor.mech.chuo-u.ac.jp

＜要約＞ 距離画像を領域分割（セグメンテーション）した情報と，特徴量ベースの人物検出器を組み合わせることで，高精度な人物検出とリアルタイム処理を実現する手法を提案する．提案手法では，あらかじめ Mean Shift Clustering により距離画像をセグメンテーションし，人物の探索範囲を限定して処理時間の削減と誤検出の低減を実現している．また，セグメンテーション結果から人物同士の遮蔽（オクルージョン）を検出し，遮蔽部における識別器の寄与を抑制して検出精度を向上している．人物検出には Joint HOG 特徴を用いており，遮蔽を考慮した検出ウィンドウの統合により未検出を減らしている．評価実験の結果，従来の Joint HOG 特徴を用いた人物検出手法と比較して検出精度が向上し，約 11[fps]の処理速度を実現した．
＜キーワード＞人物検出，セグメンテーション，Joint HOG 特徴，オクルージョン，ステレオカメラ

1. はじめに

近年，監視カメラの映像から不審者の検出や人の行動解析などを行う技術が期待されており，画像中から人物を自動的に検出することが求められている．近年の人物検出では，HOG 特徴[1]のような輝度勾配に基づいた局所特徴量を Boosting や SVM によって学習・識別する手法が盛んに研究されている．しかし，このような局所特徴量に基づく人物検出は複雑な背景やオクルージョンによって検出精度が落ちる場合がある．また，画像中の人物の大きさが未知なため，検出ウィンドウのスケールを変化させながら複数回スキャンする必要があり，リアルタイムでの検出が困難である．

そこで，本稿ではステレオカメラから取得した距離画像を利用して，オクルージョンを考慮したリアルタイムでの人物検出手法を提案する．提案手法では特徴量算出の前処理として，距離画像を Mean Shift Clustering により領域分割する．これにより，検出ウィンドウの走査領域を限定し，距離情報からウィンドウサイズを推定できるため，処理時間を削減することができる．また，分割された領域間の距離を比較することでオクルージョンを検出し，検出精度の向上を実現する．

2. 距離画像セグメンテーション

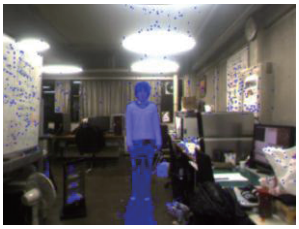
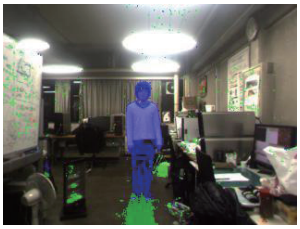
2.1 前景検出

画像中で人物が存在する領域を限定するため，背景差分により前景を抽出する．検出された前景領域はオブジェクトの影を含むため，影検出により前景から除去する．画像座標(x,y)における輝度値を $I(x,y)$ とし，背景画像における同位置の輝度値を $I'(x,y)$ とすると，影を判定する評価関数は次式で表わされる．

$$\theta_{(t+1,x,y)} = \begin{cases} \alpha\Psi_{(x,y)} + \beta\Lambda_{(x,y)} + (1-\alpha-\beta)\theta_{(t,x,y)}, \\ \qquad if\ \ \frac{I_{(x,y)}}{\eta} < I'_{(x,y)} \\ \infty, otherwise \end{cases} \quad (1)$$

ここで，θ は影と判断するためのスコアを表し，θ が閾値以下となる画素を影と判定して前景領域から除去する．Ψは前景の近傍画値と背景の近傍画素値の相違度を表し，Λ は前景の色相と背景の色相の相違度を表す．また，α，β，η はそれぞれの項に対して重みを与える定数である．今回用いた影検出手法の詳細は文献[2]にて述べられている．

前景検出結果を図 1 に示す．図中において，背景差分により抽出された領域を青，影と判定された領域を緑で表わす．

(a) 背景差分　　(b) 影検出

図 1　前景検出

2.2　前景領域のセグメンテーション

前景検出で抽出された領域からオクルージョンを検出し，重なりが生じたオブジェクトを個々に検出するため，前景領域の距離画像を領域分割する．

世界座標系 X-Y 平面(人物を俯瞰した面)にセルを構成し，前景の距離情報(図 2(c))を図 2(d)のように投影する．ただし，図 2(d)は見やすい様に 1m×1m のセルを描画しており，実際は 5cm×5cm のセルを構成する．各セルで投影された距離情報のヒストグラムを構成し，前景領域における n 個の連結成分 F_i ($i = 1,\cdots,n$)ごとにヒストグラムを構築する．

人が直立していると仮定すると，ヒストグラムのピーク周辺に人物が存在する可能性が高い．そこで，ヒストグラムの頻度に対して，ガウシアンカーネルを用いて Mean Shift Clustering を行う．任意のセル c での位置ベクトルを P_c とすると，重心位置 v における Mean Shift ベクトル $m(v)$は次式で表わされる．

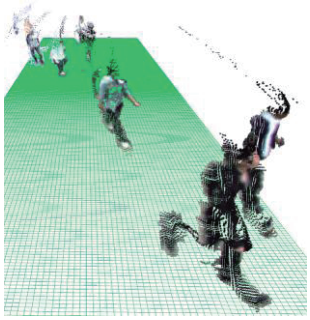
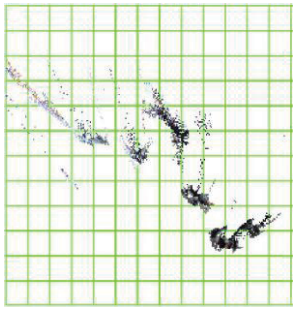
$$m(v) = \frac{\sum_c P_c H(c) g\left(\left\|\frac{v - P_c}{\sigma}\right\|^2\right)}{\sum_c H(c) g\left(\left\|\frac{v - P_c}{\sigma}\right\|^2\right)} - v \qquad (2)$$

ここで，$H(c)$は任意のセルにおけるヒストグラムの頻度，g はカーネル関数，σ はガウシアンカーネルで用いる標準偏差を表す．この Mean Shift ベクトル $m(v)$を用い，下記のステップで距離画像をクラスタリングする．

A) 投影面の大きさに応じてカーネルの初期位置，配置数を推定
B) 各カーネルを式(2)の反復計算により移動させ，ヒストグラムのピーク位置を推定
C) 近傍のカーネルを統合し，重心位置より一定範囲内のセルを同じクラスタとする

分類されたクラスタごとのセルに含まれる投影点を画像に逆投影することで，図 2(b)のように画像中のオブジェクトを分割する．


(a) 前景領域 F_i　　(b) 分割された領域 $SF_{i,j}$

(c) 前景の三次元点群　　(d) 三次元点群の俯瞰図

図 2　距離画像セグメンテーション

3.　オクルージョンを考慮した人物検出

距離画像セグメンテーションの結果を利用し，特徴量の算出時間を削減してリアルタイムでの人物検出手法を構築する．また，セグメンテーション結果からオクルージョンを検出し，遮蔽部における識別器の寄与を抑制することで検出率の向上を図る．

3.1　Joint HOG 特徴を用いた人物検出

HOG 特徴[1]は検出ウィンドウ内をセルに分割し，各セルにおける輝度勾配を勾配方向ごとにヒストグラム化することで特徴量を得る．単一の HOG 特徴では人の対称的な形状や連続的な形状を表現することが困難なため，本研究では複数の HOG 特徴の共起[3]を表現して組み合わせた Joint HOG 特徴[4]を用いる．

共起を表現した特徴量[3]を全セルの組み合わせに対し求め，1 段階目の Real AdaBoost[5]により識別に有効な特徴の組み合わせを選択し，Joint HOG 特徴を生成する．その後，生成された Joint HOG 特徴から 2 段階目の Real AdaBoost により識別に有効な特徴量のみを選択し，強識別器 $H(X)$を構築する．

$$H(X) = \sum_{t=1}^{T} h_t(X) \qquad (3)$$

ここで，Xは選択された Joint HOG 特徴，Tは 2 段階目の学習回数，$h_t(X)$は 1 段階目の学習から得られる強識別器を表す．

図 3 領域別走査

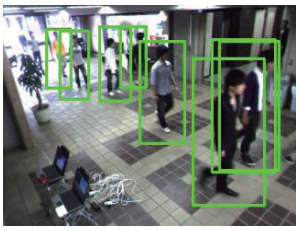


図 4 ウィンドウの統合

3.2 検出ウィンドウの走査・統合

本研究では，検出ウィンドウの走査を分割された領域 $SF_{i,j}(j=1,\cdots,m_i)$ごとに行う．ここで，m_iは前景領域 F_i が分割された数である．また，領域ごとの距離情報をもとに，ウィンドウサイズを動的に変化させることで走査回数を削減できる．

人物の大きさはカメラからの距離と反比例の関係にあるので，比例定数 k_h，k_wを用いてサイズを推定する．また，カメラの仰角による人物の見えの変化が擬似透視投影に従うと仮定し，下記の式でウィンドウサイズ R_h（高さ），R_w（幅）を推定する．

$$R_h = \frac{k_h}{W_Y(i,j)}(cos\,\theta - y\,sin\,\theta) \qquad (4)$$

$$R_w = \frac{k_w}{C_Z(i,j)} \qquad (5)$$

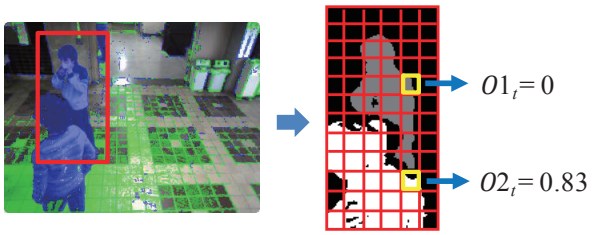
ここで，$W_y(i,\ j)$は世界座標系でのカメラから領域 $SF_{i,j}$ までの代表距離，$C_z(i,j)$はカメラ座標系でのカメラから領域 $SF_{i,j}$ までの代表距離，θはカメラの仰角，y は画像の縦幅を正規化した時の画像座標を表す．

図 2(b)での領域 $SF_{i,j}$ ごとに提案手法を用いて走査した結果を図 3 に示す．特徴量算出の時間を削減すると共に，背景からの誤検出を低減している．

特徴量算出後，人物と識別された検出ウィンドウは，近傍のウィンドウと統合されて検出結果となる．図 3 での検出ウィンドウの統合結果を図 4 に示す．ウィンドウの統合を分割された領域 $SF_{i,j}$ ごとに行うことで，3 次元空間で離れたウィンドウは統合されない．すなわち，図 3 で異なる色の検出ウィンドウ間では統合が行われない．これにより，オクルージョンによる未検出を減らしている．

3.3 オクルージョン検出

画像中で人物同士が重なる場合，遮蔽により後方人物の見えが捉えられない．そこで，検出ウィンドウ内に複数の領域 $SF_{i,j}$ が含まれるとき，各領域の距離 $W_y(i,j)$を比較してオクルージョンを検出する．図 5 に示すように，走査対象となる領域(灰色)より手前にある領域(白色)をオクルージョンとして検出する．



(a) 検出ウィンドウ　(b) 各セルの遮蔽割合

図 5 オクルージョン検出

(a) ポジティブ　(b) ネガティブ

図 6 学習に用いたサンプル

Joint HOG 特徴は組み合わせた 2 つのセル内の特徴量から弱識別器 $h_t(X)$の出力をもとめる．そこで，各セルでオクルージョンを含む割合(図 5(b)で白い領域を含む割合)を算出する．弱識別器 $h_t(X)$で使用される 2 つのセルのオクルージョンの割合をそれぞれ $O1_t$，$O2_t$ とすると，オクルージョンの割合に応じて，下記の式より最終識別器 $H'(X)$を得る．

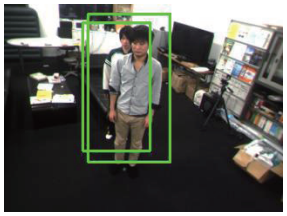
$$H'(X) = \sum_{t=1}^{T}\{h_t(X)\cdot(1-O1_t)\cdot(1-O2_t)\} \qquad (6)$$

オクルージョン割合が大きい程，弱識別器の出力が小さくなることから，遮蔽部の識別器の出力を抑制することができる．この最終識別器 $H'(X)$に対し閾値を設け，人物か否かを識別する．

4. 評価実験

4.1 実験条件

学習には NICTA Pedestrian Dataset[6]を使用し，ポジティブサンプル 7,892 枚，ネガティブサンプル 30,000 枚を用いた．使用したサンプル例を図 6 に示す．学習は 1 段階目で 10 回，2 段階目で 300 回行い，識別器を構築した．実験時のステレオカメラは Bumblebee2(Point Gray Research)を使用し，処理には Intel Core 2 Duo CPU(3.06GHz)を用いた．また，評価結果の T. Pos.(True Positive)は正しい検出，F. Neg.(False Negative)は未検出，F. Pos.(False Positive)は誤検出を表す．



● 識別器$H(X)$の平均出力
$H(x) = -4.70$

● 識別器$H'(X)$の平均出力
$H'(x) = -2.71$

(a) 実験シーン　(b) 後方人物の識別器出力

図 7　遮蔽を考慮した人物検出

4.2　オクルージョンを考慮した検出精度の評価

3.3 節で述べたオクルージョン検出を用いた人物検出手法の有用性を検証するため，式(3)，式(6)それぞれの識別器を用いて検出精度を比較する．識別に用いる閾値は誤検出がでない限界の値を設定した．実験環境は図 7(a)に示すように，オクルージョンが発生するシーン 400 フレームで評価を行った．表 1 に評価結果を示す．

表 1 より，オクルージョンを考慮することで検出精度が向上できていることが見てとれる．これは，図 7(b)に示す様に，遮蔽部の識別器出力を抑制し，最終識別器の出力が高くなったことが要因として考えられる．

4.3　提案手法の評価

提案手法の有用性を検証するために，画像全体で検出ウィンドウをスキャンする手法（従来手法）との精度を比較する．閾値は手法ごとに適当な値を実験的に求めた．評価には学習データセットと異なる単純な背景（図 8）と複雑な背景（図 9）の動画をそれぞれ 2,000 フレーム用いた．単純な背景での検出結果の例と評価結果をそれぞれ図 8, 表 2 に示す．また，複雑な背景での検出結果の例と評価結果をそれぞれ図 9，表 3 に示す．

表 2，表 3 の結果より従来手法では背景の複雑化により検出精度が低下しているのがわかる．これは図 9(b)に示すように，背景に人物の形状に似た形状（図の中心にある十字模様など）が映ると誤検出を誘発してしまうことが原因である．それに対し，提案手法では処理領域を限定することで誤検出を低減している．また，オクルージョンを考慮するため，図 8(b)，図 9(b)においてオクルージョンにより未検出になっている人物も，図 8(a)，図 9(a)では正しく検出していることが見てとれる．

表 1　遮蔽考慮による検出精度の比較

Classifier	T. Pos.	F. Neg.	F. Pos.
Eq. (3) : $H(x)$	71.3 %	28.7 %	0.0 %
Eq. (6) : $H'(x)$	89.1 %	10.9 %	0.0 %

表 2　単純な背景のシーンにおける評価結果

Method	T. Pos.	F. Neg.	F. Pos.
Proposed	80.0 %	20.0 %	1.3 %
Conventional	73.0 %	27.0 %	9.8 %

表 3　複雑な背景のシーンにおける評価結果

Method	T. Pos.	F. Neg.	F. Pos.
Proposed	83.1 %	16.9 %	3.1 %
Conventional	63.5 %	36.5 %	65.9 %

表 4　処理速度

Function	Proposed method [ms]	Conventional method [ms]
Capture	16.3	16.3
Subtraction	0.4	-
Stereo	18.8	-
Shadow	6.7	-
Segmentation	11.5	-
Joint HOG	30.4	502.2
Others	4.0	-
Total	88.1	518.5

4.4　処理速度

表 4 に各処理プロセスにおける計算時間を示す．画面全体をスキャンする従来手法と比較し，処理領域を限定することで処理速度が大幅に向上していることがわかる．また，提案手法では約 11[fps]で動作することから，リアルタイムでの人物検出が可能である．

5.　おわりに

本稿では，距離画像セグメンテーションにより特徴量の算出時間を削減し，リアルタイムでの人物検出手法を提案した．また，オクルージョンを考慮した検出ウィンドウの統合により未検出を減らし，オクルージョン割合に応じた識別器出力の抑制により検出精度を向上させた．

今後は，人物のパーツごとの検出を体全体の検出結果と組み合わせ，人物の見えの変化により対応できる識別器を構築していく予定である．

(a) 提案手法を用いた検出結果の例

(b) 従来手法を用いた検出結果の例

図 8　単純な背景のシーンにおける実験

(a) 提案手法を用いた検出結果の例

(b) 従来手法を用いた検出結果の例

図 9　複雑な背景のシーンにおける実験

参考文献

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” In *Proc. CVPR*, CA, USA, pp. 886-893, 2005.

[2] A. Moro, et al., “Auto-adaptive threshold and shadow detection approaches for pedestrians detection,” In *Proc. AWSVCI,* pp. 9-12, 2009.

[3] T. Mita, T. Kaneko, B. Stenger, O. Hori, “Discriminative Feature Co-occurrence Selection for Object Detection,” *Trans. on IEEE Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1257-1269, 2008.

[4] 尾崎貴洋, 山内悠嗣, 藤吉弘亘, “Joint HOG 特徴を用いた 2 段階 AdaBoost による車両検出”, 動的画像処理実利用化ワークショップ(DIA2008), I1-13, 2008.

[5] R. E. Schapire and Y. Singer, “Improved Boosting Algorithm Using Confidence-rated Predictions,” *Machine Learning*, No. 37, pp. 297-336, 1999.

[6] G. Overett, L. Petersson, N. Brewer, L. Andersson, N. Pettersson, “A new pedestrian dataset for supervised learning,” In *Proc. IEEE Intelligent Vehicle Symposium*, pp. 373-378, 2008.