

DPMのスコア回帰を用いたオクルージョン対応による 人検出性能の高精度化

宇佐見 剛司^{1,a)} 山内 悠嗣¹ 山下 隆義¹ 藤吉 弘亘^{1,b)}

概要: パーツベースの物体検出法である Deformable Part Models(DPM) は、姿勢の変動に頑健な物体検出法として知られている。DPM は全てのパーツが見えていることが前提となっており、オクルージョンの発生により検出対象のパーツが隠れると、検出が困難になるという問題がある。そこで、本研究では検出スコアとオクルージョンの発生率からオクルージョンの影響を低減したスコアを回帰することで、オクルージョンに頑健な人検出法を提案する。提案手法は、距離画像から判定したオクルージョンの発生割合と DPM の各スコアを説明変数として回帰に用いる。評価実験により、従来の DPM の検出性能と比較し、提案手法では、約 20%の精度向上を確認した。

1. はじめに

2005 年に Dalal 等が提案した Histograms of Oriented Gradients(HOG) 特徴量と Support Vector Machine(SVM) を組み合わせた手法 [1] は、人検出だけでなく他のカテゴリの物体検出問題にも適用され、幅広く利用されている。HOG 特徴量は輝度の勾配に着目した局所特徴量であり、照明の変動や局所的な位置ずれを吸収するという特長がある。そのため、高精度化を目的とした HOG 特徴量ベースの物体検出法が数多く提案されている [2][3][4]。中でも、Felzenszwalb 等が提案した Deformable Part Models(DPM)[5] は、人全体のアピランス特徴のみでなく、手や脚部等の部位毎にアピランス特徴を捉えるパーツベースのアプローチである。DPM は各部位の位置関係を学習することで、姿勢変動に対する頑健性を獲得し、高精度な人検出を実現した。しかし、全てのパーツが観測できることを前提にモデルを学習しているため、識別時に人体の一部にオクルージョンが発生すると、DPM では検出が困難になるという問題がある。

オクルージョンに対応した手法として、オクルージョン領域を判定し、オクルージョン領域に応じて識別を調整する手法が提案されている [4][6][7]。Wang 等はオクルージョン領域をクラスタリングし、オクルージョン領域に応じて適用するパーツ検出器を切り替える手法を提案した [4]。Enzweiler 等は距離情報と動き情報からオクルージョン領

域を検知し、パーツベースの識別器に対してオクルージョンの発生度合いを用いて重み付けする手法を提案した [6]。池村等はオクルージョン領域の抽出後、Real AdaBoost の弱識別器に対して重み付けする手法を提案した [7]。しかし、これらの手法は識別器のスコアに対して重み付けを行うため、オクルージョンの発生率が高くなると全体のスコアが小さくなり、正しく検出できない場合がある。

本研究では、算出したオクルージョン発生率と各スコアを説明変数とし、Support Vector Regression(SVR)[8] を用いて回帰することにより、オクルージョンの影響を低減したスコア算出法を提案する。提案手法は、距離画像から判定したオクルージョンの発生割合と DPM の各スコアを説明変数として回帰に用いる。これにより、人物の多くの領域にオクルージョンが発生しても、本来のスコアに近い出力を得ることができる。

2. Deformable Part Models[5]

DPM はパーツベースの物体検出法であり、姿勢変動に対応した手法である。本章では、DPM に用いられる識別関数とオクルージョンによる問題について述べる。

2.1 DPM における識別関数

DPM は、図 1 に示すようにルートフィルタ、パーツフィルタ、空間モデルにより構成されている。ルートフィルタは、図 1(a) のように人の全身のアピランス特徴を捉えるモデルである。図中の明暗は、HOG 特徴量を可視化したものである。パーツフィルタは、図 1(b) のように頭部や脚部等の人の特徴的な部位のアピランス特徴を捉えるモ

¹ 中部大学

Chubu University

a) usami915@vision.cs.chubu.ac.jp

b) hf@cs.chubu.ac.jp

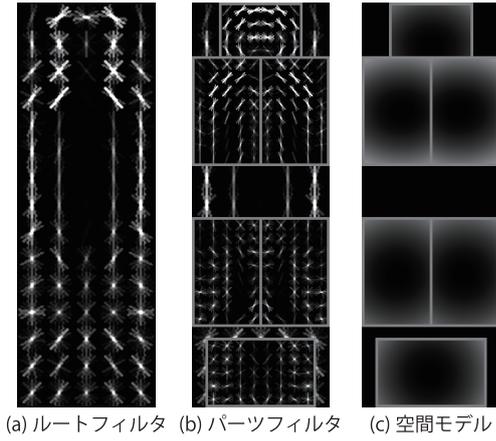


図1 DPMのモデルの可視化

デルである。図中の矩形は学習したパーツフィルタの基準位置を表し、明暗はHOG特徴量を可視化したものである。空間モデルは、図1(c)のようにパーツの位置変動に応じて重み付けするモデルである。空間モデルにより与えられる変形コストは、基準となる位置から離れるほど大きくなる。図中の明暗は変形コストを可視化したものである。

DPMの検出スコアは、上記の3種類のモデルから式(1)に示す識別関数より求める。

$$\text{score}(p_0) = F'_0 \cdot \phi(H, p_0) + \sum_{i=1}^n \max_{x_i, y_i} (F'_i \cdot \phi(H, p_i) - d_i \cdot \phi_d(dx_i, dy_i)) + b \quad (1)$$

第1項はルートフィルタのスコア、第2項はパーツフィルタのスコア、第3項はバイアスである。 F'_i ($i = 0, \dots, n$)は各フィルタの重みベクトルであり、 $\phi(H, p_i)$ ($i = 0, \dots, n$)は各フィルタに対応する検出ウィンドウ内の特徴ベクトルである。これら2つのベクトルの内積 $F'_i \cdot \phi(H, p_i)$ ($i = 0, \dots, n$)が各フィルタにおけるスコアとなる。 $i = 0$ の場合はルートフィルタであり、 $i = 1, \dots, n$ の場合はパーツフィルタである。 n はパーツ数である。 (dx_i, dy_i) は各パーツの基準位置からの移動量であり、式(2)により求める。

$$(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i) \quad (2)$$

また、 $\phi_d(dx_i, dy_i)$ はパーツの移動方向と移動量を表した2次関数であり、式(3)となる。

$$\phi_d(dx_i, dy_i) = (dx_i, dy_i, dx_i^2, dy_i^2) \quad (3)$$

この $\phi_d(dx_i, dy_i)$ は変形コスト $d_i \cdot \phi_d(dx_i, dy_i)$ の算出に用いられる。

2.2 オクルージョンによる問題点

DPMは全てのパーツが観測できることを前提にモデルを学習するため、人体の一部にオクルージョンが発生した場合、検出が困難になるという問題がある。そこで、まずDPMにおけるオクルージョンの影響について調査す

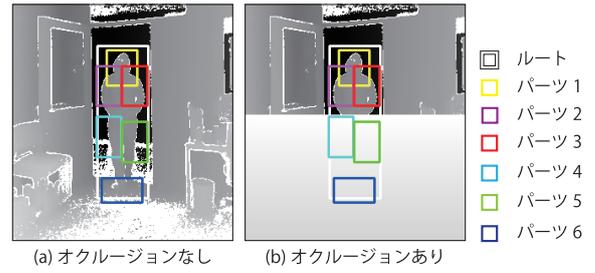


図2 オクルージョンの疑似的な生成

る。図2(a)はDPMにより検出した人のルートフィルタとパーツフィルタの位置を示す。このとき、DPMの検出スコアは、1.519である。次に、図2(b)のように疑似的にオクルージョンを発生させた際の検出スコアは、-0.361である。これらの検出スコアを含めた各パーツのスコアを表1に示す。表1より、オクルージョンが発生したルートのスコアとパーツ4、5、6のスコアが大幅に低下し、その結果、検出スコアが低くなることがわかる。このように、DPMはパーツ領域の全体にオクルージョンが発生すると、著しく検出性能が低下する。

3. 提案手法

本研究は、DPMのスコアとオクルージョン発生率を説明変数とした回帰によるスコア算出法を提案する。図3に提案手法の流れを示す。提案手法は、学習によりDPMモデルとSupport Vector Regression(SVR)の重みとバイアスを学習する。検出時には、入力距離画像から説明変数を算出し、SVRによりオクルージョンの影響を低減したスコアを求める。

3.1 DPMとSVRの学習

提案手法の学習では、DPMによるモデルとSVRによる重みとバイアスを求める。

3.1.1 DPMの学習

DPMは、各フィルタの重みベクトル F'_i ($i = 0, \dots, n$)と変形コストを定義する4次元ベクトル d_i ($i = 1, \dots, n$)とバイアス b を同時に学習する必要があるため、Latent SVMを用いて学習を行う。Latent SVMは目的関数 $L_{D(Z)}(\beta)$ が最小となる β を求める。目的関数 $L_{D(Z)}(\beta)$ は式(4)で定義される。

$$L_{D(Z)}(\beta) = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i f_\beta(x_i)) \quad (4)$$

式(4)の第1項は最適化項であり、第2項は損失関数である。 β は、式(5)に示す学習により得られるパラメータである。

$$\beta = (F'_0, \dots, F'_n, d_1, \dots, d_n, b) \quad (5)$$

$D(Z)$ は学習サンプル (x_i, y_i) の集合である。

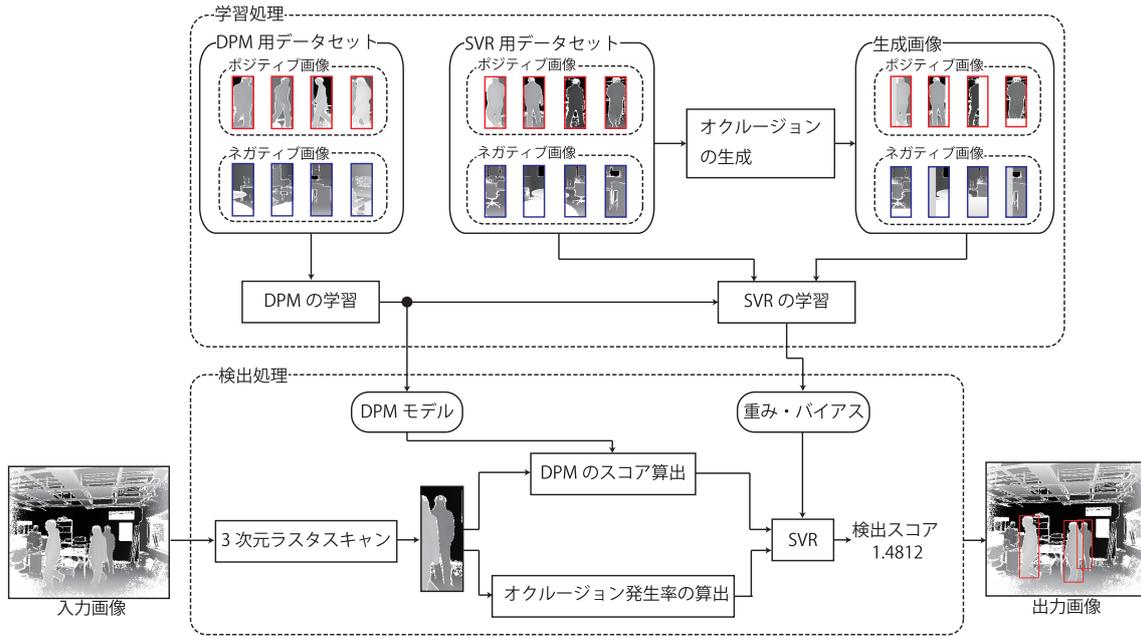


図 3 提案手法の流れ

表 1 オクルージョンの発生による DPM のスコアの変化

	検出スコア	ルート	パーツ 1 (頭部)	パーツ 2 (右肩)	パーツ 3 (左肩)	パーツ 4 (右足)	パーツ 5 (左足)	パーツ 6 (足元)
オクルージョン発生なし	1.519	1.535	0.984	0.494	0.343	0.265	0.313	0.340
オクルージョン発生あり	-0.361	0.651	0.984	0.494	0.343	-0.018	-0.025	-0.057

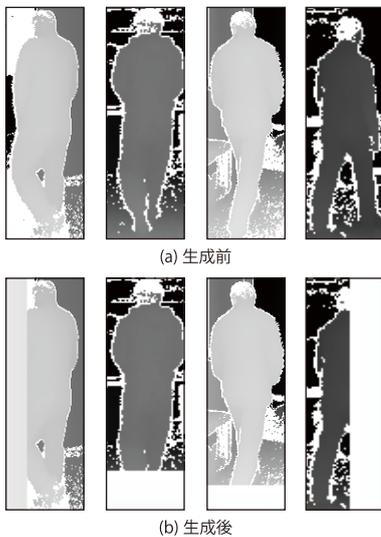


図 4 オクルージョン領域を生成した画像

3.1.2 SVR の学習

提案手法では、オクルージョン発生前の検出スコアを目的変数とし、オクルージョン発生後のスコアとオクルージョン発生率を説明変数に用いる。そのため、SVR の学習では、オクルージョンありとなしの2つのスコアが必要となる。そこで、図 4 に示すように、擬似的にオクルージョンを付与して学習サンプルを生成する。説明変数 \mathbf{x} には、DPM により求められた検出スコア、ルートスコア、パーツスコアと、パーツごとのオクルージョン発生率を用いる。

DPM のパーツ数が 6 個の場合における説明変数を式 (6) に示す。

$$\begin{aligned} \mathbf{x} = & (\text{score}(p_0), F'_0 \cdot \phi(H, p_0), \\ & F'_1 \cdot \phi(H, p_1) - d_1 \cdot \phi_d(dx_1, dy_1), \dots, \\ & F'_6 \cdot \phi(H, p_6) - d_6 \cdot \phi_d(dx_6, dy_6), \\ & O(p_1) * (F'_1 \cdot \phi(H, p_1) - d_1 \cdot \phi_d(dx_1, dy_1)), \dots, \\ & O(p_6) * (F'_6 \cdot \phi(H, p_6) - d_6 \cdot \phi_d(dx_6, dy_6))) \end{aligned} \quad (6)$$

p_i はパーツフィルタの左上角の座標 (x_i, y_i) とサイズ $(width_i, height_i)$ を内包する変数である。式 (7) によりパーツ i のオクルージョン発生率 $O(p_i)$ を求める。

$$O(p_i) = \frac{\sum_{k=y_i}^{y_i+height_i} \sum_{l=x_i}^{x_i+width_i} \alpha(k, l)}{width_i * height_i} \quad (7)$$

$\alpha(k, l)$ は、座標 (k, l) におけるオクルージョン発生の有無を表す関数であり、オクルージョンが発生していれば 1 を、発生していなければ 0 を出力する。

SVR の学習は、目的関数が最小となる重み \mathbf{w} とバイアス b を求めることである。式 (8) に示す目的関数を最小化する \mathbf{w} と b を学習により求める [9]。

$$\arg \min_{\mathbf{w}, b} C \sum_{i=1}^N E(t_i - f(\mathbf{x}_i)) + \frac{1}{2} |\mathbf{w}|^2 \quad (8)$$

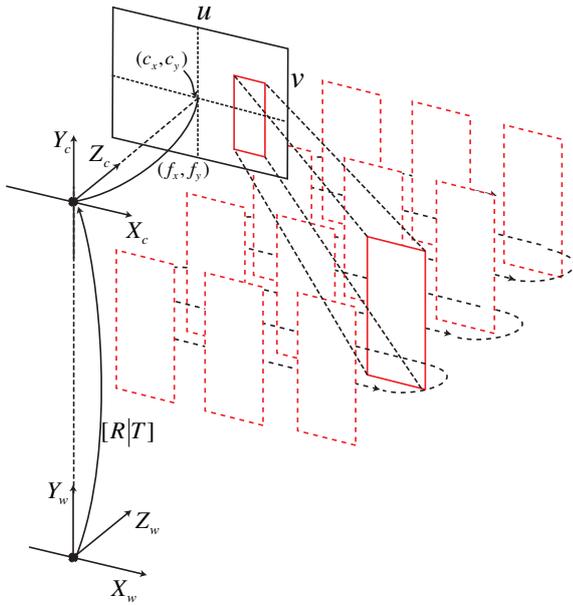


図5 3次元ラスタスキャン

$E(t_i - f(\mathbf{x}_i))$ は、式 (9) に示す誤差関数である。

$$E(\alpha) = \begin{cases} 0 & (\alpha \leq th) \\ \alpha - th & (\alpha > th) \end{cases} \quad (9)$$

th は許容する誤差の閾値を表した値である。

3.2 3次元空間におけるラスタスキャン [7]

本研究では距離画像を対象とするため、検出時にはラスタスキャンを世界座標における3次元空間にて行う。図5に示すように、3次元空間上に検出ウィンドウを配置し、床面に沿って ($Y_w = 0$) ラスタスキャンする。3次元空間におけるラスタスキャンは、床面に接地した検出ウィンドウをラスタスキャンするため、効率的な探索とともに不自然な位置に対する検出を行わない。これにより、検出精度の向上が期待できる。

3次元空間におけるラスタスキャンでは、3次元空間上に配置した検出ウィンドウを画像面に投影する必要がある。世界座標 (X_w, Y_w, Z_w) から画像座標 (u, v) への変換を式 (10) に示す。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A}[\mathbf{R}|\mathbf{T}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (10)$$

$[\mathbf{R}|\mathbf{T}]$ はカメラの外部パラメータであり、 \mathbf{R} は回転行列、 \mathbf{T} は平行移動ベクトルである。本研究では、カメラの姿勢と高さは固定とする。カメラの姿勢は、世界座標とカメラ座標が平行になるように固定する。そのため、回転行列は単位ベクトルとなる。また、世界座標の原点をカメラ位置の床面とするため、平行移動ベクトルは $[0, -\text{カメラの高さ}, 0]^T$ となる。従って、本研究でのカメラの外部パラメータ

は式 (11) となる。

$$[\mathbf{R}|\mathbf{T}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1.4 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (11)$$

\mathbf{A} はカメラの内部パラメータである。内部パラメータは、ピクセル単位で表される焦点距離 (f_x, f_y) とカメラの中心座標 (c_x, c_y) であり、式 (12) となる。

$$\mathbf{A} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

式 (10) に式 (11)、式 (12) を代入すると、次式となる。

$$\begin{cases} u = \frac{X_w}{Z_w} f_x + c_x \\ v = \frac{Y_w - 1.4}{Z_w} f_y + c_y \end{cases} \quad (13)$$

これにより、検出ウィンドウを任意の世界座標で設定し、その世界座標に対応する画像座標を求めることができる。

3.3 オクルージョンの発生判定と発生率の算出

検出対象にオクルージョンが発生している場合、検出対象よりも手前に障害物が存在することになる。従って、注目画素の距離値が検出ウィンドウまでの距離より小さいと手前に障害物があると判定することができる。注目画素の座標 (k, l) におけるオクルージョン判定は次式となる。

$$\alpha(k, l) = \begin{cases} 1 & (Z_w - z(k, l)) > t \\ 0 & (Z_w - z(k, l)) \leq t \end{cases} \quad (14)$$

$z(k, l)$ は座標 (k, l) の距離値、 Z_w はカメラから検出ウィンドウまでの距離値を表す。 t は閾値であり、本研究では約30cm手前に存在するものをオクルージョンと判定している。オクルージョン発生率は、学習時と同じく式 (7) により求める。

3.4 検出時のSVRによる回帰

検出時のSVRに入力する説明変数 \mathbf{x} は、学習時と同様に式 (6) に示す14次元のパラメータである。オクルージョン発生率は、3次元空間におけるラスタスキャンにおいて各検出ウィンドウごとに求める。オクルージョンの影響を低減したスコアは、式 (15) の識別関数 $f(\mathbf{x})$ により求められる。

$$f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w} + b \quad (15)$$

\mathbf{w} はSVRの重みベクトルであり、 b はSVRのバイアス項である。以上より、オクルージョンの影響を低減したスコアを回帰により求める。

4. 評価実験

評価実験では、従来法のDPMと提案手法を比較する。

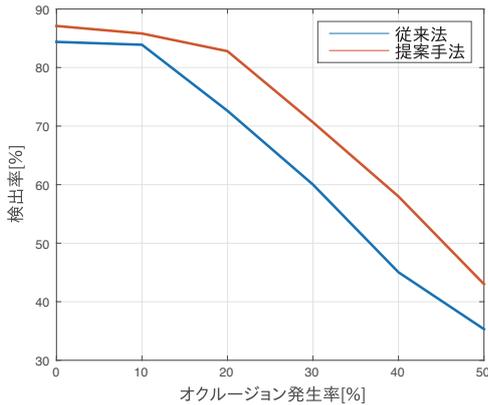


図 6 オクルージョン領域の増加に伴う検出率の比較

擬似的にオクルージョンを生成した画像と実画像を用いた 2 つの実験から提案手法の有効性を評価する。

4.1 オクルージョンに対する頑健性の評価

提案手法と従来法のオクルージョンに対する性能について評価する。

4.1.1 実験概要

本実験では、オクルージョン率が変化した場合の検出率を評価する。検出に用いる閾値は DPM の学習により得られた値を用いる。オクルージョン率を設定して擬似的にオクルージョンを付与した画像データを評価に用いる。

本実験では、KinectV2 により取得した距離画像を対象とする。DPM の学習用にポジティブサンプルに 905 枚、ネガティブサンプルに 1,008 枚を用いる。ポジティブサンプルはオクルージョンが発生していない人領域を切り出した距離画像である。SVR の学習時には、擬似的にオクルージョンを生成し、ポジティブサンプルに 1,209 枚、ネガティブサンプルに 1300 枚を用いる。オクルージョンの生成は参考文献 [10] のオクルージョンの発生パターンと頻度を用いる。評価用データは、ポジティブサンプル 800 枚に指定した割合で擬似的にオクルージョンを付与したものをを用いる。

4.1.2 実験結果

各オクルージョン発生率における検出率の比較結果を図 6 に示す。オクルージョン発生率が 10% までは両手法の検出率に大きな差は見られない。オクルージョン発生率が 20% 以上のとき、提案手法は従来法より 10% 高い検出率を得た。

4.2 実画像における検出性能の評価

オクルージョンが発生した実画像における評価を行う。

4.2.1 実験概要

本実験では、実際にオクルージョンが発生している人を含む評価用データセットを用い、検出率と誤検出率を比較する。閾値は、DPM の学習により得られた値を用いる。

表 2 検出性能の比較

	検出率 [%]	誤検出率 [%]
従来法	46.67	0.31
提案手法	65.62	4.47

DPM の学習用と SVR の学習用は、4.1.1 節のデータセットを使用する。評価用データには、KinectV2 により取得した距離画像 649 枚を用いる。画像中に含まれる人領域は 1,088 個ある。

4.2.2 実験結果

検出率と誤検出率の比較結果を表 2 に示す。提案手法は従来法に比べ誤検出率が多少増加しているが、検出率を約 20% 向上させることができた。

図 7 に、従来法と提案手法の 3 次元ラスタスキャンした際の検出例を示す。図 7(a) より、従来法で検出できない人体を提案手法は検出できていることがわかる。ただし、図 7(b) より、提案手法は従来法に比べ誤検出が多く発生する。誤検出が発生している座標のスコアを確認すると設定した閾値よりは高い値であったが、正しく検出が行われている座標のスコアと比較するとスコアが低い傾向にある。そのため、閾値の変更により誤検出の低減が可能であると考えられる。

5. おわりに

本研究では、回帰を行うことによるオクルージョンの影響を低減したスコアの算出法を提案した。提案手法は、説明変数に DPM より得られるスコアだけでなく、オクルージョン発生率も用いることで、オクルージョンに頑健な人検出が可能となった。今後は、学習サンプルやパラメータを見直すことにより、検出性能のさらなる高精度化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", CVPR, vol.1, pp.886-893, 2005.
- [2] A. Bosch, A. Zisserman and X. Munoz, "Representing shape with a spatial pyramid kernel", ICIVR, 2007.
- [3] P. Ott and M. Everingham, "Implicit color segmentation features for pedestrian and object detection", ICCV, 2007.
- [4] X. Wang, H. X. tomy and Y. Shuicheng, "An HOG-LBP human detector with partial occlusion handling", ICCV, pp.32-39, 2009.
- [5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object detection with discriminatively trained part based models", PAMI, vol.32, no.9, pp.1627-1645, 2010.
- [6] M. Enzweiler, A. Eigenstetter, B. Schiele and D. M. Gavrila, "Multi-cue pedestrian classification with partial occlusion handling", CVPR, pp.990-997, 2010.
- [7] S. Ikemura and H. Fujiyoshi, "Real-time human detection using relational depth similarity features", ACCV, pp.25-38, 2011.



(a) 提案手法の成功例



(b) 提案手法の誤検出例

図 7 従来法と提案手法の検出例

- [8] B. Debasish, P. Srimanta and P. D. Candra, "Support vector regression", NIP, vol.11, no.10, pp.201-224, 2007
- [9] C. Chih-Chung and L. Chih-Jen, "A library for support vector machines", ACM TIST, vol.2, no.2, pp.27:1-27:27, 2011.
- [10] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian detection: an evaluation of the state of the art." PAMI, vol.34, no.4, pp.743-761, 2012.