

直線成分に基づく HOL 特徴量による誤検出の抑制

山内 悠嗣 † 宮下 卓也 † 山下 隆義 † 藤吉 弘亘 †
† 中部大学

E-mail: yuu@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract

画像中から人を検出する手法として局所特徴量と統計的学習手法を組み合わせた方法が提案されている。局所特徴量により人の形状を捉えることで画像中から人を検出できるが、背景画像の直線成分と人の形状の違いを明瞭に捉えることが難しいため、誤検出を引き起こす場合がある。そこで、本稿では背景画像の直線成分と人の形状の違いを捉え、誤検出を抑制する Histograms of Oriented Lines(HOL) 特徴量を提案する。提案手法は、画像中から直線を検出し、局所領域において直線の方向ヒストグラムを作成し、特徴量として利用する。ヒストグラム作成の際、直線の長さに応じて重み付けしてヒストグラムに投票することで、検出ウィンドウを縦断するような長い直線とそれ以外の短い線分の違いを明瞭に捉える。評価実験により、直線成分の長さを捉えることができる HOL 特徴量を用いることで誤検出を抑制できたことを確認した。

1 はじめに

安心・安全のための技術として、画像から人を検出する研究が盛んに行われている。多くの人検出法は、画像に対してウィンドウを仮定し、ウィンドウ内の画像より抽出した特徴量に基づき識別器により人と背景を判別する。この処理をウィンドウの位置と大きさを網羅的に変えながら、同様の処理を行うことで画像中の人を検出する。

現在までに高精度な人検出を実現するために様々な手法が提案されており、その中の一つのアプローチとして、特徴量を如何に設計するかについて着目している研究が提案されている。人検出に用いられる代表的な特徴量としては、輝度の勾配に基づく特徴量として Histogram of Oriented Gradients(HOG) 特徴量 [1] がある。HOG 特徴量は局所的な領域において勾配方向ヒストグラムを作成するため、ウィンドウの位置ずれや大きさの違いに対する頑健性を得ることができる。また、近接する幾つかの勾配方向ヒストグラムを用いて、局所領域において正規化するため、局所的な明るさの変

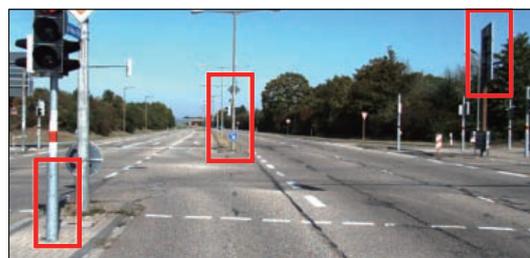


図 1 従来法 [1] で発生する誤検出例

動に対しても頑健となる。HOG 特徴量は簡単な計算ながらも優れた性能を発揮するため、HOG 特徴量をベースにした人検出法 [2, 3] や改良した手法 [4, 5, 6] が数多く提案されている。

しかしながら、これら HOG 特徴量をベースにした特徴量を用いた場合、図 1 に示すような誤検出が発生する場合がある。図 1 の誤検出した領域を観察すると、ウィンドウを縦断するような長い直線を含むような画像に対して誤検出していることがわかる。この原因は、HOG 特徴量が背景画像から観測される縦の直線と、人の肩部から脚部にかけての輪郭の違いを捉えることができないためである。この両者の違いを特徴量によって捉えることができれば、このような誤検出の抑制が期待できる。

そこで、上記のような背景画像から観測される長い直線と人から観測される輪郭の違いを捉えることができる Histograms of Oriented Lines(HOL) 特徴量を提案する。提案手法は、画像中から直線を検出し、局所領域において直線の方向ヒストグラムを作成し特徴量として利用する。ヒストグラム作成の際、直線の長さに応じて重み付けしてヒストグラムに投票することで、ウィンドウを縦断するような長い直線とそれ以外の短い線分の違いを明瞭に捉えることができる。これにより、ウィンドウを縦断するような直線を含む背景画像の誤検出の抑制が期待できる。

本稿では、まず 2 章にて従来の人検出法における誤検出の傾向について整理する。そして、3 章にて誤検出を抑制するための提案手法について述べる。4 章にて提案手法の有効性を確認するための評価実験について述べ、最後に 6 章にてまとめる。



図 2 誤検出したサンプルの例

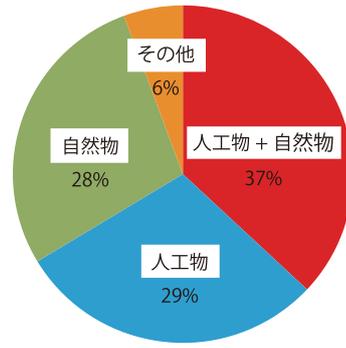


図 4 誤検出した背景画像の内訳

2 従来法における誤検出の傾向調査

本章では、従来の人検出法がどのような背景画像に対して誤検出を引き起こすのかについて調査する。

2.1 調査概要

従来の人検出法がどのような背景画像を誤検出するか調査する。人検出器として HOG 特徴量と SVM による人検出法 [1] を用いる。画像データベースとして車載カメラにより撮影された KITTI Vision Benchmark Suite[7] を用いる。

従来の人検出法がどのような背景画像を誤検出するか調査するために、まず人が存在しない背景画像から人を検出する。そして、誤検出した 30,341 枚の背景画像に対して属性を付与する。図 2 に誤検出した背景画像の一部を示す。図 2 から誤検出するような背景画像には、複雑なテクスチャを持つものが多く、特に人の形状と似ている長い直線を含むような画像が多いことがわかる。そこで、本研究では下記の属性を人手で付与する。

- 物体の性質
 - 自然物
 - * 樹木 / 芝, 葉 / その他
 - 人工物
 - * 建物 / 車両 / 電柱 / 標識 / その他
 - その他
- 直線の長さの割合
 - 0 ~ 30% / 31 ~ 60% / 61 ~ 100%



- 物体の性質
 - ・ 人工物
 - ・ 標識
 - ・ 自然物
 - ・ 樹木
- 直線の長さの割合
 - ・ 71% ~ 100%

(a) 誤検出サンプル (b) 属性

図 3 誤検出したサンプルと付与する属性の例

なお、誤検出した背景画像には物体の性質に関しては属性の重複も含むこととする。図 3 に誤検出した背景画像に対して付与した属性の例を示す。図 3(a) には、木の幹と標識のポール部分が観測されている。そのため、物体の性質としては人工物の標識、自然物の樹木の 2 つの属性を付与する。

2.2 傾向調査の結果

本節では、誤検出した背景画像に付与した属性の傾向について整理する。

物体の性質

図 4 に誤検出した背景画像に含まれる物体の性質の属性の内訳を示す。まず、誤検出したサンプル 30,341 枚のうち、人工物 8,911 枚、自然物 8,501 枚、人工物と自然物 11,218 枚、その他 1711 枚であった。

次に、各物体の性質の内訳を図 5 に示す。人工物の内訳 (図 5(a)) を見ると、最も多く誤検出している画像は建物であることがわかる。誤検出した建物の画像の一部を図 6 に示す。誤検出した建物の画像を見ると、多くの画像において縦の直線が含まれていることがわかる。同様に、誤検出が多い車両においても同様の傾向が見られた。自然物の内訳 (図 5(b)) を見ると、芝や葉などのテクスチャが複雑な画像の割合が圧倒的に多い。

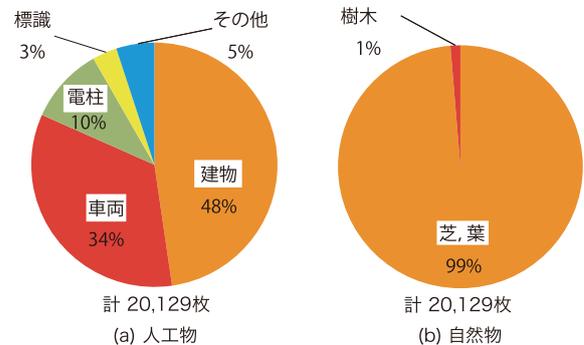


図 5 誤検出した背景画像の物体の属性の内訳

直線の長さの割合

図 7 に誤検出した背景画像のウィンドウに対する直線の長さの割合の内訳を示す。図 7 から、誤検出した



図 6 建物の属性を持つ誤検出したサンプル

背景画像からは短い直線から長い直線まで観測されていることがわかる。

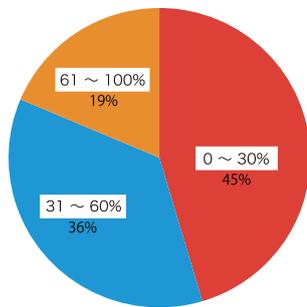
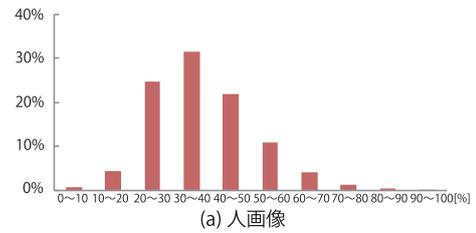


図 7 誤検出した背景画像のウィンドウに対する直線の長さの割合の内訳

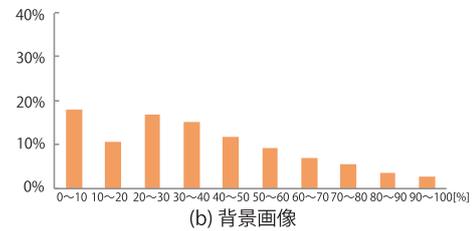
次に、人画像と誤検出した背景画像の直線の長さの割合を比較する。図 8 に人画像と誤検出した背景画像における直線の長さの割合のグラフを示す。これは、ウィンドウの縦の画素数に対し、観測された直線の長さの割合を表す。人画像においては、肩部から脚部にかけての形状が直線として観測できるため、ウィンドウの縦の画素数に対して 5 割程度の長さを持つ。一方、背景画像は人画像とは異なり、短い線分や長い線分が占める割合も多かった。人工物の画像は縦の長い直線を多く含み、自然物の画像は複雑な細かなテクスチャが多いため、短い直線を多く含む傾向がある。

2.3 考察

以上の調査結果を整理する。物体の性質については、建物のように縦の直線が多く観測される背景画像や、テクスチャが複雑な葉っぱや芝を含む背景画像を多く誤検出していることを確認した。直線の長さの割合については、人画像と背景画像では傾向が異なることを確認した。人画像は、肩部から脚部にかけての形状が直線として観測でき、ウィンドウの縦の画素数に対して 5 割程度の直線の長さをもつ。一方、背景画像は人画像とは異なり、短い線分や長い線分が占める割合も多かった。



(a) 人画像



(b) 背景画像

図 8 直線の長さの割合の比較

以上より、人画像と誤検出した背景画像の違いとして、直線の長さの傾向が異なることを確認した。今回の調査で用いた HOG 特徴量は、局所領域における勾配をヒストグラム化した特徴量であるため、両者の違いを捉えることができず誤検出した可能性がある。例えば、ウィンドウを縦断するような直線を含む画像は明らかに背景画像だと言える。しかし、HOG 特徴量は、ある局所領域において、特定の方向の勾配の頻度が多いというように捉え、直線性に関する情報を表現しない。また、同様に背景画像に短い直線が含まれるような背景画像においても、HOG 特徴量ではある局所領域において、特定の方向の勾配の頻度が少ないというように捉える。このように HOG 特徴量では、勾配情報をヒストグラムでまとめるため、直線に関する情報が失われてしまう。これは HOG 特徴量単体では表現能力に限界があることを表し、誤検出を少なくするためには両者の違いを捉えることができる特徴量が必要といえる。

3 HOL 特徴量

2 章にて誤検出した背景画像の傾向を調査し、人画像と誤検出した背景画像には、直線の長さに関して違いがあることを確認した。また、HOG 特徴量では両者の違いを捉えることができないために、誤検出している画像が存在する可能性について述べた。これらの違いを捉えることができる特徴量を用いることで、背景画像の誤検出の抑制が期待できる。本章では、これらの違いを捉えるために、画像から検出した直線に基づく特徴量として Histograms of Oriented Lines(HOL) 特徴量について述べる。

3.1 直線検出

HOL 特徴量を算出するために、まず画像データから直線を検出する。直線検出は古くから研究されてい

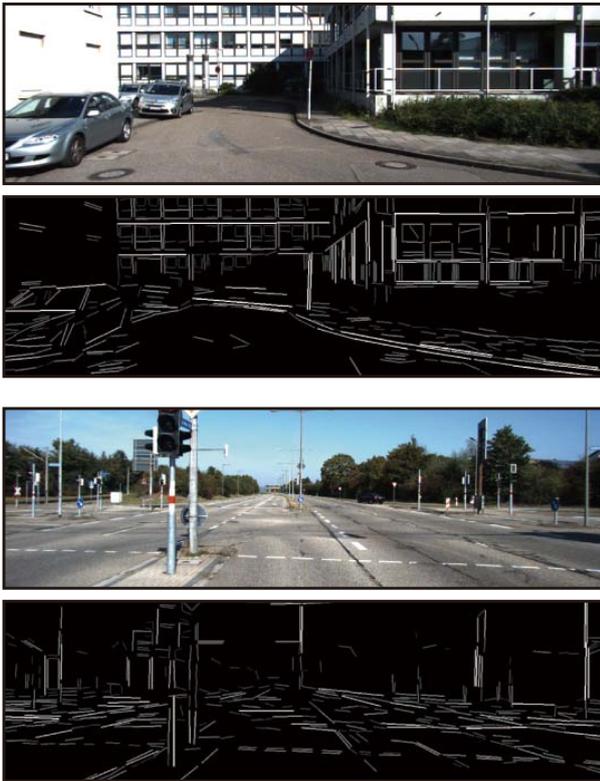


図9 LSDによる直線検出の結果の例

るトピックであり, 様々な手法 [9, 10, 11] が提案されているが, 本研究では Line Segment Detector(LSD)[8]を用いる. LSDは計算コストが少なく, パラメータの調整が比較的簡単, エッジの多い画像に対しても良好な結果が得られる特長を持つ. 図9にLSDを用いた直線検出の結果を示す. 直線の輝度の違いは線分の長さを表す.

3.2 HOL 特徴量の算出

直線検出結果から特徴量を計算する. 人画像と誤検出した背景画像の違いを捉えるために, 提案する特徴量では線分の長さを考慮する.

提案するHOL特徴量は, 人検出に有効なHOG特徴量と似た計算過程で抽出される. まず, 直線検出の結果画像から直線の角度 θ と直線の長さ l を計算する. そして, 式(1)に示すようにHOG特徴量と同様にセル c_i 領域毎に8方向に量子化した直線ヒストグラム $V_{c_i} = \{v_{c_i}(1), v_{c_i}(2), \dots, v_{c_i}(8)\}$ を作成する.

$$v_c(n) = \sum_x \sum_y l'(x, y) \delta[n, \theta'(x, y)] \quad (1)$$

ここで, $\theta'(x, y)$ は $\theta(x, y)$ を8方向に量子化した角度, $\delta[\cdot]$ はクロネッカーのデルタ関数を表す. l' はシグモイド関数により非線形変換した重み付きの長さを表す. これにより, ウィンドウを縦断するような長い直線は大きな重みが与えられ, より人画像と背景画像の直線性の違いを捉えることができる.

図10に人画像と背景画像から抽出されるHOG特徴量とHOL特徴量の違いを示す. ある局所領域だけに着目した際, HOG特徴量の場合では似たような特徴量が得られる. 一方, HOL特徴量の場合では人画像と背景画像から異なる特徴量が得られていることがわかる.

人画像と長い直線を含む背景画像から得られる特徴量は全く異なる値となる. また, 葉や芝などの細かなテクスチャを多く含む画像からは直線が検出されないため, これら背景画像と人画像の特徴量は異なる値となる.

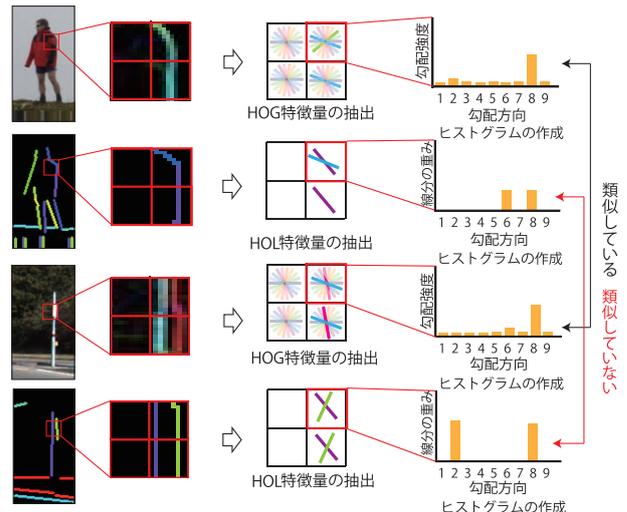


図10 HOG特徴量とHOL特徴量の抽出の流れ

4 評価実験

提案手法であるHOL特徴量の有効性を評価するために評価実験を行う.

4.1 実験概要

提案手法を評価するために, 以下の特徴量の検出性能を比較する.

- 従来法: HOG特徴量
- 提案手法: HOG特徴量 + HOL特徴量

HOG特徴量のパラメータはセルを 8×8 ピクセル, ブロックサイズを 2×2 セル, 量子化数を9とし, HOL特徴量のパラメータはセルを 16×16 ピクセル, 量子化数を9とする. 学習法としては線形SVMを用いる. 評価にはDetection Error Trade-off(DET)カーブを用いる.

4.2 データセット

本実験では, KITTI Vision Benchmark Suiteを用いる. 学習用サンプルを用いて識別器を学習した後, KITTI Vision Benchmark Suiteの学習用背景画像から人検出を行う. 誤検出した背景画像30,341枚のうち,

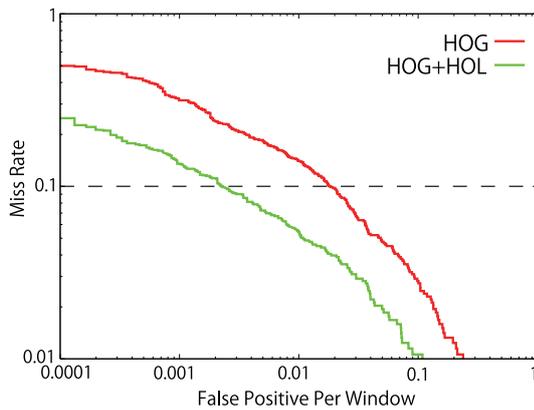


図 11 DET カーブによる結果の比較

10,000 枚を学習用背景画像に加え, 再度識別器を学習する. 評価には, 残りの誤検出画像 20,341 枚を使用する.

4.3 実験結果

HOG 特徴量と HOG+HOL 特徴量での評価実験の結果を図 11 に示す. HOG 特徴量単体よりも, HOL 特徴量を追加した方が性能が良いことがわかる.

図 12 に検出結果の例を示す. HOG 特徴量のみでは縦の直線が人の形状に似ているため, 誤検出しているが, HOL 特徴量を併用することで誤検出を抑制できていることがわかる.

4.4 考察

HOL 特徴量を併用することで検出性能が誤検出を抑制できていることを確認した. ここでは, どのような背景画像の誤検出を抑制できているかについて検証する.

表 1 に FPPW が 0.01 の際の HOG 特徴量のみを用いた識別器と, HOL 特徴量を併用した識別器の誤検出した背景画像の数を示す. まず, HOG 特徴量のみを用いた場合では, ウィンドウの縦の画素数に対する直線の長さの割合が長い背景画像が苦手な傾向があることがわかる. これは, HOG 特徴量が背景画像から観測される直線と人の形状の違いを捉えることができないため, 誤検出した可能性が高い.

一方, HOL 特徴量を併用すると直線の長さに依らず, 誤検出を抑制できていることがわかる. 特に, HOL 特徴量を加えることで, HOG 特徴量では苦手であった長い直線を含む背景画像の誤検出を抑制できていることが確認できる. これは, HOL 特徴量を計算する際, 直線の長さに応じて重み付けし, ヒストグラムに投票しているためだと考えられる.

最後に SVM により線形 SVM を学習した際に得られる重みベクトルに着目する. 図 13 に重みベクトルを可視化した例を示す. ここでは, 重みベクトル \mathbf{w} が正 ($\mathbf{w} > 0$) であればポジティブ, 負 ($\mathbf{w} < 0$) であればネガティブとして画像により表現した. 明るさは重みベクトル \mathbf{w} の絶対値を求めた際の大きさを表している.

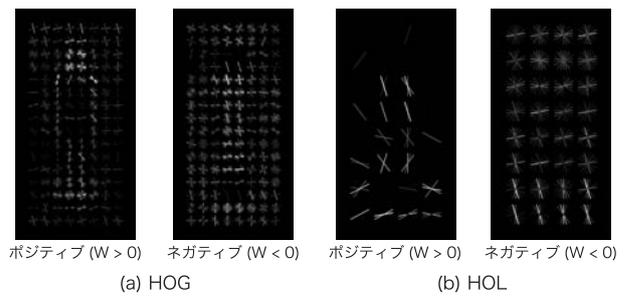


図 13 SVM の重みベクトルの可視化

図 13(a) のポジティブを見ると, 人の輪郭に沿って重みベクトルが大きな値を持っていることがわかる. 図 13(a) のネガティブでは, 人の輪郭の方向とは逆の方向と腹部の重みベクトルが大きな値を持っている.

一方, 図 13(b) のポジティブを見ると, 人の体幹及び足下の形状に沿って重みベクトルが大きな値を持っている. 図 13(b) のネガティブを見ると, 人の体幹及び足下の形状の逆の方向の重みベクトルが大きな値を持っている. また, 人の周囲の重みベクトルが大きな値を持っている. これは, 人が存在する場合には, 人の周囲には直線が観測されないことを表している. 直線検出法である LSD は, 短すぎる直線は検出しないアルゴリズムとなっているため, 背景から観測される直線が人によって隠れるために, 観測できなくなったためだと考えられる.

5 おわりに

本稿では, 誤検出の抑制のために背景から観測される長い直線と人から観測される輪郭の違いを捉えることができる Histograms of Oriented Lines(HOL) 特徴量を提案した. 提案手法は, 画像中から直線を検出し, 局所領域において直線の方向ヒストグラムを作成し, 特徴量として利用する. ヒストグラム作成の際, 直線の長さに応じて重み付けしてヒストグラムに投票することで, 検出ウィンドウを縦断するような長い直線とそれ以外の短い線分の違いを明瞭に捉える. 評価実験により, 直線成分の長さを捉えることができる HOL 特徴量を用いることで誤検出を抑制できたことを確認した. 今後は, より広域にわたる直線成分の関連性を捉えるために HOL 特徴量の共起によ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," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. 886-893, 2005.

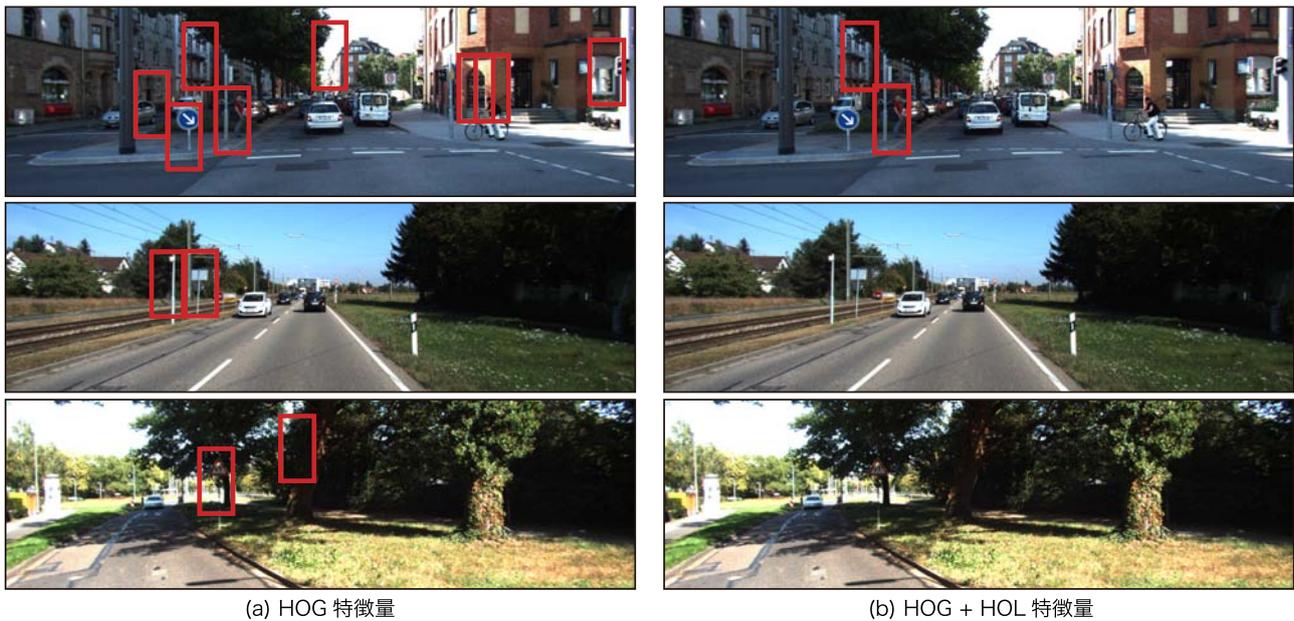


図 12 検出結果の比較

表 1 誤検出した背景サンプル数 [枚]. () は評価サンプルの数に対する割合を表す.

直線の長さの割合 [%]	評価サンプル	HOG	HOG + HOL
0 ~ 30	13,773	3,558 (25.8%)	3,127 (22.7%)
31 ~ 60	10,915	4,815 (44.1%)	2,677 (24.5%)
61 ~ 100	5,653	3,234 (57.2%)	1,419 (25.1%)

- [2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.32, no. 9, pp.1627-1645, 2009.
- [3] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP Human Detector with Partial Occlusion Handling," *IEEE International Conference on Computer Vision*, 2009.
- [4] C. Hou, H. Ai, and S. Lao, "Multiview Pedestrian Detection Based on Vector Boosting," *Asian Conference on Computer Vision*, pp.210-219, 2007.
- [5] T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection," *Pacific-Rim Symposium on Image and Video Technology*, pp.37-47, 2009.
- [6] P. Ott and M. Everingham, "Implicit Color Segmentation Features for Pedestrian and Object Detection," *IEEE International Conference on Computer Vision*, 2009.
- [7] G. Andreas, P. Lenz, and R. Urtasun, "Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.
- [8] R.G. von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall, "LSD: A Fast Line Segment Detector with a False Detection Control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.32, No. 4 pp. 722-732, 2010.
- [9] J. B. Burns, A. R. Hanson, and E.M. Riseman, "Extracting Straight Lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 4, pp. 425-455, 1986.
- [10] A. Etemadi, "Robust Segmentation of Edge Data," *International Conference on Image Processing and its Applications*, pp. 311-314, 1992.
- [11] J. Matas, C. Galambos, and J. Kittler, "Robust Detection of Lines Using the Progressive Probabilistic Hough Transform," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 119-137, 2000.