

HOG 特徴量と差分ステレオによる人物検出

Human Detection Using Subtraction Stereo with HOG Features

○学 有江 誠 (中央大) モロ アレッサンドロ (University of Trieste /JST CREST)
 星川 佑磨 (中央大/JST CREST) 学 生形 徹 (中央大)
 正 寺林 賢司 (中央大/JST CREST) 正 梅田 和昇 (中央大/JST CREST)

Makoto ARIE, Chuo University, arie@sensor.mech.chuo-u.ac.jp
 Alessandro MORO, University of Trieste / CREST, JST, alessandro.moro@stud.units.it
 Yuma HOSHIKAWA, Chuo University / CREST, JST, hoshika@sensor.mech.chuo-u.ac.jp
 Toru UBUKATA, Chuo University, ubukata@sensor.mech.chuo-u.ac.jp
 Kenji TERABAYASHI, Chuo University / CREST, JST, terabayashi@mech.chuo-u.ac.jp
 Kazunori UMEDA, Chuo University / CREST, JST, umeda@mech.chuo-u.ac.jp

In this paper, we propose a fast and stable human detection method using a stereo camera system called “Subtraction Stereo” with a Histograms of Oriented Gradients (HOG) feature. The proposed method performs scanning detection window using the distance of foreground regions from the subtraction image. Fast and stable human detection is achieved by the HOG feature of foreground regions with on Real AdaBoost classifier. And the experimentally demonstrated that our solution compared normal image with subtraction images. In addition, when we computed the HOG feature, we verified the size of an important cell.

Key Words: Stereo Camera, Human Detection, HOG features, Surveillance Camera

1. 序論

近年、監視映像からの不審者検出や人流計測のために、カメラ画像から自動的に人物をリアルタイムで検出することが期待されている。近年では、Graphics Processing Unit(GPU)を用いた高速な画像処理での人物検出法が提案されている[1]。GPUを用いた実装する、画像から得られる特徴量をデバイス間で通信する必要があるが、人物検出でよく用いられる Edge of Orientation Histograms(EOH)特徴量[2]や Histograms of Oriented Gradients(HOG)特徴量[3][4][5]は、特徴量数が多く、莫大な転送時間がかかる。また、GPUを用いない場合には、特徴量算出に大量の計算時間が必要なため、リアルタイムでの人物検出は困難である。これらの特徴量は、入力画像から取得するためにウィンドウを画像全体に対して密に走査するため、背景情報による誤検出と計算時間の増加につながる。また、画像内での人物の距離に応じた見えに対応するために、異なるサイズのウィンドウを複数回走査している。以上より入力画像から HOG 特徴量を算出するために、処理時間がかかるという問題がある。

本手法では、特徴量を算出する際に画面内の情報を前景領域に限定し、ウィンドウの走査範囲及び走査回数を削減することで HOG 特徴量を算出する際の処理時間を削減している。本研究では、ステレオカメラを用いた“差分ステレオ(Subtraction Stereo)”による前景領域を抽出する手法を提案している[6]。本論文では差分ステレオより得られる前景領域と、その領域の三次元座標や高さ、幅の情報を用いて、HOG 特徴量による高速で高精度な人物検出を行う手法を提案する。

2. 差分ステレオの概要

2.1 基本アルゴリズム

差分ステレオの基本アルゴリズムを図 1 に示し、差分ステレオより得られる視差画像を図 2 示す。通常のステレオでは、左右カメラの画像をマッチングすることで示唆画像を得る。これに対し、差分ステレオではまず左右カメラそれぞれで差分処理によって前景領域を抽出し、その後抽出された前景領域をマッチングすることで距離情報を得る。この手法は、

差分で得られた前景領域しか計測できないという欠点がある反面、対応点探索を左右カメラの前景領域に限定することができ、計算量の削減とともに、マッチングをロバスト化することができる。

2.2 前景領域の三次元計測

差分ステレオより得られる視差画像から前景領域を抽出し、各領域の三次元座標と高さ、幅の情報を取得することができる。また、ステレオカメラを用いることでカメラの地面からの高さや仰角がある場合でも、三次元情報を計測することができる。

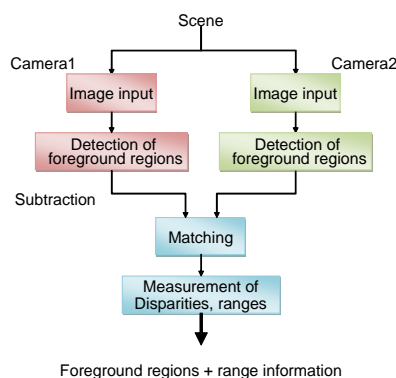
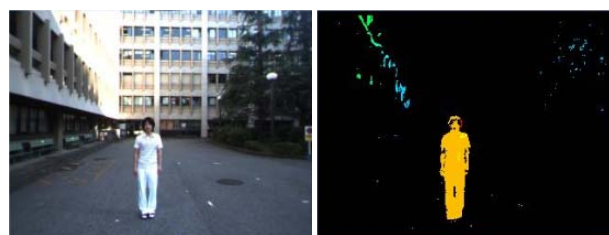


Fig. 1 Flow of Subtraction Stereo



(a) Input image

(b) Output disparity image

Fig. 2 Disparity image obtained by Subtraction Stereo

3. 人物検出

差分ステレオを用いたウィンドウの走査手法でHOG特徴量を算出する、高速で高精度な人物検出の一連の流れを以下に示す。

まず、差分ステレオを用いて前景領域を抽出し、その領域での距離情報に応じて走査範囲と走査するウィンドウのサイズを動的に変化させる。そして、ウィンドウ毎にHOG特徴量を算出し、統計的手法であるReal AdaBoost識別器にかけることで、ウィンドウ内の前景領域の人物判定を行う。

3.1 ウィンドウの走査手法

本研究では、差分ステレオを用いたウィンドウの走査手法とウィンドウサイズの動的変化手法を提案し、HOG特徴量を求めるための計算時間と誤検出の問題を解決する。

一般的な人物検出においてDalalは、入力画像からHOG特徴量を算出するのに、ウィンドウを画像全体に走査し、人物検出を行っている。また、画像内での人物の距離に応じた見えに対応するために、異なるサイズのウィンドウを複数回走査している[3]。この手法では背景領域のような人物ではない領域に対してウィンドウを走査しHOG特徴量を求めるため、大量の計算時間がかかり、誤検出も増加する。この問題に対し、差分ステレオから得られる前景領域に走査領域を絞り、その領域の距離情報よりウィンドウのサイズを動的に変化させることで解決を試みる。

例えば、カメラの高さが1.6m、仰角が 0° での人物(身長1.68m)のカメラからの各距離における大きさ(高さと幅)を図3に示す。図3より人物のカメラからの距離と大きさは反比例の関係にあるため、人物の高さと幅の比例定数 k_h, k_w を算出する。また、カメラの高さと仰角がある場合には、画面内での人物の位置によって、近くでは小さくなり、遠くでは大きくなるというような見え方が異なる。本手法では、人物の高さと幅の比例定数 k_h, k_w と差分ステレオより得られる前景領域の距離情報を用いて、擬似透視投影を仮定した式(1)、(2)よりウィンドウサイズの補正を行う。

$$height = \frac{k_h}{W_Y} (\cos \theta - y \sin \theta) \quad (1)$$

$$width = \frac{k_w}{C_Z} \quad (2)$$

ここで、 W_Y は世界座標系でのカメラから前景領域までの距離、 C_Z はカメラ座標系でのカメラから前景領域までの距離、 θ はカメラの仰角、 y は焦点距離を1に正規化した画像座標である。

以上の手法を用いることで、画面内の前景領域に対してウィンドウを密に走査し、走査回数を1回にすることができる。

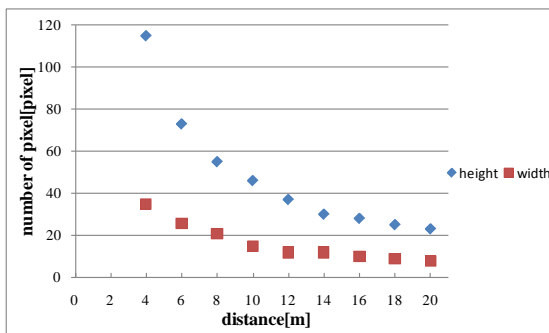


Fig. 3 Relation between distance and size of human region

3.2 HOG 特徴量算出

差分ステレオを用いたウィンドウの走査手法とウィンドウサイズの動的変化手法に基づく走査ウィンドウに対してHOG特徴量を算出する。HOG特徴量とは、局所領域における輝度勾配の方向と強度をヒストグラム化した特徴量であり、物体の大きな形状を表現することが可能なものである。そのため、HOG特徴量は高精度な物体や人物検出が可能であり、勾配情報を取得するための優秀な特徴量の1つである[3]。しかしながら、物体や人物検出を行うには画像全体から大量の特徴量を算出しなければならないため大量な計算時間がかかるという問題がある。以下にHOG特徴量算出アルゴリズムを述べる。

まず、注目画素と隣接した画素である左右、上下それぞれで差分を行うことで式(3)、(4)より輝度勾配の強度と方向を算出する。

$$magnitude(x, y) = \sqrt{I_x^2(x, y) + I_y^2(x, y)} \quad (3)$$

$$orientation(x, y) = \tan^{-1} (I_y^2(x, y) / I_x^2(x, y)) \quad (4)$$

ここで、横方向の勾配の $I_x^2(x, y)$ 、縦方向の勾配の $I_y^2(x, y)$ は $[-1, 0, 1]$ のフィルタを用いて算出したものである。そして、ウィンドウ内を格子状にし、各格子をセルとする。セル内のピクセル数は、提案するウィンドウサイズの動的変化手法を用いるため各ウィンドウごとで異なるが、それぞれのセルごとで勾配方向数を横軸、勾配強度を縦軸としたヒストグラムを作成する。本研究では、 $0^\circ \sim 180^\circ$ を 20° ずつに分割するため、9方向の勾配方向のヒストグラムとなる。

最後に、各セルで作成した輝度の勾配方向ヒストグラムを 3×3 セルを1つのブロックとして式(5)より正規化を行うことでウィンドウ内のHOG特徴量を算出する。

$$f = \frac{V}{\sqrt{\|V\|^2 + \varepsilon^2}} \quad (5)$$

ここで、 V は正規化前のHOG特徴量、 ε は正規化定数である。

提案するウィンドウの走査手法を用いるため、各ウィンドウの大きさは異なるが、セルとブロックの個数は同じである。

3.3 Real AdaBoost 識別器構築

本研究では、Real AdaBoost識別器を用いて人物検出を行う[7]。本手法は、Dalalが用いているSVMよりも計算量の少ないReal AdaBoost識別器を用いている[3]。

Boostingは、弱識別器を線形結合し強識別器を構築する手法である。弱識別器とは、人物と認識できるが人物以外の物も認識する可能性が高い複数の識別器である。以下にBoostingの詳細な学習アルゴリズムを述べる。まず、検出対象である人物(Positiveクラス)と非検出対象である車や木の葉の揺れ、照明環境の影響で前景領域と認識された横断歩道(Negativeクラス)などの学習用画像 $\{(x_1, y_1), \dots, (x_N, y_N)\}$ をN個用意する。ここで、PositiveクラスかNegativeクラスの2つのラベル $y_i \in \{+1, -1\}$ を用意する。そして、学習用画像の重み $\{D_t\}_{t=1 \dots T}$ (T は反復させる学習回数)を設け、初期値の重みを D_1 は $1/N$ とする。弱識別器 $\{h_m\}_{m=1 \dots M}$ (M は弱識別器の番号)は $D_t(i)$ に

より作成される確率密度分布 W_p^j (j は確率密度分布のBIN, p は学習用画像のラベル) から構築される. そして, 式(6)が最小になるような h_t を選択する. $D_t(i)$ を更新し正規化することで T 個の識別器を得ることができる.

$$Z_m = 2 \sum_j \sqrt{W_+^j W_-^j} \quad (6)$$

最後に, これらの識別器を以下の式(7)より線形結合することで強識別器 $H(x)$ を構築する.

$$H(x) = \text{sign} \left(\sum_{t=1}^T h_t(x) \right) \quad (7)$$

$H(x)$ が正ならば人物, 負ならば人物以外と識別する. 以上の強識別器構築の一連の流れを図4に示す.

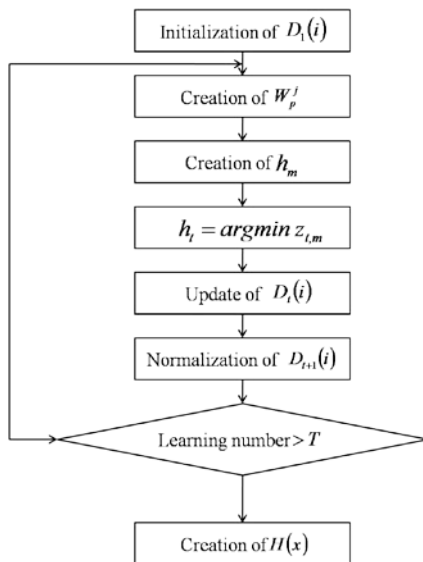


Fig. 4 Real AdaBoost flow

4. 実験

提案する HOG 特徴量と差分ステレオによる人物検出手法の有効性を検証するため, 二つの精度評価実験を行った. 一つ目は, 誤検出と処理時間の違いを比較するため, 通常画像と差分画像を用いた際の人物検出精度の検証を行い, 二つ目は, 人物検出精度が HOG 特徴量を求める際のウィンドウ内でのセルの個数に依存するため, セルの個数を変え人物検出精度の検証を行った. 人物検出結果の検証には, 以下の 4 つの評価指標を用いた.

- True Positive : 真値の人数の内, システムが検出した人数
- False Positive : システムが人物ではないものを人物として検出した数
- False Negative : システムが人物を検出できなかった数
- Speed : 処理速度

検証には, ステレオカメラから得られる差分画像より評価用画像 1000 フレームを用いた. 画像サイズは 320×240 である. また, Real AdaBoost 識別器構築用の学習用画像は Fig. 6 のような人物である positive クラスと人物でない negative クラスそ

れぞれ 2021, 2231 枚用いた. 一つ目の実験では, 通常画像で人物検出精度を検証するため, 通常画像より評価用画像と学習用画像を同じ枚数用いた. ステレオカメラは Bumblebee2(Point Grey Research. カラー, $f=3.8\text{mm}$, $b=120\text{mm}$, VGA, 48fps), PC は DELL XPS420, Core2 Duo 3.00GHz を使用した.

4.1 通常画像と差分画像での人物検出精度の比較

通常画像で人物検出を行う際に様々なサイズのウィンドウを複数回走査するが, 通常画像(1)では 30×60 , 通常画像(2)では 60×120 でウィンドウのサイズを固定し走査回数を 1 回にしている. 差分画像では, 3.1 節で述べた提案する前景領域とその領域の距離情報に応じたウィンドウの走査手法を用いている.

通常画像と差分画像での人物検出精度の結果を表 1 に示し, 人物検出結果を図 5 に示す. 通常画像(1),(2)と差分画像はともに HOG 特徴量を用いた Real AdaBoost 識別器による人物検出を行っているため, True Positive に関しては精度の違いは表われなかった. しかし, 差分画像での人物検出ではウィンドウの走査領域を前景領域に絞り, 大きさを動的に変化させたため通常画像のものより False Positive と Speed の性能が向上している. また, 画像 1 枚当たりでのウィンドウの数が大幅に減り, HOG 特徴量の数が少なくなるため, リアルタイムでの人物検出が可能だと考えられる. 人物を検出できなかった原因は, 人物同士の重なりによって生じるオクルージョンによるものであると考えられる.

Table 1 Experimental Results

	True positive rate (%)	False positive rate (%)	False negative rate (%)	Speed (ms)
Normal image (1)	87.2	3.7	12.7	158.3
Normal image (2)	79.4	3.7	20.5	67.6
Subtraction image	87.5	0.8	12.4	42.8

4.2 異なるパラメータでの比較

人物検出精度は, HOG 特徴量を求める際のウィンドウ内でのセルの個数に依存する. そこで本節では, ウィンドウ内でのセルの個数を人物のカメラまでの距離の関係より 3×6 , 6×12 , 8×16 の 3 パターンに分け人物検出精度の比較を行った. 検証する際のパラメータは, 4.1 節と同様である.

セルの個数を 3 パターンに分けたときの人物検出精度の結果を表 2 に示す. セルの個数を増やすことで良いヒストグラムを作成することができ, 高い精度の人物検出を行うことができる. しかし, ブロックで特徴量を正規化した後に HOG 特徴量が他のものよりも多くなるため, Speed の性能が低下している. 逆にセルの個数が少ないときは, True Positive と False Positive の性能が低下している. これらの結果より, 4 つの評価指標を考慮すると, ウィンドウ内でのセルの適した個数は 6×12 だと考えられる.

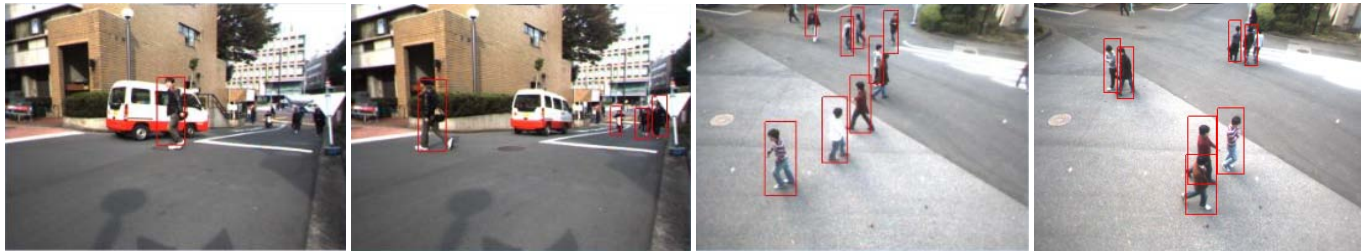


Fig. 5 Sample dense detections on images by the HOG features with proposing method

Table 2 Experimental Results

	True positive rate (%)	False positive rate (%)	False negative rate (%)	Speed (ms)
Number of cells 3×6	82.7	1.0	17.2	40.5
Number of cells 6×12	87.5	0.8	12.4	42.8
Number of cells 8×16	88.3	0.8	11.6	49.8

5. 結論と今後の展望

差分ステレオから得られる前景領域に走査領域を絞り、その領域の距離情報よりウィンドウサイズを動的に変化させる手法で HOG 特徴量を算出することにより、高速で高精度な人物検出を提案した。また、提案手法の有効性を通常画像での人物検出精度の比較実験より検証し、HOG 特徴量を求める際の適したウィンドウ内でのセルの個数を検証した。

今後は Fig. 7 に示すようなオクルージョンが生じる複雑なシーンでの人物検出精度の向上のために、歩行時の人物の動きが少ない上半身の HOG 特徴量を活用した手法を検討する。



(a) Positive class



(b) Negative class

Fig. 6 Learning images for Real AdaBoost classifier



Fig. 7 Examled of missed detection

文 献

- [1] L. Zhang and R. Nevatia, "Efficient Scan-Window Based Object Detection using GPGPU," Computer Vision and Pattern Recognition, 2004.
- [2] K. Levi and Y. Weiss, "Learning object detection from a small number of examples: The Importance of good feature," IEEE Computer Vision and Pattern Recognition, vol.2, pp.53-60, 2004.
- [3] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Computer Vision and Pattern Recognition, pp.886-893, 2005.
- [4] F. Suard and A. Broggi, "Pedestrian Detection using Infrared images and Histograms of Oriented Gradients," IEEE Symposium on Intelligent Vehicle, pp.206-212, Jun, 2006.
- [5] Q. Zhu, S. Avidan, M. The and K. Cheng, "Fast Human Detection Using Cascade of Histograms of Oriented Gradients," IEEE Computer Vision and Pattern Recognition, Vol.2, pp.1491-1498, Jun, 2006.
- [6] K. Umeda, et al., "Subtraction Stereo - A Stereo Camera System That Focuses on Moving Regions-," Proc. of SPIE-IS&T Electronic Imaging, vol.7239 Three-Dimensional Imaging Metrology, 723908, 2009.
- [7] R.E. Schapite, Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Mach. Learn, no.37, pp.297-339, 1999.