

HOG 特徴量と差分ステレオによる複数識別器を用いた人物検出

中央大学/JST CREST ○有江誠, モロアレッサンドロ, 星川佑磨, 生形徹, 寺林賢司, 梅田和昇

Human Detection Using Multiple Classifiers Based on Subtraction Stereo with HOG Features

Chuo Univ. / JST CREST Makoto ARIE, Alessandro MORO, Yuma HOSHIKAWA,
Toru UBUKATA, Kenji TERABAYASHI, Kazunori UMEDA

It is important to detect humans from an image automatically for the measurement of human flow with a surveillance camera. We have proposed a stereo vision method called "Subtraction Stereo" which focuses distant measurement on foreground regions. We construct a human detection method with little incorrect detection using multiple detectors. A fast and stable human detection is achieved using the subtraction stereo with Histograms of Oriented Gradients (HOG) features. The effectiveness of our method is verified by experiments.

1. 序論

監視カメラからの人流計測のために、カメラ画像から自動的に人物を検出することが重要である。これまでに我々は、前景領域のみに距離計測を限定する差分ステレオを提案している[1]。また、人物の大まかな形を表現することが可能な Histograms of Oriented Gradients (HOG) 特徴量[2]を差分ステレオより得られる前景領域のみから算出し、前景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高速なウィンドウの走査手法を提案している[3]。本研究では、HOG 特徴量を用いた識別器を複数構築し、誤検出の少ない人物検出手法を行うことを目的とする。

2. 差分ステレオの概要

差分ステレオの基本アルゴリズムを Fig.1 に示す。通常のステレオでは、左右のカメラの画像をマッチングすることで視差画像を得る。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、その後に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。

3. 人物検出

HOG 特徴量と差分ステレオを用いた人物検出の一連の流れを以下に示す。

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領域と走査するウィンドウの大きさを動的に変化させる。ウィンドウ毎に HOG 特徴量を算出する。また、統計的手法である RealAdaBoost 識別器を全身と上半身の 2 つに対して構築し、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

Dalal 等は、HOG 特徴量を算出するウィンドウを様々なサイズで走査し、画像全体から人物検出を行っている[2]。しかし、この手法では、大量の計算時間が必要となりリアルタイムでの人物検出は難しく、誤検出も多くなる。

本研究では、差分ステレオより得られる前景領域のみに対し、ウィンドウを走査する。また、カメラから前景領域までの距離を三角測量の原理より算出し、距離値に応じてウィンドウのサイズを決定する。ウィンドウの高さと幅は、擬似透視投影を仮定した以下の式より求める。

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

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

ここで、 k_h, k_w は人物の高さと幅の定数、 W_y は世界座標系での距離、 C_z はカメラ座標系での距離、 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。前景領域に対しウィンドウサイズを動的に変化させることでウィンドウの走査回数を 1 回にすることができる。以上より、誤検出の少ないリアルタイムでの人物検出が可能になる。

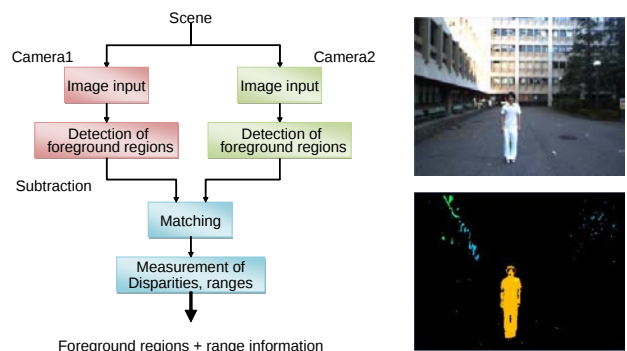


Fig.1 Schema of subtraction stereo

3. 2 HOG 特徴量の算出

局所特徴量である HOG 特徴量は、局所領域における輝度の勾配方向をヒストグラム化した特徴量であり、ある一定領域に対する特徴量の記述を行う。そのため、大まかな物体形状を表現可能である。また、照明の変化を受けにくい。算出方法は[2]に示されている。

3. 3 複数識別器の構築

Boosting は、次式に示すように、検出対象であるかを出力する性能の良くない T 個の弱識別器 $h_t(x)$ を線形結合することで、検出性能の良い強識別器 $H(x)$ を学習する。

$$H(x) = \text{sign} \left(\sum_{t=1}^T h_t(x) \right) \quad (3)$$

出力値は Boosting の手法により異なるが、本研究では Real AdaBoost[4]を用いる。

本論文では、全身と上半身の識別器を Boosting により構築し、組み合わせた誤検出の少ない人物検出手法を提案する。全身識別器のみでは、ドアや建物のような輝度勾配の直線成分を含む特徴量を人物と認識する誤検出が増加する。そこで、人物の頭や肩といった特徴的な形を学習させた上半身識別器を用いて誤検出を減少させる。

HOG 特徴量を用いた際の Boosting による学習の流れは次の通りである。まず検出対象と非検出対象の学習サンプルを多数用意し、各サンプルに重みを設定する。初期時はすべて同じ重みとする。検出対象の学習サンプルは Fig.2 のような全身と上半身の画像をそれぞれ用意する。そして、サンプルの中から人物検出に最も有効な HOG 特徴量を弱識別器として選択する。人物として正しく識別できるサンプルの重みは小さくし、誤って識別したサンプルの重みを大きくする。これらの処理を繰り返し行い、強識別器を学習する。このように、Boosting を用いることで多数の特徴量候補の中から人物検出に有効な特徴量を学習することができる。この学習方法により、全身と上半身の識別器をそれぞれ構築する。詳細な Real AdaBoost の学習アルゴリズムは[4]に示されている。

4. 複数識別器を用いた人物検出実験

3.3 節で示した複数識別器を用いて人物検出の有効性を検証するため、全身識別器、上半身識別器、全身と上半身識別器の組み合わせの3つのパターンで1000フレーム分に対して検出精度評価実験を行った。Real AdaBoost 識別器用の学習用画像は人物である positive クラスが全身と上半身画像をそれぞれ17,394枚、人物でない negative クラスを21,890枚である。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。それぞれの識別器での検出率(TP), 誤検出率(FP), 適合率($TP/(TP+FP)$)を比較したものを Table 1 に示す。また、出力画像の例を Fig.3 に示す。

Table 1 より、全身と上半身の識別器を組み合わせることで、FP が減少しているのが分かる。これは、全身識別器でしばしば検出される、輝度勾配の直線方向を含む特徴量での誤検出が上半身識別器では検出されないからである。そのため、全

身と上半身識別器を組み合わせることで、全身識別器と比較して検出率は低くなったが、適合率が高くなっているのが分かる。また、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。

5. 結論と今後の展望

差分ステレオと HOG 特徴量による全身と上半身の識別器を構築し、組み合わせることで誤検出の少ない人物検出手法を提案した。また、提案手法の有効性をそれぞれの識別器と比較することで検証した。

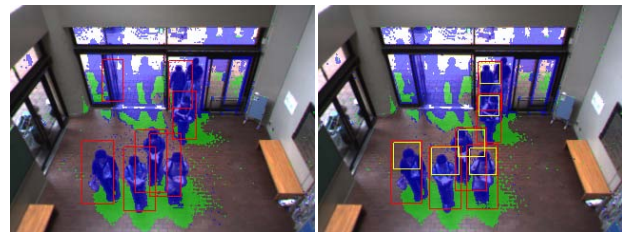
今後は人物が多く複雑なシーンで人物検出を行うために、全身と上半身の識別器の人物検出精度を向上させる。



(a) Total body images (b) Upper body images
Fig.2 Sample of the human images

Table 1 Experimental results

	TP (%)	FP (%)	Precision (%)
Total body	78.5	3.2	96.1
Upper body	86.2	14.4	85.7
Total + Upper	77.8	0.9	99.0



(a) Total body detector (b) Total and Upper body detector

Fig.3 Human detection results

参考文献

- [1] 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.
- [2] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", *International Conference on Computer Vision and Pattern Recognition, Vol.2*, pp 886-893, June 2005.
- [3] 有江誠, モロアレッサンドロ, 星川佑磨, 生形徹, 寺林賢司, 梅田和昇: "HOG 特徴量と差分ステレオによる人物検出", 日本機械学会ロボティクス・メカトロニクス講演会 2010 講演会論文集, 2P1-D17, 2010.
- [4] R.E. Schapire, Y. Singer, "Improved boosting algorithms using confidence-rated predictions", *Mach. Learn.*, no.37, pp.297-336, 1999.