

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*

有江 誠** モロアレサンドロ*** 寺林賢司*** 梅田和昇***

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

Makoto ARIE, Alessandro MORO, Kenji TERABAYASHI and Kazunori UMEDA

In this paper, we propose a fast and stable human detection based on “Subtraction Stereo” which can measure distance information of foreground regions. Scanning an input image by detection windows is controlled about their window sizes and number using the distance information obtained from subtraction stereo. This control can skip a large number of detection windows and leads to reduce the computational time and false detection for fast and stable human detection. Additionally, we construct a human detection method with little incorrect detection using multiple detectors. The effectiveness of our method is verified by experiments.

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。しかし、人物の様々な外観やオクルージョンの問題から人物検出は非常に困難である。人物検出精度の高い手法の一つとして、N. Dalal and B. Triggs¹⁾によって提案された HOG 特徴量に基づく人物検出がある。HOG 特徴量はエッジ情報を用いるため照明変化に頑強である。また、オブジェクト検出のための様々な特徴量は、P. Viola and J. Jones²⁾, K. Levi and Y. Weiss³⁾, and B. Wu and R. Nevatia⁴⁾によって提案されている。

N. Dalal and B. Triggs¹⁾では、人物検出を行う際に検出ウィンドウを画像全体に対して密に走査するため、多くの計算時間を必要とする。また、画像内での人物の大きさは、カメラから人物までの距離と画像内の人物の位置によって異なるため、様々なサイズの検出ウィンドウを複数回走査しなければならない。よって、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。

人物検出全体で多くの計算時間を必要とするのは、(1) HOG 特徴量の算出、(2) 入力画像全体に検出ウィンドウを走査、(3) 異なるサイズの検出ウィンドウを複数回走査の過程である。(2) 検出ウィンドウの画像全体への走査は、計算コストの問題に加え、検出ウィンドウの数が増えることで人物に形状の似た背景領域での誤検出が増加し、検出精度の低下をまねく。処理速度と検出精度を向上させるためには、背景領域を走査する検出ウィンドウを削減する必要がある。また、(3) 大きさを変えた検出ウィンドウの複数回走査は、画像中における人物の大きさが分からないために必要となる過程で、遠近距離の人物が同時に存在する監視カメラ用途などでは様々な大きさを試みる必要があり、処理速度の低下を引き起こす。検出ウィンドウの大きさを予め

知ることができれば、計算量の削減が期待できる。

そこで本論文では、前景領域の距離情報を少ない計算量で取得することのできる差分ステレオ⁵⁾を用いて、検出ウィンドウの効率的な走査方法を提案する。差分ステレオにより得られる距離情報に基づき、画像中の各人物に適した検出ウィンドウの大きさを決定する。

ステレオ処理により得られた距離情報を活かし、限られた領域と検出ウィンドウのサイズのみを走査する手法は鈴木⁶⁾によって提案されている。しかし、カメラの地面からの高さや仰角によって画像内での人物の見え方が異なる為、検出ウィンドウの高さと幅を補正する透視投影モデルを仮定する必要がある。また、TOF カメラより得られた距離情報を用いた効率のよい検出ウィンドウの走査手法は、池村⁷⁾によって提案されている。しかし、人物検出とは関係のない背景領域も検出ウィンドウを走査する為、処理速度の低下や誤検出の増加を引き起こす可能性がある。

そこで本論文では、擬似透視投影モデルを仮定して検出ウィンドウの高さと幅を補正する。また、検出ウィンドウを走査する領域を背景差分法により前景領域に絞ることで、処理速度の増加と誤検出の減少が期待できる。人物が存在する可能性のある前景領域にのみ一度走査することで、高速かつ高精度な人物検出を目指す。

また、人物の全身の学習サンプルから構築される全身識別器では、下半身の特徴量の影響で誤検出が多くなる傾向がある。そこで人物検出精度の向上のために、人物の特徴(頭、肩)を表現しやすい上半身の学習サンプルを用いた上半身識別器を組み合わせる。そして、組み合わせた新たな識別器を複数識別器と定義する。全身と上半身の識別器は、Real AdaBoost²⁾によって構築される。一方、人物検出の最近の研究では、単眼視^{1), 8)~11)}、ステレオ視^{12)~16)}および LIDAR センサ¹⁷⁾を用いたものがある。人物検出の様々なアプローチは、¹⁸⁾に示されている。HOG 特徴量は、SIFT 特徴量¹⁹⁾と同様に勾配方向ベースの特徴抽出法であり、画像内の人物の特徴を表すことができる。HOG 特徴量は、SIFT とは異なり、画像を格子状に分割したものをセルと呼び、複数のセルをブロックとしてセル単位で重複させながら走査することで画像内のエッジ情報を密に表

* 原稿受付 平成 23 年 10 月 21 日

** 中央大学大学院 (東京都文京区春日 1-13-27)

*** 中央大学

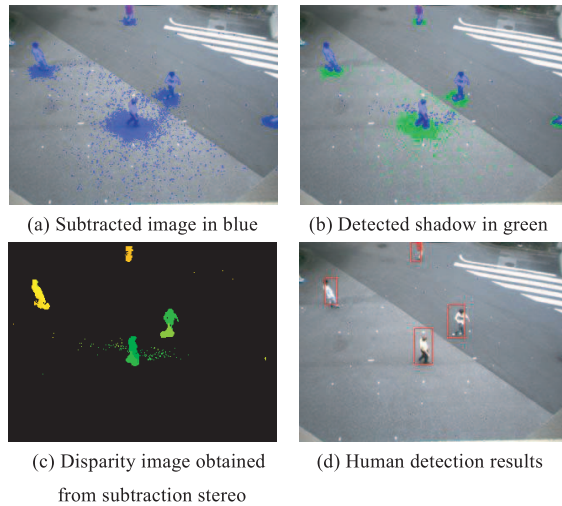


Fig.1 Overview of the proposed human detection based on subtraction stereo

現できるものである。Zhu²⁰⁾は、識別器を複数のステージに分けて検出の効率を向上させるために、カスケード型の識別器を用いている。

リアルタイムの人物検出手法として、Shashua⁸⁾によって提案された単眼の車載カメラから人物検出を行うものがある。これは、人物の検出ウィンドウ候補を迅速に検出する手法を用いている。その検出ウィンドウ候補は、2段階の識別器を用いることで人物か否かを識別する。

Ess¹⁴⁾⁻¹⁶⁾は、外観に基づくオブジェクト検出、ステレオ視より得られる三次元情報と地面検出、拡張カルマンフィルタを用いたオブジェクト追跡を統合した人物検出と追跡の精度の高い人流計測手法を提案している。人物が多くオクルージョンが生じる環境に対応するために、人物の前後フレームの情報を用いて視覚的に見えない部分の追跡を補足している。しかし、各要素ごとに多くの計算時間を必要とし、加えて三次元情報を用いているため、リアルタイムでの人物検出は困難であると考えられる。

Bajacharya¹³⁾は、非常に混雑した環境で 40m までリアルタイムで人物を検出することが可能なステレオ視での人物検出を提案している。しかし、5Hz での人物検出のため、人物追跡手法と組み合わせたときに追跡精度が低下する可能性がある。我々の提案している人物追跡手法²⁶⁾は、前後フレームの情報を用いるため 10Hz 以上の処理速度を必要とする。

Tuzel¹¹⁾では、共分散行列の特徴量²¹⁾に基づく人物検出を INRIA データセット上で行い、HOG 特徴量¹⁾よりも人物検出の性能が良いこと検証している。しかし、Paisikriangkara²²⁾は、異なるデータセットである DaimlerChrysler の人物のデータセット²³⁾を用いると、共分散行列の特徴量よりも HOG 特徴量を用いた人物検出の方が、精度が高くなることを検証している。

本論文では、HOG 特徴量を用いた人物検出¹⁾を従来手法と位置付け、より高速で誤検出の少ない人物検出手法を提案する。

以下では、まず、2 章で差分ステレオを用いた前景領域の距離計測を説明する。次に、3 章で前景領域の距離情報に基づく検出ウィンドウのサイズおよび数を調整する手法を提案する。そして、

4 章で人物の全身と上半身の識別器を組み合わせた複数識別器の構築手法を提案する。最後に、5 章で提案手法の有効性を確認するために、処理速度と人物検出精度を¹⁾と比較する。

2. 差分ステレオによる前景領域の距離計測

本論文では、差分ステレオを用いた高速で誤検出の少ない検出ウィンドウの走査手法を提案する。人物検出を行うとき、入力画像の前景領域にのみ検出ウィンドウを走査する。そして、前景領域の距離情報に基づいて、検出ウィンドウの大きさを適切に決めることにより、ウィンドウサイズを変えた複数回の走査をせずに、検出精度の向上を狙う。よって、背景領域に対して検出ウィンドウを走査しないため、計算時間と誤検出を減少することができる。また、図 1 (b) に示す人物や物体の影を除去することで人物検出の精度は更に向上する。前景領域の抽出手法と影検出手法を以下に示す。

2.1 差分ステレオ

エッジは人物検出のための多くの情報を含んでいる。そのため、差分ステレオ⁵⁾によって背景情報を除去することで効率的にエッジを算出できる。差分ステレオは背景差分法によって入力画像から前景領域を抽出し、左右カメラの前景領域をマッチングすることで距離情報を得る。また、前景領域の実際の高さや幅を計測することができるという利点がある。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 影検出

影検出²⁴⁾は、前景から人物や物体の影を除去するために使われる。差分ステレオを用いて得られた画像は、影によって影響を受けた雑音を含んでいる。この雑音は人物ではない領域となるため、人物検出精度に影響を及ぼす可能性がある。2 次元画像上の位置 (x, y) のピクセルの輝度を $I(x, y)$ 、そして、背景のピクセルの輝度を $I'(x, y)$ と定義するとき、影の評価のための式を以下に示す。

$$\theta_{(t+1, x, y)} =$$

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

ここで、 θ は影の評価に相当し、閾値によって影か否かを決定する。 Ψ と Λ の関数は、ピクセル間の色の不変性を示す。 α, β, η は、テクスチャの重み、色、輝度である。影検出の結果を図 1 (b) に示す。また、影除去後の差分ステレオを用いた視差画像を図 1 (c) に示す。

3. 人物検出

差分ステレオを用いた前景領域の抽出と距離計測後の HOG 特徴量に基づく人物検出の流れは、(1) 検出ウィンドウのサイズを動的に変化、(2) 検出ウィンドウ内から HOG 特徴量の算出、(3) Real AdaBoost 識別器を用いた人物判定、(4) 人物と認識された検出ウィンドウの統合である。

3.1 検出ウィンドウサイズの動的変化

本論文では、差分ステレオを用いた検出ウィンドウのサイズを前景領域に合わせて動的に変化させる手法を提案する。提案手法を用いることで、HOG 特徴量の算出時間と誤

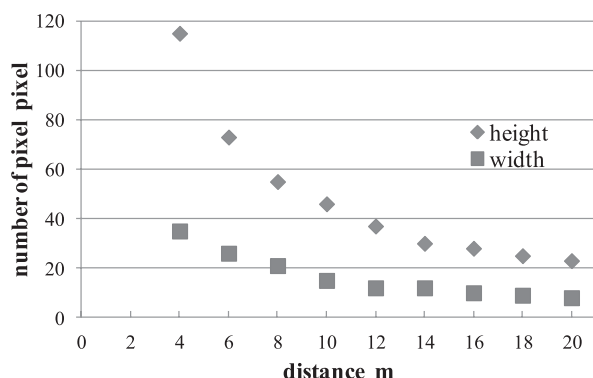


Fig.2 Relation between distance and size of human region

検出が増加する問題を解決することができると考えられる。従来手法¹⁾では、入力画像から HOG 特徴量を算出する際に、検出ウィンドウを画像の端から端まで密に走査する必要がある。よって、計算時間が多くなり、リアルタイムでの人物検出は困難である。また、画像内の背景領域も検出ウィンドウを走査するため、誤検出が増加する。

この計算時間と誤検出の増加は、差分ステレオから得られる前景領域に走査領域を絞ることで解決する。なぜなら、人物の候補である領域にのみ検出ウィンドウを走査でき、計算時間と検出精度の性能が向上すると考えられるからである。

また、画像内の距離による人物の大きさ、位置に対応するため、異なるサイズの検出ウィンドウを複数回走査し、計算時間が増加する問題が生じる。

この問題に対して、画面内の前景領域に走査する検出ウィンドウのサイズを距離やカメラの仰角の情報から動的に求め解決する。まず、走査する検出ウィンドウのサイズを前景領域の距離情報から求める。次に、擬似透視投影を仮定し、検出ウィンドウの高さを補正する。様々な透視投影法があるが、擬似透視投影を仮定している理由としては、弱透視投影は、距離情報のみで検出ウィンドウの高さ補正を行うが、擬似透視投影は、カメラの高さと仰角、画像内の人物の位置を考慮して検出ウィンドウの高さ補正をしているからである。

例えば、カメラの高さが 1.6m、仰角が 0°の時、人物（身長は 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 2 に示す。図 2 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(2)、(3)より検出ウィンドウのサイズの補正を行う。

$$height = \frac{k_h}{W_Y} (\cos \theta - y \sin \theta) \quad (2)$$

$$width = \frac{k_w}{C_Z} \quad (3)$$

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



(a) Before clustering (b) After clustering

Fig.3 Detection windows recognized as a human unified using mean shift clustering

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検出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。

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検出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多大な計算時間がかかるという問題点がある。以下に HOG 特徴量算出アルゴリズムを述べる。

HOG 特徴量を算出する際にガウシアンフィルタによって平滑化を行うことでノイズを軽減し、輝度勾配を算出しやすくする。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(4)、(5)より輝度勾配の方向と強度を算出する。

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

$$magnitude(x, y) = \sqrt{I_x^2(x, y) + I_y^2(x, y)} \quad (5)$$

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、各ウィンドウによって異なるが、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°~180°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の 3×3 セルを 1 つのブロックとして式 (6) 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。

$$f = \frac{V}{\sqrt{\|V\|^2 + \varepsilon^2}} \quad (6)$$

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、セルの個数は 5×10 である。

3.3 検出ウィンドウ統合

人物検出のために画像内の前景領域に対して、検出ウィンドウを走査する。検出ウィンドウ内から HOG 特徴量を算

出し、識別器を用いて人物か否かを判別する。しかし、検出ウィンドウを密に走査するため、図 3 (a) のように 1 人の人物に対して複数のウィンドウが検出される。これらの複数のウィンドウから人物の数と正確な位置を得るために検出後の後処理が必要である。そのため、人物と認識された検出ウィンドウを Mean shift クラスタリング²⁵⁾を行う。Mean shift クラスタリングを行うことによって、図 3 (b) のように画像内の人数と正確な位置を得ることができる。

4. 識別器構築

本論文では、検出ウィンドウ内の前景領域が人物か否かを判別するための識別器の構築に Real AdaBoost²⁾を用いる。しかし、図 4 (b) より人物の下半身には、縦方向の直線成分である輝度勾配を多く含む傾向があることがわかる。そのため、人物の全身の学習サンプルのみで構築した全身識別器では、自動ドアや建物等を誤検出する問題が生じる。そこで、人物の上半身の学習サンプルから構築される上半身識別器を併用し、誤検出が減少する人物判別手法を提案する。

4.1 Real AdaBoost 識別器

本論文では、識別器の学習手法に Real AdaBoost²⁾を用いる。この手法は、高次元の大量の学習サンプルを学習するのに有効で効率的な学習アルゴリズムである。Support Vector Machine (SVM)等の統計学的手法では、学習サンプルから抽出されたすべての特徴量から単一の強力な判別関数を学習するが、Real AdaBoost アルゴリズムでは、学習サンプルに重みを付けて学習することで、学習回数分の弱識別器を構築する。そのため、人物を表現しやすい特徴(頭、肩、腕等)毎の弱識別器を構築できる利点がある。弱識別器は識別精度が良くないが、それぞれの弱識別器を線形結合することで高精度な人物検出が可能な強識別器を形成する。式 (7) を用いて強識別器を構築する。

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

ここで、 $H(x)$ は強識別器、 T は学習回数(弱識別器の個数)、 h_t は弱識別器、 x_i は学習サンプルである。学習サンプルに重みを付けているため、それぞれの弱識別器ごとに人物検出の性能が異なる。よって、人物検出の精度を補完し合って線形結合する強識別器は、学習回数が多いほど人物検出の精度が良くなる。カメラの仰角によって人物の見え方が異なるため、仰角ごとの学習サンプルを数種類用いている。学習サンプルの詳細を 5.1 節に述べる。

4.2 複数識別器構築

本論文では、人物の全身と上半身の学習サンプルを用いた人物検出に有効な複数の識別器を提案する。全身識別器を構築するには、検出対象である人物全身のサンプルである Positive サンプルと車や木の葉の揺れ、照明環境の影響で前景領域と認識された建物をまとめたサンプルである Negative サンプルを用いる。そして、それぞれのサンプルに重みを付けて 4.1 節に示す Real AdaBoost アルゴリズムを用いることで全身識別器を構築する。

人物検出を行う際には、構築した全身識別器を前景領域に対して走査する。しかし全身識別器のみでは、ドアや建物のような輝度勾配の縦方向の直線成分を多く含む物を人物と誤検出しやすい。なぜならば、人物の下半身には図 4 (b) のように縦方向の直線成分である輝度勾配が多く含まれているからである。図 4 (a), (b) でのヒストグラムの縦軸と横軸は、それぞれ輝度勾配の強度と方向を示している。人物の上半身は、図 4 (a) に示すセルで区切ら

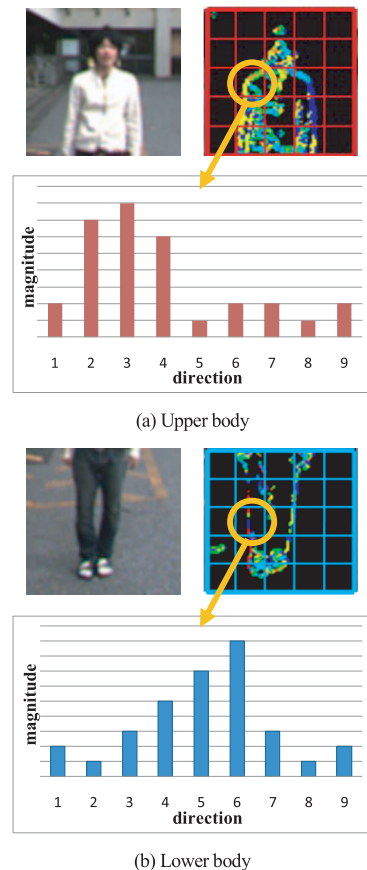


Fig.4 Comparison of the intensity gradients between the upper body and the lower body

れた領域内の輝度勾配から頭や肩を表現することが可能である。そのため、上半身の学習サンプルから識別器を新たに構築する必要がある。上半身識別器は、Positive サンプルに人物の上半身画像を用い、全身識別器で用いたものと同様の Negative サンプルを用いて構築する。上半身のサンプルから HOG 特徴量を算出する際は、セルの個数を 5×5 にし、セル内のピクセル数やブロックの構築は全身の学習サンプルのものと同様である。

5. 実験

提案手法の有効性を検証するため、二つの実験を行った。一つ目の実験では、通常画像と前景領域を抽出した差分画像を用いた際の人物検出精度を比較した。差分ステレオによる検出ウィンドウの削減とウィンドウサイズの動的変化手法を用いる。二つ目の実験では、様々な識別器での人物検出精度を比較した。全身と上半身の学習サンプルから構築した全身識別器、上半身識別器、全身と上半身を走査する複数識別器を用いる。

5.1 実験条件

本論文では、独自に構築したデータセットを用いて提案手法の有効性を検証した。独自のデータセットを用いている理由としては、カメラの仰角により画像内での人物の見え方が異なるからである。また、仰角が大きくなる程、人物全体を映すことが困難であるという問題がある。これらの問題から人物を検出されない場合があるので、仰角ごとに図 5 (a) のような Positive サンプルを用意している。データセットの Positive サンプルの画像の枚数は、カメラの仰角が $0 \sim 50^\circ$ の画像から 10° 毎に 3,500 枚ずつ用いて、計

17,500 枚である。ここで、Positive サンプルは、背景とは分離されていない人物画像を示す。また、図 5 (b) のような Negative サンプルの数は、20,000 枚でありカメラの仰角は考慮していない。すべてのサンプル画像のサイズは 64×128 [pixel] である。

監視カメラ用途を想定しているため、様々なシーンに対応した人物検出を行わなければならない。しかし、INRIA のデータセット¹⁾では、カメラの仰角を考慮していないため、学習サンプルとして用いていない。

実験には、ステレオカメラ Bumblebee2 (Point Grey Research, カラー, f=3.8mm, b=120mm, VGA, 48fps), PC は DELL XPS420, Core2 Duo 3.00GHZ, メモリ 4GB を使用した。

5.2 通常画像と差分画像での人物検出精度の比較

一つ目の実験では、5.1 節で述べたデータセットを用いて、通常画像と差分画像で人物検出を行った際の人物検出精度と処理速度の性能を比較した。本手法では、差分画像は検出ウィンドウの走査領域とウィンドウサイズを決定するのに用い、検出ウィンドウ内から HOG 特徴量を求める際には、通常画像を用いている。それゆえに、通常画像と差分画像で人物検出を行う際は、同じデータセットで構築した識別器を用いる。この実験で用いる識別器は人物の全身の学習サンプルから構築した全身識別器のみである。

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 (TP) : 真値の人数のうち、システムが人物を検出した割合
- False Positive (FP) : システムが人物ではない物を人物として検出する割合
- Precision : システムが検出した物が人物である割合
- Processing Time (PT) : 画像 1 枚にかかる処理時間

通常画像で人物検出を行う際には、検出ウィンドウのサイズを 30×60, 60×120 [pixel] で固定し、それぞれのウィンドウを入力画像に対して走査回数を 1 回にしている。差分画像で人物検出を行う際は、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は、1000 フレーム分に対して人物検出を行った。

通常画像と差分画像での人物検出精度の結果を表 1 に示す。表 1 より差分画像を用いた人物検出では、通常画像での人物検出よりも FP と PT の性能が向上したのが分かる。なぜならば、差分画像を用いることで人物検出には関係のない背景情報を除去できたからである。また、前景領域の距離情報とカメラの仰角を考慮する擬似透視投影を仮定することで、検出ウィンドウのサイズを走査する前景領域のみに対して算出することができるため、検出ウィンドウを複数回走査する必要がない。そして、入力画像に対する検出ウィンドウの数が大幅に減少し、計算量の多い HOG 特徴量の算出が少なくなる。そのため、リアルタイムでの人物検出が可能となる。図 6 (a) のようにカメラの高さが低く、仰角が小さい場合、検出ウィンドウのサイズは人物からカメラまでの距離によって大きく変化する。したがって、処理速度は検出ウィンドウのサイズを前景領域に対して動的に変化させることで減少したと考えられる。逆に、図 6 (b) のようにカメラから人物までの距離にかかわらず、検出ウィンドウのサイズの変化が少ない場合、処理速度は検出ウ

ィンドウを前景領域にのみ走査することで減少したと考えられる。以上から、様々なシーンに対して、処理速度を減少することができる。人物を検出できなかった原因は、図 6 (b) のような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。

5.3 様々な識別器での人物検出精度の比較

二つ目の実験では、一つ目の実験と同様に 5.1 節で述べたデータセットを用いて、人物の全身の学習サンプルを用いた全身識別器と上半身の学習サンプルを用いた上半身識別器とそれらの識別器の両方を走査する複数識別器の人物検出精度を比較した。ここで、全身識別器を構築するための人物の全身の学習サンプルは、カメラの仰角ごとに分けている。上半身識別器では、カメラの仰角ごとの識別器は構築せずに、17,500 枚の上半身の学習サンプル画像から上半身識別器を構築している。また、上半身の学習サンプルの画像のサイズは 64×64 pixel である。

人物検出結果の検証には、TP, FP, Precision の評価指標を用いた。人物検出に用いる画像は、差分画像である。全身識別器と上半身識別器、複数識別器は共に、1000 フレーム分で人物検出を行った。複数識別器を用いた人物検出の結果を図 7 に示す。

全身識別器と上半身識別器、複数識別器での人物検出の結果を表 2 に示す。表 2 より複数識別器は、TP の性能で他の識別機よりも性能が劣るものの、本来の目的である FP の性能を大幅に向上できていることが分かる。よって、Precision の性能も向上した。全身識別器のみでは、縦方向の直線成分の輝度勾配を多く含む物を検出する傾向があるため、図 6 (b) のようにドアや外壁を人物として誤検出してしまい FP の性能が低下したと考えられる。また、上半身識別器のみでは、TP の性能は高いが、FP の性能が著しく低下してしまう。全身識別器と上半身識別器の誤検出の傾向が異なるため、これらの識別器を組み合わせた複数識別器を用いることで、人物検出精度の高い人物検出を行うことが可能となった。人物を検出できなかった原因は一つ目の実験と同様に人物同士の重なりによって生じるオクルージョンであると考えられる。

Table 1 Performance comparison between the proposal and reference [1]

	TP %	FP %	Precision %	PT ms
Reference [1] (30×60)	77.2	10.5	88.0	423.4
Reference [1] (60×120)	71.2	7.8	90.1	223.8
Proposal with subtraction stereo	78.5	3.2	96.1	58.3

Table 2 Performance comparison between the various classifiers

	TP %	FP %	Precision %
Full body classifier	78.5	3.2	96.1
Upper body classifier	86.5	15.2	85.1
Multiple classifier	77.9	0.8	99.1

6. 結 論

本論文では、HOG 特徴量と差分ステレオを用いた高速で高精度な人物検出手法を提案した。差分ステレオより抽出される前景領域にのみ検出ウィンドウを走査することで、人物検出精度と処理速度が向上した。また、前景領域の距離情報を用いて、検出ウィンドウの大きさを画像中の人物サイズに適切に合わせることで、大きさを変えた複数回の走査をなくし、処理速度が向上した。そして、人物の全身識別器と上半身識別器を組み合わせた複数識別器を用いることで、人物検出精度が更に向上した。

しかし、人物同士の重なりによって生じるオクルージョンや人物からカメラまでの距離が遠いときに人物検出の精度が低下するという問題が生じるため、改善が必要である。また、ロバストな人物追跡手法²⁶⁾と統合した高精度な人流計測を行う予定である。



Fig.5 Sample of training image

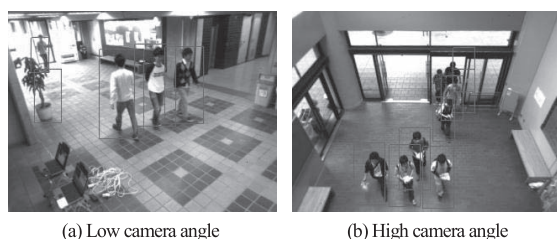


Fig.6 Example of false detection using full body detector

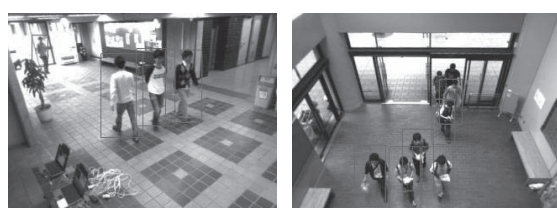


Fig.7 Example of human detection using multiple detectors

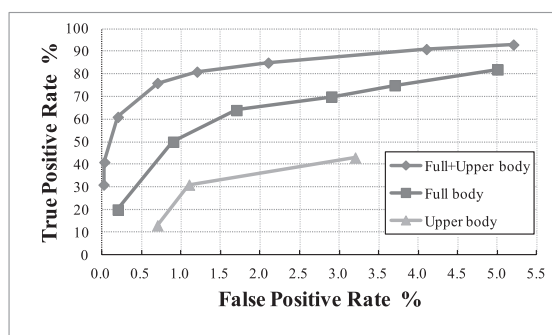


Fig.8 Performance comparison of each classifier

参 考 文 献

- 1) N. Dalal, B. Triggs : Histograms of oriented gradients for human detection, International Conference on Computer Vision and Pattern Recognition, **2**, (2005) 886.
- 2) P. Viola, J. Jones : Rapid object detection using a boosted cascade of simple features, International Conference on Computer Vision and Pattern Recognition, **1**, (2001) 511.
- 3) K. Levi, Y. Weiss : Learning object detection from a small number of example: The importance of good feature, International Conference on Computer Vision and Pattern Recognition, **2**, (2004) 53.
- 4) B. Wu, R. Nevatia : Detection of multiple, partially occluded human in a single image by bayesian combination of edgelet part detectors, International Conference on Computer Vision, **1**, (2005) 90.
- 5) 梅田和昇, 寺林賢司, 橋本優希, 中西達也, 入江耕太 : 差分ステレオ運動領域に注目したステレオ視—の提案, 精密工学会誌, **76**, **1**, (2010) 128.
- 6) 鈴木一生, 呉海元, 和田俊和 : ステレオカメラを用いた顔領域の高速化, 情報処理学会研究報告 CVIM, (2008) 107.
- 7) 池村翔, 藤吉弘亘, “距離情報に基づく局所特徴量によるリアルタイム人検出”, 電子情報通信学会論文誌, **J93-D**, **3**, (2010) 355.
- 8) A. Shashua, Y. Gbalyahu, and G. Hayun. : Pedestrian detection for driver assistance systems: Single-frame classification and system level performance, IEEE Intelligent Vehicle Symposium, (2004) 1.
- 9) P. Viola, M. Jones, and D. Snow : Detection pedestrian using patterns of motion and appearance, International Conference on Computer Vision, (2003) 734.
- 10) P. Sabzmeydani, G. Mori : Detecting pedestrians by learning shapelet features, International Conference on Computer Vision and Pattern Recognition, (2007) 1.
- 11) O. Tuzel, F. Porinki, and P. Meer : Human detection via classification on riemannian manifolds, International Conference on Computer Vision and Pattern Recognition, (2007) 1.
- 12) D. M. Gavrila, S. Munder : Multi-cue pedestrian detection and tracking from a moving vehicle, IJCV, **73**, **1**, (2007) 41.
- 13) M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. H. Matthies : Results from a real-time stereo-based pedestrian detection system on a moving vehicle, in IEEE Workshop on People Detection and Tracking at ICRA, (2009).
- 14) A. Ess, B. Leibe, and L. Van. Gool : Depth and appearance for mobile scene analysis, International Conference on Computer Vision, (2007) 1.
- 15) A. Ess, B. Leibe, K. Schindler, and L. V. Gool : A mobile vision system for robust multi-person tracking, International Conference on Computer Vision and Pattern Recognition, (2008) 734.
- 16) A. Ess, B. Leibe, K. Schindler, and L. Van. Gool : Moving obstacle detection in highly dynamic scenes, in Proceedings of ICRA, (2009) 56.
- 17) K. Fuerstenberg, K. Dietmayer, and V. Willhoef : Pedestrian recognition in urban traffic using a vehicle based multilayer laserscanner, IEEE Intelligent Vehicle Symposium, **1**, (2002) 31.
- 18) P. Dollar, C. Wojek, B. Schiele, and P. Perona : Pedestrian detection: A benchmark, International Conference on Computer Vision and Pattern Recognition, (2009) 304.
- 19) D. G. Lowe : Distinctive image feature from scale-invariant key-points, IJCV, **60**, **2**, (2004) 91.
- 20) Q. Zhu, S. Avidan, M. Yeh, and K. Cheng : Fast human detection using a cascade of histograms of oriented gradients, International Conference on Computer Vision and Pattern Recognition, (2006) 1491.
- 21) O. Tuzel, F. Porinki, and P. Meer : Region covariance: A fast descriptor for detection and classification, International Conference on Computer Vision and Pattern Recognition, **3952**, (2006) 589.
- 22) S. Paisitkriangkrai, C. Shen, and J. Zhang : An experimental study on pedestrian classification using local features, In IEEE Inter. Symp. on Circuit and System, (2008) 2741.
- 23) S. Munder, D. Gavrila : An experimental study on pedestrian classification, Trans. PAMI, **28**, **11**, (2006) 1863.
- 24) A. Moro, et al. : Auto-adaptive threshold and shadow detection approaches for pedestrian detection, In Proc. AWSVCI, **1371**, (2009) 307.
- 25) D. Comaniciu, P. Meer : Mean Shift Analysis and Applications, International Conference on Computer Vision, **2**, (1999) 1197.
- 26) Y. Hoshikawa, et al. : Human Tracking Using Subtraction Stereo and Color Information, In Proc. AWSVCI, (2009) 5.