

視覚障害者歩行支援のため の 単眼カメラを用いた歩行者 検出システム

A Monocular Pedestrian Detection system for the Visually Impaired

岸野 嵩久¹, Ruggero Micheletto¹

ABSTRACT

This paper presents a pedestrian detection system for visually impaired. In our detection method, Histograms of Oriented Gradients (HOG) are used as image features. Due to increase the false detection rate, accuracy of objects recognition is reduced. Therefore our system deals with multiple parts detector. Since our system performs three phase detection by the filter of each part which is deformable and movable, it can raise the accuracy. Additionally, we introduced scan area prediction process using particle filter for reducing processing time. And our system estimates the distant to pedestrian in front with single camera. The distance is calculated using the perspective model. In evaluation experiment, our system improves the problem of false detection. It is about 18.9% better than the standard HOG detector. And scan area prediction process was able to reduce processing time more than half. Experimental results also

show that our system can estimate the distance to pedestrian.

Keyword: 視覚障害者, 歩行者検出, パーティクルフィルタ, Histograms of oriented gradients

¹ 横浜市立大学大学院生命ナノシステム科学研究科知覚情報科学研究室

1. はじめに

現在、国内の視覚障害者は31万人を上回る[1]。そのため、視覚障害者の歩行を支援するシステムの開発は福祉的に大きな意義がある。

視覚障害者の歩行を支援するシステムに求められる要素は大きく分類すると、経路のナビゲーションと移動中の安全確保の2つになる。視覚障害者歩行支援のための経路ナビゲーションシステムは、GPSを利用した座標即位や特定のランドマーク認識による位置推定などのこれまで多く提案されてきた[2]。しかし、これらのシステムでは接近する歩行者や車などの移動物体のことは考慮されていないことが多く、衝突回避の喚起など視覚障害者の歩行中の安全確保について課題が存在している。また、視覚障害者が外出する際に最も不安に感じることの1つが人や車などの存在と歩行中に衝突することであることがアンケート調査によって示されている[1]。以上の背景から、視覚障害者が歩行中に接近する他の歩行者を察知して衝突の回避を可能にするシステムは、歩行中の安全確保のためにも必要なものである。

歩行支援システムを実装するデバイスは使用者が携帯可能な大きさや重量でなければならない。また、普及のしやすさも考慮すると、小型カメラなどの身近にある既存の携帯型デバイスに標準的に装備されているセンサーを用いることが望ましい。一般的に、歩行者などの障害物を検出するシステムはIRセンサーなどの距離を測定するのに特化したセンサー、もしくは複数のカメラによるマ

ルチビジョンを用いる。しかし、距離測定用のセンサーや複数のカメラは我々が生活の中で使用している携帯デバイスには標準的に装備されていることは少ない。これらの理由から、我々は携帯電話やスマートフォンといったモバイルデバイスに注目した。これらのデバイスは一般に入手しやすいことに加えてカメラを標準装備している。

以上のような背景から、本研究では単眼カメラで撮影された映像から前方の歩行者をリアルタイムに検出し、その距離を推定することで衝突の危険性を察知できるようなシステムの開発を目的としている。

2. システム構成

本システムは大きく分けて、歩行者検出と距離推定の2つのフェーズから構成される。歩行者検出は Histogram of Oriented Gradients(HOG) 特徴量に基づく人物検出法を行う[3]。そして、検出した対象の座標から投影法を用いてユーザーとの距離の推定を行う。

2.1 歩行者検出

2.1.1 HOG 特徴量

HOG 特徴量とは、局所領域の輝度値の勾配とその方向をヒストグラム化して特徴量としたものである。勾配の強度 m とその方向 θ は以下の式によって算出される。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2}$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)}$$

ここで、 $f_x(x, y)$, $f_y(x, y)$ は各画素における輝度値 L の差分であり、以下の式で与えられる。

$$f_x(x, y) = L(x+1, y) - L(x-1, y)$$

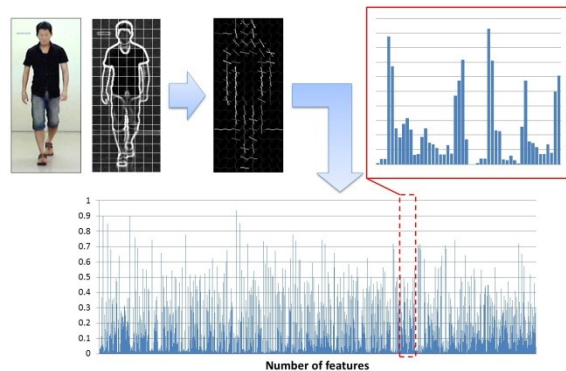
$$f_y(x, y) = L(x, y+1) - L(x, y-1)$$

各ピクセルで算出した θ を 0° から 180° の範囲に変換し、 20° ごとの9方向の勾配ヒストグラムに分割する。そして、画像を $n \times n$ ピクセルのセルと呼ばれる領域に分割し、セルごとの勾配ヒストグラムを作成する(Fig.1)。セルは $l \times l$ セルごとに正規化される。これにより、局所領域の明るさの変化にも影響されにくくなる。

算出された特徴量は機械学習により統計的に識別に有効な特徴が抽出され、対象の検出に使用する識別器が作成される。本研究では、機械学習に Real AdaBoost を使用した[4]。

2.1.2 Multiple parts HOG detector

Fig.1 HOG 特徴量の算出過程



HOG 特徴量に基づく人物検出はこれまでの特徴量の方法に比べて高い精度の検出が可能であるが、背景が複雑な場合や変化が大きい場合には誤検出が多いという課題があった。そこで、本研究では HOG 特徴量を用い複数の識別器を組み合わせたカスケード型 Multiple parts HOG detector を用いた検出法を提案する。Multiple parts HOG detector は 2 種類の HOG 特徴量を用いる。1 つは全身の特徴を表現する粗い解像度の特徴量、もう 1 つは体の部分を表現する細かい解像度の特徴量である。全身を粗い解像度、部位を細かい解像度で表現することで効率的な検出を行う。今回

検出する部位は、頭と脚の2ヶ所とした。

Fig.2 Multiple parts HOG detector の構成

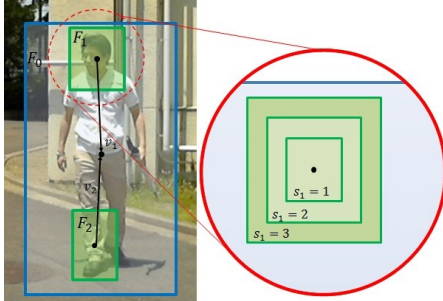


Fig.2 に Multiple parts HOG detector の構成図を示す。人モデルはベースフィルタ F_0 と2つのパーツモデル (P_1, P_2) から構成される。

各パーツモデルは以下のように与えられる。

$$P_i = (F_i, v_i, s_i)$$

ここで、 i は部位の番号(頭=1,脚=2)であり、

F_i は部位の識別器パートフィルタである。

v_i はベースフィルタと部位の相対位置を表すベクトル、 s_i はパートフィルタのサイズを表す。パートフィルタのサイズは各パーツ3段階用意されている。また、各パートフィルタは可動式であり、ベースフィルタの範囲内をパートフィルタの面積の50%以上が被るように位置を変えられる。これにより、姿勢の変化にも対応することが出来る。

2.1.3 パーティクルフィルタによる位置予測

Multiple parts HOG detector による検出は複数の識別器が組み合わさっている分、計算時間が掛かる。そのため、本研究では次フレーム画像上の歩行者の位置を予測し、その領域に走査範囲を限定することで処理時間を削減させる。予測には時系列フィルタの一種であるパーティクルフィルタを利用する[5]。

パーティクルフィルタは時刻 t までの観測データ Z_t を用いて、確率変数 X_t の確率分布 $p(X_t | Z_t)$ を N 個の重み付き標本サンプル

$\{s_t^{(n)}, w_t^{(n)}; n=1, \dots, N\}$ で近似表現する。

$s_t^{(n)}$ は X_t 上でのサンプル、 $w_t^{(n)}$ はその重みであり、サンプルの総数は $N=500$ 個とした。各サンプルの重みは観測された尤度によって与えられ、重みに比例して復元抽出が行われる。これにより、尤度の高い付近にサンプルが存在することになる。Multiple parts HOG detector は学習した人の特徴とどの程度類似しているか信頼度を計算する。本研究では、検出器が算出した信頼度をサンプルの重みを決定付ける尤度として設定した。Fig.3 に実際に検出器の算出した信頼度による画像上の尤度分布を示す。各サンプルは定められた状態遷移をし、次フレームの位置を予測する。今回、状態遷移には等速直線運動を用いた。状態遷移をしたサンプル群の重心を予測位置として、次フレームの走査領域を移動する。

2.2 単眼カメラでの距離推定

単眼カメラで撮影した画像から、現実での対象の位置を推定するためには投影法を利用する必要がある。投影法から距離を算出する方法は大きく2つあり、画像上の対象の大きさを利用する方法と画像上の対象の位置を利用する方法が挙げられる。大きさによる方法は歩行者の身長によって画像上での大きさに個人差が生じるため、距離のみに依存した安定な値を得にくいことが考えられる。従って、位置を利用した方法によって歩行者の位置を推定する方法が有効であると考えられる。本研究では、地面の接地点である脚の座標を基に、ピンホールカメラモデルによる投影法を用いて歩行者の距離の推定を行う。

Fig.4 に画像上の座標を距離の幾何関係を示す。今回は簡単のためにカメラの高さ h を固定し、地面とカメラは平行であると仮定した。カメラから歩行者までの距離を d とすると、画像上の

y 座標は以下の式で求められる。

$$y = \frac{fh}{d} + \frac{H}{2}$$

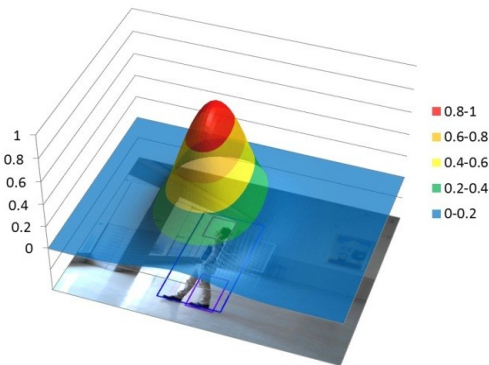
ここで、 f はピクセル単位での焦点距離、 H は画像の縦幅である。上式を変換し、以下の式によって y 座標から距離 d を算出する。

$$d = fh \frac{2}{2y - H}$$

Table.1 検出精度(%)

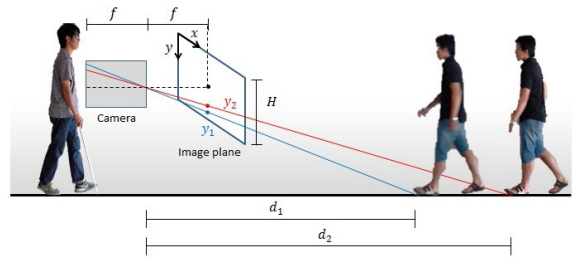
	Precision	Miss	False
HOG	91.90	9.10	25.21
Multiple	88.70	11.30	6.29
HOG			

Fig.3 検出器の信頼度による尤度分布



カメラの高さを $h=1.4[m]$ に固定して予備実験を行い $f=540[pixel]$ であることを求めた。また、接近の程度を判断する目安として、距離を3段階のレンジレベルに分類した：Near-range (4m以下)、Medium-range (4m~8m)、Far-range (8m以上)。

Fig.4 画像と歩行者までの距離との幾何関係



3. 検出評価

3.1 データベース作成

評価実験のために学習用画像データベースを作成した。このデータベースは特徴量を学習し、識別器を構築するのに使用される。使用した画像は様々な場所で撮影しているため、背景や照明が異なったデータになっている。学習用の画像は撮影画像から一部分を切り出したものを使用した。使用枚数は人物が写った画像を4000枚、背景のみの画像を6258枚である。

3.2 精度評価

検出精度の評価のために、学習用とは別に評価用のデータベースを作成した。このデータベースは学習用データベースと同様に撮影された画像から一部を切り出したものであるが、学習用に使用した画像とは異なる画像を用意した。評価用データベースは人物画像1000枚、背景画像1000枚から構成されている。評価用データベースに対して従来のHOG特徴量による検出法(HOG)とMultiple parts HOG detectorによる検出法(Multiple HOG)を適用して検出結果を比較した。なお、両手法の検出するための検出器の閾値は同じ値に設定した。Table.1に実験結果を示す。従来のHOGによる手法は検出率がMultiple HOGに比べて3.2%高くなった。一方で、Multiple HOGは人以外の物を誤検出する確率が従来法に対して18.9%削減されている。検出率の減少が見られたものの、誤検出についてはその約6倍の精度向上が確認された。この結果から、Multiple

HOG による検出法は、実環境のような複雑な背景の場合においても歩行者のみを検出できることが示唆されている。

3.3 撮影画像への検出実施

従来の HOG 特徴による手法と Multiple parts HOG による手法の両手法で、複数の場所で実際に撮影した映像から歩行者の検出を行った。なお、撮影時にカメラは 1.4m の高さに固定してある。

Fig.5 HOG(左)と Multiple parts HOG(右)の検出例

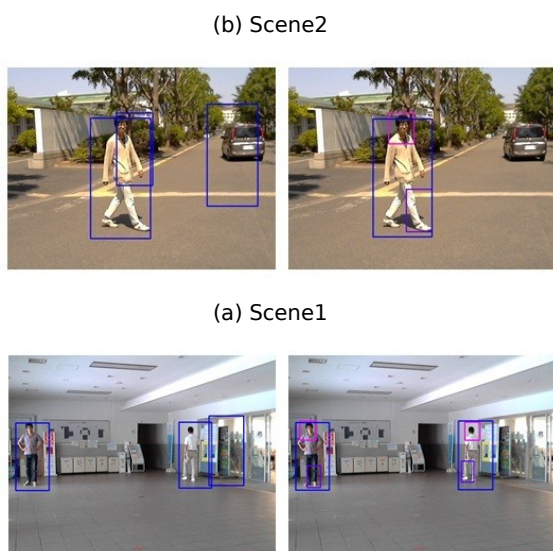


Table.2 平均処理時間(msec)

		Time
HOG	Scene1	1689.92
	Scene2	1924.52
	Average	1807.22
HOG + Particle Filter	Scene1	680.51
	Scene2	545.38
	Average	612.95

Fig.5 に歩行者の検出例を示す。両シーンとも左側は従来の HOG 特徴による検出、右側が

Multiple parts HOG による検出例である。従来法は車などの付近に誤検出(図中矢印)が見られるが、Multiple parts HOG ではこれらの誤検出は解消されている。この結果から、複雑な背景環境にあっても従来よりも強固な検出を行うことができると言える。また、Fig.6 に実際の検出結果から歩行者の位置を推定した結果を示す。推定結果は歩行者の位置関係、及びレンジレベルを正確に表すことができた。

Fig.6 検出結果に基づく距離算出例

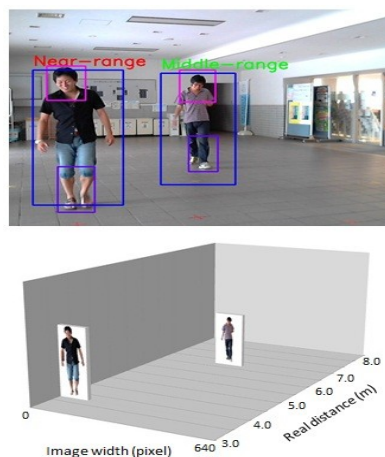


Table.2 に Intel Core i7 CPU 2.93GHz を用いた際の 1 フレームの平均処理時間を示す。パーティクルフィルタによる走査範囲の予測を利用した場合、平均処理時間は 612.95msec であった。こ

れは、範囲予測を用いない場合の処理時間 1807.22msec の約 33% の処理時間に相当する。この結果から、パーティクルフィルタによる走査範囲の予測は処理時間の削減に効果的であったと言える。

4. まとめと今後の展望

視覚障害者の自立歩行を支援するシステムの開発のために歩行者検出法を提案し、検出結果から歩行者との距離の推定を行った。複数の識別器を組み合わせた Multiple parts HOG detector は、

通常の HOG による検出では誤認識されてしまった背景を正確に区別することが可能であった。また、パーティクルフィルタによる検出器の次フレームの走査範囲予測を行うことで、1 フレームに掛かる処理時間を 65%以上削減することが出来た。そして、検出した画像上の座標を基に前方の歩行者の距離を推定することができた。一方で、リアルタイムで安定に処理を行うためには更なる処理時間の削減も必要になっていると考えられる。

今後は、更なる処理時間の削減を試みることに加えて、スマートフォンやタブレットなどのデバイスへ試験的に実装し、カメラが平行でない場合にも対応することも検討している。

参考文献

- [1]障害者福祉研究会, “障害者のための福祉 2004,” pp. 7-9, 2004.
- [2]A.Helal, S. Moore, and B. Ramachandran. Dridhti, “An integrated navigation system for the visually impaired and disabled,” *Fifth International Symposium on Wearable Computers (ISWC01)*, pp.149-156, 2001.
- [3]N.Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *CVPR*, vol.1, pp.886-893, 2005.
- [4]R.E. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” *Mach. Learn.*, no.37, pp.297-336, 1999.
- [5]E. Cuevas, D. Zaladiver, and R. Rojas, “Particle filter for vision tracking,” *Technical Report B, Fachbereich Mathematikund Informatik, Freie Universitat Berlin*, 2005.