

屋外 PMR のための LRF と小型全方位カメラを併用した 広範囲からの人検出

斉藤 学[†] 山崎 公俊[†] 畑尾 直孝[†] 花井 亮[†] 岡田 慧[†]
稲葉 雅幸[†]

[†] 東京大学情報理工学系研究科
東京都文京区本郷 7-3-1

E-mail: †{saito,yamazaki,hatao,hanai,k-okada,inaba}@jsk.t.u-tokyo.ac.jp

あらまし 本論文では LRF と小型全方位カメラを併用することで広範囲からの人検出手法を提案する。LRF から得られる情報だけでは人であるか否かを検出するには弱く、画像からの識別だけでは計算コストが大きい。そこで LRF により人の可能性のある対象を追跡し、その間に画像の HOG 特徴量を用いた識別による人検出を連続して行い、識別結果を統合することで最終的な識別を行う。また、本提案手法をパーソナルモビリティロボットに搭載することで、屋外において合図をする人を乗客と判断しタクシーアプリを実現させた。

キーワード 人検出, 自律移動ロボット, 全方位カメラ, タクシーサービス, レーザレンジファインダ, HOG 特徴量

Human Detection using LRF and Small Omniview Camera for Outdoor Personal Mobility Robot

Manabu SAITO[†], Kimitoshi YAMAZAKI[†], Naotaka HATAO[†], Ryo HANAI[†], Kei OKADA[†], and
Masayuki INABA[†]

[†] School of Information Science and Technology, The University of Tokyo
Hongo, Bunkyo-ku, Japan

E-mail: †{saito,yamazaki,hatao,hanai,k-okada,inaba}@jsk.t.u-tokyo.ac.jp

Abstract In this paper, we propose a human detection method using a LRF and a small omniview camera. The resolutions of LRFs are too low to recognize human reliably, and human recognition by high resolution camera images requires high calculation cost. Tracking human candidate segments and human detection from clipped camera image using HOG feature can reduce false positive human detection. By combining this human detection method and hand wave detection on a personal mobility robot, the robot moved as a taxi service robot.

Key words Human Detection, Autonomous Mobility Robot, Omniview Camera, Taxi Service, LRF, HOG Feature

1. はじめに

パーソナルモビリティロボットとは、一人乗りの自律移動型ロボットであるが、周辺を認識し人の検出が可能であれば、ユーザに対し能動的に働きかけられる。しかし屋外環境で人を検出するためには、遠方の人を低解像度の画像を利用して発見できる手法が必要となる。

本研究では屋外パーソナルモビリティロボットのための LRF と小型全方位カメラの併用による頑健な人発見手法を提案する。この手法は LRF の情報を元に人検出処理を行うべき領域を限定し、並行して追跡することで、屋外環境で活動するロボットが得た遠方にいる低解像度の人画像についても誤検出を減らすことが可能となって

いる。また、それを用いたパーソナルモビリティロボットの乗客の発見行動を含む、人間の手振りによる呼びかけに応じて迎えに来るタクシーロボットシステムの実現についても実験する。

2. LRF と全方位カメラを併用した人検出

2.1 従来研究

本研究では 10 ~ 20[m] までの遠距離に存在する人を広範囲から検出する必要があるが、これまでに多くの人検出手法が提案されている。

一般的な物体認識手法の応用として、人間の全身画像の輪郭を用いた手法が提案されている。例えば Edgelet 特徴量による認識は、画像から抽出された輪郭のうち、



図 1 Human size in the panoramic image

人の輪郭画像に特徴的な形状を捉えて認識している [1] . 一般形状認識の可能な特徴量である HOG(Histogram of Oriented Gradients) [2] による人の特徴検出はその有用性が示されている . しかし , これら画像処理による手法のみで画像全体から特定の物体を検出するには大きな計算コストが必要である .

また , 通常のカメラ画像以外を利用する手法としては , 赤外線カメラ画像による人検出 [3] や , 複数の高さの異なる LRF を用いた検出方法 [4] が提案されている .

画像と LRF の併用による人検出の先行研究として , Hao ら [5] は LRF とカメラのキャリブレーション手法を提案し , 両センサを用いて人物の検出を行った . LRF の検出した領域の画像を切り出して , そのエッジ画像の曲率の大きさによって人と柱を識別している . この手法はコンパクトな構成で誤識別の数が少ない人検出手法であるが , 柱以外の一般の物体に対して識別できていない .

2.2 本研究のアプローチ

モビリティロボットに搭載可能なコンパクトなシステム構成で遠距離かつ広範囲からも人検出を行いたいので , そのために LRF と小型全方位カメラを用いる . LRF は , その得られる距離情報から人体のスケールに合致する領域を抜きだし , 追跡を行うために用い , 全方位カメラは得られた領域にある物体が人であることを画像から認識するために用いる .

全方位カメラは周囲の画像を一度に取得できることから人検出機能の応用の幅を広げるものと考えられるが , 一般に低解像度になってしまうという問題があり , 本研究で用いるカメラも 1280[pixel] に周囲 360[deg] の画像が収まるので , 人の視力に換算すると 0.06 程度である . 図 1 は人が写っているパノラマ画像の例である . 人を赤丸で囲ってあるが , たとえば人の肩幅を 500[mm] とすると , 10[m] 離れた人の幅はおよそ 10[pixel] となっている .

そこで低解像度画像を用いた人の検出方法として , 大まかな形状を識別できる HOG 特徴量を SVM によって識別する方法を用いる . その手順は次の 3 ステップであ

る . 画像処理の領域を限定するために , LRF の点群の情報によって人の可能性のある領域を追跡しておく . 次に毎フレームの画像の候補領域に対応する部分の HOG 特徴量を計算することで , 様々なスケールで画像全体から検出することに比べ効率的に検出する . そして追跡している情報と HOG による検出結果を紐付けて , 時系列の識別結果を統合することで識別性能の向上を図る . 以上の各ステップについて次章以降で順に説明する .

3. LRF による人候補追跡

この章では LRF の情報から人の可能性のある領域を追跡する方法を述べる .

始めに , LRF の点群を物体ごとに切り分ける , セグメンテーションを行う必要がある . 本研究では Nunez ら [6] の手法を参考に行った . その後 , 得られた各セグメントの長さが 200[mm] 以上 1000[mm] 以下のものを人の可能性のある領域として処理する . しかし背景となっているセグメントは , 他の物体に遮られて全体が見えず , 画像処理によって正しく検出することが困難であると考えられるため , 追跡対象からは取り除いた .

抽出された人候補領域はフレームごとに位置や形状が変化するため , 確率的な追跡手法が必要であり , Vermaak ら [7] の提案したミクスチャパーティクルフィルタを用いた . 現在の状態の確率分布に対して , 観測結果を元に計算した尤度を用いて , 状態分布を更新することで揺らぎのある観測をもとに状態を推定することができる (式 1 , 式 2) 手法である . ミクスチャパーティクルフィルタは総数が変化する複数の対象に対して追跡を行うことができる拡張手法である .

$$\text{Predict} : p(x_t|y^{t-1}) = \int D(x_t|x_{t-1})p(x_{t-1}|y^{t-1})dx_{t-1} \quad (1)$$

$$\text{Update} : p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{\int p(y_t|s_t)p(s_t|y^{t-1})ds_t} \quad (2)$$

ただし , x_t は時刻 t における d 次元状態ベクトル , $y^t = \{y_1, \dots, y_t\}$ は時刻 t までに得られたすべての観測結果とする . 状態ベクトル x_t の時間発展の確率関数は , $D(x_t|x_{t-1})$ で表す .

追跡の様子を , その時のカメラ画像と合わせてに図 2 に示す . カメラ画像は中央が進行方向正面で , LRF の計測範囲と同じ左右 135 度の部分を切り出したものである . LRF のデータを表示した右の画面は , 緑色が追跡対象のセグメント , 白色の点はパーティクルの分布 , 白い線は追跡セグメントの軌跡であり , 中心やや左の白丸は LRF の位置である . また , それぞれの数値は画像のフレーム番号である . 複数の人間と同時に , 樹木や自動車の一部なども追跡対象となっていることがわかる .

4. 画像の切り出しと HOG 特徴による識別

LRF を用いて人の可能性のある領域を追跡している

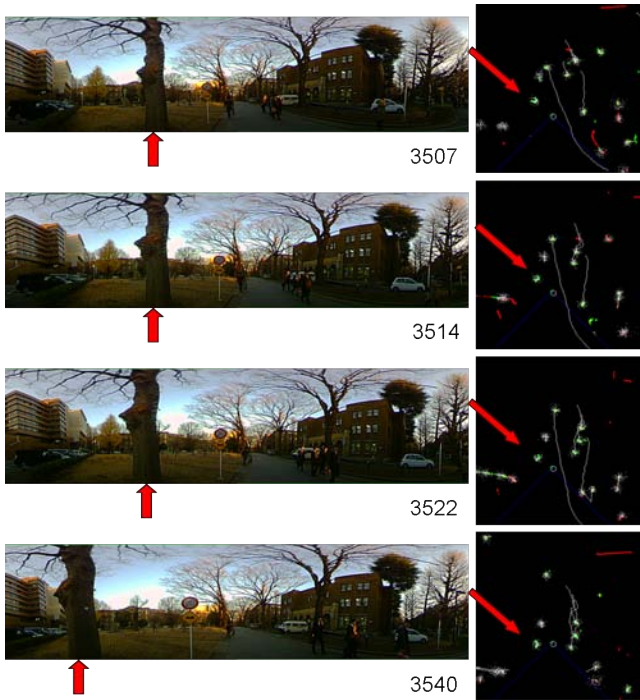


図 2 Tracking segments using mixture particle filter



図 3 Panoramic image and rectangles to clip

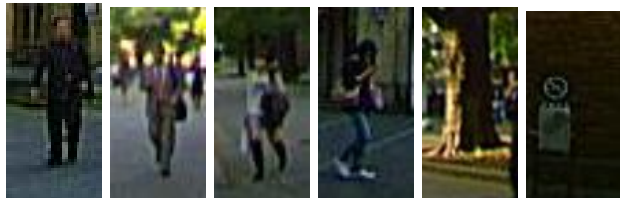


図 4 Automatically clipped images from panoramic images

ので、全方位カメラ画像からその方向と距離に対応する領域を切り出すことで画像処理を行う領域を限定する。また人の検出という課題に対しては、画面全体ではなく予め制限した領域に対してのみ人発見の処理を行うことで、誤検出の絶対数を減少させることができる。

図 3 は LRF の情報を元に全方位カメラの画像の切り出す領域を赤い枠で囲ったものである。対象までの距離は LRF により既知であるので、地面から $-400[\text{mm}] \sim 2000[\text{mm}]$ の高さに対応する領域を切り出すこととした。その際、画像の横幅は高さの半分の画素数となるようにした。図 4 は切り出された領域を同じ大きさに拡大した画像である。

次に、全方位画像から切り出した人候補領域の画像の HOG 特徴量を計算し、対象の画像が人であるか否かの識別率を調べる。HOG 特徴量を計算するため入力画像



(a) Clipped images from camera images (distances: 5m, 7.5m and 10m) (b) INRIA Person Dataset

図 5 Part of our and INRIA person dataset

を 64×32 にリサイズし、その中央部分の 30×60 画素を用いた。HOG 特徴量ベクトルの算出のためのパラメータは、セルサイズは 5×5 、輝度勾配のヒストグラム化のビン数は 9、正規化を行うブロックサイズは 3×3 とした。よって画像から得られる特徴量ベクトルの次元は 3240 次元となる。

図 5 (a) の画像は、切り出して学習に用いたデータセットの一部である。走行時に取得した画像から生成したデータセットは全身が写っている人画像が 510 枚、人以外の画像が 515 枚である。人以外の画像は、全方位カメラ画像からランダムに切り出したもの 300 枚と、移動中に LRF で切り出した画像のうち、人と誤認識しそうな看板や樹木の画像 215 枚となっている。

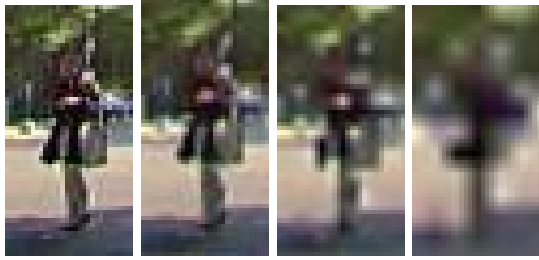
距離の異なる対象を画像から切り出すので、複数の解像度が混在している。そこで解像度の影響を調べるために、INRIA Person Dataset [8] の画像を用いて入力画像の解像度とその識別率の関係を調べる。人画像 920 枚と、人以外の画像はランダムに切り出された 1000 枚を取り出して用いる図 5 (b)。

本研究で SVM の実装は LIBSVM [9] を用い、RBF カーネルを用いた C-SVM を利用した。以下の各識別実験では、5 交差確認法を用いて識別率の評価を行った。交差確認法とは、ラベルが既知のデータセットを用いて識別率を推定する方法で、データセットを n 分割し、そのうち 1 個を抜き出し $n-1$ 個のデータで学習した識別器を使っての識別率を調べることを、分割した各集合に対して行いその識別率平均を求める方法である。SVM のパラメータを変化させた最適な値での識別率は、INRIA のデータセットで 99.8%、本研究で得たデータセットでは 94.5% となった。

5. 複数の解像度の画像を用いた HOG 特徴量による識別性能

5.1 解像度の異なるデータに対する予測性能

LRF を基に切り出した画像の解像度は低く、7[m] の距離の対象を切り出した画像は、HOG 特徴量の入力画像サイズ (32×64) とほぼ同じサイズになる。それ以上に距離の離れた対象の画像はさらに低解像度になるため、そのような画像に対する HOG 特徴による識別性能



(a) 30pixel (b) 20pixel (c) 12pixel (d) 6pixel

図 6 Scaled-up images after being reduced once

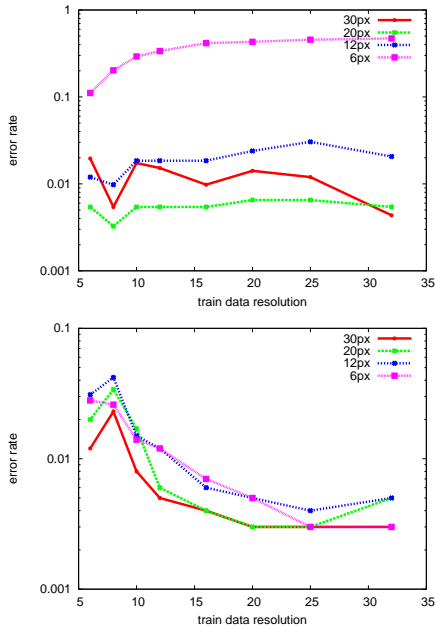


図 7 The performance of SVM varying train data and test data resolution (INRIA). (top:false negative, bottom:false positive)

についての評価が必要である。

山内ら [10] は空間的・時間的に異なるさまざまな HOG 特徴量を組み合わせて識別を行っているが、その中で解像度の異なる同じ画像から算出した HOG 特徴量について考察がなされ、体の部分ごとに識別率の高くなる解像度が異なることが示されている。

以上のことから、教師データとテストデータの解像度が変化した場合に HOG 特徴量による識別性能がどのように変化するかを調べる実験を行った。事前作成したデータセットは、そもそも解像度の異なる画像が混ざっており比較実験ができないので、INRIA から取り出したデータセットを用いた。まず、HOG 特徴量の計算に用いる領域の解像度がそれぞれ幅 30,20,12,6[pixel] になるよう一度縮小してから拡大し (図 6)、それらを訓練データとテストデータに分ける。訓練データの解像度を横軸として変化させたときの、テストデータの解像度ごとの識別率の変化を図 7 に示す。

識別される訓練データの解像度を横軸に取った識別率のグラフを図 7 に示す。特に減らしたい誤検出の率で

ある false positive のグラフは、訓練データの解像度が 15[pixel] よりも高ければ値に大きな変化はないので、一定以内の距離の画像であれば訓練データの解像度によらず識別が可能であると言える。本研究で用いるカメラであれば 15m ほどになる。

false negative の値も訓練データの解像度による変化は少なく、むしろ低解像度の 6[pixel] の画像に対する識別性能は向上しているといえる。これは高解像度の positive 画像は背景に対して中央部にエッジが出やすい特徴的な画像であり、一方 negative 画像は背景や物体の一部など平均すると特徴のないデータであるため、低解像度の画像はエッジが少なくなると考えられるので、もともとの画像によらずにその多くが negative として識別されてしまうためだと考えられる。

5.2 解像度の異なる訓練データによる識別実験

訓練データとテストデータの解像度の異なる場合の汎化性能について評価したが、学習データには切り出した画像を拡大するため様々な解像度のものが含まれる。そこで、複数の解像度の画像を混合させて訓練データとした場合に、識別性能がどのように変化するかを調べた。交差確認法のために訓練データとテストデータを分けた後、訓練データを 4 種類の解像度に画像を変換し、それら全てを使って学習を行った。

幅 30~21[pixel] の高解像度な 4 種類を混合させたデータセットと、30~15[pixel]、30~6[pixel] までの低解像なものそれぞれを用意して学習させ、30~10[pixel] までの各解像度の画像に対する識別性能を評価した。比較対象として最も高解像度である 30[pixel] のデータで学習させた識別器による識別結果も並べる。INRIA のデータセットを用いて上述の評価を行うと、結果は図 8 のグラフのようになった。

未検出率 false positive は、複数の解像度のデータを混合して学習した識別器では改善する傾向がある。特に低解像度の画像を識別する際には、訓練データに低解像度の画像を含むほど性能が良くなった。一方、誤検出率は 6[pixel] の低解像度の画像を訓練データセットに含む場合には性能が悪化した。それ以外のデータセットでは目立った変化は無かった。

以上の点から、複数の異なる解像度の画像をまとめて訓練データとして学習を行うことは、識別性能を悪化させないことがわかった。

6. 提案する人検出手法を用いたタクシロボットの試験

パーソナルモビリティロボットによる移動支援アプリとして提案する、タクシーサービスの実現について述べる。これまでに述べた人検出の提案手法は屋外において人の可能性のある領域を列挙して追跡するものであるが、これと手振りの検出を組み合わせるとパーソナルモビ

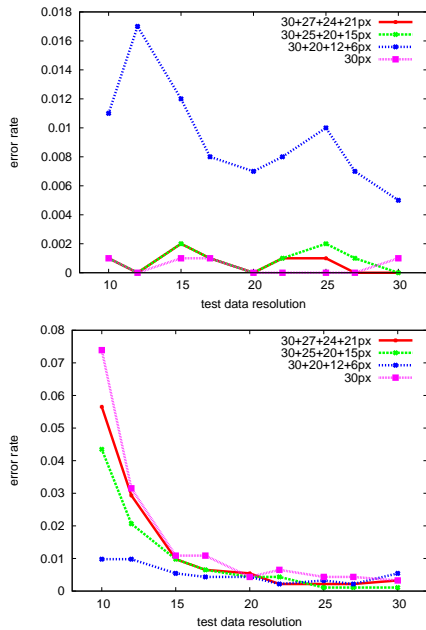


図 8 The performance of SVM using mixed multi-resolution images as train data(INRIA),(top:false negative, bottom:false positive)

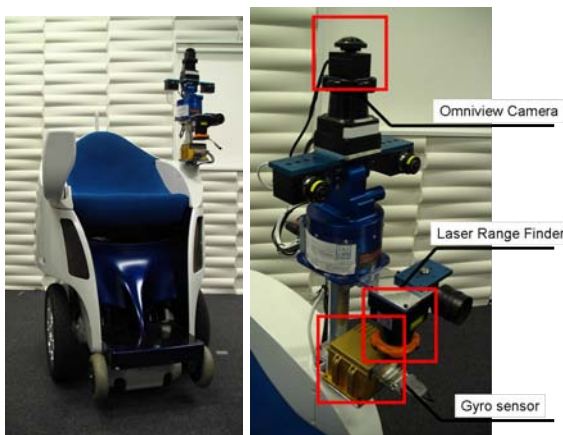


図 9 Robot and sensor positions

リティロボットの行動生成実験を行う。

今回実験に用いたロボットである PMR [11] は、トヨタ自動車の MOBIRO を元にセンサや計算機を増強した、一人乗りの倒立二輪型ロボットである (図 9 左). LRF は HOKUYO の UTM-30LX を、全方位カメラは NM33 に俯角 17° の魚眼レンズをつけたものを上向きに使用する。カメラの高さは 130cm であるので、約 4m 離れた対象の足元を撮影できる。これらのセンサと揺れ補正のためのジャイロセンサはロボットに固定した (図 9 右)。

PMR に搭載した計算機はノート PC は 2 台であり、lenovo の T60p(CPU:Intel(R) Core(TM)2 2.33GHz, Memory:2GB) と、Apple の MacBook Pro(CPU:Inten(R) Core(TM)2 Duo 2.60GHz, Memory:3GB) である。T60p に LRF とカメラを接続し人候補領域の追跡と画像処理及び識別の全てを行い、MacBook は経路生成などのシステム用 PC とした。

表 1 Classification result of images collected by moving robot

dist [m]	4 ~	6 ~	8 ~	10 ~	12 ~	14 ~	16 ~	18 ~	20 ~
true pos	146	129	72	45	21	12	5	4	0
false pos	10	15	24	59	25	22	28	26	9
false neg	61	75	47	36	38	18	6	2	0
true neg	72	64	132	235	96	210	110	96	29

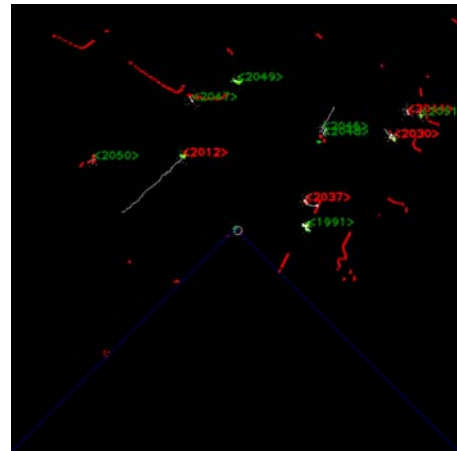


図 10 The result of the proposed human detection method: red clusters are judged as human and green ones are non-human.

6.1 人検出と追跡

ロボットが屋外を移動中に切り出した画像 1979 枚を、識別器で分類した識別結果を距離ごとに分けたものを表 1 に示す。識別器を作るための訓練データは、大学の構内で撮影された画像であるが、テスト環境とは異なる場所で撮影されたものであるため、識別対象の画像は訓練データには含まれていない。また、解像度が低いため人にどうかを判断できなかった画像は negative として分類した。

LRF で追跡中の一つの対象に対して得られる複数の識別結果を統合して、その時点において人であるかどうかの認識結果を求める。識別結果が positive の回数を p , negative を n とし、 $p^2/(p^2 + n)$ の値で識別結果を決めた。 p が 2 乗であるのは、positive な識別結果の絶対数が多くなった対象を、単純に線形に positive と negative の比率で評価するのでは、識別率が低かったからである。経験的にこの値が 0.9 以上で人と認識させた。

図 10 は、LRF による追跡領域を全方位カメラ画像から切り出して識別を行い、その結果を統合して判定している結果の画像である。 < > で囲まれた数字は追跡対象の通し番号であり、赤字が人であると判定されているものである。2017 番, 2037 番は人であり正しく判定できているが、誤検出も 2 つ存在している。

追跡と識別結果の評価は、10Hz で 5000 フレーム行い、計 2156 のクラスタを追跡した。人に由来するものは 130, 全クラスタの平均寿命は 36 フレームであった。センサと追跡対象の間を車が通ると、クラスタ追跡が途切れるため、クラスタの平均寿命は短くなっている。追跡



(a) Human detection result (red rectangles mean positive, black ones mean negative)



(b) Hand wave detection



(c) Approaching the user and go to the selected destination automatically

図 11 Experiment of taxi service application

したクラスタのうち人だと判定したものは 116 個で、うち 38 個が true positive, 78 個が false positive であった。

今回の追跡実験において人セグメントのほとんどが歩行者や自転車であり、立ち止まっている人はわずかで、移動したり向きを変えたりすることでオクルージョンが生じ追跡が途切れたりするので、立ち止まっている人に比べ検出できるセグメントの割合は減少するため、true positive の個数が 38 個と少なくなっている。また、誤識別率は $78/2026 = 3.8\%$ となった。全体として positive な検出数自体が少ないので、false positive の数も少なくなっているが、この 3.8% という値は HOG 画像単独での識別率に比べると低くなっているため、ノイズの大きな屋外環境においても有効性があると言える。

6.2 タクシーアプリ

以上の人検出に手振り動作の認識と目的地までの移動を組み合わせることでタクシーアプリを実現する。

タクシーサービスに必要となる手振り動作の認識を、人間のジェスチャー認識として解析することは難しい。しかし手振りに対しては、画像中の輝度の変化の中から手振りに特有の周波数を FFT によって抽出する手法 [12] が提案されている。実験ではこの手法を参考に、手振りによって生じる輝度変化の特性を捉えて認識する。

対象に対しズームを行って画像を取得、Haarlike 特徴による上半身検出により顔位置の推定を行う。次にロボットの移動を考慮し、テンプレートマッチによるフレーム間の移動の補償を行う。推定される顔位置の左右においてフレーム間差分を計算し大きな変動があるかどうかを調べる。顔の片側で大きな変動が一定時間以上続くことを手振りの特徴として検出を行った。

図 11 のように、巡回中に手を振っているユーザを見つけ (a,b)、目の前に停車する。ユーザはロボットに乗り、備え付けられた端末を操作することで目的地まで移

動することができる (c)。

7. おわりに

LRF を用いた前処理により HOG 特徴量の計算時間が減少したため、同一対象に対する複数回の識別結果を取得でき、解像度の低い画像を用いた識別であってもある程度の精度で人の検出が可能であることを示した。

全方位カメラの画像を用いた人認識の手法により遠距離に存在する人を検出できることにより、広範囲から発見した人に対して手振りの認識などを統合することで、PMR によるタクシーサービスという応用例を示した。

今後の課題としては、対象の距離に応じて変化する識別性能を考慮した識別結果の統合方法を考える必要がある。また、HOG 特徴を用いているため人の検出のみではなく一般の物体に対しても適用可能な手法であるので、大きさや断面形状が既知の物体を探す課題の応用例を見つけ適用していきたい。

文 献

- [1] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. *ICCV '05*, pp. 90–97. Citeseer, 2005.
- [2] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. *Lecture Notes in Computer Science*, Vol. 3952, p. 428, 2006.
- [3] R. Miezianko and D. Pokrajac. People detection in low resolution infrared videos. In *In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPR Workshops 2008)*, pp. 1–6, 2008.
- [4] O.M. Mozos, R. Kurazume, and T. Hasegawa. Multi-Layer People Detection using 2D Range Data. In *In Proc. of IEEE International Conference on Robotics and Automation Workshop (ICRA Workshops 2009)*, 2009.
- [5] LI Hao, Y. Ming, and Q. Huijia. Camera and Laser Scanner Co-detection of Pedestrians. In *In Proc. of IEEE International Conference on Robotics and Automation Workshop (ICRA Workshops 2009)*, 2009.
- [6] P. Nú nez, R.Vázquez-Martin, A. Bandera, and F. Sandoval. Feature extraction from laser scan data based on curvature estimation for mobile robotics. In *In Proc. of IEEE International Conference on Robotics and Automation (ICRA 2006)*, Vol. 1, pp. 1166–1172, 2006.
- [7] J.Vermaak, A.Doucet, and P.Perez. Maintaining multi-modality through mixture tracking. In *In Proc. of International Conference on Computer Vision(ICCV 2003)*, Vol. 2, pp. 1110–1116, 2003.
- [8] N.Dalal. INRIA Person Dataset. <http://pascal.inrialpes.fr/data/human/>.
- [9] Chih-Chung Chang and Chih-Jen Lin. LIB-SVM – A Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [10] 山内悠嗣, 山下隆義, 藤吉弘巨. Boosting に基づく特徴量の共起表現による人検出. *電子情報通信学会論文誌 D*, Vol. 92, No. 8, pp. 1125–1134, 2009.
- [11] 山岡正明. モビリティロボット. *自動車技術会 2008 年春季大会フォーラムテキスト*, 2008.
- [12] 入江耕太, 梅田和昇. 濃淡値の時系列変化を利用した画像からの手振りの検出. *日本ロボット学会誌*, Vol. 21, No. 8, pp. 923–931, 2003.