

人物の挙動認識に基づく自律移動型ロボット制御の研究

集美輝^{†1} 池田亮^{†1} 鹿嶋雅之^{†1} 佐藤公則^{†1} 渡邊睦^{†1}

ロボットによる日常生活の支援において自律移動ロボットの人物の挙動認識は重要であり、中でも人物追跡は特に重要である。本研究では、Kinectによる距離計測と人物認識を利用して挙動を認識する。追跡開始時に対象の衣服の情報を学習し、テンプレートマッチングを用いて追跡対象人物の特定を行う。しかし、カメラ画像全体にマッチングを行うと処理時間が膨大になるので、距離情報を用いたマッチング範囲の限定を行う。これらを組み合わせることにより、ロバストな人物追跡を実現する手法を提案する。また、ジェスチャー認識を加え、人間とロボットがコミュニケーションを行えるようにした。屋内の廊下での実験により、提案手法の有効性を確認した。

Research of The Autonomous Movement Type Robotic Control Based on Action Recognition of A Person

MIKI ATSUMARI^{†1} RYO IKEDA^{†1}
MASAYUKI KASHIMA^{†1} KIMINORI SATO^{†1} MUTSUMI WATANABE^{†1}

In support of daily life by a robot, the autonomous mobile robot to recognize the behavior of the person is important and especially human tracking is important. In this paper, we used the person recognition and distance measurement by the Kinect. First, the information on clothes is learned simultaneously with a pursuit start. Next, tracked person is specified using template matching. At this time, the range which matches is limited using distance information. By adding gesture recognition, human and a robot enabled it to communicate. Experimental results in indoor hallway, we have confirmed the effectiveness of the proposed method.

1. はじめに

ロボット開発は、主に人間の出来ないことやロボット単体で活動を行うことを目的に、工場ロボットや宇宙開発など特殊な環境のものを想定して行われてきた。近年では、警備用ロボット、公共施設用ロボット、家庭用ロボットなどの開発が盛んに行われている。これらは主に人間の仕事や生活を補助し、協力することで、日常生活の支援を行うことが目的である。日常生活の支援において自律移動ロボットの人物の挙動を認識することは重要であり、その中でも人物追跡は最も重要な認識の内の一つである。また、日常生活の中にロボットが浸透するには人とのコミュニケーションが必要であり、その上様々な状況下でも臨機応変な対応が求められる[1]。

我々は、ロボットによる人物追跡を実現させるために単眼カメラでの画像処理とソナーセンサーを用いたロボットでの研究を行ってきた[2]。しかし、ソナーセンサーでは得られる情報が少なく、追跡が不安定であり、人物のコミュニケーションも行われていない。そこで、本研究では画像処理に加え、さらに多くの情報を得るために RGB-D カメラ (Microsoft 社の Kinect[3]) を用いた距離計測を利用することで、より安定した人物追跡とコミュニケーションを実現することを目的とした人物の挙動認識について述べる。

2. 従来研究との比較

ロボットの人物追跡の第一段階として人物を検出する必要がある。例えば、ステレオビジョンを用いて得られた人物シルエットから複数人物の検出および重なりを検出を実現する研究が行われているが、特定人物の追跡は行われていない[4]。

また、近年はロボットのための人物追跡の研究も行われている。中でも、HOG 特徴量とレーザレンジファインダ (LRF) を組み合わせた手法を用いて安定的な人物追跡を行う研究がいくつかみられる。しかし、それだけでは特定人物の追跡が難しいため、さらにパーティクルフィルタを組み合わせる手法も提案されている。中には高速な人物追跡が行われているが研究もあるが、実際にロボットを動かしての検証が行われていない[5]。また、実際にロボットに実装して実験を行っているものもあるが、人物とのコミュニケーション方法を考慮しておらず、日常生活での共存を想定していないのが現状である [6]。共存を意識した人物追跡ロボットの開発も行われているが、あらかじめユーザーを登録し、服の情報を記憶させる必要があるため手間がかかることが問題としてあげられる[7]。

本研究の人物検出は RGB-D カメラ (Kinect) で行う。Kinect を用いたロボットによる人物追跡を行う手法 [8] が提案されているが、Kinect のスケルトントラッキングのみを使用して追跡を行っている。これでは特定の人物追跡は困難である。

^{†1} 鹿児島大学工学部
Kagoshima University Faculty of Engineering

そこで、本研究では Kinect による人物検出と距離計測を行い、対象人物の衣服の情報を追跡開始と同時に学習することで特定を行う。これによりロバストな人物追跡を実現する。また、ジェスチャー認識を行うことで人物とロボットの簡易的なコミュニケーションを実現する。

3. 研究の概要

本研究では、移動ロボットの上部に RGB-D カメラとして Kinect を設置する。ソナーセンサーだけでは情報量が少量に限られ、かつ不安定であるが、RGB-D カメラを用いることで、より多くの情報を利用出来る。Kinect は、API 群の OpenNI を利用することで距離画像の差分を元に人物を抽出することが可能である。今回は、そこから得られるユーザー情報を利用し、ロボット制御の実装を行う。

本研究の流れは、追跡する人物の学習し、その情報を元に人物の追跡を行い、終了の合図である「バイバイ」のジェスチャーを検出することで終了とする。その間、コミュニケーションの一環としてロボットの簡易的な発話も行う。

4. 追跡する人物の学習

Kinect は人間のサンプル画像を学習させておくことで人物部位ベースの識別器を用意している。これを用いることでリアルタイムに人物の頭、首、胸、腰、両膝、両肩、両肘、両手首、両手、両指先、両膝、両足首、両足の 24 か所の部位を認識することが可能である。

今回は、Kinect の骨格認識（スケルトントラッキング）の際に行うキャリブレーションポーズを追跡開始の合図として用いる。キャリブレーションポーズとは、ギリシャ文字の ϕ （プサイ）に似たポーズのことである。

本研究では、追跡する人物の衣服の情報をテンプレートとして切り出し、テンプレートマッチングを行うことで特定人物の追跡を行う。テンプレートを切り出す範囲は、胴（TORSO）を中心とした一定の範囲とする。取得するテンプレートの範囲は、利用目的などにより変化すると考えられるので、実験的に変更する必要がある。テンプレート取得の様子を図 1 に示す。これは、Kinect のユーザー認識を用いてカメラ画像のマスク処理を施した画面である。テン

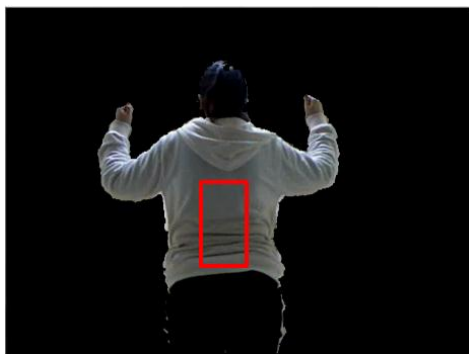


図 1 テンプレート取得画面

プレートを取得した場所は図 1 の赤枠の部分である。図 1 では、人物の後ろ姿のテンプレートを取得しているが、人物を見失った場合の為に正面のテンプレートも取得する。なお、人物の骨格認識は、フレーム毎の処理が膨大になるので、認識後に中断する。

また、テンプレート取得と同時にユーザーの距離情報を保持しておく。これにより、マッチング範囲を限定する。詳細は次章で説明する。

5. 人物追跡

Kinect からユーザー情報を取得し、その情報を利用することでテンプレートマッチングの範囲を限定する。マッチングの結果から追跡対象を識別し、ロボットによる追跡を行う。追跡の処理の流れを図 2 に示す。

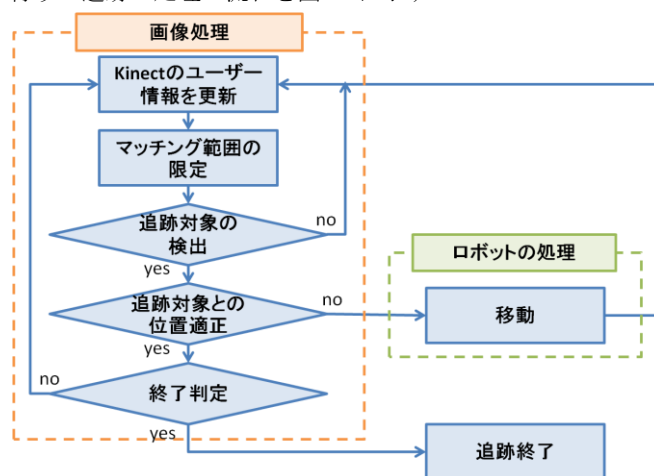


図 2 追跡処理の流れ

5.1 ユーザー情報の更新（人物検出）

Kinect のユーザー認識は、距離情報と動き（差分）から人物を検出する。本研究では、これを用いて人物検出とする。ただし、今回 Kinect はロボットに搭載している。本来、Kinect は固定である事が前提なので、Kinect 自身が動いてしまうことで背景の一部もユーザーとして認識してしまう場合が発生する。しかし、ある程度背景との分離は行えるので、条件を設けることでこの問題を回避し利用する。

5.2 マッチング範囲の限定

テンプレートマッチングによって特定の人物を検出し、追跡を行う。しかし、毎フレーム画面全体（640x480）にマッチングを行うと処理時間が膨大になり、実時間での処理は不可能となる。そこで、前フレームとのユーザー情報の差分からマッチングを行う範囲を限定する。

最初に、取得したユーザー情報を利用してカメラ画像のマスク画像を生成する。生成した画面を図 3 に示す。マッチング範囲の限定のために得られたユーザーごとの『①距離（ユーザー範囲の距離情報の平均）、②重心、③ピクセル数』を計算する。前フレームの①から③の情報を保持しておき、その差分からマッチングを行うユーザーを限定する。

本研究では、これらの条件を実験的に決定する。条件を満たしたユーザーは、図 3 の紫色の枠で囲われているユーザーである。奥に居る人物は、距離の条件を満たしていないため、マッチング対象外になっていることが分かる。

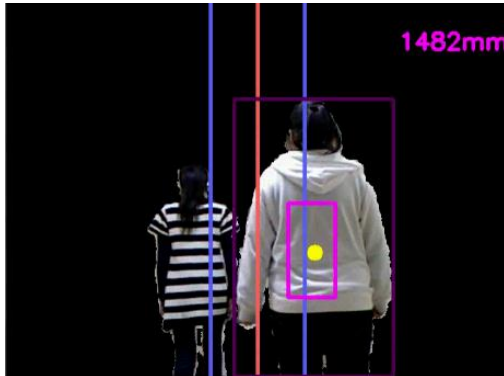


図 3 マッチングシーン

5.3 テンプレートマッチング

次に限定したユーザーの範囲内でテンプレートマッチングを行う。今回用いた比較方法は ZNCC (Zero-mean Normalized Cross-Correlation) である[9]。この手法は、テンプレートおよびマッチング画像の輝度値の平均値をそれぞれの値から引いて計算することで、明るさの変動があっても安定的に類似度を計算することができる。マッチング結果 (類似度) R を式(1)により求める。テンプレートの輝度値を $T(i, j)$ 、画像の輝度値を $I(i, j)$ とする。座標の (i, j) は、テンプレートの幅を m 画素、高さを n 画素としたとき、左上を $(0, 0)$ 、右下 $(m - 1, n - 1)$ とする。

$$R = \frac{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} ((I(i, j) - \bar{I})(T(i, j) - \bar{T}))}{\sqrt{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (I(i, j) - \bar{I})^2 \times \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (T(i, j) - \bar{T})^2}} \quad (1)$$

式(1)で得られた類似度 R の最大値を持つユーザーを追跡する人物とする。

また、追跡する人物を見失った際に別のユーザーを追跡することを防止するため、スコアに閾値を設定する。今回は最大スコアが 0.35 以上のユーザーを追跡する。

5.4 追跡対象の検出

5.3 でマッチングを行い、追跡対象の検出が成功すると特定したユーザーの重心 (図 3 の黄色の丸) を中心とした範囲を再度テンプレートとして取得し、更新する。これは、開始時に取得したテンプレートを各フレームで使用していると、人物が回転する動作の際に類似度が著しく下がってしまうためである。これにより、安定して類似度の高いマッチングを行う。

追跡対象が検出されなかった場合は、テンプレートの更新を行わず、再度ユーザー情報の更新を行い、同様の処理を行う。10 フレーム繰り返して検出できなかった場合は、『見失った (LOST)』状態と認識する。LOST 状態のときは、更新したテンプレートではなく、開始時に取得した、

人物の正面の衣服のテンプレートを使用してテンプレートマッチングを行う。

5.5 ロボットの移動

追跡する人物とロボットとの距離によってロボットの速さを変化させる比例制御 (P 制御) を行う。今回使用するロボットは 2 輪独立駆動のロボットであるので、左右の車輪にそれぞれ速度 v_R 、 v_L を与えることで制御する。

前後方向の移動を次の式(2)で行う。 L_{H-R} を人間とロボットの適正距離、 K_{p1} を比例定数、 L を人物とロボットの距離とする。図 3 の右上の表示が追跡対象との距離である。人物が接近すると速度 v_R 、 v_L は負の値になるので、ロボットは後退する。

$$v_R = v_L = K_{p1}(L - L_{H-R}) \quad (2)$$

また、ロボットは追跡する人物を常に中心に捉えておく必要がある。処理画面の幅は 640 ピクセルであるので画面の中心から ± 60 ピクセルの範囲を適正範囲とし、目標値を 320 ピクセルとする。図 3 の中心の橙色の直線が画面の中心、左右の青色の直線が ± 60 ピクセルを示している。回転移動を式(3)で行う。 K_{p2} を比例定数、 P_c 人物の重心の X 座標とする。

$$v_R = -K_{p2}(P_c - 320), \quad v_L = K_{p2}(P_c - 320) \quad (3)$$

ここで、距離が離れている、かつ回転の適正範囲を超えている場合は、式(2)、(3)の各値をそれぞれ足す。ただし、安全を考慮して K_{p1} の値を低めに再設定する。

5.6 終了判定

手を出して手を左右に振る『バイバイ』のジェスチャーを検出することで追跡終了の合図とする。検出の処理の流れは以下ようになる。ただし、ここで決定した数値は実験的に決定したものである。模式図を図 4 に示す。

- (1) 追跡するユーザーの平均距離から 100mm 前にあるピクセルを抽出する
- (2) (1)で抽出したピクセルの距離情報を小さい順に 2000 ピクセル抽出する
- (3) (2)で抽出したピクセルの重心を求める (これを掌とし、図 4 の緑色の丸である)
- (4) 最初に検出した重心を中心に上下 100 ピクセル以内で、15 ピクセル以上左右に振る

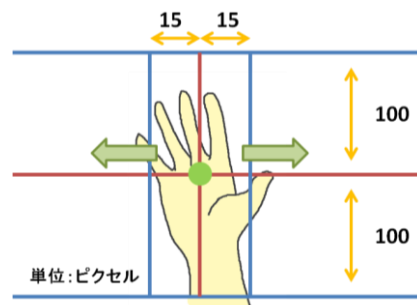


図 4 ジェスチャー検出

5.7 ロボットの発話

本研究では、人とのコミュニケーション方法の一つとして、ロボットの発話を行う。開始や終了時の挨拶だけでなく、ロボット自身がどのような状態になっているかなども被追跡者に知らせる。そこで、今回の実験の中で、以下のような状況のときにロボットが発話を行うようにした。

- 追跡開始
- ロボットと人物の距離が近い、または遠いとき
- 人物を見失ったとき、また、その後再発見したとき
- 終了の合図を検出したとき

これにより、ロボットとのインタラクションを高めていく。

6. 実験

6.1 実験環境

本研究で使用する移動ロボットは MobileRobots 社製の Pioneer3-DX である。地面からおよそ 90cm の位置に Kinect を搭載した。外観を図 5 に示す。

実験環境は図 6 に示すような段差のない屋内の廊下と踊り場で行い、また、室内での実験も行った。



図 5 移動ロボット



図 6 実験環境

6.2 予備実験

ここでは、テンプレートの大きさ、マッチング範囲の限定条件と人物とロボットの適正距離を決定するための予備実験を行う。

(1) テンプレートの大きさ

追跡する人物の服の情報を十分得るために、取得するテンプレートの大きさを出来る限り大きく設定する必要がある。今回は、60x120 ピクセルをテンプレートの大きさとした。これ範囲より小さく設定してしまうと情報量が少なすぎ、また、大きく設定すると人物が遠くに離れた際に人物の範囲に収まらない恐れがあるためである。

(2) マッチング範囲の限定条件

5.2 で示したマッチング範囲の限定条件『①距離（ユーザー範囲の距離情報の平均）、②重心、③ピクセル数』を予備実験から以下のように決定した。

- ① 距離情報の平均が±300mm 以内

- ② ユーザーの重心の X 座標の差が±300 ピクセル以内

- ③ ユーザーの全ピクセル数が±30000 ピクセル以内

③のピクセル数の決定は、人物がカメラの視野に収まらない場合を考慮した結果から設定した。

(3) ロボットとの適正距離

Kinect の距離の測定範囲が 800mm であり、かつ人物が急に止まった際に安全に停止できる距離を考慮する。また、適正距離の範囲を狭く設定すると発散する恐れがあるので適正範囲を広めに設定する。

したがって、今回は適正距離を 850mm から 1100mm とし、ロボットの移動の際に用いる目標値 L_{H-R} を 975mm とする。

6.3 実験結果

廊下の直線と緩やかなカーブでの追跡は、開始から終了のジェスチャーまでを一連の流れとして実験を行い、安定して追跡することができた。追跡の様子を図 7 に示す。左画面がカメラ画像をユーザー認識した部分でマスク処理したもの、右の画面がカメラ画像上でユーザー認識した部分を 3 色で塗りつぶしたものである。桃色の枠がテンプレートマッチングの結果で最も類似度の高かった箇所である。A フレームが開始の合図である。B のフレームの処理画面で左側の壁をユーザーと認識しているが、マッチング範囲を限定しているため、テンプレートマッチングを行っていない。C から E フレームでカーブしており、F フレームでロボット向き合いジェスチャー認識を行い、終了となる。

別の人物が向かい側から歩いて来てすれ違った場合の実験も行った。すれ違いの様子を図 8 に示す。結果から分かるようにすれ違いが起こった場合でも対象を追跡することが出来た。

また、追跡している人物とロボットの間を別の人物が横切った場合の実験も行った（図 9）。B フレームで、横切った第三者が完全に追跡対象を遮ってしまっているが、マッチング範囲の限定によって人物は LOST した状態になっている。C フレームで再び人物を検出できている。

定量的な実験は、本稿執筆時には間に合わなかったため、発表の際に改めて報告する。

7. おわりに

本研究では、RGB-D カメラ（Kinect）による人物検出と距離計測を行い、さらに対象人物の衣服の情報を追跡開始時に学習することで特定人物の追跡を実現する方法を提案し、ロボットでの実験を行った。また、人物の挙動を認識することでコミュニケーションを行うことができた。

実験結果から特定の人物を安定して追跡を行うことができた。また、第三者がすれ違うシーンにおいても衣服の情報をを用い、マッチングの範囲を限定することで追跡することができた。

今後の課題は、追跡対象の衣服がすれ違う人物や背景と

似ていると見失う場合があるので、テンプレートマッチングだけでなくカラーヒストグラムや特徴点を用いるなどの対策が必要である。また、対象の位置予測を利用する必要がある。

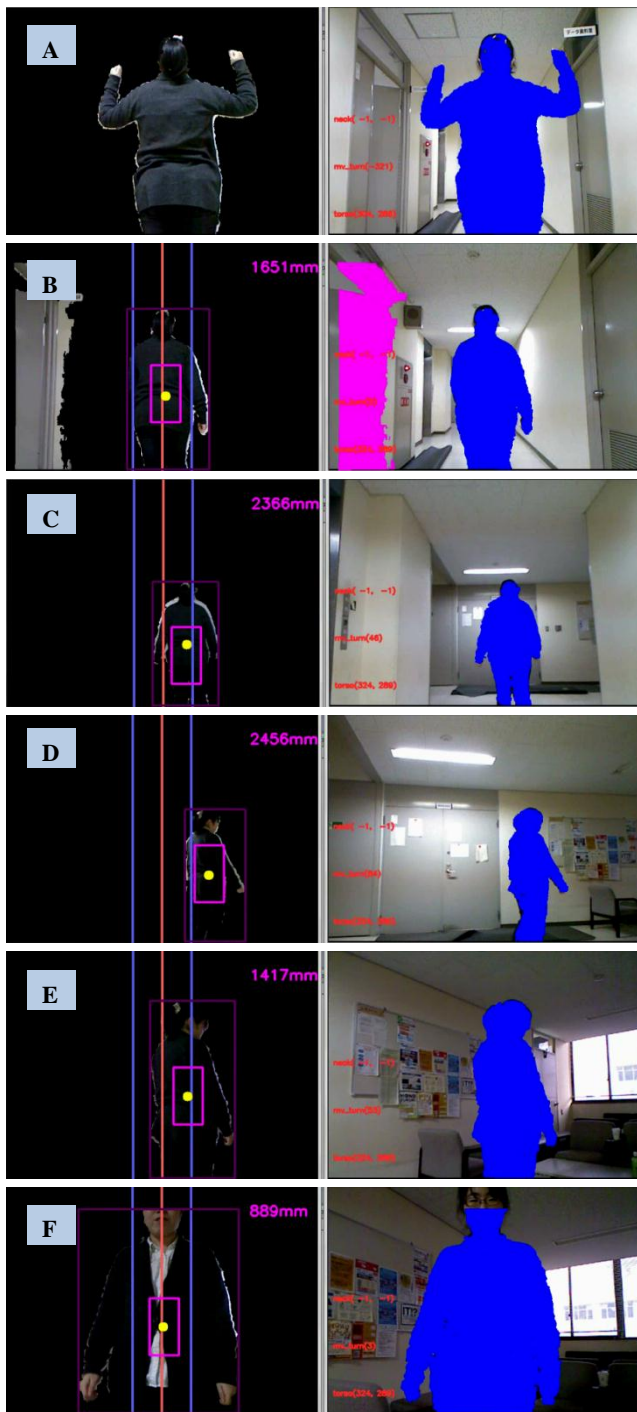


図 7 追跡の様子

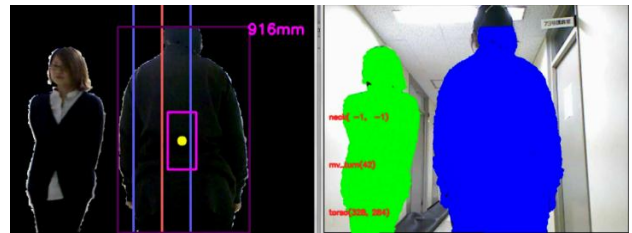


図 8 すれ違った場合

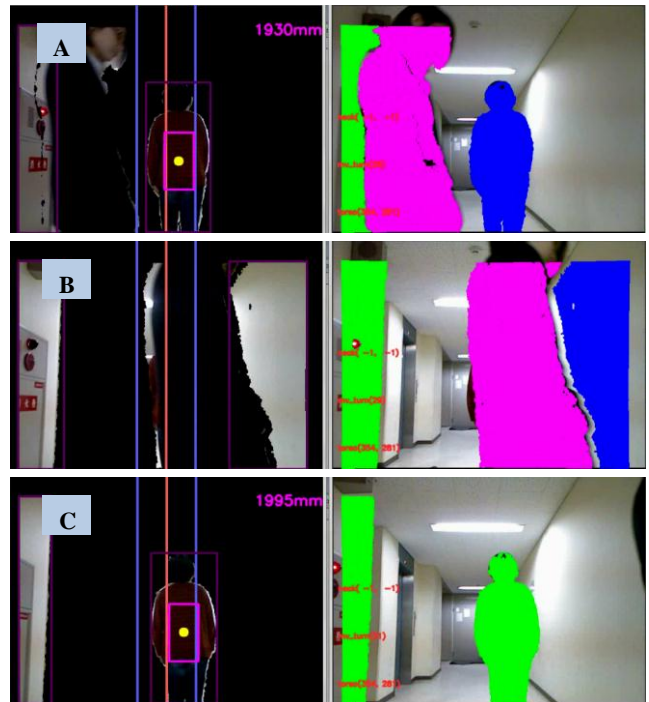


図 9 横切った場合

参考文献

- [1]松日楽信人, 小川秀樹, 吉見卓, “人と共存する生活支援ロボット”, 東芝レビュー, Vol.60, No.7, pp.112-115, (2005)
- [2]中野広樹, 下脇克友, 片山明伯, 渡邊睦, “カルマンフィルタを用いた足位置予測に基づく人物追跡自律移動ロボットの研究”, 情報処理学会コンピュータビジョンとイメージメディア (CVIM) 研究会報告, Vol.2004-CVIM, No.146, pp.9-16, (2004)
- [3]KINECT for Windows
<http://www.microsoft.com/en-us/kinectforwindows/>
- [4]佐竹純二, 三浦 純, “複数人物のシルエットの重なりを用いたステレオビジョン人物追跡”, 画像の認識・理解シンポジウム (MIRU2012), IS2-60, (2012)
- [5]小林祐輔, 監物建秀, 渡辺寛望, 小谷信司, “移動ロボットによる奥行き方向の変化に対応した人検出”, 日本ロボット学会学術講演会予稿集 (CD-ROM), Vol.30, ROMBUNNO.2J3-4, (2012)
- [6]栗井真志, 清水隆史, 山下 淳, 金子 透, 浅間 一, “カメラ・LRF 搭載移動ロボットによる HOG 特徴量を利用した人物追跡と生成地図を用いた自律帰還”, 第 24 回自律分散システム・シンポジウム資料, pp.47-52, (2012)
- [7]溝口博, 菊池隆司, 伊藤啓太, 松日楽信人, 吉見卓, 中本秀一,

西山学, “マルチモーダルインタラクションにより対人追従するロボット(お供ロボット)の研究開発”, 日本機械学会ロボティクス・メカトロニクス講演会'05 講演論文集, ALL-N-022, (2005)

[8]石川祐輝, 渡辺寛望, 小谷信司, “Kinect の奥行き画像を利用した自律移動ロボットによる特定人物の追従”, 日本ロボット学会学術講演会予稿集 (CD-ROM), Vol.30, ROMBUNNO.2J3-5, (2012)

[9]OpenCv.jp, “物体検出”

http://opencv.jp/opencv-2.2/c/imgproc_object_detection.html