

タイトル	3台のKinect を搭載した移動ロボットによる特定物体のハンドリング
著者	健名，裕希；郡司，佑輔；深谷，健一；Kenmei, Yuuki; Gunzi, Yuusuke; Fukaya, Ken-ich
引用	工学研究：北海学園大学大学院工学研究科紀要(13): 25-31
発行日	2013-09-30

研究論文

3 台の Kinect を搭載した移動ロボットによる特定物体のハンドリング

健 名 裕 希*・郡 司 佑 輔**・深 谷 健 一***

Specific object handling by the mobile robot with three sets of Kinect device

Yuuki Kenmei*, Yuusuke Gunzi** and Ken-ichi Fukaya***

概要

ロボット技術を介護や家事などの生活分野に適用した生活支援ロボットの実用化が少子高齢化社会への大きな支援になると期待されており、「指定された特定の物体を取ってくる機能」が必要とされている^{1)~3)}。人間が自然な動作で音声と指差しの指示を出すと、移動ロボットが指定された特定の物体を取ってくるシステムを構築した。RGB カメラ、深度センサ、マルチアレイ・マイクロフォンを内蔵する Kinect センサを 3 台移動ロボットに搭載し、それぞれのセンサが音声認識、指差し認識、特定物体認識を分担処理し、認識結果を統合処理することで目標とする機能を実現した。

1. はじめに

Kinect (Microsoft 社, RGB カメラ・深度センサ・マルチアレイマイクロフォン搭載デバイス) を 3 台移動ロボットに搭載することにより、人間の指差し動作、音声といった自然な動作を認識させ、指示した特定物体を移動ロボットにハンドリングさせるシステムを構築し、実験を行った結果を報告する。

2. システム構成

実験システムのハードウェア外観を図 1 に、システムの構成を図 2 に示す。ハードウェアは移動ロボット (Mobile Robotics 社 Pioneer3-DX)、ノートパソコン (Windows7, 32bit)、Kinect 3 台、12 V 鉛蓄電池、DC-AC 変換器である。指示する人間とハンドリングする物体の Kinect による計測

可能な位置関係が指差し認識、音声認識、特定物体認識で異なるため、それぞれの役割を分担する 3 台の Kinect を用意する。3 段の棚からなる台を移動ロボットに搭載し、3 台の Kinect をそれぞれの段に 1 台ずつ設置する。最上段の Kinect1 には指差し認識、2 段の Kinect2 には音声認識、最下段の Kinect3 には特定物体認識を担わせる。最下段には Kinect と一緒にノートパソコン、12 V 鉛蓄電池、DC-AC 変換器 (Kinect 駆動用交流電源供給) を搭載する。3 台の Kinect とノートパソコンを USB 接続し、ノートパソコンと移動ロボットを RC-232C 接続することで 1 台のノートパソコンから 3 台の Kinect と移動ロボットを制御する。ソフトウェアとしては移動ロボットを制御する Aria と、Kinect を制御する KinectSDK を使用し、さらに後述の画像処理のために OpenCV を用いる。プログラミング言語は C++ と C# を使用する。

* 北海学園大学工学部電子情報工学科 (現在: 東京コンピュータサービス株式会社)

Faculty of Engineering (Electronics and Information Eng.), Hokkai-Gakuen University (present: Tokyo Computer Service Co., Ltd.)

** 北海学園大学工学部電子情報工学科 (現在: 株式会社札幌システムサイエンス)

Faculty of Engineering (Electronics and Information Eng.), Hokkai-Gakuen University (present: Sapporo System Science Co., Ltd.)

*** 北海学園大学大学院工学研究科電子情報工学専攻

Graduate School of Engineering (Electronics and Information Eng.), Hokkai-Gakuen University

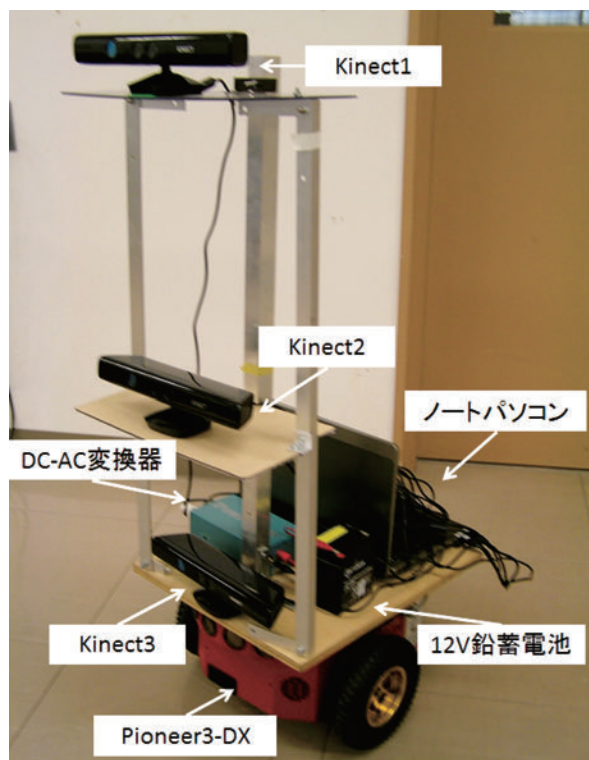


図1 実験システムのハードウェア外観

3. 音声の認識と出力

Kinectには、マルチアレイ・マイクロフォンが搭載されており、あらかじめ認識させたい単語を登録しておけばKinectに向かってしゃべった音声を登録してある単語と比較し、特定単語として認識させることができる。今回は特定物体名称を「コーラ」、「カフェオレ」、「カフェモカ」、「コーヒー」、「ガラナ」、「オレンジ」の飲み物6種類、移動ロボットに対する動作指示として移動ロボットの移動開始指示「取って」、緊急停止指示「待って」の2種類、合計8種類の単語を登録した。これにより移動ロボットに対してハンドリングする特定物体の指定、指定された指差し座標の特定物体ハンドリングの開始、指差し座標への移動中の緊急停止を音声認識により指示することができる。

また、音声出力も実装し、音声合成ソフト「VOICEROID+民安ともえ」で作成した音声をノートパソコンのスピーカから出力する。音声認識により特定物体の指定がされた場合には「〇〇（特定物体名称）ですか？」、移動ロボットへ移動開始の指示が行われると「〇〇（特定物体名称）を取ってきます」と出力され、緊急停止の指示では「停止します」と出力する。これにより音声指示が正確にKinectに認識されているかどうかを

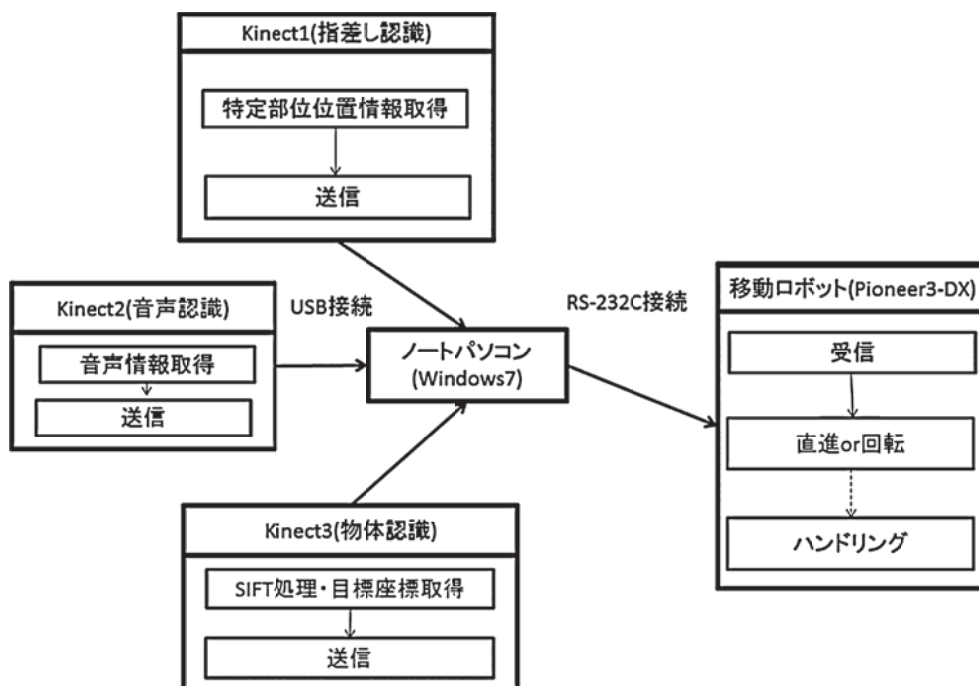


図2 システムの構成

出力された音声によって確認できる。

4. 指差し認識

Kinect は、人間の頭や手、足などの各部位の位置データをリアルタイムに最大 20 点取得するスケルトントラッキング機能を有する。システム起動時からスケルトントラッキングが開始されており、移動ロボットへの移動開始指示の音声を確認すると、各部位の中から掌 $A(x_A, y_A, z_A)$ と手首 $B(x_B, y_B, z_B)$ の 2 点の部位の位置座標を取得する。図 3 にスケルトントラッキング機能による掌と手首の検出状況を示す。図 4 に示すように床面に直交座標系を設定し、掌と手首の座標を通る直線の延長線と床面から高さ h の平面が交差する点が指差し座標 $C(x_C, y_C, h)$ となる。 (x_C, y_C) 座標は次式で求められる (座標導出は付録に示す)

$$x_C = \frac{1}{z_B - z_A} \{x_A(z_B - h) - x_B(z_A - h)\}$$

$$y_C = \frac{1}{z_B - z_A} \{y_A(z_B - h) - y_B(z_A - h)\}$$

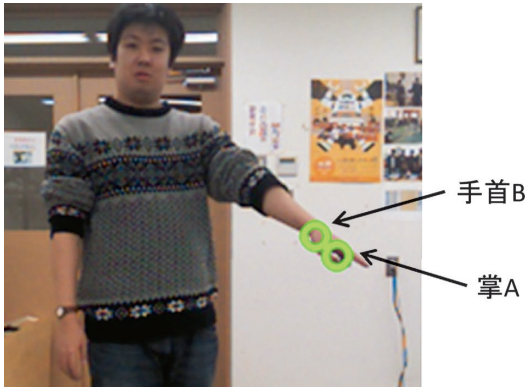


図 3 スケルトントラッキングによる掌と手首の検出

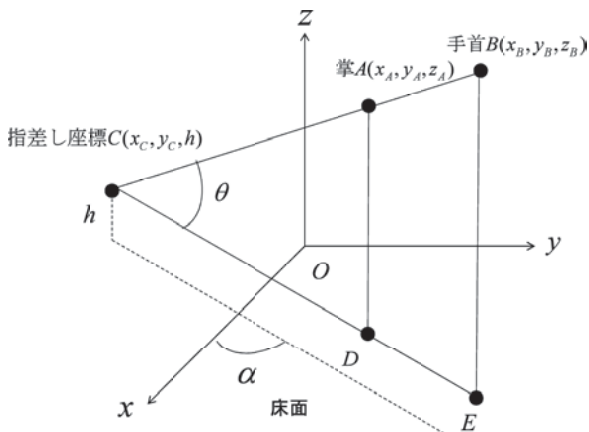


図 4 掌・手首座標と指差し座標との関係

ここで h は特定物体が置かれた台の床面からの高さとする。得られた指差し座標から移動ロボットの移動距離、回転角度を計算し、ロボットに送信する。今回の実験システムでは特定物体を床面に置いたので高さ $h=0$ となる。

5. 特定物体認識と物体ハンドリング

移動開始指示を受け、移動ロボットが指差し座標へ移動し、その過程で正面に物体があると検知した場合、次に特定物体の認識、物体のハンドリングへ移る。特定物体認識のための手法として OpenCV にも実装されている SIFT (Scale-Invariant Feature Transform) を用いる。

5.1 SIFT

SIFT は、画像中の特徴点の検出と特徴量の記述を行うアルゴリズムであり⁴⁾、検出した特徴点に対して、画像のスケール変化・回転・照明変化等にロバストな特徴量を記述する。そのため、物体認識や物体検出、画像マッチングなどに用いられる⁵⁾。

5.2 特定物体ハンドリング

移動ロボットが物体群の手前で停止した後は、SIFT を行う前準備、SIFT による特定物体認識さらに物体ハンドリングの順に処理を進める。

SIFT を行う前準備として、Kinect から取得される RGB 画像から、物体の存在する部分を切り抜き、さらに物体の縦横比を計測し、SIFT を実行する順番を決定する。これらの処理には Kinect の距離センサから取得される深度データを用いる。

Kinect に映る移動ロボット正面の空間を一定距離毎に 4 つに区分し、Kinect から取得した深度データを走査し、各区分の距離に対応する深度データが一定以上連続して存在する箇所を発見する。そのデータ群が発見された場合、その場所に対応する RGB 画像内の範囲を切り抜くための RGB 画像座標データを取得する。

こうして SIFT を実行する画像に物体背景や他の物体を含めず、一つの物体だけを最小限に収めることにより、SIFT の実行結果をより正確にする。実際に Kinect から取得された RGB 画像を

図5に、先の手順で物体の部分だけ抜き出した画像を図6に示す。さらにこのとき、発見した深度データ群の最も左端と右端、上端の座標を実距離で取得しておくことで発見した物体の縦横比を計測する。そうして得た縦横比をあらかじめ登録しておいた特定物体の縦横比の情報と照らし合わせ、近いものからSIFTを実行するようにソートする。SIFTの実行には多少の時間がかかり、無作為にSIFTを実行すると無駄な時間が頻繁に発生することになるが、ソートすることでSIFTを実行する回数を最小限にする。これらの作業を全ての物体に適用するが、この際に全ての物体の位置情報を取得し保存する。この位置情報は物体ハンドリングのために用いる。

ソートした順にSIFTを実行し、最後まで指定された物体がなかったときは移動ロボットが初期位置に戻る。指定された物体が発見された場合はそのハンドリングを行う。

先に取得した全物体の位置情報より、移動ロボットと指定された物体との間に他の物体はなく、障害がないと判断された場合は指定物体まで直進してハンドリングを行う。しかし、移動ロボットと指定物体の間に他の物体が存在し、障害があると判断された場合は全物体の位置に応じ迂回し

て指定物体のハンドリングを行う。移動ロボットの迂回は前方から、左方から、右方から、後方からの順でどの方向から指定物体へ向かうかの優先順位を決定する。それぞれの方向から見て、指定物体と左右にある物体間の距離を求め、移動ロボットが入り込める余裕があるのなら、ハンドリングに向かう。

以上の一連の手続きをフローチャートとして図7に示す。

6. 実験結果

今回は指差し認識の精度を確認するため、最大半径3mの円周上に20点のマーカを配置し、全てのマーカを1度ずつ指差しすることを1セットとし、実験を行う。マーカ配置を図8に示す。3名の被験者が各人3セット指差しを行い、1セット毎の取得座標データの実測値と理論値の誤差の平均を求め、おおよその精度を計測する。全ての指差しの誤差平均は x 座標が19cm、 y 座標が46cmとなった(表1参照)。指差しに対する誤差は個人差があるが、大きくはない。さらに1名の詳しい実験結果を表2に示す。指差した位置の角度が同じでも距離が離れていくにつれて誤差は



図5 Kinectにより取得されたRGB画像（反転表示画像）



図6 抽出された物体画像

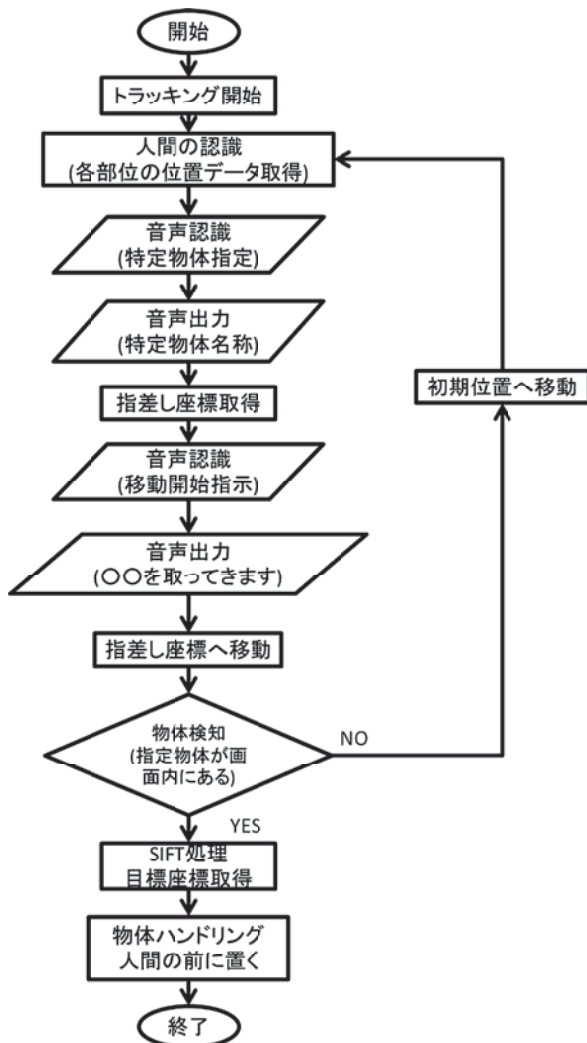


図7 処理のフローチャート

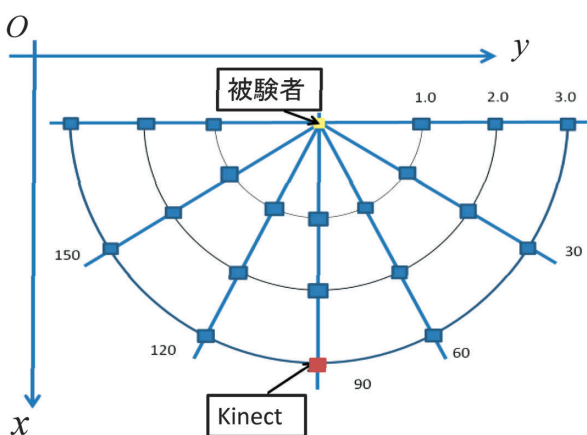


図8 指差し実験マーカ配置図

大きくなっている。特に顕著に表れているのは距離が3mの場合である。この傾向はどの参加者、どの角度においてもほぼ同様であった。これは人間が遠くを指差した場合、腕全体が Kinect と平

表1 指差し座標の誤差

	誤差平均(cm)	
	x 座標	y 座標
被験者 1	20	42
被験者 2	16	38
被験者 3	19	56
平均	19	46

表2 詳細指差し座標の誤差

被験者位置		誤差[cm]	
角度[度]	距離[m]	x 座標	y 座標
30	1	3	11
30	2	7	16
30	3	16	92
150	1	9	18
150	2	7	44
150	3	43	65

行に近づき、手首が手に隠れることにより Kinect が手と手首を判別し位置を割り出すことが困難になっていること、さらに人間がはっきりとその位置を指差することができていないことが原因と考えられる。

一方、特定物体認識では、Kinect の距離センサから取得される深度データが正確に得られず設置された物体を発見できないことがあるが、発見さえできていれば抜き出した画像からの SIFT による特定物体認識はほぼ成功する。先の指差しの精度がさほど高くないため、移動ロボットの移動位置にずれが生じ、Kinect の視野角内の物体群全てが映らない問題も発生した。

特定物体認識後の物体ハンドリングについては、移動ロボット正面の物体全てを発見できていれば、問題なく指定物体をハンドリングできていたが、一部物体を発見できていない場合は指定物体のハンドリングはできても、他の物体を倒してしまうことがあった。その後、人間の前まで物体を運び、初期位置に戻るのは概ね問題はなかった。移動ロボットへ指示した移動距離、角度と実際に移動する距離、角度には若干の誤差があるため、最終的に初期位置に戻った時に数 cm の誤差が発生した。

以上の一連の動作の状況を図9に示す。

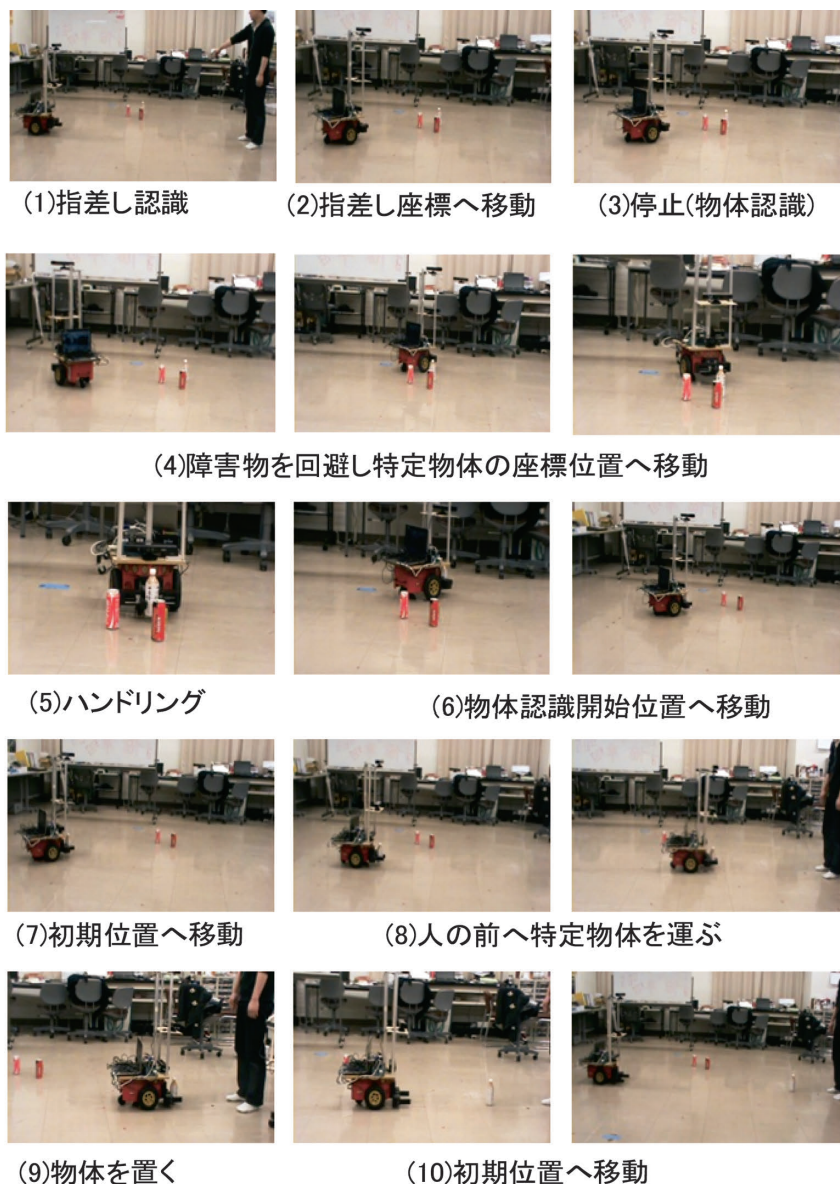


図9 システム稼働状況

7. 結論

3台の Kinect を搭載した移動ロボットによる特定物体ハンドリングの実験システムを構築した。人間の自然な動作で指示を出し、移動ロボットに特定物体をハンドリングさせる動作を可能とした。しかしその精度はあまり高くなく、特に指差しに関しては人間とロボットの距離関係の影響もあるが、現在設定している 3 m では人間が意図した場所に正確に移動させるには至らなかった。指差し座標の取得では今後さらに正確な値を導くためのデータの扱いを検討すべきである。

特定物体認識に関しては Kinect の最大解像度は 1280×960 であるが、SIFT の実行には大きな

差し障りはなかった。しかし物体の種類が増えた時の影響は今後画像処理に関して考慮が必要と思われる。Kinect に全ての物体が映らないという問題は、Kinect をさらに増やし視野角を広げることによって解決可能である。

物体ハンドリングでは、現在のシステムで他の物体を迂回しながら指定物体を取得することはできているが、そのための移動範囲が広いため、効率的な移動の改良が必要である。

人間が特殊な機器を持つ、特殊な動作をするなどの制約を必要とせず、人間の自然な動作により特定物体を指示し、移動ロボットに「指定された特定の物体を取ってくる機能」を実装できた。

付録 指差し座標の導出 (図 4 参照)

B , A , C を結ぶ線分を斜辺とする直角三角形 BCE を想定する。

$$\tan \theta = \frac{z_B - z_A}{\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}}$$

$$\begin{aligned} CD &= \frac{z_A - h}{\tan \theta} \\ &= \frac{z_A - h}{z_B - z_A} \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \end{aligned}$$

$$\cos \alpha = \frac{x_B - x_A}{\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}}$$

$$\sin \alpha = \frac{y_B - y_A}{\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}}$$

$$\begin{aligned} x_C &= x_A - CD \cos \alpha \\ &= \frac{1}{z_B - z_A} \{x_A(z_B - h) - x_B(z_A - h)\} \end{aligned}$$

$$\begin{aligned} y_C &= y_A - CD \sin \alpha \\ &= \frac{1}{z_B - z_A} \{y_A(z_B - h) - y_B(z_A - h)\} \end{aligned}$$

謝辞

本研究は、文部科学省私立大学戦略的研究基盤形成支援事業 (平成 20 年～平成 25 年) の支援を受けて行われた。

【参考文献】

- 1) 永田和之, 脇田優仁, 小野栄一: 人が指示した物を取ってくれる日常生活支援ロボット, 日本機械学会福祉工学シンポジウム 2007, MG313, pp.235-238, 2007.
- 2) 山本貴史, 齋藤史倫, 橋本国松 他: 生活支援ロボット HSR の開発, 日本ロボット学会第 30 回記念学術講演会, 3C2-1, 2012.
- 3) 塩濱教幸, 深谷健一: SIFT を用いた特定物体認識の高速化と自律移動ロボットによる物体ハンドリングへの応用, 第 4 回日本ロボット学会北海道ロボット技術研究専門委員会学術講演会, pp.120-122, 2012.
- 4) D. G. Lowe: Object recognition from local scale-invariant features., Proc. of IEEE International Conference on Computer Vision (IJCV), 60(2), pp.363-370, 1984.
- 5) 藤吉弘亘: Gradient ベースの特徴抽出 — SIFT と HOG —, 情報処理学会 研究報告 CVIM 160, pp. 211-224, 2007.
- 6) 黒田尚考, 廣井富, 三宅真司: ロボットアバタを用いた指差し行為の移動ロボットへの実装, 日本ロボット学会第 30 回記念学術講演会, 1H2-3, 2012.