

ビンピッキングのためのRGB-Dカメラを用いた三次元位置姿勢推定，および把持可能性を考慮したスコアリング手法

西 卓郎^{1,a)} 吉見 隆¹ 高瀬 竜一¹ 原田 研介¹ 永田 和之¹ 新良貴 陽平² 河井 良浩¹

概要：ホビー用として普及している安価なRGB-Dカメラより得られる距離画像を対象として，大量の三次元特徴点を用いることでバラ積み物体の個々の位置姿勢を安定して検出する手法を提案する．ランダムサンプリングに加えCADモデルからあらかじめ類似姿勢を計算しておくことで，大局的最適解を推定可能となった．また，CADモデルに把持点の情報を付加することにより，対象物近辺のマニピュレータの干渉空間を計算することで複数の対象物の中から実際に把持可能な対象物の位置姿勢，把持点の位置，エンドエフェクタの姿勢を決定することが可能となり，複雑な協調制御を行うことなしに実用的なビンピッキングシステムが構築できることを示した．

1. はじめに

製造現場におけるビンピッキングでは，対象物の位置姿勢の検出を視覚部に要求されることが多い．これに応えるため，現在までに，対象物の境界線の特徴量として用いる手法 [1][2] や，点群の局所特徴量を用いる手法 [3][4][5][6] などが提唱され，一定の成果を上げている．しかしこれらの手法の多くは，シーンの三次元形状がある程度安定して得られることを前提としており，複雑で高価な撮像装置，設備の導入や環境の整備が必要となつて，導入，運用にかかるコストが大きくなる問題がある．

また，実際のシステムにおいては，物体の位置姿勢に加え，それが安定して把持可能であるかどうかの検証が必要とされる．これに関しては現在までに，マニピュレーションの問題として，把持可能点の探索により動作計画を行う [7]，対象物を円筒で近似し干渉チェックを行う [8] などのアプローチがなされてきているが，現実には実装の困難さなどからヒューリスティックな手法で解決されることも多い [9][10]．しかし対象物の位置姿勢検出と同時に周囲に存在する障害物の位置情報も取得可能であることから，把持可能性についてもビジョン側で処理することでこの問題を効率的に解決できるものと期待できる．

本報では，ホビー用として普及している安価なRGB-Dカメラを用いて取得したノイズの多い距離画像から安定し

てバラ積みされた物体の個々の位置姿勢を検出できるアルゴリズムの開発し，検出した対象物の位置姿勢がマニピュレータで安定して把持可能なことを考慮したスコアリング手法を提案する．

2. タスクとシステム構成

本研究では製造現場からの要請に基づき，バラ積みの状態の5種類の形状の部品 (図 1-a.) を治具上 (図 1-b.) に組み上げ，図 1-c. の状態にすることをタスクとして設定した．

3種類のパイプおよび2種類のプレートからなる金属製の各対象部品は混在することなく別々の箱にバラ積みで入れられており，これを上方からRGB-Dカメラ (図 1-1.) によって撮像して位置姿勢の推定を行い，その結果に基づいて双腕ロボットの右腕に取り付けられた吸着パッド (図 1-2.) により部品を取り出し，左手の組み付け用三指ハンド (図 1-3.) に持ち替えた上で治具への組み付けを行う．

本報ではこれら一連のタスクのうち，特にバラ積み状態の部品の取り出しに関する工程について述べる．

3. バラ積み物体の位置姿勢推定

バラ積み状態の対象物の三次元形状は市販の安価なRGB-Dカメラを用いて計測する．しかしこの距離画像はノイズや欠落を多く含み，従来の手法ではマニピュレータによる把持を行うための精度で位置姿勢の推定を行うことが困難であった．本研究では，距離画像に対して大きなウィンドウを使用して二次曲面を当てはめることでノイズや欠落の改善を図ると共に，三次元特徴量として，シーンおよびモデルを構成する各点の三次元座標値と，観測方向に依存し

¹ 独立行政法人産業技術総合研究所
AIST, Tsukuba, Ibaraki 305-8586 Japan

² 東京都立大学大学院
Tokyo City University, Setagaya, Tokyo 158-0087 Japan

a) zophos@ni.aist.go.jp

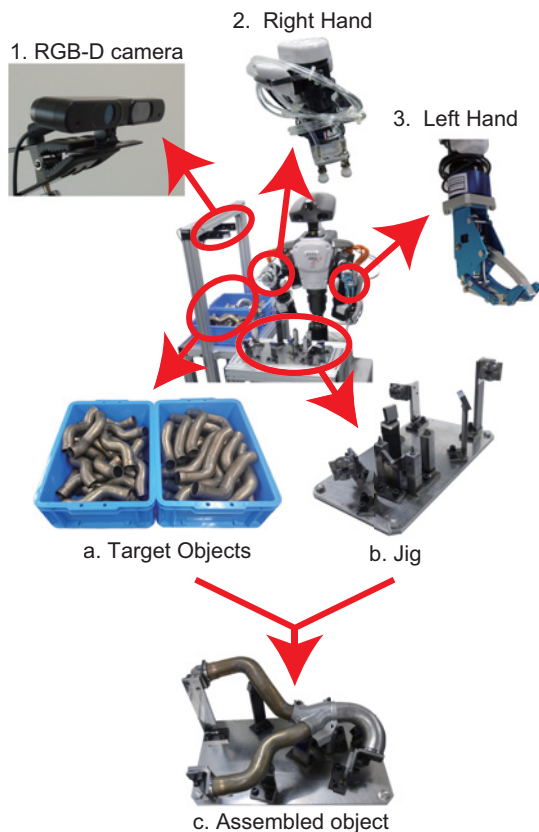


図 1 システム全体図

Fig. 1 Systems overview

ない不変特徴量として古くから知られる、主方向ベクトルおよび最大曲率、最小曲率の 14 次元のベクトルを用いることで、ノイズに対して頑強となるモデル照合を可能とする手法を開発した [11].

3.1 認識モデルの作成

STL に代表される CAD モデルは一般的に点の密度が一般様では無い。本報では、CAD モデルを一定密度の点群データとして展開し、モデルの複雑さに応じて定める半径 (量子化ピッチ) を持つウィンドウ内にある点を対象として、主成分分析を行うことで第 1 主成分 u 、第 2 主成分 v 、および u 、 v の外積 w を求めて式 (1) の曲面へのあてはめを行い、主方向ベクトルおよび最大/最小曲率の計算を行った。

$$w = Au^2 + Bv^2 + Cuv + Du + Ev + F \quad (1)$$

ウィンドウ内に存在する点が一定数より少ない、またはあてはめ誤差が一定値より大きい場合には、欠損値として照合に用いないようにした。

3.2 シーンの作成

3.2.1 特徴量の計算

シーン各点の主方向ベクトル、および最大/最小曲率は、距離画像上でウィンドウ処理を用い、式 (2) で表されるパラメトリック二次曲面をあてはめることで求めた。

$$x_i = A_i s^2 + B_i t^2 + C_i st + D_i s + E_i t + F_i \quad (2)$$

ここで、 x_i ($i: \{0, 1, 2\}$) はシーン各点の三次元座標値、 s, t は二次元画像上の col, row 座標値を表す。

この処理は、特徴量の計算と同時にシーン中のノイズや欠損を低減する効果も持つ。ただしウィンドウ処理により物体の境界付近の形状が大きく変化してしまうことを避けるため、曲面あてはめは、あらかじめ二次元画像上で距離エッジを用いて領域分割処理を行い、注目画素と同一領域に属するウィンドウ内の点だけを使って行うこととした。なおウィンドウサイズは、モデルの代表的な曲率半径と入力デバイスの特性から決定することとした。

3.2.2 シーンの領域分割

モデル照合の対象となるシーン中の点の数を制限するために、先述の距離エッジに加え、主方向ベクトルエッジおよび曲面の分類を用いてシーンの領域分割を行う。ただし曲面の分類は、一般的なガウス曲率、平均曲率を用いる方法ではなく、Shape Index[12] と呼ばれる方法に基づき、最大曲率 C_{max} 、最小曲率 C_{min} を式 (3) により極座標変換して得られる C_r および C_θ から、凸: $-3/4\pi \leq C_\theta < -\pi/4$ 、鞍型: $-\pi/2 \leq C_\theta \leq 0$ 、凹: $-\pi/4 < C_\theta \leq \pi/4$ 、平面: $0 \leq C_r \leq Th_F$ とした。ここで Th_F は入力デバイスの特性に合わせた閾値であり、概ね $0.002 \leq Th_F \leq 0.01$ 程度の値を取る。

$$\begin{cases} C_\theta = \tan^{-1} \frac{C_{min}}{C_{max}} \\ C_r = \sqrt{C_{max}^2 + C_{min}^2} \end{cases} \quad (3)$$

これにより、シーン中の 1 点は、最大 3 つの領域に同時に所属することになる。

3.3 モデル照合

三次元座標と主方向ベクトルが既知の場合、モデル上の一点 (モデル点) とシーン中の一点 (シーン点) の組み合わせ (ペア) を決定することで、モデルの位置姿勢を最大/最小曲率方向ベクトルの符号方向を除いて一意に決定できる。提案手法ではこのことを利用し、シーンの領域毎に最適なペア (シードペア) を RANSAC[13] により求めることでモデル照合を行う。位置姿勢の評価は、シードペアにより決定された位置姿勢をとるモデルの全点をシーンに投影し、a) 法線方向が視線方向に対して 90 度以内、b) 三次元座標値が投影先シーン点と一定距離以内、c) 主方向ベクトルおよび曲率の投影先シーン点とのコサイン距離が一定値以上、の三つの条件をすべて満たす点を数え上げることで行った。ただしシードペアを構成するモデル点はシーン点と同じ曲面の型を持つものから選ぶようにした。また評価値が一定値以上のシードについて、現在のシーン点の周辺で評価値がより高くなる点を再帰的に探索することで近傍の点の中から最適な姿勢を求めるようにした。

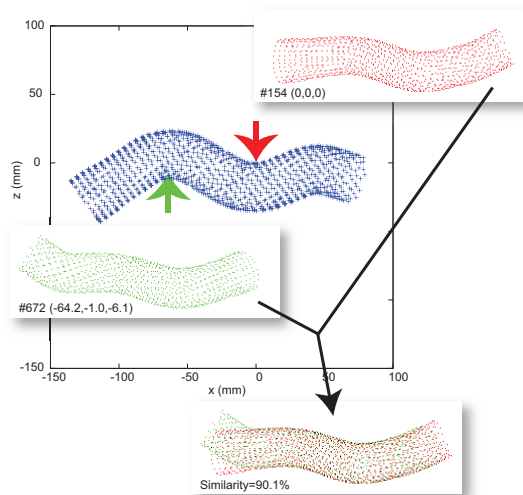


図 2 類似姿勢の例

Fig. 2 Example of similar pose

3.4 類似姿勢の探索

対象物の形状によっては、図 2 に例示するように、異なる 2 姿勢間の投影画像が似通った形状となることがあり、ここまで述べた手法のみでは、このような類似姿勢による局所解を排除することが難しい。

これを解決するため、本研究では、事前にモデルのシルエットから類似姿勢データベースを作成し、各領域の位置姿勢の推定結果についてデータベースからスコアが現在より高くなる姿勢を検索することとした。

類似姿勢データベースは、任意の二つのモデル点を中心として、それぞれの点の主方向フレームが一致するようモデルを回転、平行移動して二次元平面に投影し、姿勢のシルエットが 90%以上重複するものを、類似姿勢を取り得るモデル点としてリストアップすることで作成した。ただし、一つのモデル点の近傍の点は相互に類似する姿勢になることが多く、またモデル照合時に近傍探索を行うため、2 点間の直線距離が閾値以下または法線方向の内積が閾値以上の点については類似姿勢探索を行わないようにしている。

位置姿勢推定後の類似姿勢検索では、シードペアのモデル点をキーとしてデータベースを検索し、類似姿勢をとる点を発見した場合には現在シードペアのモデル点を発見したモデル点と置き換えて位置姿勢の評価を行い、評価値の高い方のモデル点を採用することを再帰的に繰り返すことで、大域的最適解を求めるようにした。

4. 把持可能性に基づくスコアリング

マニピュレータによる部品の把持を安定して行うためには、対象物の位置姿勢のみならず、対象物の形状と把持を行うエンドエフェクタの形状、把持方法の組み合わせにより決定される、把持動作後の対象物の姿勢が安定する「把持点」に、マニピュレータが接近(アプローチ)可能となる必要がある。また、把持後の対象物の持ち上げ(リフト

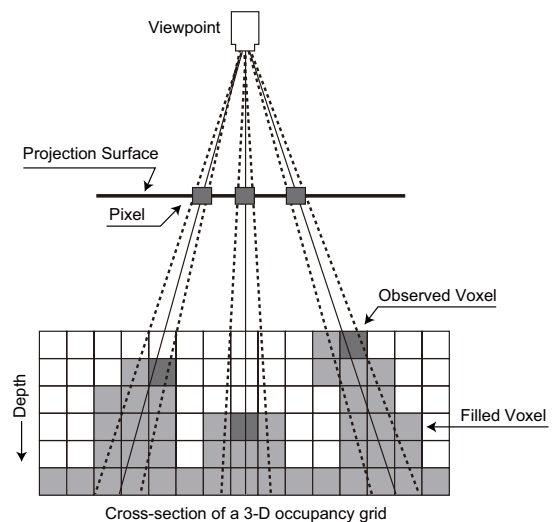


図 3 3-D occupancy grid の作成方法

Fig. 3 Building method for a 3-D occupancy grid

アップ) 動作中に他の物体が干渉しないことも必要とされる。加えて、マニピュレータが待機姿勢から把持点手前のアプローチ開始点まで、なるべく無理のない動作で移動可能であることも重要となる [14][15]。

本研究では、1) 対象物の把持点は CAD モデルからあらかじめ計算により求められている、2) リフトアップ経路は把持点から事前に一意に決定できる、3) 待機姿勢からアプローチ開始点までのマニピュレータの経路は動作計画により自動的に計算される、の 3 つを前提条件として、把持およびリフトアップが可能なが多いほど把持に適した姿勢であるとみなし、対象物の把持点へのアプローチおよびリフトアップ経路上の障害物空間を 3-D occupancy grid[16]と呼ばれる手法により作成して、アプローチおよびリフトアップの際にマニピュレータおよび対象物が干渉しない把持点の数を数え上げることでスコアリングを行うこととした。

4.1 3-D occupancy grid の作成

3-D occupancy grid は、三次元格子構造をもつ作業空間を用意し、図 3 に示すように、二次元投影画像の画素ごとに、あらかじめ設定した最遠点から画素の持つ三次元位置まで、視点からの距離を一定ピッチで分割して占有空間を計算し、ボクセルを積み上げていくことで作成する。

これにより面のみ構造である距離画像を密な構造に変換することで、単一視点からなるいわゆる 2.5 次元構造の距離画像では不可能な、視線方向に対して水平に近い面を持つ物体に対する干渉チェックが可能となる。

4.2 マニピュレータの干渉チェック

上述の方法で作成した作業空間中でアプローチおよびリフトアップ動作時のマニピュレータの干渉チェックを行うために、本研究ではエンドエフェクタの形状およびアプ

ローチ、リフトアップ動作の軌跡を円筒で近似することとした。

近似円筒はマニピュレータの手先座標系で作成され、先述の方法で推定されたモデルの位置姿勢毎に、各把持点に設定された把持姿勢にあわせて回転、平行移動を行い、カメラ座標系に座標変換して 3-D occupancy grid 内に配置される。この配置された近似円筒内にボクセルが存在する場合、マニピュレータに干渉する物体がシーン中に存在すると見なすこととした。

また、エンドエフェクタが把持対象物と接する部位(コンタクトエリア)を近似円筒内に設定し、その中にボクセルが含まれていない場合には、位置姿勢推定が誤っているものと見なし、設定した点数以上の把持点で誤推定と判定された位置姿勢については把持候補から除外するようにした。

4.3 対象物の干渉チェック

形状が比較的一様なエンドエフェクタと異なり、多様な形状を持つ対象物およびその軌跡を円筒で近似することは難しい。本研究ではリフトアップ時の干渉チェックについて、3-D occupancy grid から推定された位置姿勢を持つ対象物を取り除いた作業空間中で、モデルをリフトアップ動作の軌跡にあわせて一定幅で移動させ、モデルを構成する各点の位置での作業空間中のボクセルの数を数え上げることで行うこととした。

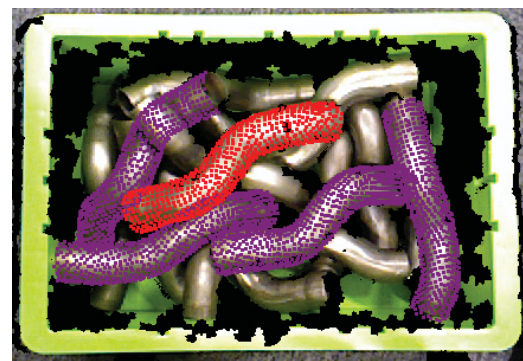
この手法は円筒近似に比較して計算コストが大きくなるため、円筒近似によるマニピュレータの干渉チェックにより把持可能とされた把持点に関してのみ対象物の干渉チェックを実行することで、計算時間を削減するようにした。

5. 実験結果および考察

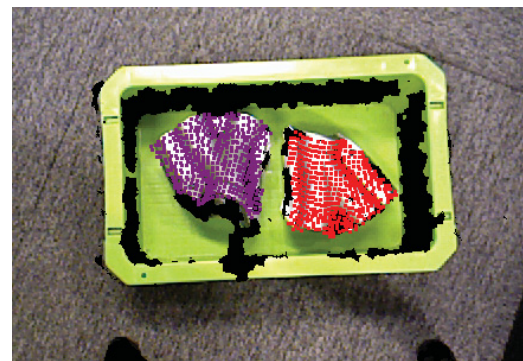
5.1 位置姿勢推定

佐藤らの手法 [17] により歪み補正済みの PrimeSense 社製 RGB-D カメラ Xtion を用いて撮像した約 650mm 先のバラ積み状態のパイプおよび凹凸のある板状部品に対して、提案手法を適用した結果を図 4 に示す。認識モデルは STL フォーマットの CAD モデルから量子化ピッチ 5mm で作成し、シーンは 31x31pixel のウィンドウサイズ(実寸で約 30mm 相当)で特徴量の計算を行った。この結果より明らかなように、本手法は、対象物の形状によらず、ノイズに対し高い頑強性を持つとすることができる。これはシーンのウィンドウ処理による効果のみならず、照合処理で位置姿勢の候補の選出、評価に大量のシーン点を用いることで、ランダムノイズを統計的に抑え込むためであると考えられる。

次いで類似姿勢探索の有効性を確認するため、図 5 に、図 4-a) の中央部の対象物の姿勢推定結果を人為的に置き



a) Pipes (found 5 objects)



b) Plates (found 2 objects)

図 4 位置姿勢推定結果

Fig. 4 Pose estimation results

換え、構築したデータベースから類似姿勢を検索した結果を示す。

類似姿勢データベースは、視線方向は常にモデル点の法線方向と一致しているものと仮定し、モデル点間の距離が 30mm 以上もしくは法線方向の角度が 30 度以上異なる点同士で作成した。この結果モデル点 1168 個のうち 338 組の組み合わせが類似姿勢となった。

モデル点 #633 を採用した誤姿勢(図 5-a)、図 5-c)の赤の点)の姿勢推定スコアは 179 であったのに対し、同点の類似姿勢検索結果であるモデル点 #655 を採用した正しい姿勢(図 5-b)、図 5-c)の緑の点)のスコアは 295 となった。類似姿勢データベース中でモデル点 #633 と #655 の姿勢類似度は 92.3%と評価されており、実際に両姿勢間の相違は三次元画像を用いても視認することが難しい。しかし、図 5 の白丸で囲んだ部分に注目すれば、モデル点 #633 を採用した図 5-a)より、モデル点 #655 を採用した図 5-b)の方が正しい姿勢を示していることがわかり、類似姿勢においてもモデル照合における姿勢評価が正しく機能していることがわかる。

このことから明らかなように、事前に類似姿勢データベースを作成し、姿勢推定結果を類似姿勢と比較する手法は、局所解として出力された誤った姿勢を全点の照合を行うことなしに修正するために有効であると言える。

ただし類似姿勢探索を用いても、位置姿勢の推定結果にパイプの長軸方向に最大 10mm 程度、垂直方向に最大 5mm

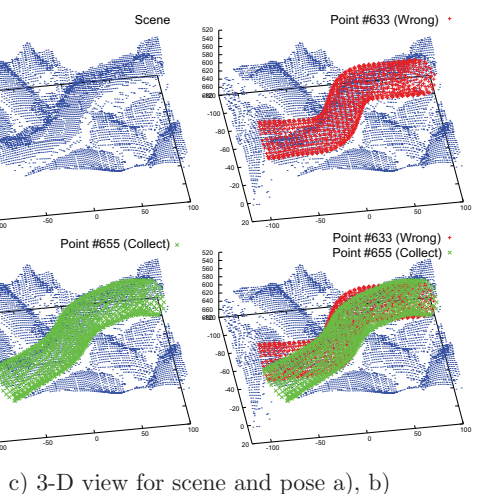
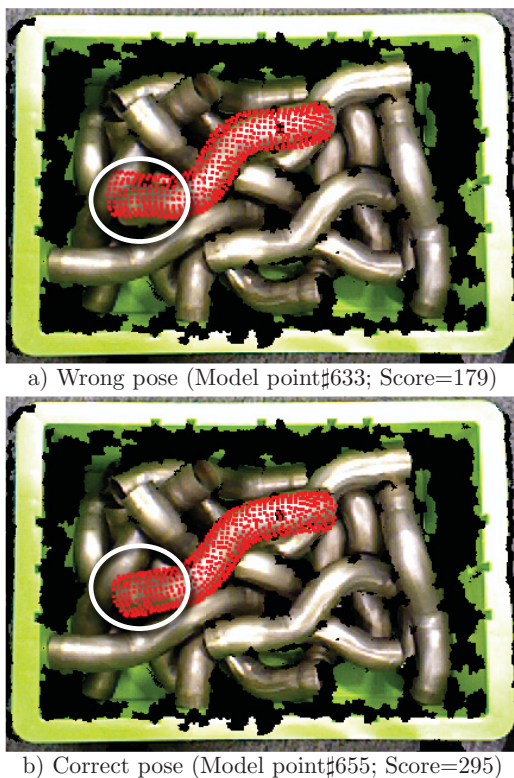


図 5 類似姿勢検索結果
 Fig. 5 Result of Pose-similarity search

程度の誤差が生じることがあった。これについては、本報での提案手法の後段に ICP[18][19] などの精密な位置合わせ手法を導入することで対処可能であると考えられる。

5.2 把持可能性に基づくスコアリング

図 6 に図 4-a) の 3-D occupancy grid を 5mm ピッチで作成した結果を、図 7 にはマニピュレータモデルと把持点情報をそれぞれ示す。

図 6 では視線方向に対してほぼ平行なる箱の壁面などが台形状に補間されていることがわかる。

図 7-a) は、厚み 80mm のアームの先端を原点として半径 7.5mm の吸着パッド 2 本を持った厚み 50mm, 直径 75mm のエンドエフェクタを装着し、手先座標系 X 軸と逆方向

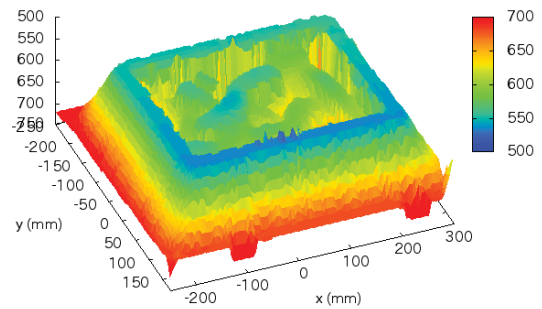


図 6 図 4-a) の 3-D occupancy grid 作成結果
 Fig. 6 3-D occupancy grid for Fig. 4-a)

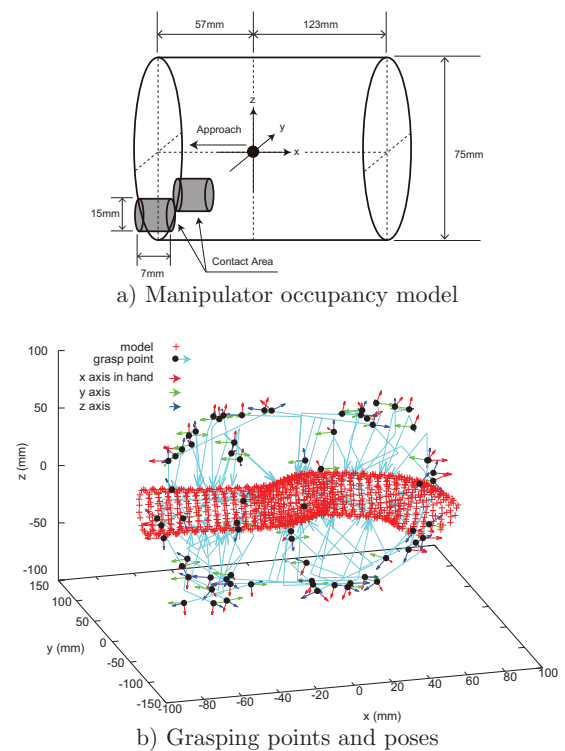


図 7 マニピュレータモデルと把持点情報
 Fig. 7 Manipulator occupancy model and grasping points

に 50mm のストロークでアプローチを行うことを示している。また把持点に到達後、吸着のために 7mm の押し込み動作を行うことから、吸着パッドに相当する 2 本の半径 7.5mm, 長さ 7mm の円筒空間をコンタクトエリアとして設定している。

図 7-b) では、エンドエフェクタと対象物の干渉およびリフトアップ動作時の対象物の姿勢の安定性を考慮し、把持点として 70 点の位置とエンドエフェクタの姿勢が計算されている [20]。またリフトアップ方向は物体および把持点の位置姿勢に関係なく、全て視線方向と一致するものとしている。

これらの情報を用いて、図 4-a) の推定結果に対し、把持可能性に基づくスコアリングをおこなった結果を図 8 に示す。なお図中のスコアは到達可能と見なされた把持点の数を表す。

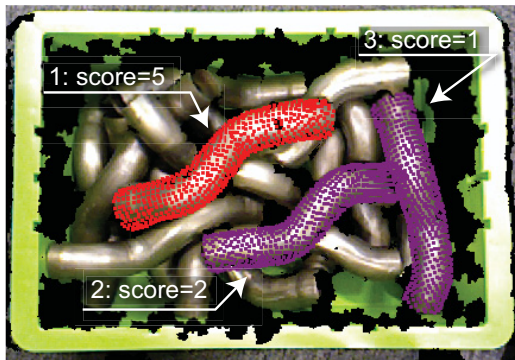


図 8 把持可能性に基づくスコアリング結果

Fig. 8 Matching results using graspability information

図 4-a) では 5 個の候補がリストアップされているが、図 8 では 3 個に減っている。これは、削除された候補について、アプローチ時に箱の側面や他の対象物にマニピュレータが干渉し、全ての把持点に到達できなかったためである。

実際の把持動作を川田工業製 HiroNX を用いて行ったところ、物体の位置姿勢および把持可能と推定された把持点の位置を動作計画プログラム Choreonoid graspPlugin[21]に渡すことで、良好な結果を得ることができた。

以上より、物体の位置姿勢推定を行う際に把持可能性を考慮することで、動作計画プログラムに二次元画像や距離画像を渡すなど複雑な協調制御を行わなくとも実用的なビンピッキングが可能になるということが言える。

6. おわりに

本報では、主方向ベクトルと曲率を用いた三次元モデル照射手法に加え、類似姿勢データベースの作成することで、対象物の形状によらずシーンのノイズや欠損に対して頑強な位置姿勢推定が可能となることを示した。また、位置姿勢推定の際に、把持点へのマニピュレータの到達可能性を考慮したスコアリングを行うことで、ヒューリスティックな処理やビジョンと動作計画プログラムの間で複雑な協調制御を行うことなしに、実用的なビンピッキングが可能となることを示した。

今後、定量的な精度評価および他の手法との比較を行うことで、本手法の実用化を進めていく予定である。

参考文献

[1] Sumi, Y., Kawai, Y., Yoshimi, T. and Tomita, F.: 3D Object Recognition in Cluttered Environments by Segment-Based Stereo Vision, *International Journal of Computer Vision*, Vol. 46, pp. 5–23 (2002).

[2] 丸山健一, 河井良浩, 富田文明: 遮蔽輪郭線を用いた物体モデルに基づく 3 次元物体位置姿勢推定, *電気学会論文誌 D*, Vol. 131, pp. 505–514 (2011).

[3] Feldmar, J. and Ayache, N.: Rigid, affine and locally affine registration of free-form surfaces, *International journal of computer vision*, Vol. 18, pp. 99–119 (1996).

[4] Johnson, A. E. and Martial, H.: Using spin images for efficient object recognition in cluttered 3D scenes, *PAMI on IEEE*, Vol. 21, pp. 433–449 (1999).

[5] Rusu, R. B., Blodow, N. and Beetz, M.: Fast point feature histograms (FPFH) for 3D registration, *ICRA '09 on. IEEE*, pp. 3212–3217 (2009).

[6] 秋月秀一, 橋本 学: 特徴的 3-D ベクトルペアを用いたばら積み部品の高速位置姿勢認識, *電気学会論文誌 C*, Vol. 133, pp. 1853–1854 (2013).

[7] C. Dupuis, D., Léonard, S., Baumann, M. A., Croft, E. A. and Little, J. J.: Two-Fingered Grasp Planning for Randomized Bin-Picking, *Robotics: Science and Systems Workshop-Robot Manipulation: Intelligence in Human Environments* (2008).

[8] Harada, K., Nagata, K., Tsuji, T., Yamanobe, N., Nakamura, A. and Kawai, Y.: Probabilistic Approach for Object Bin Picking Approximated by Cylinders, *2013 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3727–3732 (2013).

[9] Ghita, O. and Whelan, P. F.: A bin picking system based on depth from defocus, *Machine Vision and Applications*, Vol. 13, pp. 234–244 (2003).

[10] Sanz, P. J., Requena, A., esta, J. M. I. and Pobil, A. P. D.: Grasping the not-so-obvious: vision-based object handling for industrial applications, *Robotics & Automation Magazine*, Vol. 12, pp. 44–52 (2005).

[11] 西 卓郎, 吉見 隆, 原田研介, 河井良浩: ノイズに頑強な三次元モデル照射手法, 第 17 回画像の認識・理解シンポジウム (MIRU2014), pp. SS3–24 (2014).

[12] Koenderink, J. J. and van Doorn, A. J.: Surface shape and curvature scales, *Image and vision computing*, Vol. 10, pp. 557–564 (1992).

[13] Fischler, M. A. and Bolles, R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, Vol. 24, pp. 381–395 (1981).

[14] Bohg, J., Morales, A., Asfour, T. and Kragic, D.: Data-Driven Grasp Synthesis—A Survey, *IEEE Trans. on Robotics and Automation*, Vol. 30, pp. 289–301 (2014).

[15] Nagata, K., Miyasaka, T., Nenchev, D. N., Yamanobe, N., Maruyama, K., Kawabata, S. and Kawai, Y.: Picking up an Indicated Object in a Complex Environment, *Proc. of 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'10)*, pp. 2109–2116 (2010).

[16] Elfes, A.: Sonar-Based Real-World Mapping and Navigation, *IEEE Journal of Robotics and Automation*, Vol. 3, pp. 249–265 (1987).

[17] 佐藤雄隆, 岩田健司, 永見武司, 竹内啓五: RGB-D カメラから得られる Depth データの歪み補正, *ビジョン技術の実用ワークショップ (ViEW2013) 講演論文集 (CD-ROM)*, pp. IS1–F6 (2013).

[18] Chen, Y. and Medioni, G.: Object modelling by registration of multiple range images, *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2724–2729 (1991).

[19] Besl, P. J. and McKay, N. D.: Method for registration of 3-D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239–256 (1992).

[20] Harada, K., Kaneko, K. and Kanehiro, F.: Fast grasp planning for hand/arm systems based on convex model, *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 1162–1168 (2008).

[21] 辻 徳生, 原田研介: graspPlugin for Choreonoid, *日本ロボット学会誌*, Vol. 31, No. 03, pp. 232–235 (2013).