機能属性の空間配置に着目した類似形状物体の 6 自由度姿勢推定 *

秋月 秀一 **

青木 義満 **

6DoF Pose Estimation for Similar Objects Using Spatial Relationship of Part-Affordance

Shuichi AKIZUKI and Yoshimitsu AOKI

In this paper, we propose a method to estimate 6D pose (3D position and pose) of the everyday objects, especially Kitchen or DIY tools, even if there are not the same 3D model of the target object. We assume that the tools that belong to same category have common spatial relation in part-affordances, the role of each part. The proposed method uses this assumption for pose estimation. By minimizing two types of consistencies, the spatial relationship of part-affordances and the shapes, we have confirmed that the proposed method achieved higher performance compared with traditional pose estimation method.

Key words: 3D object recognition, pose estimation, affordance, point cloud processing

1. 序 論

近年、ロボットアームによる作業の自動化は産業分野のみならず、生活支援の分野においても活躍の場を広げつつある。生活支援用途のロボットに期待される重要なタスクの一つは、人間と同じように DIY や調理用途の道具を人間と同様に取り扱うことである。これを実現するためには、ロボットの目の前に置かれている対象となる道具 (以下、本論文では人間が調理用途や DIY 用途で利用する道具を指す) の持つべき位置や、叩くや注ぐ等の道具特有の機能を作用させる位置、さらにはその方向を認識する必要がある。したがって、操作対象となる道具の 6 自由度姿勢 (回転・並進のパラメータ、以下、単に姿勢と記述する。) を推定することは有効な手段の一つである。

一般的に物体の姿勢は、対象物の CAD 等の 3D 形状モデルをリファレンスとしたモデルマッチング^{1)~5)}によって推定される。モデルマッチングは現物と同一物体の 3D モデルを必要とする性質上、この条件が成り立ちやすい製造等の産業用途に利用されるケースが多かった。すなわち、製造業では取り扱う部品の CAD モデルは製品の設計時に製作されるので、モデルマッチングとの相性が良いシチュエーションであるといえる。しかしながら、対象物を道具に変更することを考えると、通常のモデルマッチングでは姿勢を推定することが困難になるという問題点が発生する。道具の 3D モデルはエンドユーザのもとに配布されることがほとんどなく、現物に対応した 3D モデルを利用することができないからである。

本研究の目的は、現物の 3D モデルが利用できない場合においても物体の姿勢を推定することである。対象物が道具であれば、同一カテゴリの道具は意匠が異なっていたとしても、各部位の役割の配置には共通性があると考えられる。提案手法ではこれを姿勢推定の手がかりとする。

提案手法では、対象物と同一カテゴリの類似物体の形状モデ

ルを利用できることを前提とし、これに対して各部位ごとに機能属性をラベル付けしておく。このデータをリファレンスとして、物体モデルとシーン中の物体の形状および機能属性同士の配置を利用したコストを最小化することによって姿勢変換パラメータを推定する。

提案手法では、1 つの 3D モデルにさえ機能属性や把持方法を関連付けておけば、そのまま現物の物体を取り扱うことが可能なので、対象物ごとにモデルデータを用意する必要がなくなるという利点がある。

以下、第 2 章では、モデルマッチングの関連研究について概観する。第 3 章で本研究で着目する機能属性の概念と、同一カテゴリ物体が持つ機能属性の空間配置の共通性に関して説明し、これを手掛かりとする新たなモデルマッチングアルゴリズムを第 4 章にて説明する。第 5 章にて提案手法の効果を検証する実験を公開データベースにて実施し、第 6 章で本研究成果をまとめる。

2. 関 連 研 究

対象物の姿勢は物体モデルと入力シーン間の誤差を最小化する姿勢パラメータを推定することによって算出される。ここで重要なことは、物体モデルと入力シーンの誤差関数の定義方法である。

最も単純な誤差関数は物体モデルと入力シーンを 3 次元の点群データで表現しておき、それらの二乗距離を最小化する手法である。これには ICP アルゴリズム¹⁾がある。

また、形状情報を特徴ベクトルに抽象化し、それらのノルムを誤差関数とする手法がある。これには、1. 特定の視点から観測した物体の見え全体を特徴量化する手法と、2. キーポイント周りに設定した局所的な形状情報を特徴量化する手法がある。まず、1. の手法には、多視点から画像としてレンダリングした物体モデルの見えのデータを多数用意し、画像ごとに勾配方向を量子化した特徴量として表現し、マッチングをおこなう²⁾³⁾がある。2. の手法の代表例は SHOT 特徴量⁴⁾である。キーポイントを中心とした球形領域を分割し、各ブロック内

* 原稿受付 平成 30 年 5 月 7 日

掲載決定 平成 30 年 9 月 30 日

** 正 会 員 慶應義塾大学 (神奈川県横浜市港北区日吉 3-14-1)

の法線方向分布を特徴量化する。いわゆる Hand-crafted な 3D 局所特徴量としては最も有名な手法である。近年では、Deep learning によって特徴表現を学習する手法も提案されている。例えば文献⁵⁾の手法では、Convolutional Auto-Encoder を利用して RGB-D のパッチデータから特徴量を学習する手法が提案された。

しかしながら、上記のモデルマッチングは物体モデルと対象物の形状が完全に一致していることを前提としている。マッチングの対象を類似物体同士にした場合には、正しい姿勢変換パラメータにおいても意匠の違いが二乗距離や特徴量間ノルムに反映されるため、必ずしも最小化されず、推定に失敗する可能性がある。

類似形状物体同士の位置合わせ手法は Aldoma らが提案した手法⁶⁾が存在する。この手法は対象物が平面に安定姿勢で存在すること前提とし、物体モデルから計算した安定平面の法線回り 1 軸の回転角と拡大率を変えながら、シルエット間の誤差が最小化される姿勢パラメータを算出する。この手法では、安定平面の法線回りの回転のみを考慮するため、完全な 6 自由度探索ができないという問題点がある。提案手法では、完全な 6 自由度の探索が可能な類似形状物体同士の位置合わせアルゴリズムを提案する。

3. 日用品の機能属性とその空間配置

3.1 機能属性

道具の形状は、その使用目的にとって都合がよくなるように設計されている。例えば、ハンマーの柄の部分は手によって握りやすい太さに設計されており、その先のヘッド部分は物体を叩くために平面で構成されている。このように道具を構成する各部位の形状と役割の間には密接な関係が存在しており、形状が持つ役割のことを機能属性と呼ぶことにする。

本論文では、機能属性の中でもとくに道具に着目したものである。Myers ら文献⁷⁾で紹介している機能属性を取り扱う。例えば、Grasp:握られる、Pound:叩く、Scoop:掬う、Contain:内部に保持する、である。これらの機能属性で共通なことは、一つの剛体として機能を発揮することである。機能属性は他にも材質由来のもの、形状を変化させることによって機能を発揮させるもの等、多岐にわたるが⁸⁾、本論文では最も基本的である剛体の状態で利用することのできる機能属性を取り扱うこととする。

ところで、Gibson らによる Affordance⁹⁾は機能属性と密接に関連している概念である。Affordance は人間と環境との間に起きるインタラクションの可能性をまとめて表現する概念であり、ハンマーの柄は持つべき場所であるかどうかという、作り手の意図は反映されていない。本論文では、手で持たれるために設計されているハンマーの柄のように、いわゆる意図して作りこまれた機能のことを元来の Affordance と区別するために機能属性と記述する。

3.2 同一カテゴリ物体の機能属性配置の共通性

我々の身の回りにはさまざまな道具が存在するが、同一名称(=同一カテゴリ)の道具であれば、意匠の違いがあっても、それを正しく使用することができる。このことは、同一カテゴリの道具の形状は、使いやすさを損なわない範囲でバリエーションを持っているとも言え換えられる。したがって、同一カテゴリの道具は、意匠の影響によって形状が異なっていた

Fig. 1 Common affordance labels on different objects in common category. Left: original images, right: affordance labels overlaid onto the original images. Green and red shows label 'grasp' and 'pound', respectively

としても、機能属性間の空間的な配置には共通性があると考えられる。図 1 に、異なる意匠のハンマーとその機能属性の空間配置の例を示す。この画像は Affordance segmentation のためのデータセット¹⁰⁾に登録されている画像とその Ground Truth としての機能属性ラベルの可視化結果である。各ハンマーは柄やヘッドの形状が異なるが、扱い方は共通であるので、各機能属性をもつパーツ間の位置関係は、異なるハンマー同士でも共通であることがわかる。したがって、この共通性を手掛かりとして位置合わせをおこなう。

4. 提案手法

4.1 前提条件

提案手法では、物体モデルと入力シーンに対して下記の条件が成り立つことを前提に、類似物体間の位置合わせをおこなう。

1. 同一カテゴリの道具の機能属性の配置は共通的であるとする。
2. 対象物のカテゴリ(コップ、ハンマーなど)は既知であり、事前に物体モデルには機能属性が付加されている。
3. 入力データ(色距離センサによる計測データ)に対して、機能属性ラベルが付加されている。本研究では、文献^{10, 13, 14)}等で提案された機能属性の認識手法を姿勢推定の前段に利用することを前提とする。また、対象物の大まかな位置は機能属性の推定結果や、YOLO¹¹⁾や SSD¹²⁾等を利用することによって検出できる。

条件 2. で用意する物体モデルは、入力シーンに存在する物体の形状と完全に同一でなくてもよい。準備するコストは高くはないと考えられる。条件 3. においても、一般物体認識アルゴリズムの高度化に伴い容易に実現可能であると考えられる。したがって、実験ではこれらが満たされた上で処理をおこなった。

4.2 概要

図 2 に提案手法の流れを示す。提案手法の入力データは入力シーンの点群データ S 、認識対象のカテゴリの物体モデル M である。物体モデルの各点 $p \in M$ は事前に付与させておいた機能属性ラベルを持つ。これは $l(p)$ と表記する。出力データは S における M の 6 自由度姿勢(回転・並進の 3 パラメータずつ)である。物体モデルの位置姿勢推定は、初期位置姿勢推定、精密位置合わせの 2 ステップで構成される。初期位置姿勢推定では、シーンに対して推定した機能属性ラベルを手掛かりとして、物体モデルの大まかな位置を決定し、これを初期位置とする。精密位置合わせは、初期位置を起点として物体モデルの姿勢を変動させながら、尤もらしい位置姿勢を推定する

Fig. 2 Overview of the proposed method

Fig. 3 Pose initialization using affordance labels. (a) A RGB image of the input scene. This image is used only visualization. (b) The object model and its affordance labels. (c) Point cloud of the input scene. Each color indicates part affordance. (d) The input scene and the object model overlay with initial pose

ステップである。次節以降ではそれぞれの処理の手順を説明する。

4.3 初期位置姿勢推定

まずは、 S の各点の機能属性を推定する。これには従来手法である文献^{10, 13, 14)}等の手法を適用する。 S から物体領域の切り出しをおこない、これを位置姿勢推定対象の点群 S' とする。 S' と M から共通的な機能属性ラベルを抽出し、それらの重心点を算出し、重心点同士を一致させる姿勢変換パラメータを初期位置 R_{ini} とする。共通する重心点が一つのみの場合は R_{ini} は平行移動成分のみから構成されるが、二つ以上存在する場合は、共通する2つの機能属性ラベルの重心点を結ぶ線分が一致する位置姿勢変換パラメータを R_{ini} とする。図3に初期姿勢の例を示す。(a) は入力シーンの外観である。シーン中にはハンマーが置かれている。(b) は物体モデルである。色はそれぞれ手動で割り当てた機能属性ラベルである。(c) は点群として表示した(a)のシーンである。(d) は(c)に対して推定した初期位置姿勢 R_{ini} を適用した(b)のモデルを重畳した結果である。機能属性ラベル Pound, Grasp が同一直線上で一致していることが確認できる。

この時、入力シーン中の物体は一方向から計測した距離データであるため、全周囲の形状データを持つ物体モデルとの重心位置は一致しない。そこで、次節で説明する処理によって平行移動・回転を含む6次元の姿勢パラメータを最適化する。

4.4 精密位置合わせ

R_{ini} を初期位置として、物体モデル M をシーン S' に対して位置合わせする変換を修正する。提案手法では形状の一致性と機能属性ラベルの一致性の両方を評価するコスト関数、

$$\tilde{\mathbf{R}} = \arg \min_{\mathbf{R}} (\omega C_s(\mathbf{R}) + (1 - \omega) C_{af}(\mathbf{R})) \quad (1)$$

を定義し、これを最小化する最適な変換パラメータ $\mathbf{R}$ を探索する最適化問題として解く。式(1)は物体モデル M に適用する姿勢変換パラメータ $\mathbf{R}$ によって変化する二つの評価関数 $C_s(\cdot)$, $C_{af}(\cdot)$ で成り立っている。 ω は二項の合成重みであり、実験では0.5を用いた。

$C_s(\cdot)$ は以下の通りであり、

$$C_s(\mathbf{R}) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{R}\mathbf{p}_i - N_p(\mathbf{R}\mathbf{p}_i, S' | l(\mathbf{p}_i))\|_2 \quad (2)$$

関数 $N_p(\mathbf{R}\mathbf{p}_i, S' | l(\mathbf{p}_i))$ は条件付きの最近傍点探索を示す。 S' から $\mathbf{p}_i$ の機能属性ラベル $l(\mathbf{p}_i)$ を持つ $\mathbf{R}\mathbf{p}_i$ の最近傍点を取り出している。

$C_{af}(\cdot)$ は、機能属性ラベルの一致性を評価するコスト関数である。

$$C_{af}(\mathbf{R}) = 1 - \frac{1}{n} \sum_{i=1}^n N_k(\mathbf{R}\mathbf{p}_i, S' | l(\mathbf{p}_i)) \quad (3)$$

関数 $N_k(\cdot)$ は k 近傍探索を示し、 k 近傍内に $l(\mathbf{p}_i)$ と位置する点が見つかったら1を返却し、見つからない場合に0を返却する。したがって、姿勢変換 $\mathbf{R}$ を適用したときに近くに一致する機能属性ラベルを持つ点の割合を評価している。

式(1)のコスト関数の最適化には粒子群最適化法(PSO)を利用する。一般的に形状間の位置合わせにはICPアルゴリズムのように勾配降下法ベースの方法が採られるが、PSOでは局所解を脱するアルゴリズムが組み込まれているので、こちらを利用した。

5. 実験と考察

一般的な位置合わせのコストとして利用される、二乗誤差の最小化をコスト関数とした手法を比較対象として、本手法の有効性を評価した。以下、実験に用いたデータを説明し、定性的な評価として、物体モデル同士の位置合わせ結果を比較し、定量的な評価として、実シーンに対する姿勢推定性能を比較した。

5.1 実験に用いたデータ

異なる意匠の物体モデル同士の位置合わせをおこなうために、複数の公開データセットから物体モデル、入力シーンを収集した。

物体モデルに関しては、ShapeNet¹⁵⁾, Thingiverse¹⁶⁾ から収集した。収集したカテゴリはマグ、ハンマー、スプーンである。図4に物体モデルの例を示す。メッシュで提供されている物体モデルを点群データに変換したのちに、各点に機能属性ラベルを付加した。各ラベルはMyersらが文献⁷⁾で紹介した機能属性である。GraspとW-grasp(Wrap-graspの略)は機能的な違いがあり、Graspは手で囲い込むように握り、その道具を利用する際に掴むハンドル部に備わる機能のことであり、W-graspは手が閉じ切らないで持つことができる部分を指す。Contain, Pound, Scoopはそれぞれ、液体を保持する、ものを叩く、ものを扱う機能を表す。また、この図では各物体の原点部分に3次元の座標系を描画している。

Fig. 4 Object models and its parts affordance labels

入力シーンには Myers らが文献⁷⁾で公開したデータセット (UMD Part-affordance dataset) を利用した。図 3(a), (c) は提供されているデータの例である。このデータでは、日用品がターンテーブルに乗せられており、これを 1 回転させながら Kinect によって RGB-D データが撮影されている。各物体につき 1 周分約 250–300 フレームのデータが用意されており、各点に対して機能属性ラベルが割り付けられている。しかしながらターンテーブルの回転角度は公開されていないので、このままでは評価に利用することができない。本実験では、初期フレームからの相対的な回転角度を手動で算出し、これを回転角度の正解データとした。

5.2 実験に用いたパラメータ

提案手法を動作させるためには、コスト関数と PSO に設定すべきパラメータがある。事前にいくつかのパラメータで処理をおこなったのちに、認識率が高いパラメータを採用した。

コスト関数においては、 $\omega = 0.5, k = 30$ とした。 k は物体の大きさに依存するパラメータであるが、本実験に用いた物体は大きさの違いが少なかったため、同一の値を用いた。

PSO においては、繰り返し回数 = 100、粒子数 = 300 とした。更新則の 3 つの変数は PSO の収束安定性を解析した文献¹⁷⁾を参考に、一般的に広く用いられている値 (慣性係数 = 0.729、位置に関する 2 つの係数 = 1.49445) を用いた。各粒子の初期位置の生成範囲は 4.3 節で説明した手順で生成した姿勢パラメータを中心に、Roll, Pitch, Yaw それぞれ $[-0.5\pi, 0.5\pi]$ 、平行移動成分は $[-3.0, 3.0]$ mm とした。

また、物体モデルは 1000 点にランダムサンプリングし、処理対象のシーンデータは 640x480 の解像度の距離画像をカメラの内部パラメータを用いて点群に変換し、Voxel Grid Filter によって 5mm 間隔にダウンサンプリングした。

5.3 物体モデルの正対化

まず、位置合わせ性能の基本的な性能評価のために、姿勢の異なる同一カテゴリ間の物体同士の位置合わせをおこなった。図 5 に結果を示す。

この実験では図中赤枠で囲われた物体を位置合わせの対象とし、残りの物体を対象物と同一の姿勢に変換する変換パラメータを推定した結果である。したがって、赤枠外の物体が赤枠内の物体と同一の向きに表示されていれば正しい結果が得られたことになる。(a), (b) がそれぞれ比較手法、提案手法による結果である。対象物がハンマーの場合はどの物体も正しく位置合わせができていた。しかしながら、(a) の中央のスプーンは上下が反転した結果が得られた。中央のスプーンは柄が太く、赤枠のスプーンを位置合わせした際、上下が反転していた場合のコストが正しい姿勢のコストとほぼ同等であったことが原因と考えられる。マグに関しては円筒部分の軸周りに推定結果がばらついていて、マグの取っ手部分の形状は全体の面積に対して小さいため、位置ずれのみを評価するコスト関数への影響度

Fig. 5 Qualitative result of alignment for three kind of objects. Object models with red bounding box are aligned with other objects. (a) Result of the baseline method. (b) Result of the proposed method

が小さい。位置合わせ対象のマグ同士は円筒部分も意匠が異なっており、これもコストに影響を及ぼすため、小さい形状である取っ手同士を位置合わせすることが困難であったと考えられる。

一方、提案手法では同一の機能属性同士で位置ずれを評価している。したがって、スプーンの例のように反転することや、マグのように小さな形状のパーツが無視されることが少なく、結果として正しい推定結果を得ることができた。

本実験で用いたハンマーの中で、図 5 の左上のものはヘッド部にくぎ抜き機能を持つ。くぎ抜き機能に対応した機能属性ラベルを付加したうえで、これらを別カテゴリの物体として取り扱うべきであると考えられるが、ものを叩く部分とくぎ抜き部分に明確な区切れが無い、くぎ抜きに対応する機能属性が用意されている機能属性ラベルに無い、という理由からヘッド部全体に Pound のラベルを割り当てた。くぎ抜き付き/くぎ抜き無しハンマーをそれぞれ別カテゴリとして扱った場合、クラス内の物体形状の多様性は小さくなるため、提案手法の位置合わせ精度は向上すると考えられる。

5.4 姿勢推定の精度評価実験

姿勢に対する正解データを付与した UMD Part-affordance dataset に対して姿勢推定精度の評価実験をおこなった。評価に使った入力データ数はマグ、ハンマー、スプーンでそれぞれ 120, 163, 182 である。各フレームの姿勢パラメータのうち、Roll, Pitch, Yaw 角の平均誤差が 15 度以内であった場合を推定成功とした。角度情報はいずれのコスト関数による姿勢推定方式と比較したときにも共通的に使える正解データであるため、これを正解の基準とした。ただし、ハンマーとスプーンについては、対称性を持っているため正解姿勢を一意に決めることができなかった。したがって、これらに関しては入力シーンに結果を重ねし、目視にて判定した。結果を表 1 に示す。認識に用いた物体モデルは図 4 の物体である。各物体の姿勢推定結果の例を図 6 に示す。この図では、各物体の姿勢の推定結果に

Fig. 6 Pose alignment results for three types of objects, mug, hammer, and spoon. The object model is overlaid onto the RGB image of input scene

基づいて物体モデルを剛体変換し、入力シーンのRGB画像に緑の点で重畳している。見やすさの観点から、重畳の際には図4の原点の3次元座標軸も同時に表示している。

なお、本研究で重要な点は機能属性を利用することによる、位置姿勢の推定精度に与える影響を調べることである。したがって、理想的な機能属性推定結果が得られることを想定して、入力シーンに付加する機能属性ラベルはUMD Part-affordance datasetにGround Truthとして与えられているものを用いた。

クラス‘マグ’において形状のみを用いた場合の推定性能が低い理由は、計測点群の全数に対する取っ手部分の点数が少ないため、その部分の位置ずれが無視されて位置合わせされてもコストに対する影響が少ないからである。同様の現象は、他のクラスにも起きた。さらにハンマーとスプーンでは、長軸方向にずれたとするとコストが変化しにくいから、位置ずれが大きくなりやすかった。

一方、提案手法では、同一の機能属性を持つ点群同士で距離計算をおこなうため、上記のような現象が起きにくかった。例えば、図6(a)の例は三種類の意匠の異なるマグを一種類のマグのみで推定した結果である。図4のマグの取っ手方向に向いている座標軸は、認識結果の重畳後も取っ手方向に向いていることからわかるように、これらの姿勢は正しく推定できたといえる。しかしながら、図6(c)の中央の物体のように、物体同士のサイズが大きく異なる場合は、推定結果が安定しにくいことが分かった。この物体に対する成功率は2.0[%]であったため、クラス‘スプーン’の認識結果が他の物体に比べて低くなった。

提案手法のコスト関数は機能属性の一致性を評価するため、形状の一致性が相対的に弱められている。位置合わせ対象が形状の異なる物体同士であるため、ある程度低くても悪影響が少ないと考えたためであり、実際に姿勢推定精度向上の効果があつた。しかしながら、実験結果を確認するといくつかの例ではスプーンが表裏を間違えて収束している事例も存在した。表裏で形状がよく似た物体や、対称性のある物体ではこのような収束結果が起きるという特性も明らかになった。

5.5 スケール変動に対する適用範囲

提案手法では、位置合わせする対象物間の大きさ（単にスケールと呼ぶ）の違いが大きい場合に認識精度が低下すること

Table 1 Reliability of pose alignment [%]

	Mug	Hammer	Spoon
Baseline	37.5	56.9	18.1
Proposed	82.5	96.3	56.6

Fig. 7 Relationship between magnification ratio of the object model and the results

がわかった。そこで、物体モデルのサイズを少しずつ変動させながら特定シーンへの姿勢推定をおこない、提案手法のスケール変動に対する適用範囲を調べた。結果を図7に示す。

実験に用いた物体モデルとシーン中の物体のスケールはほぼ同一である。したがって、拡大率1の時に同一の大きさの物体同士の位置合わせをおこなったことに相当する。プロット1点は20回姿勢推定をおこなった際の平均値である。まず、コストに関しては0.85を境に線形に値が増加した。シーン中の物体は、境界部分に計測抜けが発生するため、物体モデルの拡大率が実際よりも小さい場合に形状の一致度合いが高くなりやすく、ピークになったと考えられる。次に、姿勢誤差に関して考える。[0.65, 1.15]の範囲であれば平均誤差が10[deg]程度であり、精度が高いことがわかった。したがって、実利用時には、物体モデルと入力シーン中の対象物の大きさの違いがこの範囲内に収まっているかどうかを確認することによって本手法の適用可能性を検証することが可能である。また、対象物を切り出すことさえできれば大まかなサイズを計算可能であるため、あらかじめ物体モデルの大きさを変更した上で姿勢推定を実行すれば、精度の良い結果が得られると考えられる。

6. おわりに

本研究では、入力シーンに存在する物体と同一の3Dモデルが利用できない状況においてもその6自由度姿勢を推定する手法を提案した。提案手法では、同一カテゴリの物体が持つ機能属性の空間配置に共通性があることを利用して、これを位置合わせする。実験により、機能属性の空間的な配置の位置ずれを位置合わせ時のコストとして取り入れることによって従来手法よりも信頼性の高い姿勢推定が可能になることを確認した。しかしながら、対象物の大きさが物体モデルと大きく異なると信頼性が低下することが確認されたので、今後はこの点の改善と、ロボットによる物体操作への応用を進める予定である。

謝 辞

本研究は JSPS 科研費若手研究 (B) 17K12761 の助成を受けて実施された。

参 考 文 献

- 1) P. J. Besl and N. D. McKay, A Method for Registration of 3-D Shapes, *IEEE Trans. PAMI*, **14**, 2, (1992) 239.
- 2) S. Hinterstoisser, C. Cagniart, S. Illic, P. Sturm, N. Navab, P. Fua and V. Lepetit, Gradient Response Maps for Real-time Detection of Textureless Objects, *IEEE Trans. PAMI*, **34**, 5, (2012) 876.
- 3) Y. Konishi, Y. Hanzawa M. Kawade and M. Hashimoto, Fast 6D Pose Estimation from a Monocular Image Using Hierarchical Pose Trees, *Proc. European Conference on Computer Vision (ECCV)*, (2016) 398.
- 4) F. Tombari, S. Salti and L. Di Stefano, Unique Signatures of Histograms for Local Surface Description, In *Proc. European Conference on Computer Vision (ECCV)*, (2010) 356.
- 5) W. Kehl, F. Milletari, F. Tombari, S. Illic and N. Navab, Deep Learning of Local RGB-D Patches for 3D Object Detection and 6D Pose Estimation, *Proc. European Conference on Computer Vision (ECCV)*, (2016) 205.
- 6) A. Aldoma and M. Vincze, Pose Alignment for 3D Models and Single View Stereo Point Clouds based on Stable Planes, *3DIMPVT*, (2011) 374.
- 7) A. Myers, C.L. Teo, C. Fermüller and Y. Aloimonos, Affordance detection of tool parts from geometric features, In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, (2012) 1371.
- 8) K.M.Varadarajan and M. Vincze, AfNet: The Affordance Network, *Proc. Asian Conference on Computer Vision (ACCV)*, (2012) 512.
- 9) J. J. Gibson, *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin, 1979.
- 10) T.-T. Do, A. Nguyen and I. Reid, AffordanceNet: An End-to-End Deep Learning Approach for Object Affordance Detection, In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, (2018).
- 11) J. Redmon, S. K. Divvala, R. B. Girshick and A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016) 779.
- 12) W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Y. Fu and A. C. Berg, SSD: Single Shot MultiBox Detector, In *Proc. European Conference on Computer Vision (ECCV)*, (2016) 21.
- 13) A. Nguyen, D. Kanoulas, D. G. Caldwell and N. Tsagarakis, Object-Based Affordances Detection with Convolutional Neural Networks and Dense Conditional Random Fields, In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS)*, (2017).
- 14) 秋月秀一, 飯塚正樹, 香西健太郎, 橋本学, 局所特徴量による識別結果の尤度統合に基づく日用品の機能属性推定, *精密工学会誌*, **84**, 7, (2018) 658.
- 15) ShapeNet, <https://shapenet.org/>
- 16) Thingiverse, <https://www.thingiverse.com/>
- 17) M.Clerc and J. Kennedy, The Particle Swarm-explosion, Stability, and Convergence in a Multidimensional Complex Space, *IEEE trans. Evolutionary Computation* **6**, 1, (2002) 58.