

# 多視点 3D データに基づく信頼度付き領域統合による機能属性認識の高精度化

○ 城亮輔 †, 飯塚正樹 ‡, 橋本学 ‡

○ Ryosuke TACHI † Masaki IIZUKA ‡ Manabu HASHIMOTO ‡

†: 中京大学工学部, tachi@isl.sist.chukyo-u.ac.jp

‡: 中京大学大学院工学研究科, {iizuka, mana}@isl.sist.chukyo-u.ac.jp

<要約>生活支援ロボットが日用品を扱うためには、コップの取っ手のような日用品の部位ごとに備わっている“機能属性”の認識が求められている。本稿では、多視点から取得した点群を効果的に用いて高信頼な機能属性を付与する手法を提案する。複数視点の元データを統合した後に機能属性を認識する手法では、統合精度によって認識率が大きく低下することが懸念される。そこで提案手法では、複数の異なる視点から取得した 3D データの機能属性をいったん認識し、局所領域ごとに仮の機能属性を付与する。その後、信頼度を考慮しながら仮に付与した機能属性を統合することによって、誤認識を抑制した高信頼な機能属性を付与する。この領域の信頼度は、視線方向と領域の法線との関係や、隣接領域との滑らかさをもとに算出する。3種類の日用品（コップ、レードル、ヘラ）の 3D データを用いた実験により、提案手法の平均認識率が 69% であり、複数視点の元データを統合した後に従来法にて機能属性認識する手法よりも 13% 向上したことを確認した。

<キーワード> 機能属性認識, 日用品, アフォーダンス, 3次元点群

## 1. はじめに

近年、家庭などの生活空間において人々と共存し、我々の生活の支援をするロボットの実現が期待されている。これらのロボットが我々の生活の支援を、人間と同じようにおこなうためには、生活空間内に存在する多品種の日用品を正しく取り扱う必要がある。例として、ロボットがコップに飲み物を注ぎユーザのもとへ運ぶタスクを考える。このタスクを完遂するためには、ロボットが飲み物を注ぐための“液体を蓄える”部分とコップを運ぶための“掴む”部分を認識することが重要となる。多くの日用品は、このように“液体を蓄える”や“掴む”といった機能を意図してデザインされている。例えば、液体を蓄える部分であれば、その形状は窪んだ形状をしている。本研究では、意図して作り込まれた形状に備わる機能を「機能属性」と呼び、形状情報をとらえやすい 3D データを用いてこの機能属性の認識に取り組む。

従来さまざまな機能属性認識手法が提案されている。Deep Learning ベースの手法 [1-4] では、Convolutional Neural Network を用いて、RGB 画

像から物体の機能属性を認識する。また、Myers ら [5]、飯塚ら [6] は、奥行きや曲率などの形状情報を特徴量として用いた認識手法を提案した。

これら手法には、単視点から取得したデータから機能属性を認識しており、視点によって対象物の見え方が異なることから、セルフオクルージョンによって見えない領域のデータが取得できなかつたり、センサから取得される 3次元点群に欠落が生じていたりするという問題がある。

そこで本手法では、複数視点の 3D データを用いて、機能属性の認識に取り組む。多視点で機能属性の認識をおこなう場合、データを統合する順序が重要であり、一般的には 3次元点群を統合してから機能属性を認識する手法と複数の視点から取得したデータの機能属性をそれぞれ認識してから統合する手法の 2 パターンが考えられる。

第 1 の手法では、3次元点群の統合精度によって認識率が左右されることが懸念される。例えば、3次元点群の位置合わせの際に微小なズレが生じた場合、そのズレによって表面形状が乱れ誤認識することが考えられる。一方で、第 2 の手法は、先に機能属性を認識するため、このような統合精度による認識率

への影響が小さい。そのため、本研究では第2の手法によって機能属性を認識する手法を提案する。

## 2. 機能属性認識の信頼度

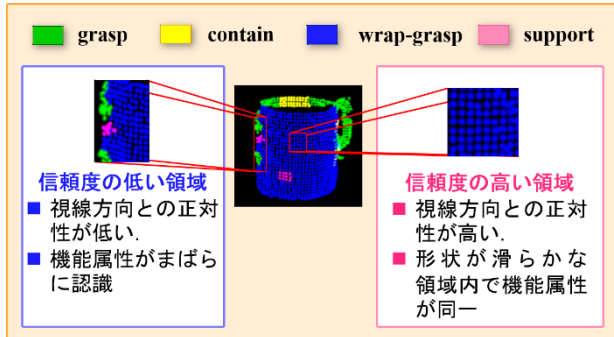


図1 機能属性の信頼度

本手法では、視点によって物体の見え方が異なることに着目した。視点によって認識しやすい領域と認識しにくい領域があることが考えられる。このことから領域によって認識結果の信頼度が異なるという仮定をおいた。この仮定に基づいて図1のように信頼度の高い領域と信頼度の低い領域を定義した。例えば、信頼度の高い領域では視線方向との正対性が高く、形状が滑らかにつながっている周辺領域と同一の機能属性をもっている。一方で、信頼度の低い領域では、エッジ部分のような視線方向との正対性が低く、周辺領域と同一の機能属性をもたないことが多い。

### 2.1 視線方向との正対性による信頼度

1つ目の条件「視線方向との正対性」(以下、条件①と呼ぶ)について説明する。一般的に3Dセンサは、センサと正対している領域は精度の高いデータの取得ができ、正対していない領域はデータの取得精度が低いデータを取得すると考えられている。この原因として、正対性が高い領域は、センサから見えている範囲が広く多くのデータが取得可能であり、正対性の低い領域は見えている範囲が狭いため、データを高精度に取得できないことが想定される。つまり、このセンサの原理上、正対性によってデータのDepth値の誤差の生じやすさが異なる。図2に条件①による信頼度のイメージを示す。注目領域と視線方向との正対性は、注目領域の法線方向によって評価している。(b)信頼度マップは、(a)の法線方向が視点に向かっている領域ほど、前述の仮定に基づき、信頼度の高い赤色となっている。反対に法線方向が視点に向かっていない領域では、信頼度の低

い青色となっている。

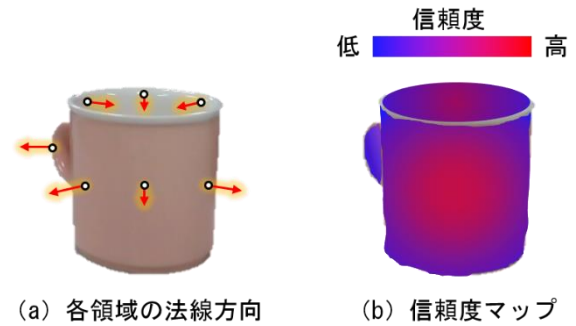


図2 条件①の信頼度イメージ

### 2.2 注目領域と隣接領域との関係性による信頼度

2つ目の条件「注目領域と隣接する領域との関係性」(以下、条件②と呼ぶ)について説明する。条件②では、対象物の形状について着目している。ここでは、コップの形状を例として挙げる。

図3に示すように、“wrap-grasp”のような同一機能属性をもつ領域間のみで形状変化を見た場合、緩やかな形状変化であることがわかる。一方で、“wrap-grasp”と“grasp”のように異なる機能属性を持つ領域間で形状変化を見た場合、機能属性が変わる境界部で大きな形状変化が見られる。そこで本手法では、形状変化が緩やかでなおかつ滑らかにつながっている領域であれば、同一の機能属性を有している可能性が高いという仮定をおいた。

この仮定に基づき注目領域の認識された機能属性と、その領域に隣接した領域の機能属性の関係性を考えると、隣接領域が同一の機能属性であるほど認識結果の信頼度が高い領域であるといえる。

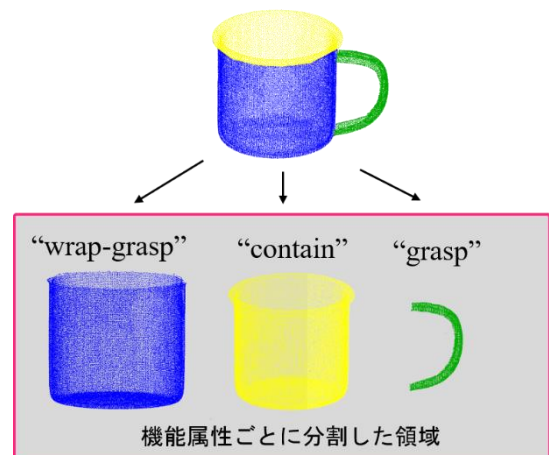


図3 コップの機能属性ごとの形状

### 3. 機能属性認識の流れ

図 4 に本手法の流れを示す。

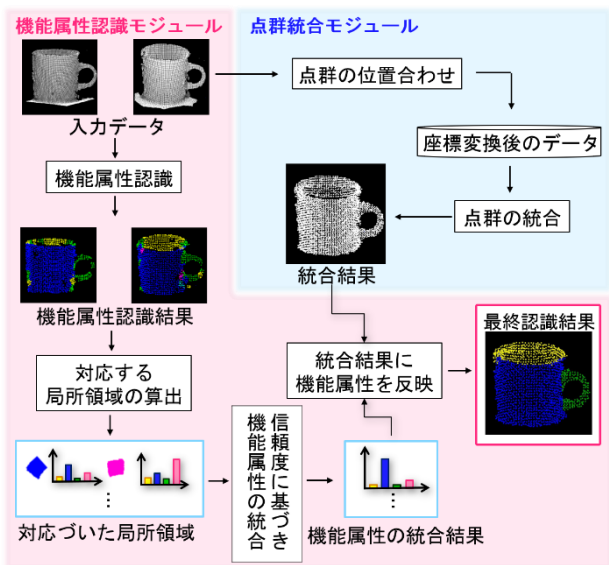


図 4 機能属性認識の流れ

本手法は、機能属性認識モジュールと点群統合モジュールの 2 つに分かれている。

点群統合モジュールでは、異なる視点から撮影した 2 つの入力データを位置合わせし、データを統合する。

機能属性認識モジュールでは、各入力データの機能属性を認識し、得られた結果を信頼度に基づいて統合する。機能属性を統合する上で、異なる視点データの対応する領域を推定する必要がある。本手法では、局所領域内に存在する 3 次元点の重心点によって対応付けをおこなう。統合した認識結果は、点群統合モジュールで作成した点群の統合データに反映し、最終的な機能属性認識結果を得る。

本研究では飯塚らの手法 [6] を用いて、各視点の機能属性を認識する。[6] では、複数の 3 次元点を 1 つの領域とした局所領域ごとに機能属性を認識し、最終的な対象物全体の認識結果を得る。また、局所領域ごとに各機能属性の尤度ヒストグラムを算出しており、機能属性の統合の際に用いる。

#### 3.1 点群の位置合わせ

異なる視点から取得した 3D データを統合するためには、センサから取得した 3D データを同一 3 次元空間中に投影し、位置合わせをおこなう必要がある。点群の位置合わせをおこなう上で、撮影時のセンサの位置姿勢を正確に推定することが重要である。本研究では、センサの位置姿勢を求める処理を簡単

化するために、センサをロボットハンド先に取り付け、ROS (Robot Operating System) によって求めることが可能なロボットハンドの外部パラメータを用いた。しかし、ロボットハンドの外部パラメータのみの位置合わせでは誤差が生じてしまうため、ロボットハンドの外部パラメータによって位置合わせされた点群座標を初期位置とし、ICP によって高精度に位置合わせした。

#### 3.2 尤度ヒストグラムの統合

本手法における尤度ヒストグラムの統合方法を図 5 に示す。

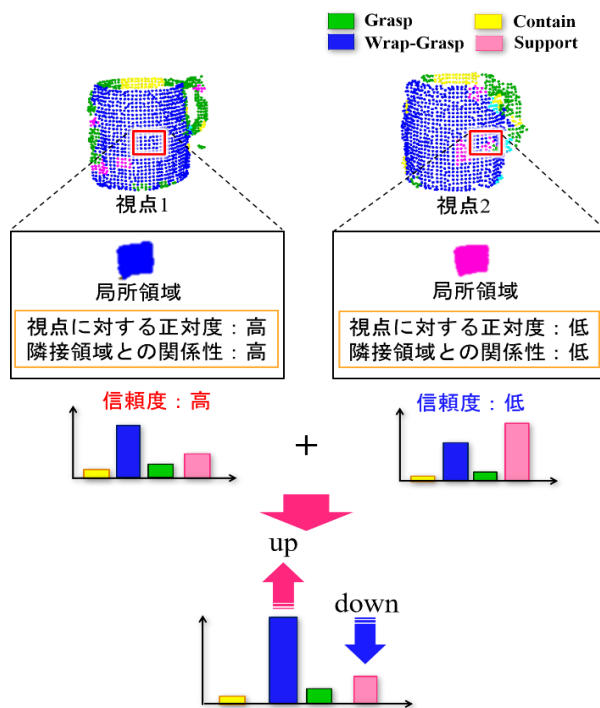


図 5 機能属性尤度ヒストグラムの統合

図 5 に示す視点 1 と視点 2 の局所領域は、それぞれ信頼度の高い領域と低い領域を表している。両領域は本来“wrap-grasp”の機能属性をもつ領域であるが、視点 2 の領域では“support”と誤認識している。そこで、この誤認識をなくすために、尤度ヒストグラムを統合する際に、2 章の信頼度に基づいて統合する。この各視点の尤度ヒストグラムの統合式を (1) 式に示す。

$$l = w_1 l_1 + w_2 l_2 \quad (1)$$

統合したヒストグラム  $l$  は信頼度に応じて比重を変化させた尤度ヒストグラム  $l_1$ ,  $l_2$  の和によって算出する。

(1) 式の重み (信頼度)  $w_1, w_2$  は以下の式で算出する.

$$w_n = \frac{ab}{|a||b|} S \quad (2)$$

$a$ : 視線ベクトル  $b$ : 法線ベクトル

$S$ : 同一機能属性を持つ周辺領域の数

(2) 式の視線ベクトルと注目領域の法線ベクトルとの内積から求められる余弦は, 条件①の信頼度を表している. 余弦の値が 1 に近いほど, その領域が視線方向に正対していることを意味している. 次に

(2) 式の  $S$  は, 条件②の信頼度を表している.  $S$  は, 注目領域と隣接している領域が同じ機能属性をもっているほど大きな値をとるようになっている.

(1) 式の重み  $w_n$  ( $n=1,2$ ) は, 条件①と条件②の信頼度の積であるため, 共に信頼度が高いほど重みが大きくなり, 低いほど小さくなる. これによって, 機能属性の誤認識が抑制されることが期待される.

### 3.3 機能属性統合結果の反映

3.2 節で統合した機能属性結果は, 点群統合モジュールで統合した結果に反映する.

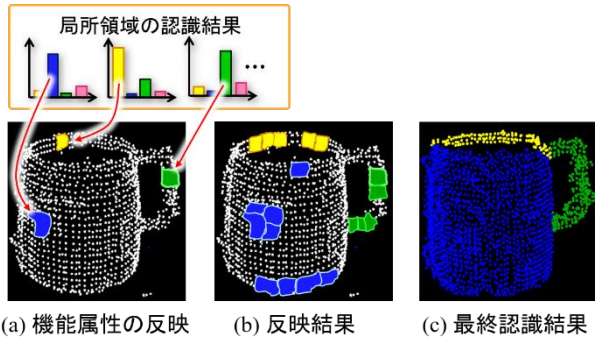


図 6 最終的な機能属性認識の流れ

機能属性の反映は, 図 6 の (a) に示すように局所領域ごとにおこなう. 機能属性を局所領域に反映した結果が (b) である. この処理を全局所領域に適用することで, (c) のような最終的な機能属性の認識結果を得る.

各視点のデータと統合データの局所領域の対応付けに, 局所領域の重心点を用いた. 0.01[m] の範囲内に存在する最も近い重心点をもつ局所領域どうしを対応しているとみなした.

## 4. 実験と考察

本研究では, 日用品の 3D データを用いて, 機能属性の認識実験をおこなった. 実験で用いた対象物を

図 7 に示す. 実験には, 3 種の日用品 (コップ, レードル, ヘラ) を異なる視点から撮影した計 25 シーンのデータを使用した. 対象物の撮影には, Intel 社製の RealSense SR300 を使用した.



図 7 実験に使用した対象物

本研究では Myers ら[5]によって定義された日用品に備わる機能属性 “grasp”, “cut”, “scoop”, “contain”, “pound”, “support”, “wrap-grasp” のうち, “cut” と “pound” を除く 5 つを認識した.

### 4.1 機能属性の認識評価実験

本実験では, 日用品のデータを異なる 2 視点から取得し, データの統合方法の違いによる機能属性認識率の比較をおこなった. 比較手法として, 元の 3 次元点群を統合してから従来の機能属性認識[6]を適用する手法と比較した. 比較した結果を表 1 と図 8 に示す.

表 1 はコップ, レードル, ヘラごとの平均認識率を示している. 認識率は, 対象物の全点群数のうち Ground Truth と同一の機能属性をもつ点群数の割合を表している.

表 1 機能属性の認識率

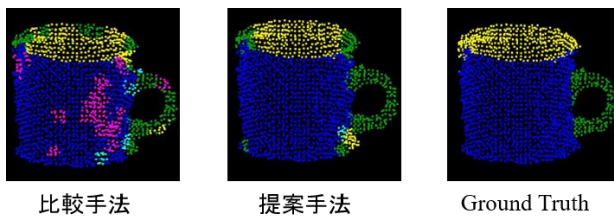
	コップ	レードル	ヘラ
元の 3 次元点群を統合してから従来法[6]を適用	0.71	0.42	0.59
提案手法	0.84	0.48	0.76



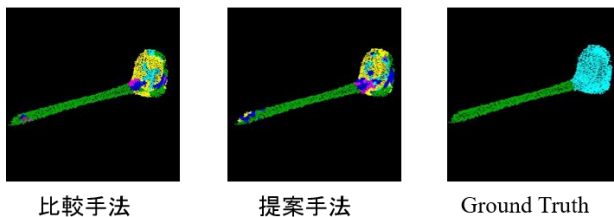
図8は各種日用品の認識結果例を示しており、(a)はコップ、(b)はレードル、(c)はヘラの認識結果である。本研究では、各機能属性を色によって表しており、機能属性と色の対応を表2に示す。

表2 機能属性と色の対応

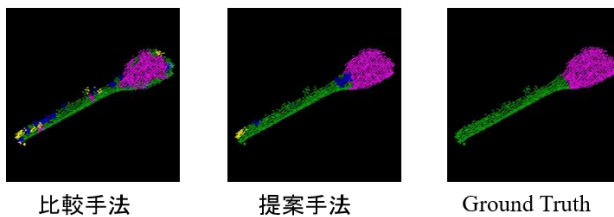
color	機能属性	機能
緑	grasp	手で握り込んでつかまれる
青	wrap-grasp	手で覆い込んでつかまれる
黄	contain	液体を蓄える
紫	support	ものを下から支える
赤	scoop	ものをすくう



(a) 機能属性の認識結果(コップ)



(b) 機能属性の認識結果(レードル)



(c) 機能属性の認識結果(ヘラ)

図8 機能属性認識結果

実験の結果、提案手法の認識率はコップが84%、レードルが48%、ヘラが76%であり、比較手法よりコップでは13%、レードルでは6%、ヘラでは17%高いことを確認した。比較手法による認識結果では、異なる視点どうしで重なり合っている領域付近に多く誤認識が生じていることが確認された。この要因として、一方の3次元点群が統合相手のノイズとして作用していることが考えられる。領域が重なり合っている部分の多くは、信頼度の高い領域とエッジ

部分のような信頼度の低い領域との組み合わせである。エッジ部分は、データの取得が困難なことからデータの取得率が低いデータである。そのため、信頼度の低い領域の3次元点群が統合相手のノイズとして作用し、滑らかであった表面形状を乱していることが想定される。機能属性の認識で用いた手法[6]では、対象物の表面形状が滑らかであることを前提としており、複数の3次元点群を1つのまとまりである局所領域ごとに機能属性を推定し、対象物全体の機能属性を認識している。そのため、信頼度の低い領域の3次元点群によって形状に乱れが生じ、局所領域の機能属性の推定が正しくおこなえていないと考えられる。

一方、提案手法による機能属性の認識は、3次元点群を統合する前に機能属性の統合をおこない、その結果を3次元点群へ反映する方法である。そのため、滑らかな形状から機能属性を推定することと同義であり、形状の乱れから生じる誤認識を無視することができている。また、信頼度によって機能属性を統合しているため、信頼度の低い領域による影響を小さくし、誤認識を抑制できていると考える。

#### 4.2 2視点の位置関係による認識率

本実験では、2視点間のセンサの位置関係に着目し、本手法の認識率の評価をおこなった。今回は、2視点間の距離が0.1[m]以上ある場合を遠い、2視点間の距離が0.1[m]以下の場合を近いとし、実験データとして、2視点間の距離を0.1[m]以上にしてコップを撮影したデータと、2視点の距離を0.1[m]以下にしてコップを撮影したデータを用いた。

図9は、(a)、(b)の2視点から取得したデータを用いて本手法で機能認識した結果を(c)に示している。

表3 機能属性の認識結果

2視点間の距離	認識率
0.1[m]以上	0.82
0.1[m]以下	0.85

表3に示すように、実験の結果2視点間の距離が近い場合のほうが遠い場合よりも認識率が約3%高いことを確認した。これは、本手法では2視点の認識結果を用いて機能属性の認識していることが関係していると考えられる。2視点間の距離が遠い場合、2視点に共通する領域が少なくなる。そのため、各視点で一度機能属性認識した時の結果に依存し、誤認識部分が多くなってしまう。一方で、2視点間の

距離が近い場合は、共通する領域が多くなり、より多くの領域で認識結果を比較し高信頼な機能属性を反映することが可能である。そのため、誤認識が抑制され認識率が高くなったと考えられる。

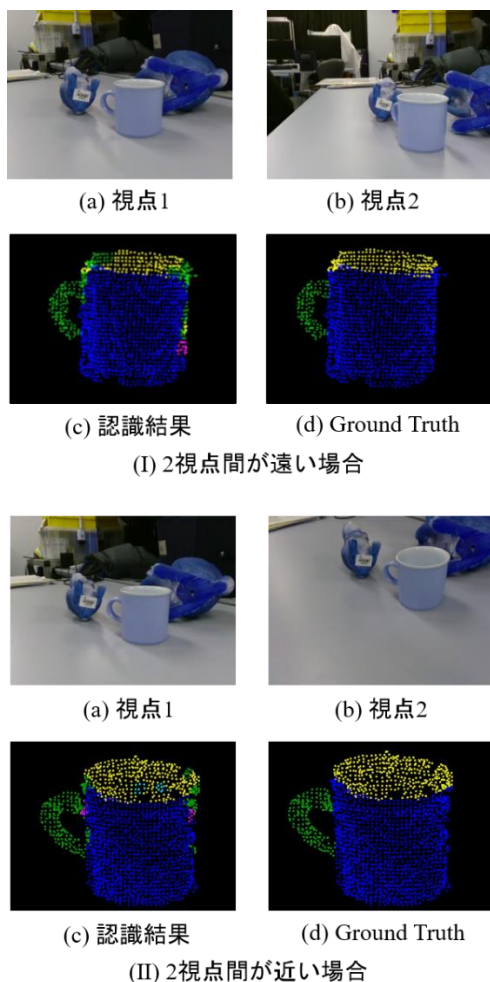


図9 2視点の位置関係による認識結果

## 5. おわりに

本研究では、複数視点の3Dデータから日用品の機能属性を認識する手法を提案した。提案手法では、各領域の認識結果の信頼度に基づいて、機能属性を統合することで高精度な機能属性認識を実現した。複数視点の3Dデータを統合したのちに機能属性認識する手法と比較した結果、提案手法の認識率はコップが84%、レードルが48%、ヘラが76%であり、比較手法よりもそれぞれ13%、6%、17%向上したことを確認した。

## 謝辞

この成果の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）からの委託業務の結果である。

## 参考文献

- [1] A. Nguyen, D. Kanoulas, D. G. Caldwell, and N. G. 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.
- [2] Thanh-Toan Do, Anh Nguyen, Ian Reid, Darwin G. Caldwell, and Nikos G. Tsagarakis, : “AffordanceNet : An End-to-End Deep Learning Approach for Object Affordance Detection”, arXiv:1709.07326 2017.
- [3] A. Roy and S. Todorovic, : “A Multi-scale CNN for Affordance Segmentation in RGB Images”, In Proc. European Conference on Computer Vision (ECCV), pp. 186-01, 2016.
- [4] J. Sawatzky, A. Srikantha, and J. Gall, “Weakly supervised affordance detection”, In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5197-5206, 2017.
- [5] A. Myers, C.L. Teo, C. Fermuller and Y. Aloimonos : “Affordance detection of tool parts from geometric features”, in Proc. IEEE International Conference on Robotics and Automation (ICRA), pp.1374-1381, 2015.
- [6] 飯塚正樹, 秋月秀一, 橋本学 : “2.5D データによる日用品の機能属性に基づいたパーツセグメンテーション”, 日本ロボット学会学術講演会, 3C1-03, 2017.