

# なぜロボットはお茶を持ってきてくれないのか

## ～ お茶会ロボットに残された課題 ～

中京大学 工学部  
橋本 学



関連資料 → <http://isl.sist.chukyo-u.ac.jp/archives3/>

Advanced Sensing & Machine Intelligence Group,  
Chukyo University

## 本日の話題

### 1. イントロダクション

- ロボットビジョンの基本課題
- 研究のモチベーション

### 2. 実現できたこと, できそうなこと

- 特定物体の認識 (位置・姿勢認識)
- 把持動作のためのロボットビジョン

### 3. “機能” 認識とその応用

- 全自動お茶会ロボットを目指して
- 人による教示動作のロボットへの転移

## 1. イントロダクション

- ロボットビジョンの基本課題
- 研究のモチベーション

## 2. 実現できたこと, できそうなこと

- 特定物体の認識 (位置・姿勢認識)
- 把持動作のためのロボットビジョン

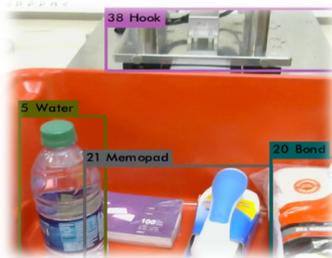
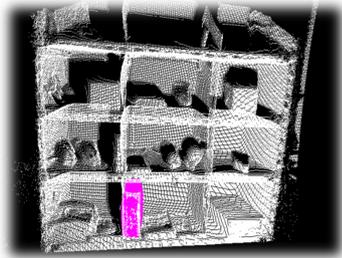
## 3. “機能” 認識とその応用

- 全自動お茶会ロボットを目指して
- 人による教示動作のロボットへの転移

# AIロボットの利用分野



どこに、なにがある？ → どう掴む？



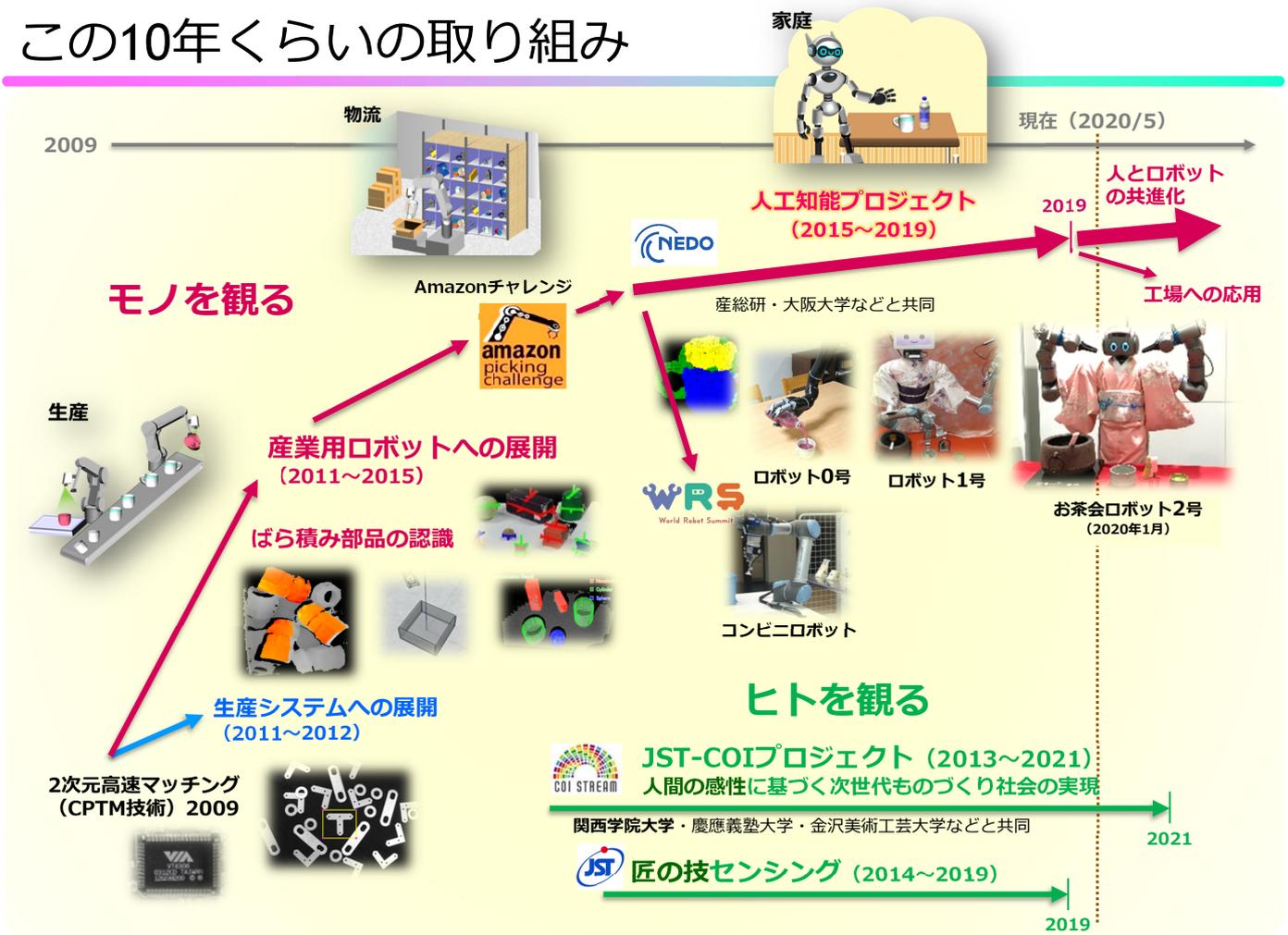
## 主要3分野におけるロボットとビジョンの課題傾向



<b>Pick and Place</b>	<b>Pick and Operation</b>
動作パターン種類 <b>少</b> (事前に生成)	動作パターン種類 <b>多</b> (その場で生成)
物体モデル利用 <b>可</b>	物体モデル利用 <b>難</b>
品種 <b>複数だがほぼ既知</b>	品種 <b>多量かつ未知</b>



# この10年くらいの取り組み

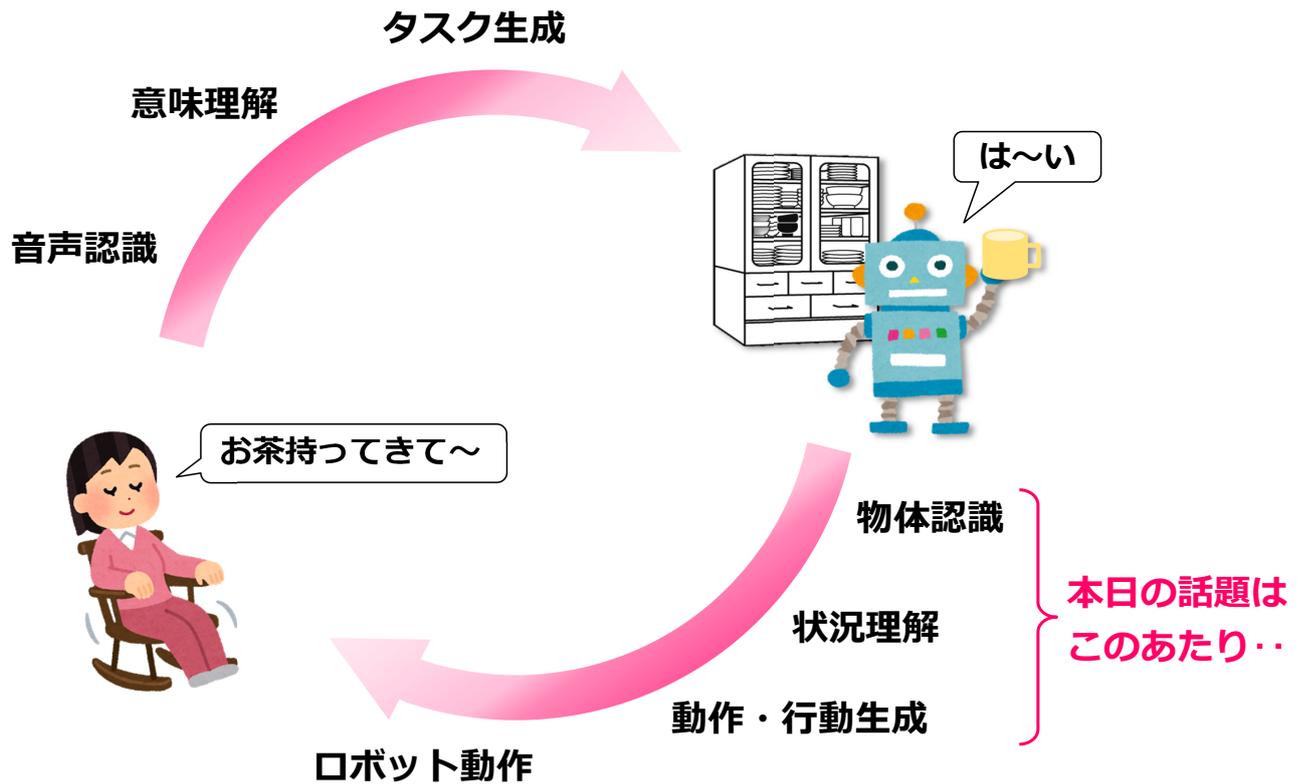


# 人間ならごく自然にできること

(26秒)

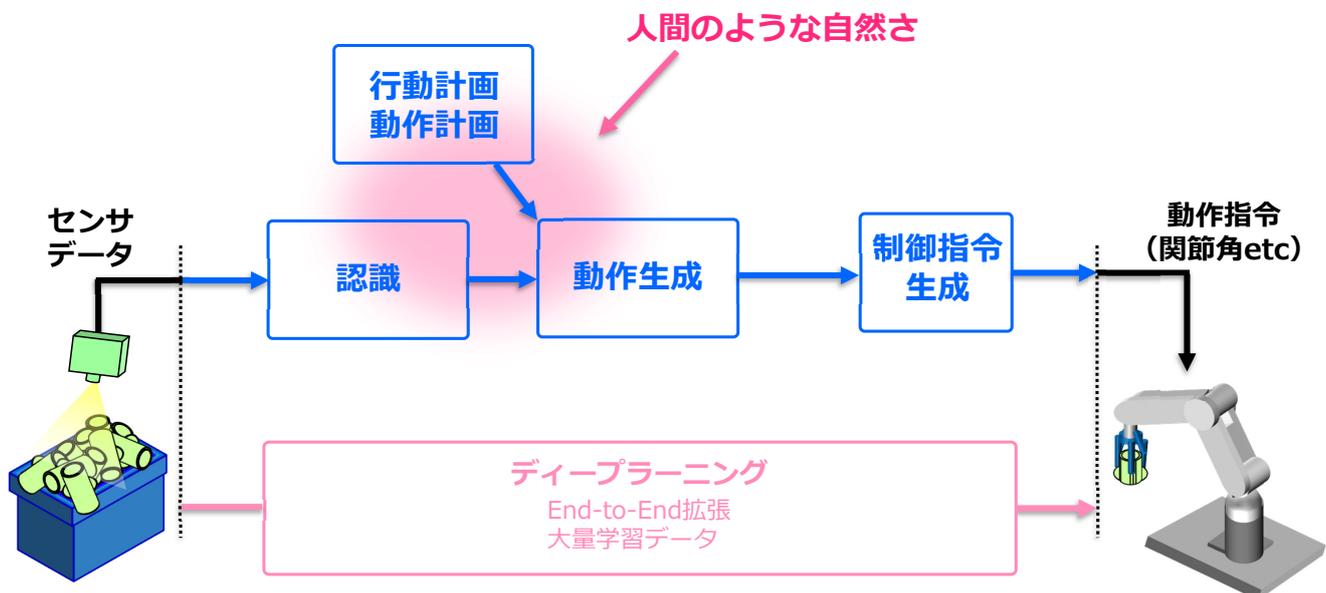


# お茶を持ってきてくれるロボットに必要な主要技術



## 最近の興味

- 物体認識とロボット動作の接点領域  
“人間味”を添えて……



## 1. イントロダクション

- ロボットビジョンの基本課題
- 研究のモチベーション

## 2. 実現できたこと, できそうなこと

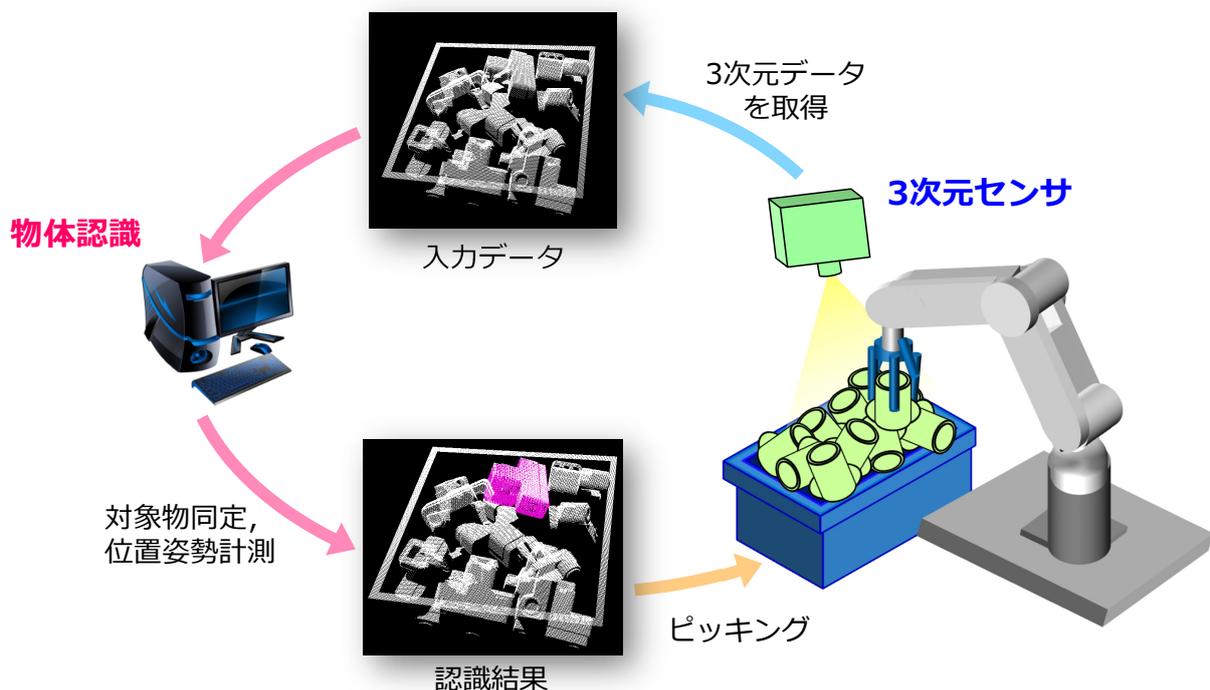
- 特定物体の認識 (位置・姿勢認識)
- 把持動作のためのロボットビジョン

## 3. “機能” 認識とその応用

- 全自動お茶会ロボットを目指して
- 人による教示動作のロボットへの転移

# ロボットビジョンの基本構成

ロボットビジョン = 3次元センサ + 物体認識 (アルゴリズム)



# 3次元センサのありがたみ

## 普通のカメラ



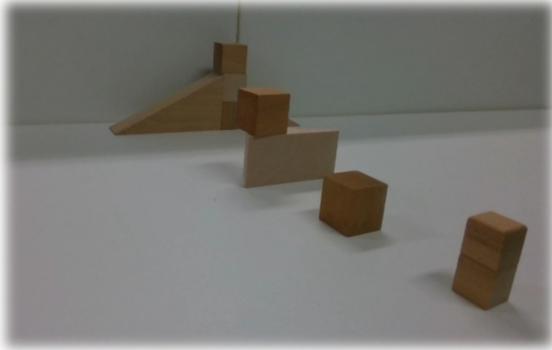
iBuffalo 200万画素カメラ

## 3次元センサ

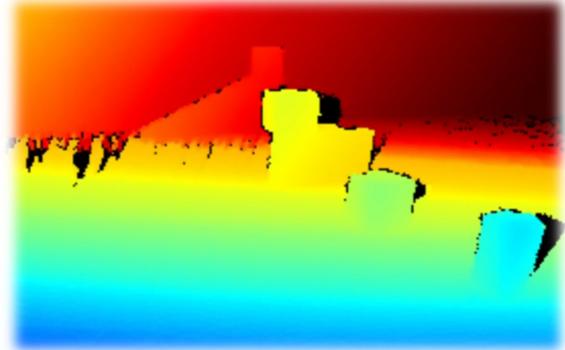


Intel RealSense SR300

## Color



## Depth



# 3次元センサの分類 (光学センサ)

青: 文献例 緑: 商品例

	パッシブ	アクティブ
三角測量	<p>ステレオ視 (2眼, 3眼, 多眼) Bumblebee 2 / XB3 (Point Grey 2006 / 2007) TVS (三次元メディア 2011) SV-M-S1 (リコー)</p> <p>マルチベースラインステレオ法 [Okutomi1993] 視体積交差法 [Matsuyama2002]</p>	<p>光切断法 (スポット光) TDS-A (パルステック 1997) 光切断法 (スリット光) VIVID9i (KONICA MINOLTA 2004)</p> <p>空間コード化法 [Posdamer1982], [Sato1985] Cartesia (SPACEVISION 2004) MELFA-3D Vision (三菱電機 2013) RV1100 (キヤノン)</p> <p>ランダムドットパターン投光法 [Hashimoto1999] Xtion PRO LIVE (ASUS 2011) Leap Motion (Leap Motion 2012) ASTORA (Orbbec 2016) SR300 (Intel 2016)</p> <p>ステレオ+パターン投光法 R200 (Intel 2015) D415 (Intel 2018) D435 (Intel 2018) ENSENSO N35 (iDS 2015)</p> <p>位相シフト法 [Halioua1989], [Zhao1994]</p>
同軸測量	<p>Shape(depth) from (De) Focus [Hiura1999] Shape from Motion</p>	<p>TOF: Time Of Flight Swiss Ranger SR3000 / SR4000 (MESA 2005 / 2008) D-Imager (パナソニック2010) DepthSense325 (SoftKinetic2012) Kinect v2 (Microsoft 2014)</p> <p>照度差ステレオ [Woodham1980]</p>

# 3次元センサの比較

販売開始年度順

	Kinect V2	ENSENSO	R200	Astra	SR300	D415	D435
							
メーカー名	Microsoft	iDS	Intel	Orbbec	Intel	Intel	Intel
型番	GT3-00005	N35-804-16-IR	82634DSB2P	Orbbec Astra	82535IVCQSPL04N	82635ASRCDVKHV	82635AWGDVKPRQ
販売時期	2014	2015	2015	2016	2016	2018	2018
概略価格	2万1578円	約80万円	約1万9000円 (開発ツール セット料金)	約1万7000円	約1万8000円	約2万2677円	約2万4400円
デプス解像度	512x424	1280x1024	628x468	640x480	640x480	1280x720	1280x720
撮影範囲 (角度・距離)	水平70°x 垂直60°  0.5~4.5m	   0.3~3.0m	   0.5~3.5m	水平60°x 垂直49.5°  0.6~8.0m (最適0.6~5.0m)	   0.2m~1.5m	水平69.4°x 垂直425.°x 斜め77°  0.3m~10m	水平91.2°x 垂直65.5°x 斜め100.6°  0.2m~10m
計測形式	Time of Flight	Stereo + Light Coding	Stereo + Light Coding	ランダムドット パターン (Light Coding)	ランダムドット パターン (Light Coding)	Stereo + Light Coding	Stereo + Light Coding
インター フェイス	USB 3.0	Ethernet	USB 3.0	USB 2.0	USB 3.0	USB 3.0	USB 3.0
電源	○	○	×	×	×	×	×

# 3次元センサの比較

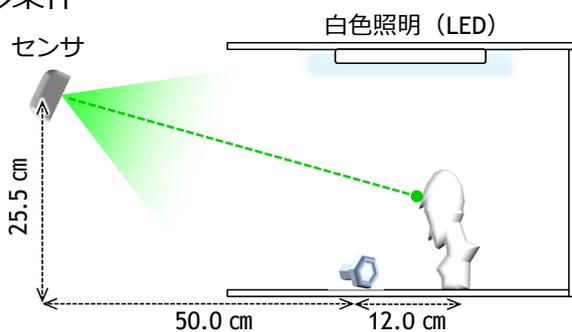
## ■ 計測データの比較

- ① コップ (鏡面加エステンレス)
- ② 蝶番 (SPCC)
- ③ 像 (石膏)
- ④ 像 (石膏・つや消し黒色塗装)
- ⑤ ビン (透明ガラス)
- ⑥ ギア (プラスチック)
- ⑦ ボルト・ナット (ステンレス)

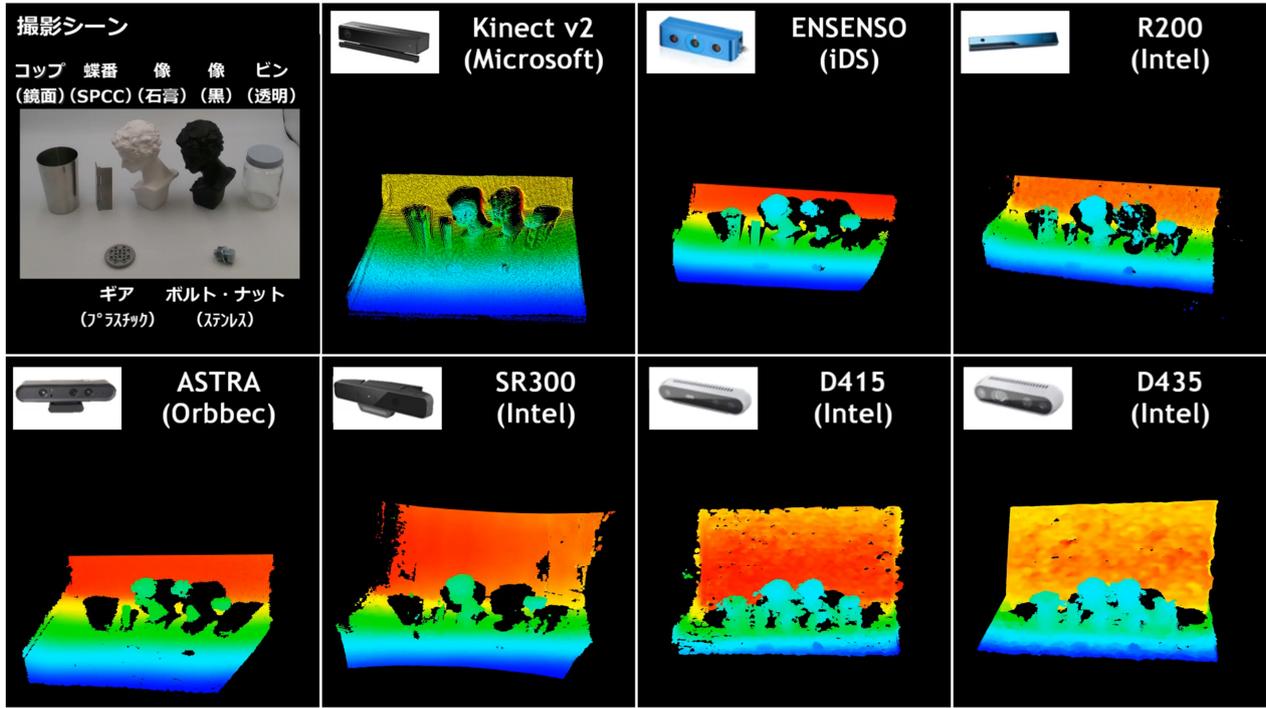


撮影シーン

撮影条件



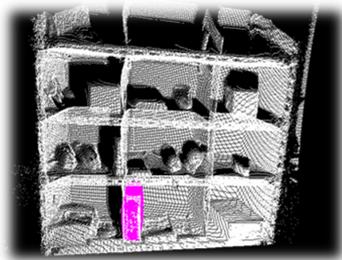
# 3次元センサの比較



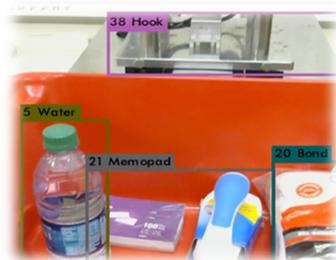
【所感】

- 拡散表面物体については ENSENSO による計測が高品質
- 全体的に情報欠落が少ないのは D415・D435, 逆に多いのはSR300 (鏡面・黒色部分) である.
- 平面安定性が高い (平面が平面として計測される) のは ENSENSO・ASTRA である. 逆に低いのは D415・D435 である.
- 黒い物体に対する計測精度は D415・D435 が高い.
- 死角が少ないのは Kinect v2・ENSENSO・SR300 である.
- **見えにくいもの ①金属 ②黒色 ③透明 ④小型**

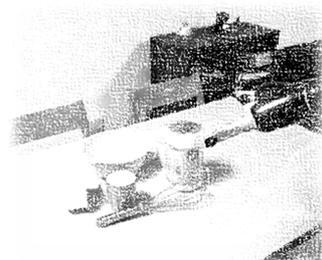
どこにある？



なにがある？



どう掴む？

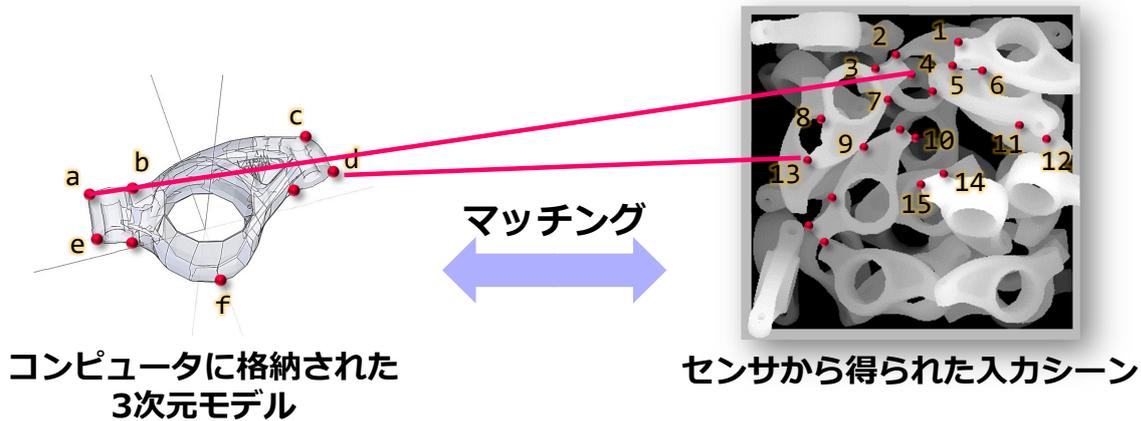


# 局所特徴量を利用した特定物体認識

## 物体認識の基本技術

Advanced Sensing & Machine Intelligence Group.  
Chukyo University

### 特定物体認識技術の例 (キーポイントマッチング)

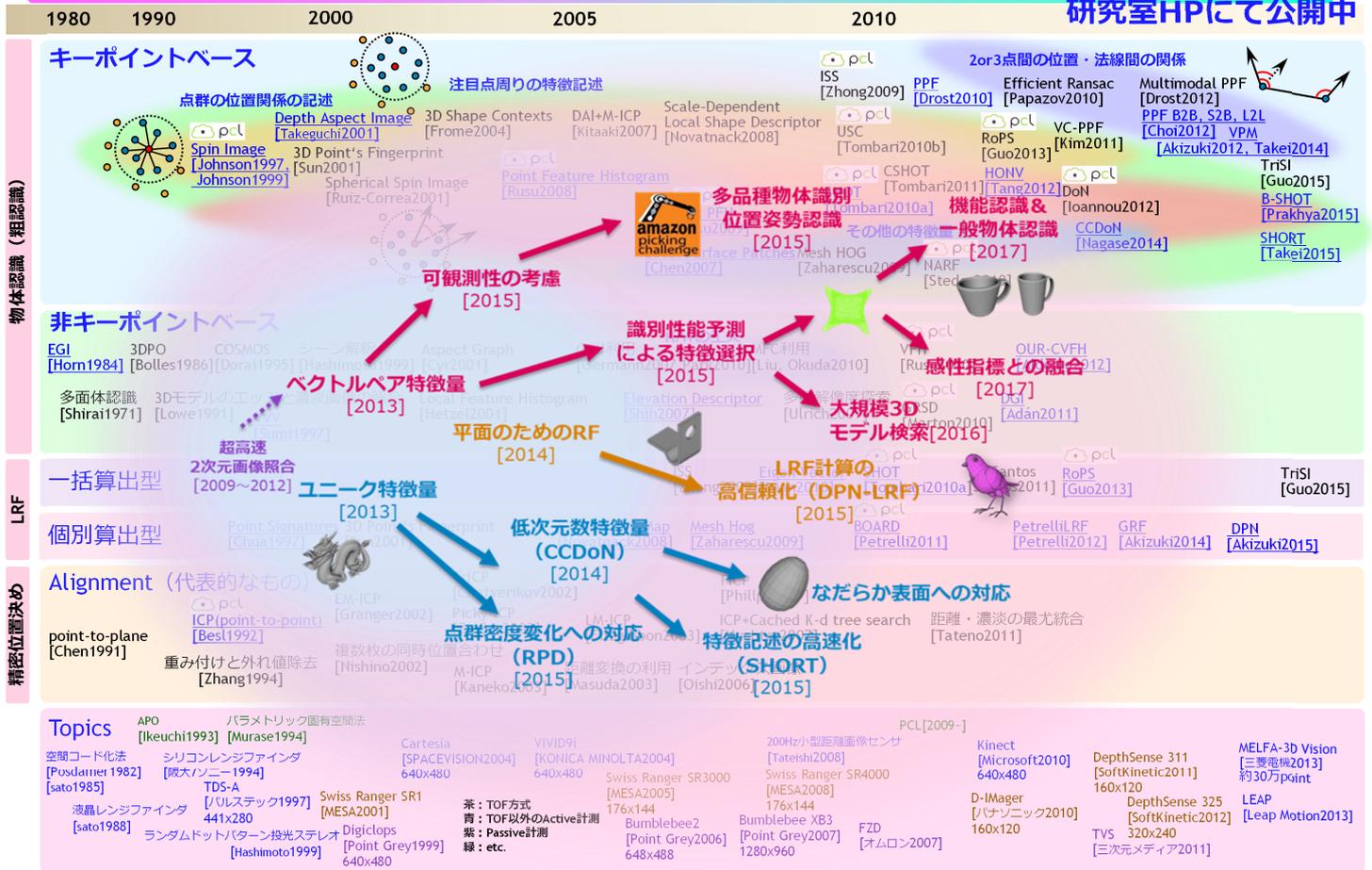


座標 (x,y,z)	点を特徴づける数値 (特徴量)	座標 (x,y,z)	点を特徴づける数値 (特徴量)
a 43.5, 50.2, 73.6	5, 2, 6, ..., 9, 20, ...	1 53.5, 55.1, 71.4	74, 36, 35, ..., 19, 76, ...
b 37.0, 25.3, 43.1	45, 2, 4, ..., 1, 3, ...	2 64.2, 60.0, 71.4	103, 2, 45, ..., 1, 34, ...
...	...	...	...
d 81.6, 25.2, 33.8	3, 52, 3, ..., 11, 4, ...	4 101.6, 35.3, 71.8	5, 2, 6, ..., 9, 20, ...
...	...	...	...
g 11.5, 9.5, 234.1	50, 32, 4, ..., 3, 13, ...	13 25.1, 220.9, 5.1	3, 52, 3, ..., 11, 4, ...
j 8.9, 794.4, 212.3	1, 95, 34, ..., 13, 5, ...	14 734.7, 5.4, 20.1	8, 34, 24, ..., 92, 44, ...
...	...	...	...
...	...	27 21.2, 7.5, 265.6	3, 52, 3, ..., 11, 4, ...
...	...	30 95.0, 26.1, 578.8	58, 29, 99, ..., 45, 320, ...
...	...	...	...

→ マッチングを成功させるためには **よい3次元特徴量** の設計が重要

# 特定物体認識（モデルベース手法）の研究マップ

研究室HPにて公開中

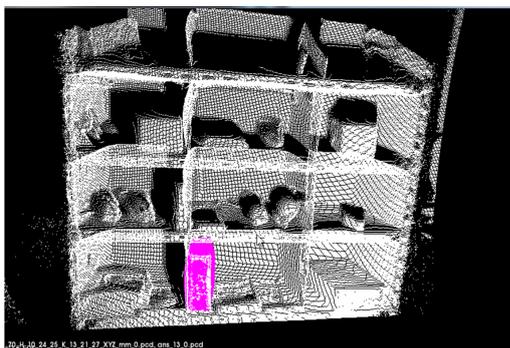


## 物体認識結果

箱状物体



入力シーン外観

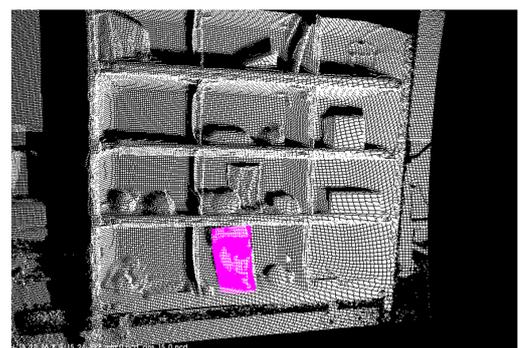


認識結果

形が変形しやすい袋状物体

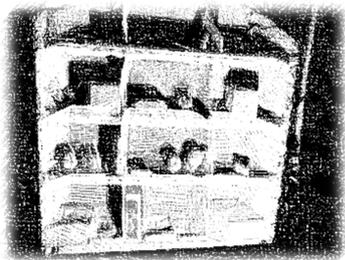


入力シーン外観

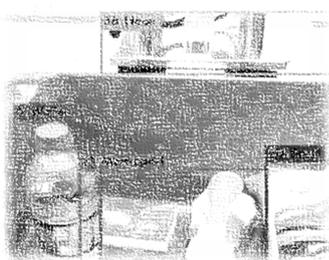


認識結果

どこにある？



なにがある？



どう掴む？

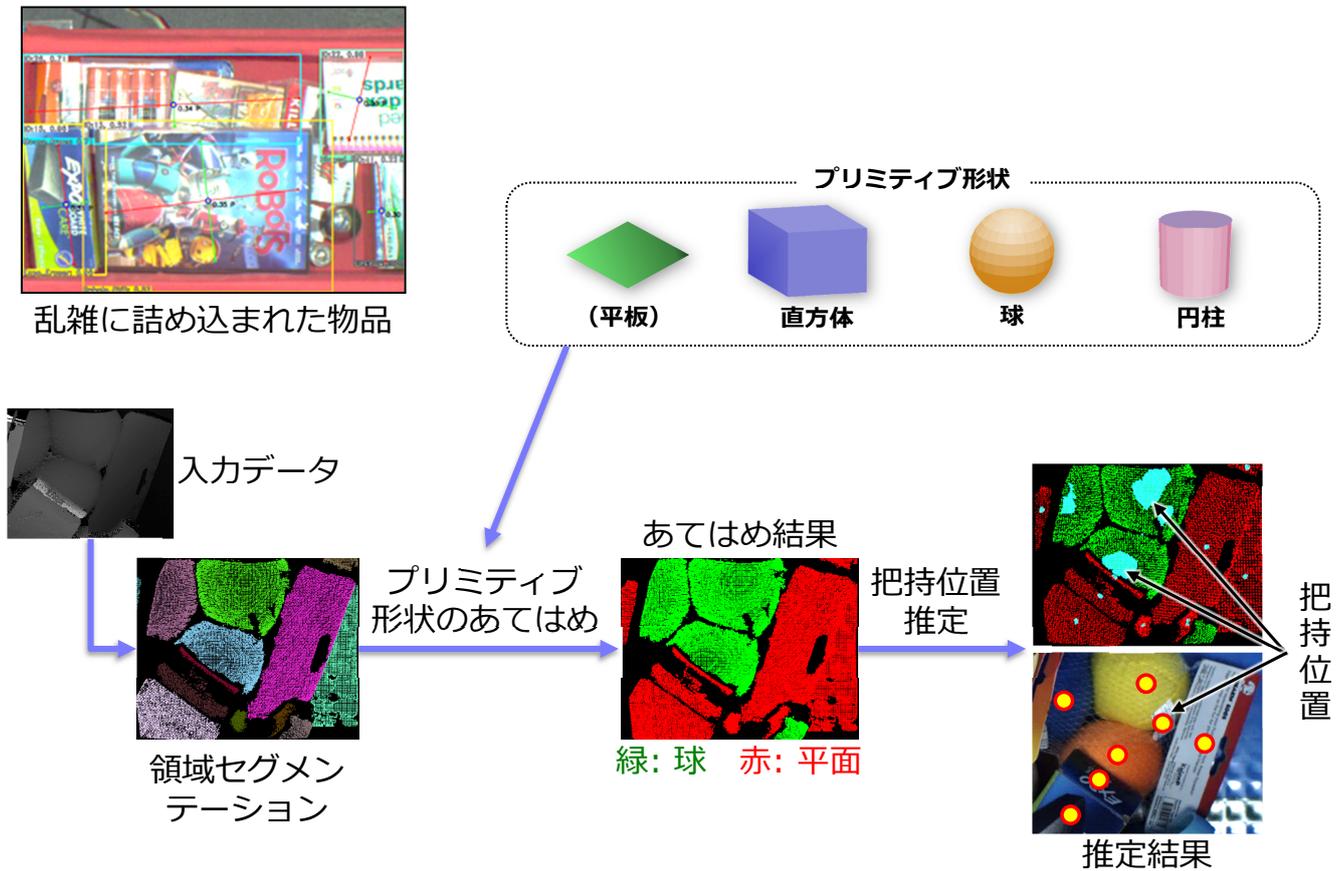


## モデルレスでの把持

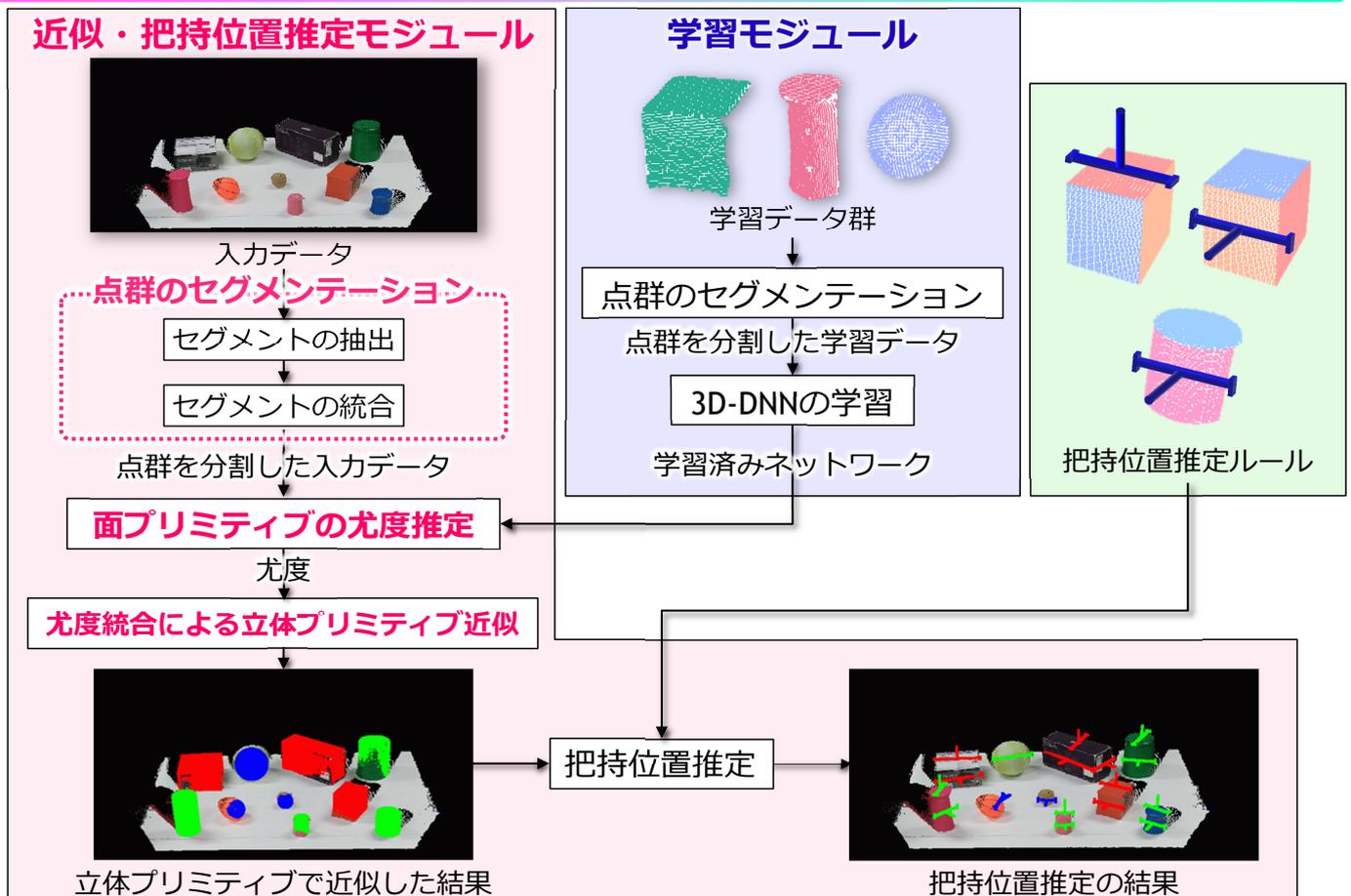
---

プリミティブ形状近似による  
把持位置決定

# アイデア



## プリミティブ形状近似に基づく把持位置推定



## 面プリミティブ・立体プリミティブの近似性能

面プリミティブの近似成功率 [%]

手法	RANSAC法	提案手法
"Rectangle"	81.7	85.3
"Disk"	0.0	69.8
"Cylindrical Surface"	52.9	92.3
"Spherical Cap"	96.0	95.1

立体プリミティブの近似成功率 [%]

手法	曲率法	End-To-End 3D DNN *	提案手法
"直方体"	92.0	83.0	94.0
"円柱"	20.0	85.0	94.0
"球"	87.0	96.0	96.0

\* : 3D ShapeNets 法



## 周囲の状況に応じた把持

ピッキングリスク最小化による  
ロボット動作パラメータ決定

# ロボットハンドによる対象物ピックアップ



成功例

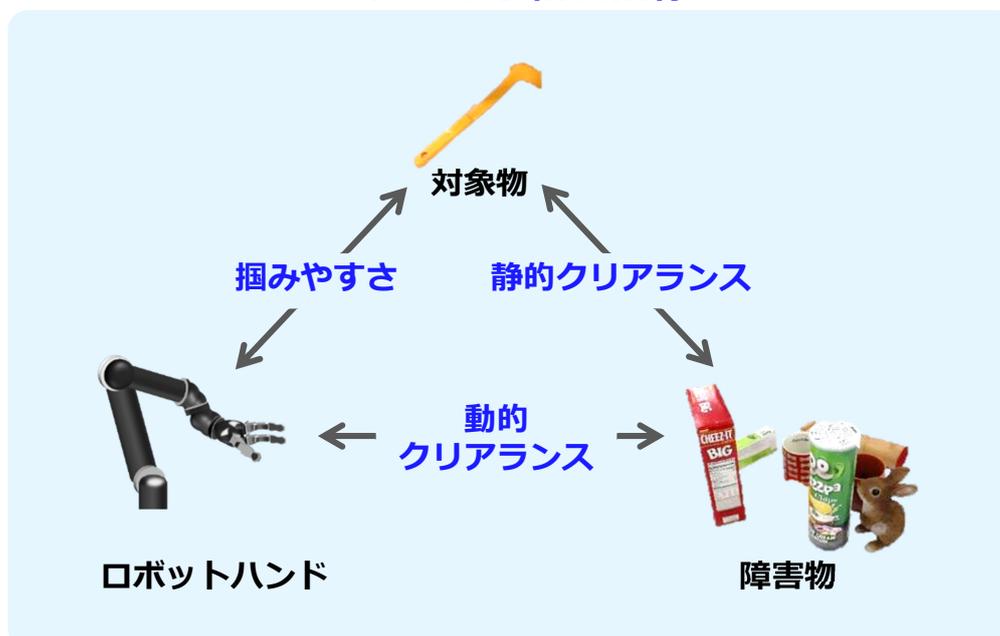


失敗例

## 基本アイデア

1. 環境内の各物体間の相対的な余裕度を数値化
2. 余裕度指標を統合し, 最大化する動作パラメータを決定

### 3種類の余裕度指標



# 把持余裕度推定モデルによる把持パラメータ推定

## 3つの余裕度指標を統合した把持余裕度推定モデル

### 把持余裕度推定モデル

$$P_S(\mathbf{p}) = P_{H-O}(\mathbf{p}) \{ \alpha P_{H-T}(\mathbf{p}) + \beta P_{T-O}(\mathbf{p}) \}$$

把持余裕度      ①ハンドと障害物の動的クリアランス      ②対象物の掴みやすさ      ③対象物と障害物の静的クリアランス

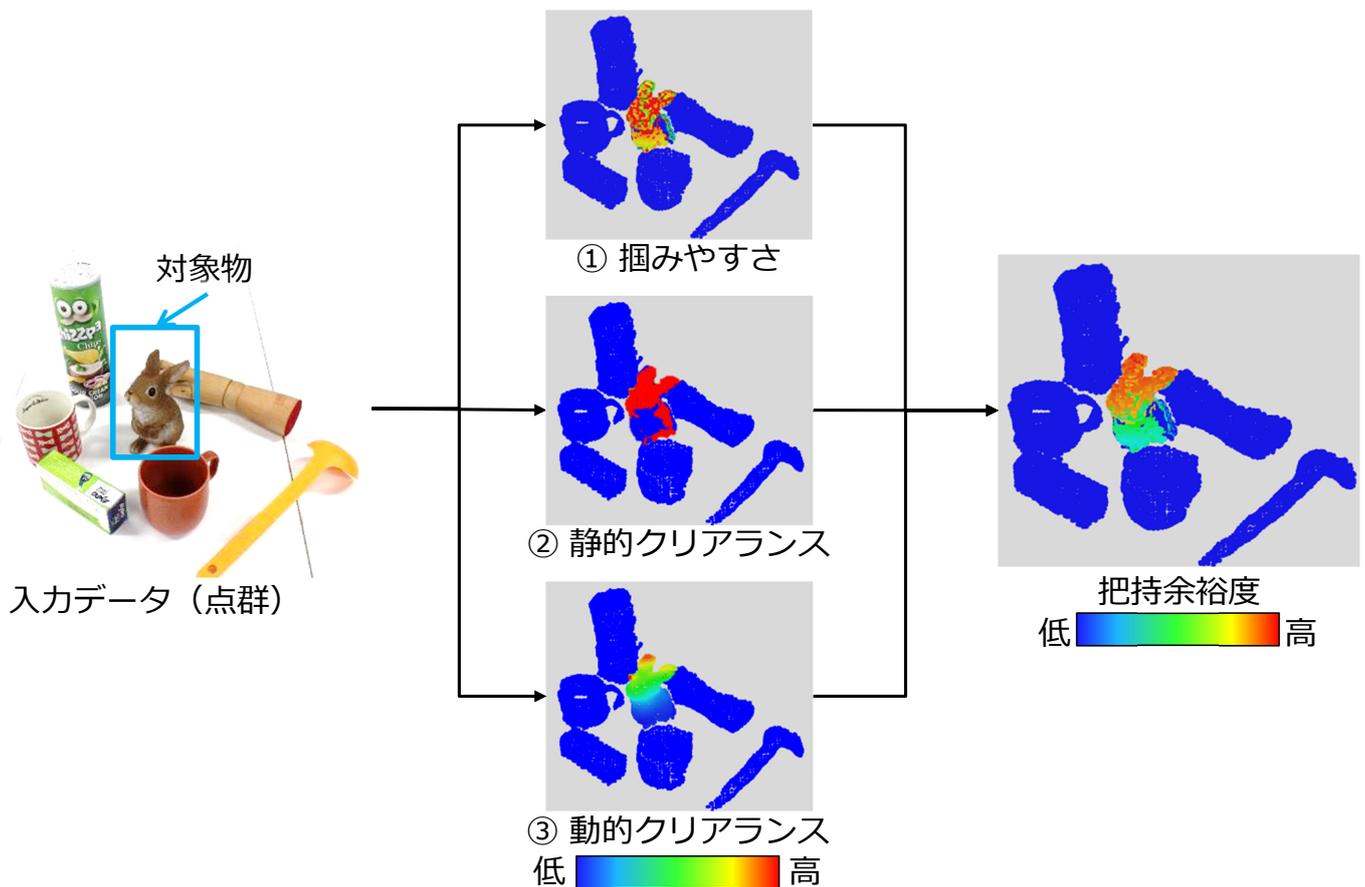
※  $\mathbf{p}$  : 把持パラメータ (把持位置, 姿勢, 開口幅)

$P_S(\mathbf{p})$  を最大化する把持パラメータ  $\tilde{\mathbf{p}}$  を推定

$$\tilde{\mathbf{p}} = \operatorname{argmax}_{\mathbf{p} \in \mathcal{S}} P_S(\mathbf{p})$$

$\mathcal{S}$  :  $\mathbf{p}$  の候補集合

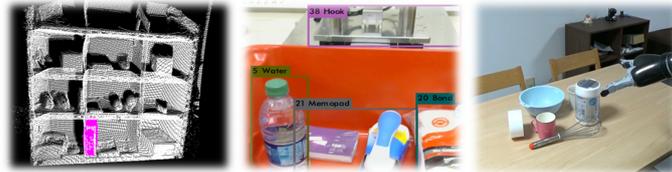
## 算出された余裕度指標



## ロボット競技大会への適用

### 認識 + ロボット制御の総合技

どこにある？    なにがある？    どう掴む？



Advanced Sensing & Machine Intelligence Group,  
Chukyo University

## Amazon チャレンジへの挑戦

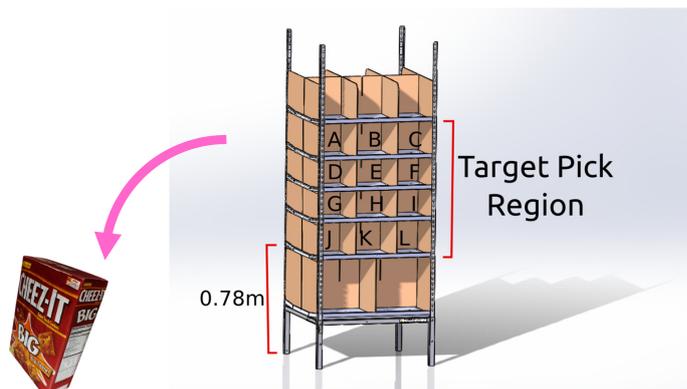
(三菱電機, 中部大とコラボ)

- 2015年 第1回大会 (シアトル)
- 2016年 第2回大会 (ライブツィヒ)
- 2017年 第3回大会 (名古屋)



2017年度 Stowタスク部門世界3位

- 制限時間内にどれだけ正確にアイテムをピックアップできたかを競う。
- 課題は年々実用的に、また、難しくなった。



## ■ 挑戦的な対象物

## + 当日配布 (未知) アイテム

競技開始の  
45 分前に配布



## 本日の話題

### 1. イントロダクション

- ロボットビジョンの基本課題
- 研究のモチベーション

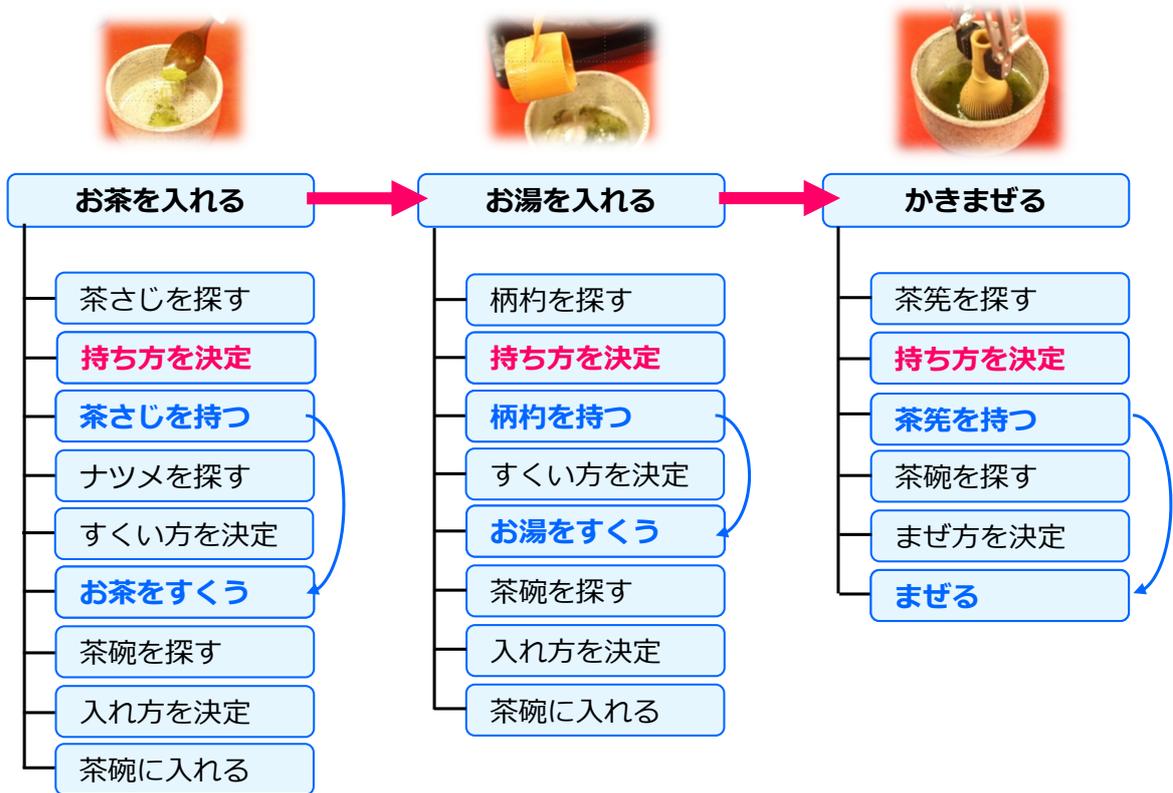
### 2. 実現できたこと, できそうなこと

- 特定物体の認識 (位置・姿勢認識)
- 把持動作のためのロボットビジョン

### 3. “機能” 認識とその応用

- 全自動お茶会ロボットを目指して
- 人による教示動作のロボットへの転移

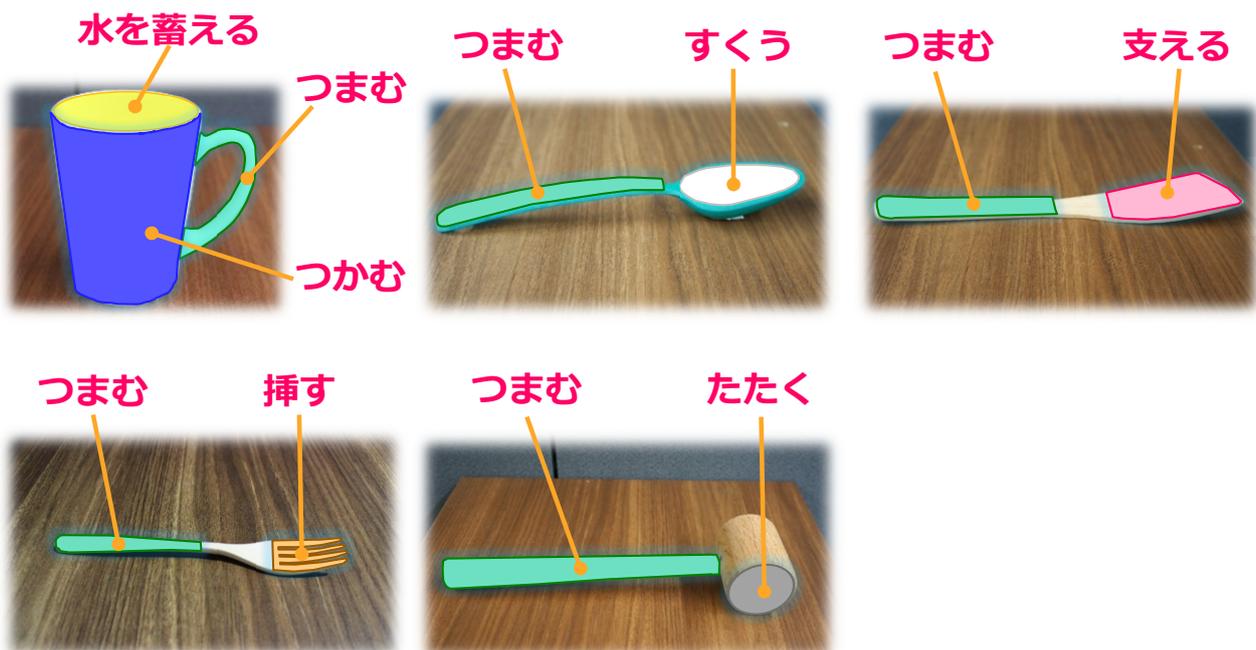
# お茶をたてる動作（ロボットの場合）



人間なら，道具の**使用目的**を想定して**持つ**ことができる。

⇒ **道具の使い方がわかるロボット**

# 日用品は“形”に“機能”を持っている



# 機能の発現に影響を与える主要要因

## 形状



## 動き (強さ・方向)



## 材質

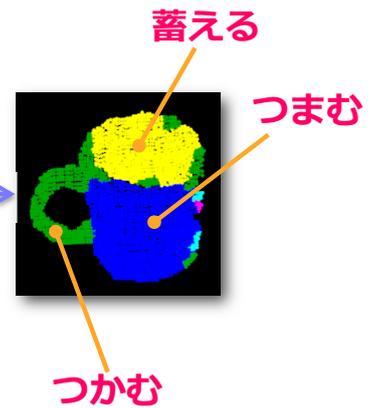
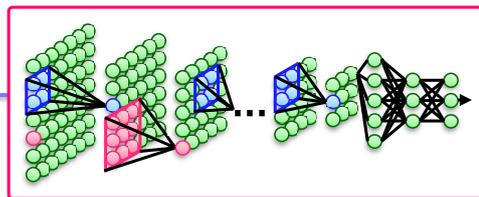


# 機械学習による“機能”認識

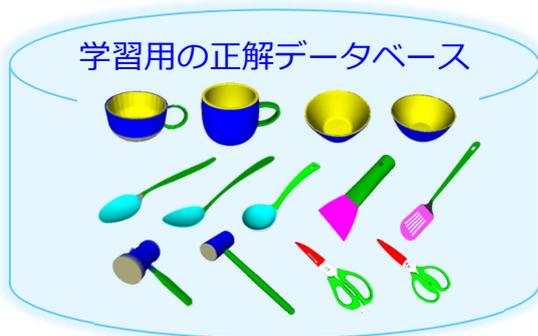
Depth or RGB

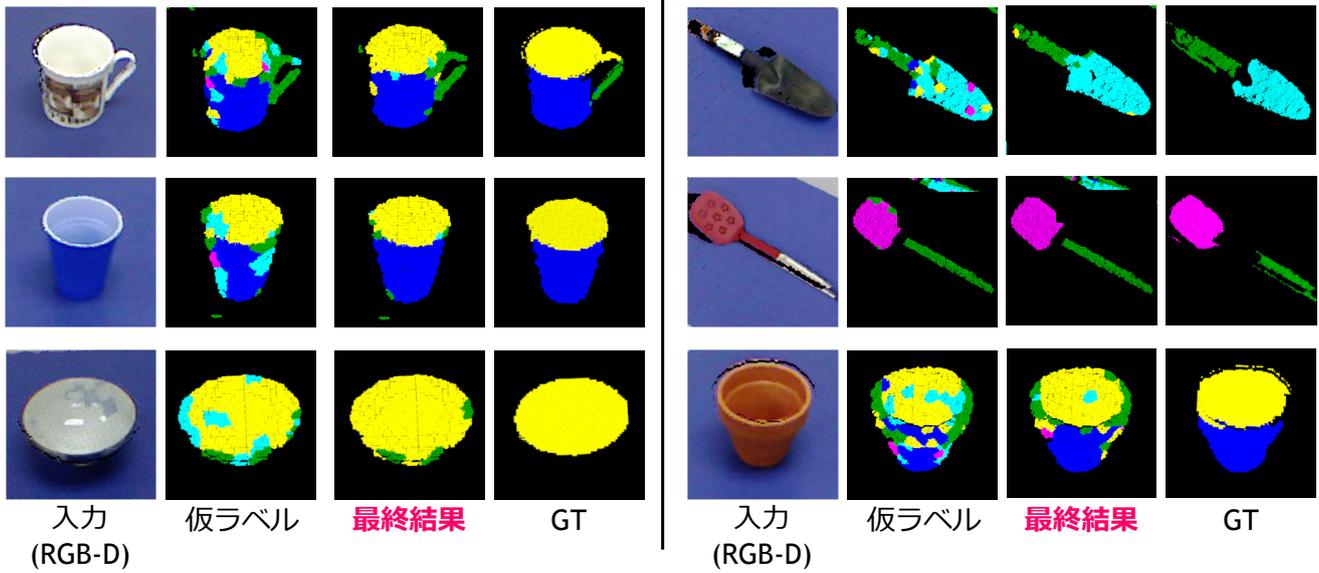


Deep Learning (他手法も検討)



学習用の正解データベース





■ Contain   
 ■ Grasp   
 ■ Wrap-Grasp   
 ■ Scoop   
 ■ Support

## 機能属性ラベル付きデータセットの公開

(2019年5月24日公開)

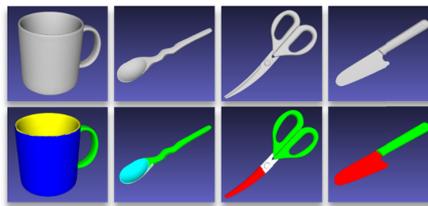


<http://isl.sist.chukyo-u.ac.jp/archives/nedopro/>

### a. 日用品 全周囲 3D モデル [モデル16品種321個, 機能9クラス]



日用品の例



機能属性ラベル付き 3D モデルの例

**データ形式**

3Dモデル	ply 形式 obj 形式 stl 形式
機能属性ラベル	ply 形式

**使用例**

全周囲データを用いたセグメンテーション [PointNet など]

### b. 日用品シーン (2.5D) [モデル10品種11014枚, 機能9クラス]



日用品シーンと機能属性ラベルの例

**データ形式**

RGB画像	png 形式
2.5Dデータ	pcd 形式
機能属性ラベル	png 形式

**使用例**

画像・点群ベースの機能認識 (セグメンテーション) [SegNet, U-Net など]

a. のデータを用いたモデルマッチング

### c. 機能属性セグメンテーション用シーン (2D) [モデル6品種2799枚, 機能6クラス]



茶道具シーンと機能属性ラベルの例

**データ形式**

RGB画像	jpg 形式
機能属性ラベル	png 形式

**使用例**

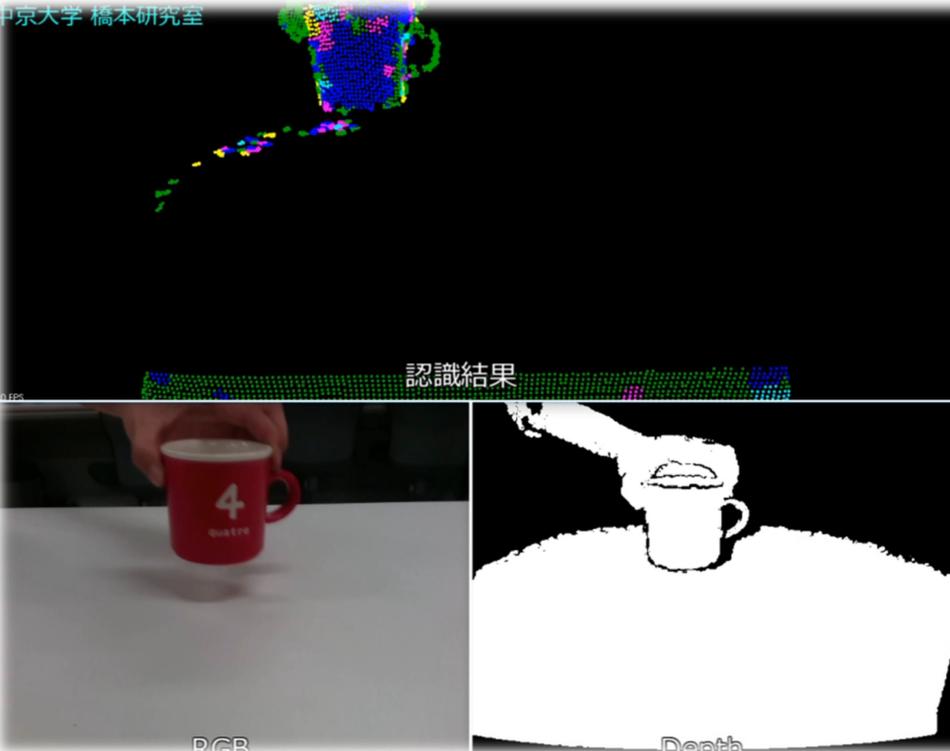
画像ベースの機能認識 (セグメンテーション) [SegNet, U-Net など]

# 機能を認識しているようす (10fps)

(44秒)

中京大学 橋本研究室

x1



機能属性

- 液体を蓄える
- つまむ
- つかむ
- 支える
- すくう

実行環境  
CPU: intel(R) core-i7(TM) 2.5GHz  
RAM: 8GB  
GPU: none

## “機能” から対象物の名前（種類）を知ることができる

さまざまな機能属性の含有比率を利用してカテゴリ名を推定



「コップ」カテゴリ

「スプーン」カテゴリ

「ハンマー」カテゴリ

水を蓄える つかむ つまむ すくう たたく 該当無し

入力データ

(34秒)

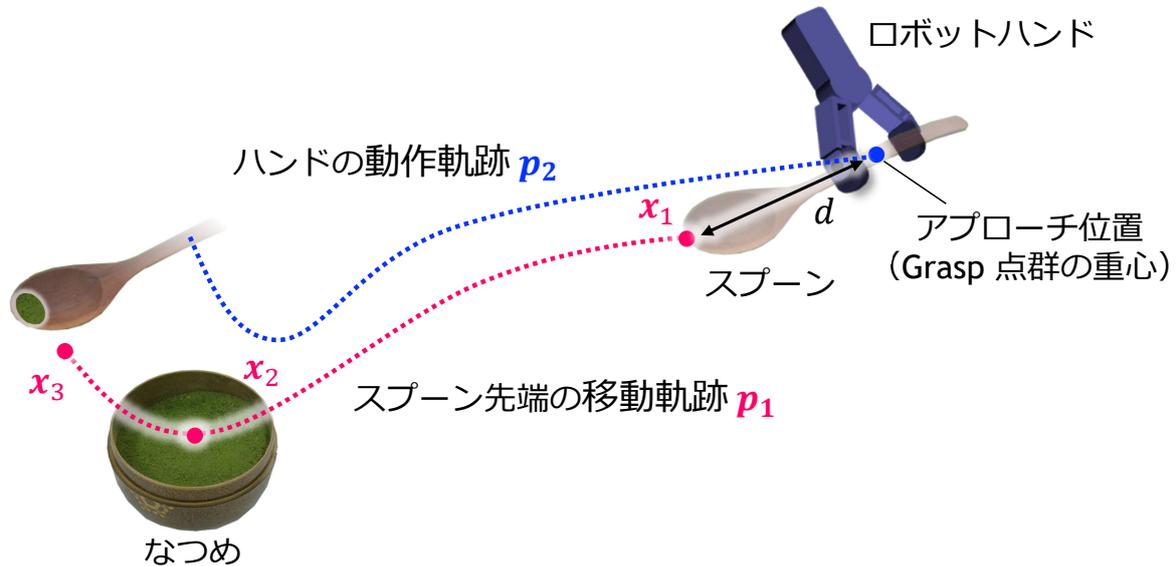


機能属性ラベル

水を蓄える つかむ つまむ たたく すくう 支える

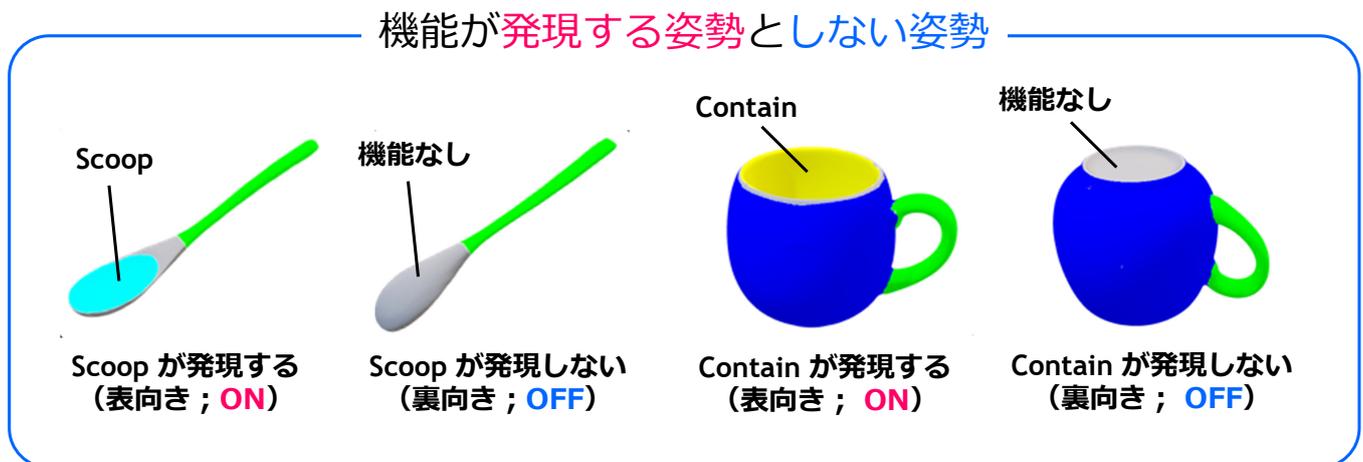
## “機能” がわかると何がうれしいか？ ①

初めて見るスプーンでも、適切なすくい動作（パラメータ）を決定できる。



## “機能” がわかると何がうれしいか？ ②

道具の姿勢によらず、動作を作ることができる。



機能情報をもとに、

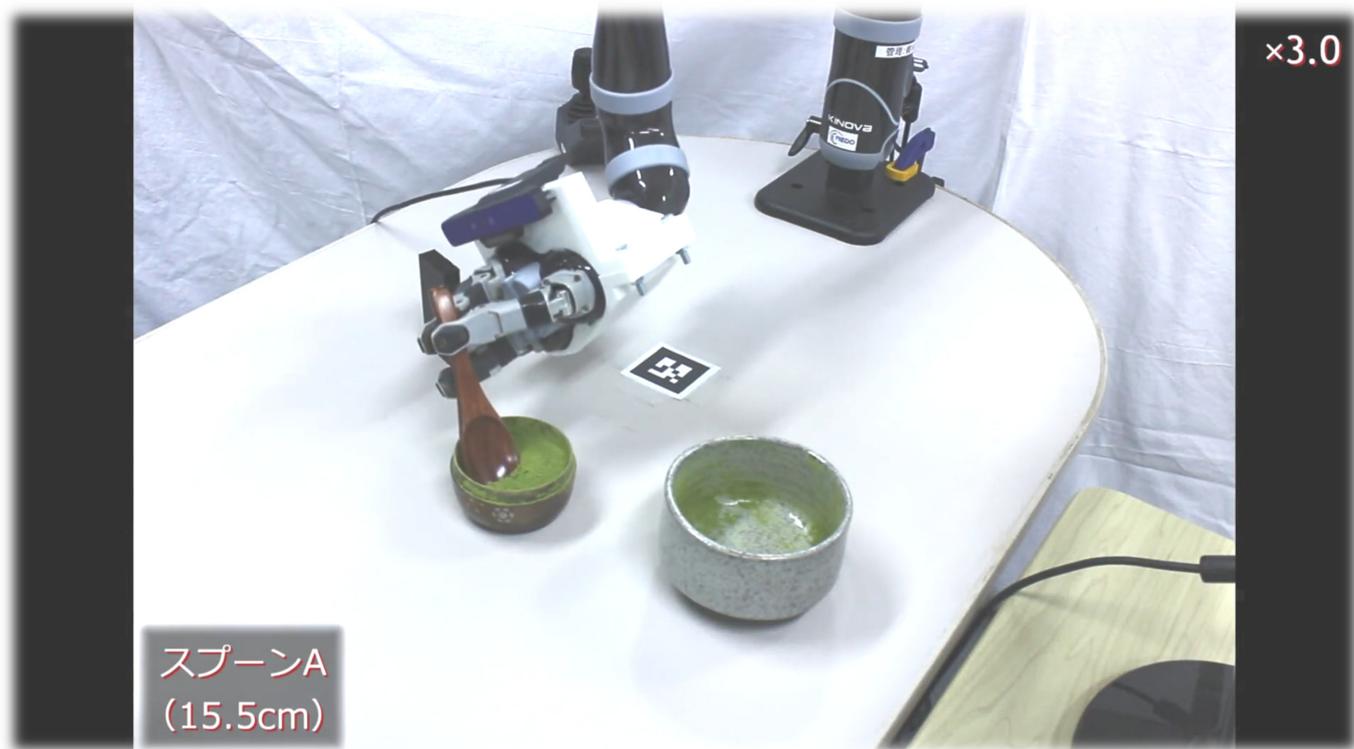
機能 = ON ならば そのまま動作

機能 = OFF ならば ONにしてから動作

# ①と②の実証

(22秒)

×3.0



スプーンA  
(15.5cm)

## お茶会ロボット2号

(2020年1月16日完成, YouTube公開中)

(3分50秒)



NEDO 産総研 中京大学 金沢大学

# お茶会ロボットにできないことの一例

「お茶を持ってきて」……この実現に、まだ多くの難題が残っている！

そもそも、この言葉を、細かいタスクに分解できない

- ・必要ならお湯を沸かす
- ・ナツメに蓋があったら開ける
- ・足りない道具があったら取ってくる
- ・
- ・

## 現在のお茶会ロボットにできないこと



## 人による動作教示のロボットへの転移

# 組立て作業で扱われている代表的な部品



(a) リンク



(b) コネクティングロッド



(c) コネクタ



(d) ボルト



(e) ナット

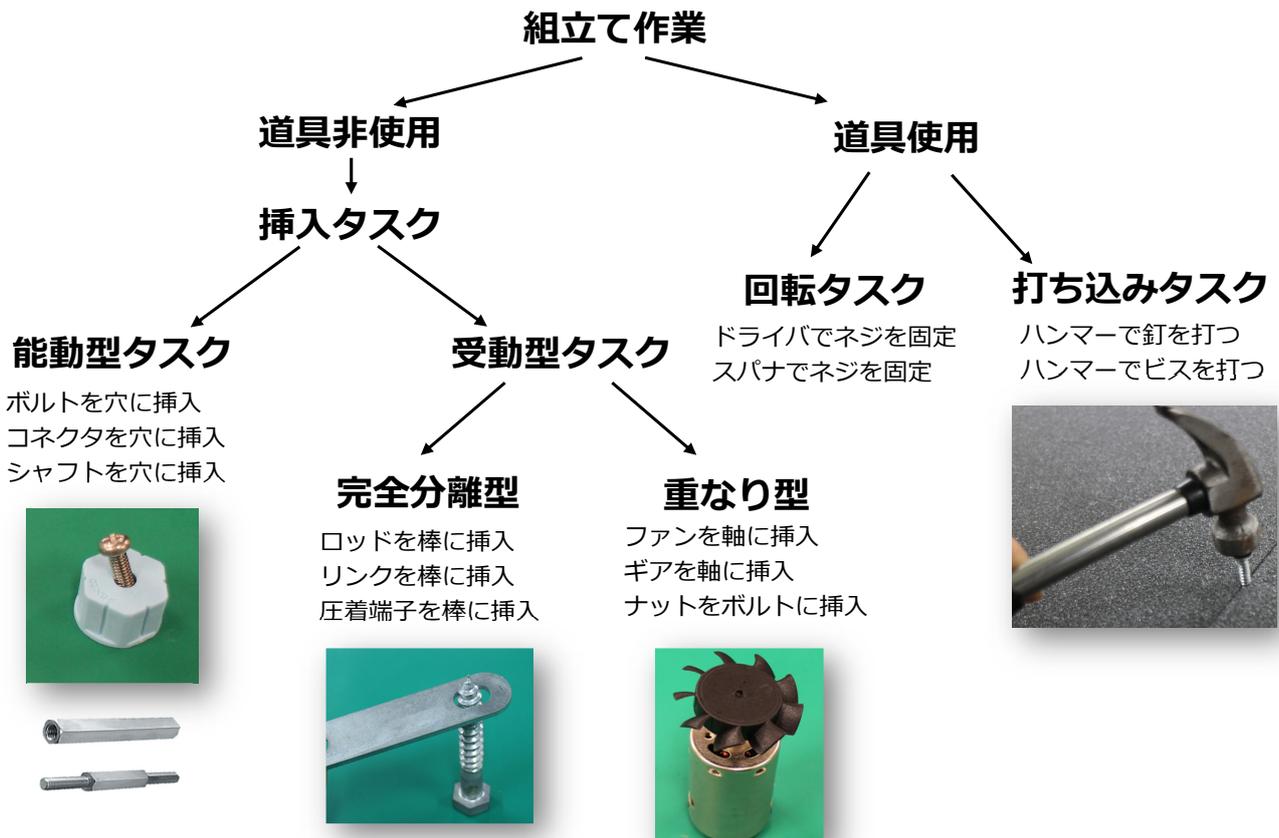


(f) ギア

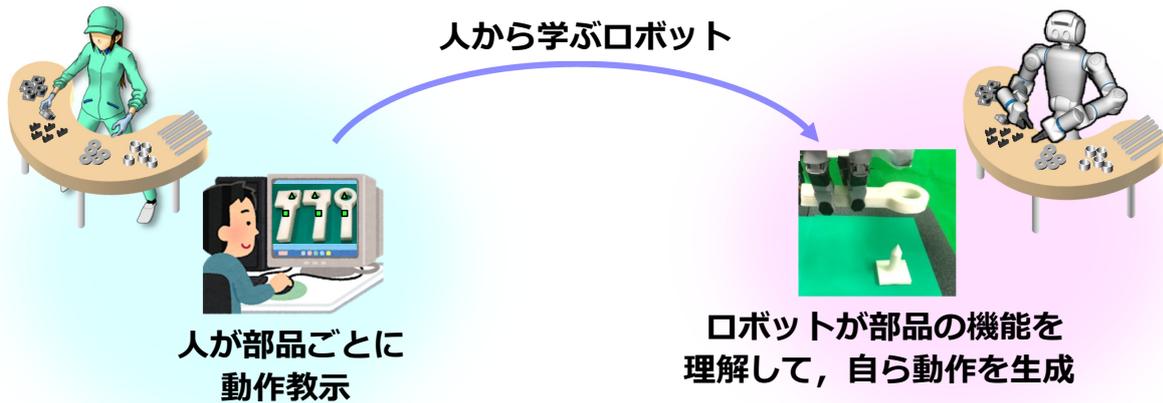
出典：ミスミ総合カタログ, <https://jp.misumi-ec.com/>

株式会社ゴーシュー, <http://www.gohsyu.com/product/>

# 代表的な組立て作業

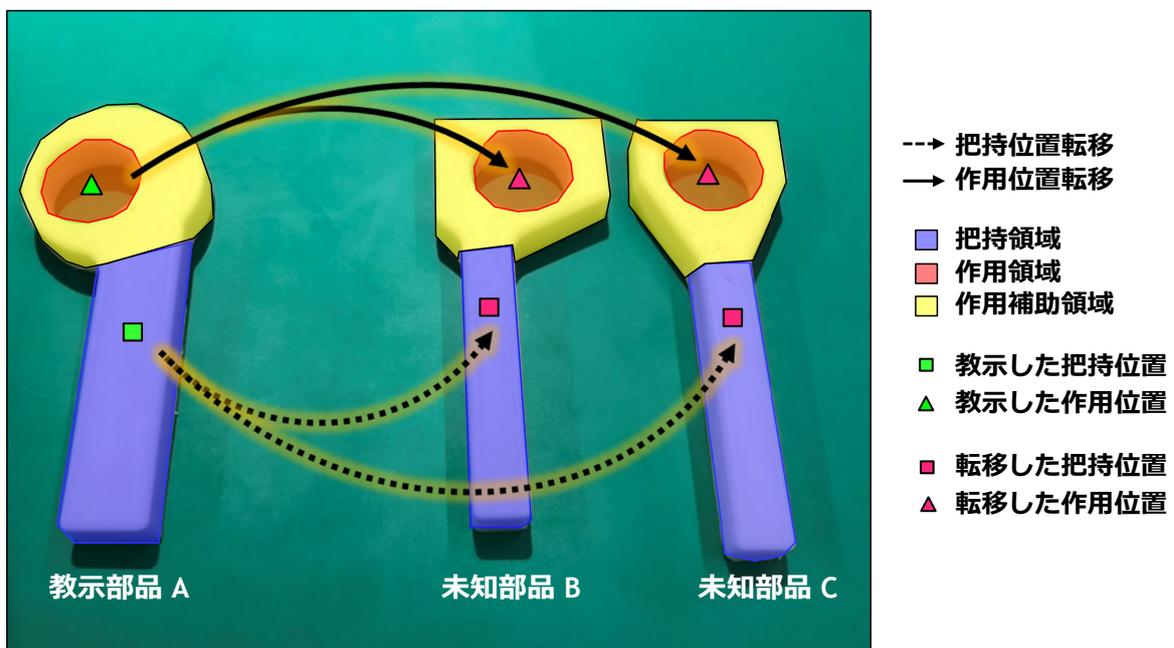


# ロボットによる組み立て作業の動作生成

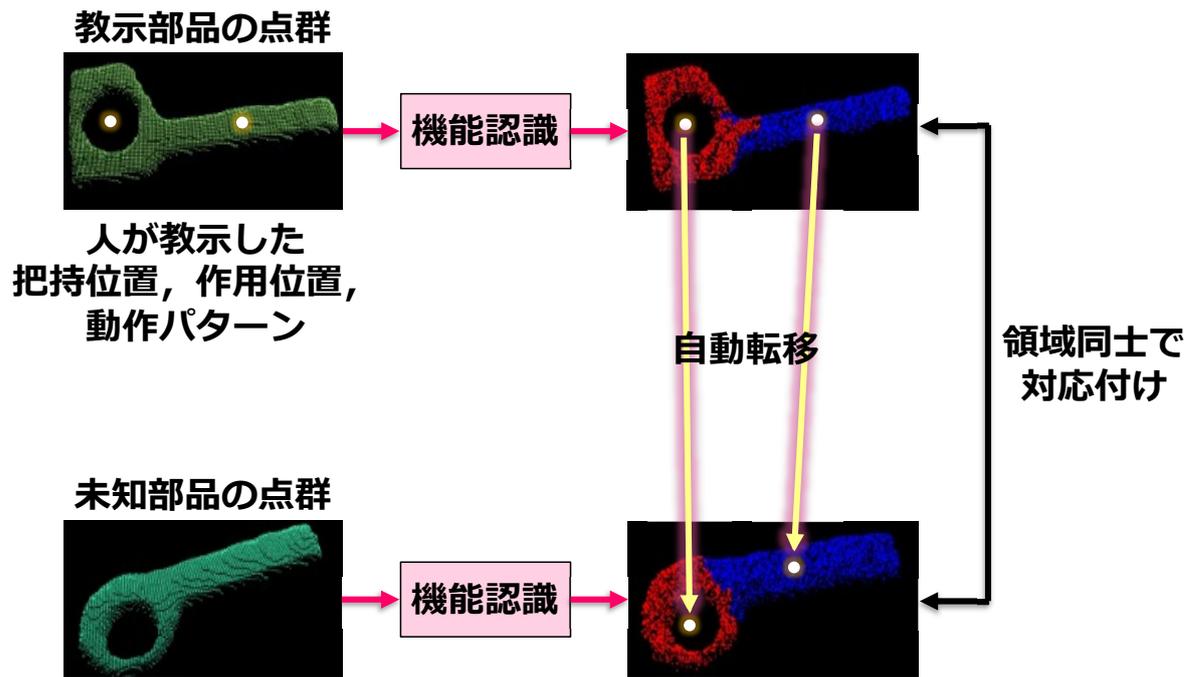


## 基本アイデア

目的 人間が教示したロボットパラメータを、未知部品に転移  
ロボットパラメータとは、把持位置と作用位置



# 提案手法の流れ



# ロボットによるロッド挿入実験



中京大学  
橋本研究室

ヒトがロボットパラメータを  
教示した場合の動作

*Thank you for your attention!*