配布用抜粋資料

2D/3D アプローチに基づく ロボットビジョン研究の概観と最新事例

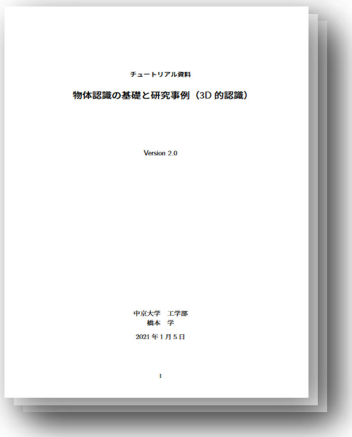
～ ロボットビジョン関連技術の紹介 ～

中京大学　工学部
橋本 学

関連情報

橋本研究室　技術文献のページ
http://isl.sist.chukyo-u.ac.jp/archives3/

3次元物体認識に関するチュートリアル資料
http://isl.sist.chukyo-u.ac.jp/wp-content/uploads/2021/05/ObjectRecognition-byHashimotoLab-3D.pdf



本日の話題

1．イントロダクション：　課題整理とロボットビジョンの概要

2．モノを視る‥‥２次元的アプローチ

戦略的画素選択に基づくテンプレートマッチングの高度化

（照明変動ロバスト化，識別性能向上，回転変動対応‥‥）

3．モノを視る‥‥３次元的アプローチ

ロボット動作生成にかかわる諸技術

（機能認識，動作教示簡素化，言語指示‥‥）

異なるセンサ間のギャップの低減

4．ヒトを観る

匠の技の分析　←時間があれば

5．まとめ

✓ **基礎的なチュートリアルは割愛させていただき，最新の研究事例を中心に．**

本日の話題

1．イントロダクション：　課題整理とロボットビジョンの概要

2．モノを視る‥‥２次元的アプローチ

戦略的画素選択に基づくテンプレートマッチングの高度化

（照明変動ロバスト化，識別性能向上，回転変動対応‥‥）

3．モノを視る‥‥３次元的アプローチ

ロボット動作生成にかかわる諸技術

（機能認識，動作教示簡素化，言語指示‥‥）

異なるセンサ間のギャップの低減

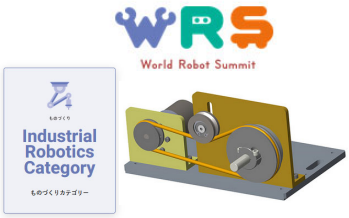
4．ヒトを観る

匠の技の分析

5．まとめ

主要３分野におけるロボット/ビジョンの課題整理

ロボット タスク	Pick and Place	Pick and Operation
	動作パターン種類 少（事前に生成）	動作パターン種類 多（その場で生成）
ロボット ビジョン	物体モデル利用 可	物体モデル利用 難
	対象物の品種 複数だがほぼ既知	品種 多量かつ未知

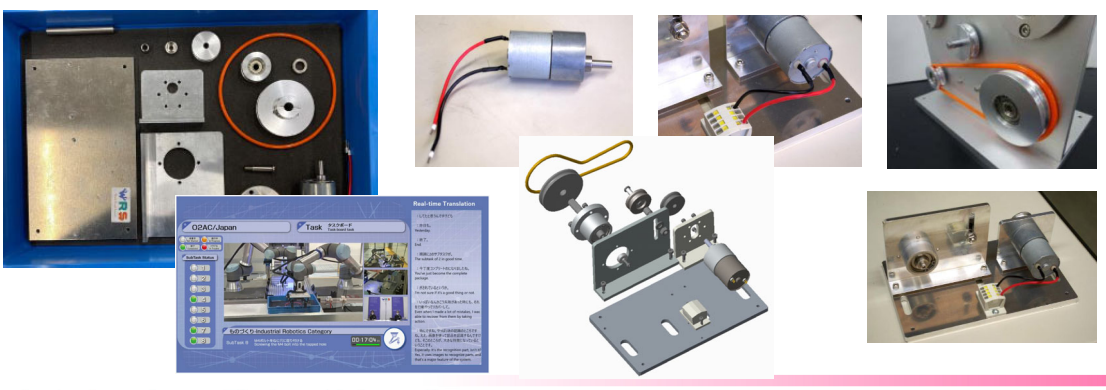


・オムロン・阪大・産総研・中京大
・プーリーなどの部品組み付け

・三菱電機・中部大・中京大
・商品取り出し，補充

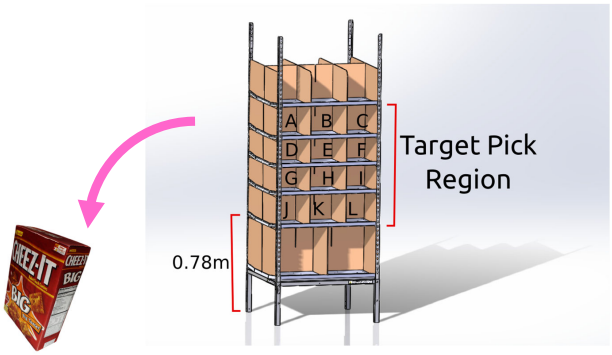
WRS：World Robot Summit

- 2021年9月9日～12日　愛知国際展示場
- ものづくりカテゴリー　Assembly Challenge（製品組立チャレンジ）
- チーム編成：　オムロン，大阪大学，産総研，中京大学
- 課題
  - タスクボードタスク
    - **プーリーなどの部品をタスクボード上の指定位置に組み付ける．**
  - 組み立てタスク（サプライズ製品有/無）
    - **トレイ上の複数の部品を使ってベルトドライブユニットを組み立てる．**
    - 直前に指定された製品（サプライズ製品）の組み立てを行う．
- 結果：７チーム中３位入賞（＋人工知能学会賞）



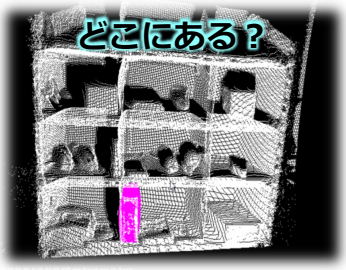
ARC：Amazon Robotics Challenge

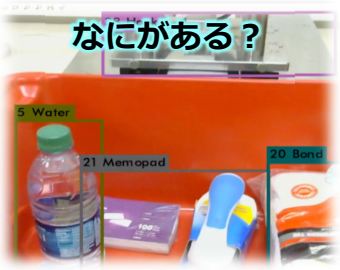
- 2015年～2017年　USA，Germany，Japan
- チーム編成：　三菱電機，中部大学，中京大学
- 課題
  - Pickingタスク
    - **棚から指定されたアイテム（商品）を取り出す．**
  - Stowタスク
    - **棚にアイテムを補充する．**
    - 2017年大会では，競技開始直前にアイテムが追加された．
- 結果：Stowタスクで16チーム中３位



ロボットビジョンの共通課題

Pick → **Operation**

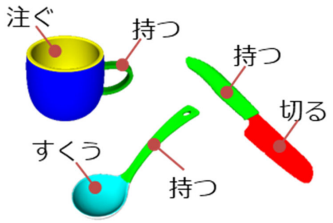




特定物体認識
（位置姿勢認識）

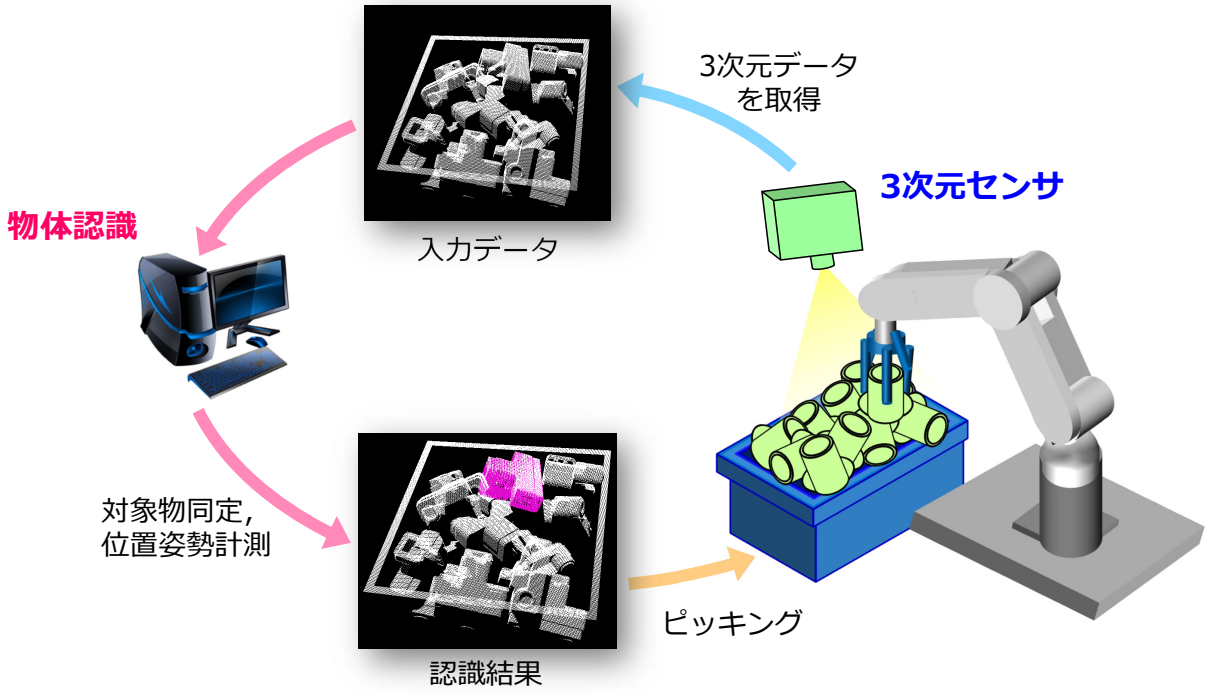
一般物体認識
（モノの名前の認識）

機能認識
（使い方の認識）



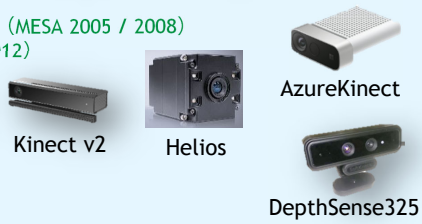
ロボットビジョンの基本構成（ピッキングロボットへの適用例）

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	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） Bumblebee2　TVS　SV-M-S1 マルチベースラインステレオ法 [Okutomi1993] 視体積交差法 [Matsuyama2002]	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） 空間コード化法 Cartesia（SPACEVISION 2004） MELFA-3D Vision（三菱電機 2013） RV1100（キヤノン） 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, 305（Intel 2016, 2019） ステレオ+パターン投光法 R200（Intel 2015），SR300，SR305 D415（Intel 2018），D435（Intel 2018） ENSENSO N35（iDS 2015） Structure core（Occipital 2018） 位相シフト法 [Halioua1989], [Zhao1994] MELFA-3D　Cartesia　RV1100　ENSENSO N35　Xtion PRO LIVE　R200　SR300, 305　ASTRA　D415　D435　Structure core
同軸測量	Shape(depth) from (De) Focus [Hiura1999] Shape from Motion	TOF: Time Of Flight Swiss Ranger SR3000 / SR4000（MESA 2005 / 2008） DepthSense325（SoftKinetic2012） Kinect v2（Microsoft 2014） Helios（LUCID 2019） AzureKinect（Microsoft 2019） 照度差ステレオ [Woodham1980]  AzureKinect　Kinect v2　Helios　DepthSense325

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°× 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	○	○	×	×	×	×	×

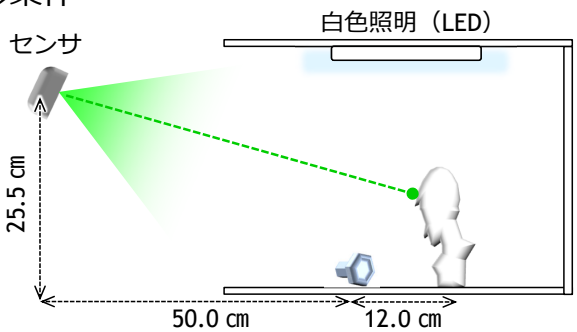
3次元センサの比較

■ 計測対象物

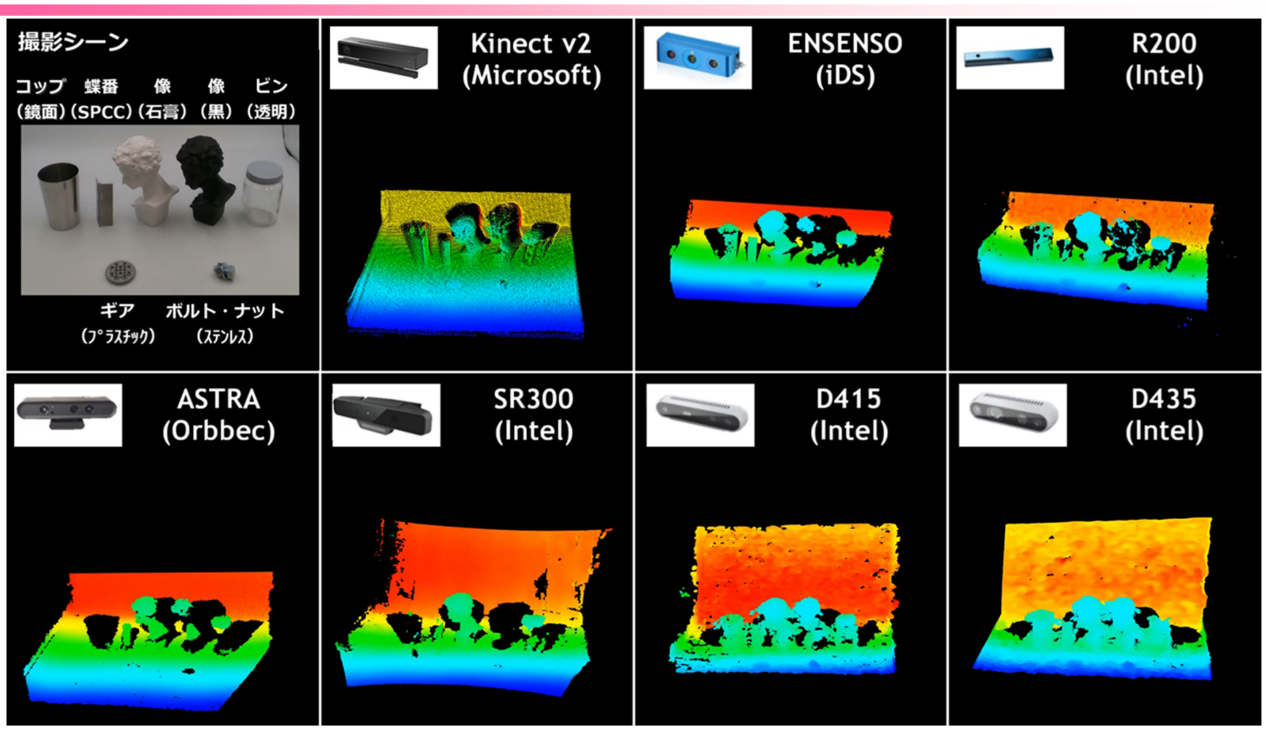
① コップ（鏡面加工ステンレス）
② 蝶番（SPCC）
③ 像（石膏）
④ 像（石膏・つや消し黒色塗装）
⑤ ビン（透明ガラス）
⑥ ギア（プラスチック）
⑦ ボルト・ナット（ステンレス）

撮影シーン

撮影条件



3次元センサの比較



【所感】

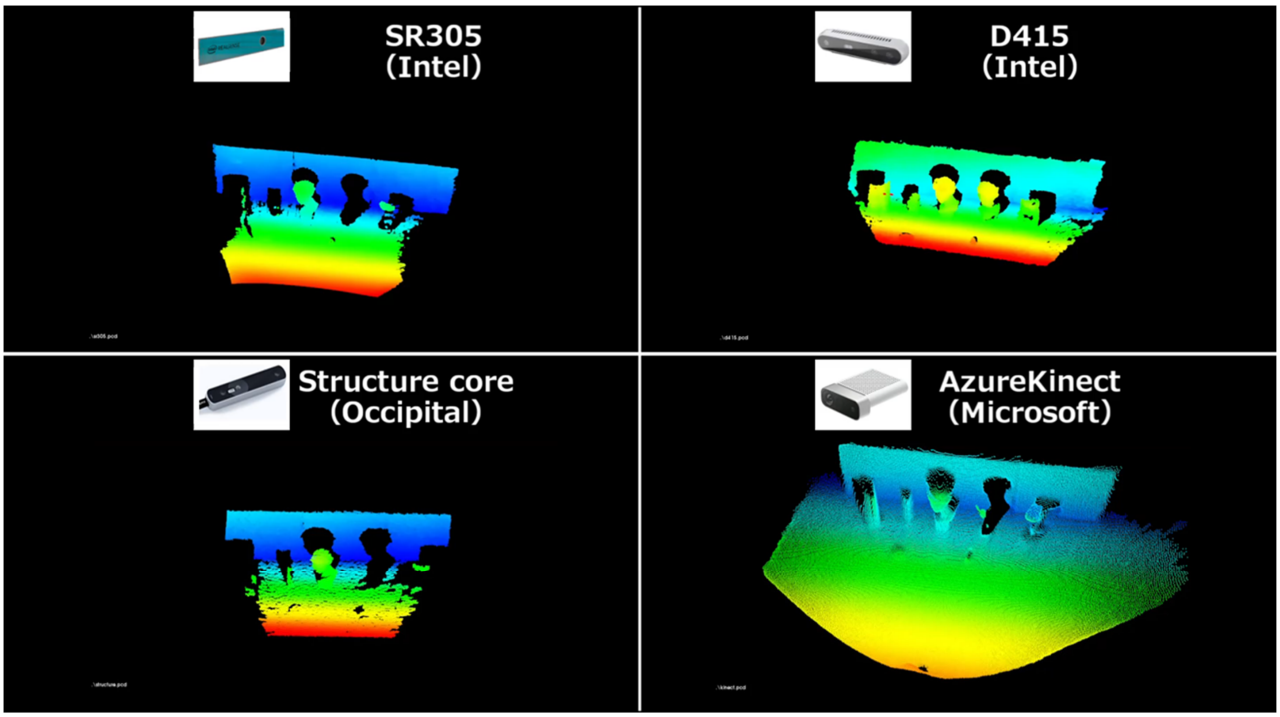
- 拡散表面物体については ENSENSO による計測が高品質
- 全体的に情報欠落が少ないのは D415・D435，逆に多いのはSR300（鏡面・黒色部分）である．
- 平面安定性が高い（平面が平面として計測される）のは ENSENSO・ASTRA である．逆に低いのは D415・D435 である．
- 黒い物体に対する計測制度は D415・D435 が高い．
- 死角が少ないのは Kinect v2・ENSENSO・SR300 である．
- **見えにくいもの　①金属　②黒色　③透明　④小型**

【追加】代替機種の性能評価

販売開始年度順

	D415	Structure core	SR305	AzureKinect
メーカ名	Intel	Occipital	Intel	Microsoft
型番	82635ASRCDVKHV	ST02D-C	82535IVSR305	1880
販売時期	2018	2018	2019	2019
概略価格	約2万2677円	約5万9800円	約1万6000円	約4万7025円
デプス解像度	1280x720	1280x960	640x480	1024x1024
撮影範囲（角度・距離）	水平69.4°x 垂直425.°x 斜め77° 0.3m～10m	水平59°x 垂直46°x 斜め70° 0.3m～10m	水平68.0°x 垂直41.5°x 0.2m～1.5m	水平120x 垂直120 0.25m～5.46m
計測形式	Stereo + Light Coding	Stereo + Light Coding	ランダムドットパターン (Light Coding)	Time of Flight
インターフェイス	USB 3.0	USB 3.0	USB 3.0	USB 3.0
電源	×	×	×	○

【追加】比較実験



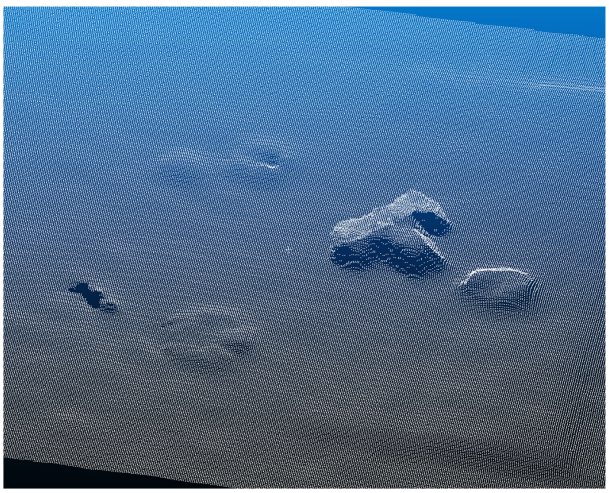
【所感】

- SR305，Structure core は，鏡面・黒色・透明物体に弱い．
- AzureKinectは，鏡面・透明部分は撮れているものの，精度は低い．
- D415 は，鏡面・黒色・透明部分の点群も取得しており，点群の欠落が少ない．
- 平面の計測精度は，AzureKinect が高く，D415 が低い．
- 小型部品の計測精度は，比較的 SR305，D415 が高い．
- **依然として見えにくい　①金属　②黒色　③透明　④小型**

2次元画像（RGB）と3次元画像（点群データ）

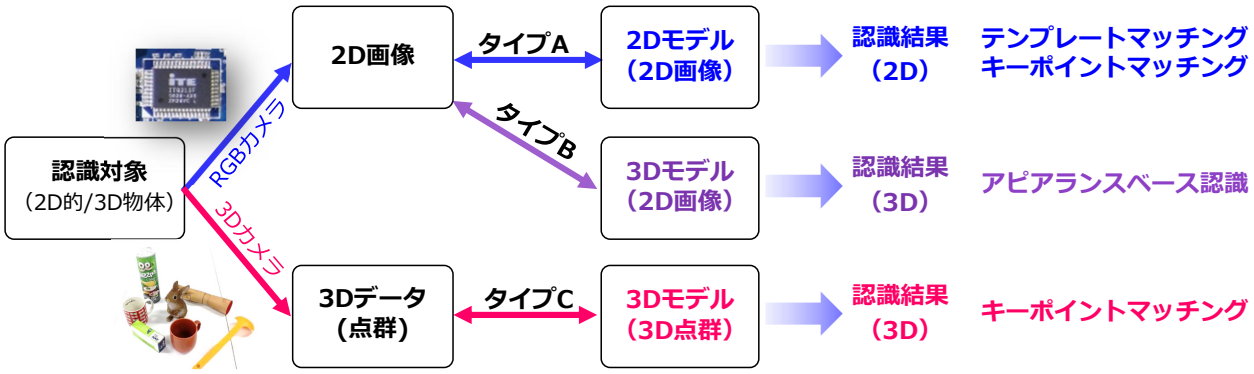
- 3次元センサは必ずしも万能ではない．

2D画像（RGB）



3D点群（Depth）

3次元対象物に対する3タイプの認識手法



	対象	使用センサ	利点	課題
A	2D的	2D	一般的なカメラの利用が可能．小型，軽量，高解像度，高速取得が期待できる．センサに起因する情報欠落などは少なく，過去のアルゴリズム資産が多い．	カメラ性能に起因して，ダイナミックレンジやパースペクティブの問題に注意が必要．ロボット適用に際しては，2Dから3Dへの情報変換や，追加処理が必要．
B	3D的	2D	タイプAと同様．	2Dから3Dへの情報復元が必須であり，曖昧さが生じやすい．対象物に関するモデルの持ち方と効率的検索が主課題．
C	3D的	3D	対象物の3D（2.5D）情報を直接的に利用できる．点群が取得できさえすれば，キーポイントマッチング等の手法が利用できる．	3Dセンサで点群化できることが前提であり，小型，黒色，金属，透明物体などは適用困難． アルゴリズムとしても，起伏の小さい対象物は苦手．

本日の話題

1．イントロダクション： 課題整理とロボットビジョンの概要

2．モノを視る‥‥2次元的アプローチ

戦略的画素選択に基づくテンプレートマッチングの高度化

（照明変動ロバスト化，識別性能向上，回転変動対応‥‥）

3．モノを視る‥‥3次元的アプローチ

ロボット動作生成にかかわる諸技術

（機能認識，動作教示簡素化，言語指示‥‥）

異なるセンサ間のギャップの低減

4．ヒトを観る

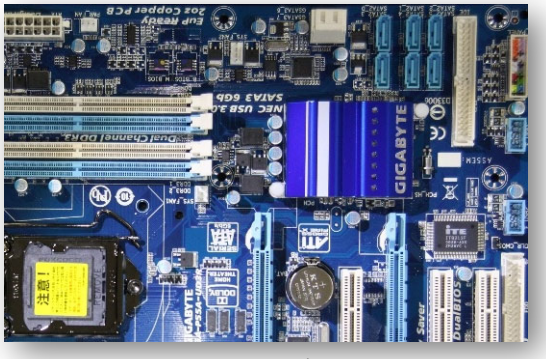
匠の技の分析

5．まとめ

２Ｄ的アプローチによる物体認識

- 実問題　　入力画像から目的の物体を探す
- 画像処理としての問題　　入力画像とテンプレート画像を照合する

テンプレート画像



入力画像

基本の３種類

① 画像の画素そのものを利用‥‥‥‥‥‥‥‥‥‥テンプレートマッチングなど

② 画像中のキーポイント特徴量を利用‥‥SIFT法など

③ 図形の輪郭や幾何学的モデルを利用‥‥Merlin法など

画素選択型テンプレートマッチング

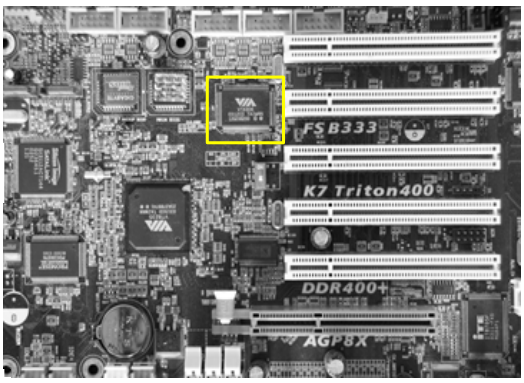
戦略に応じた少数の画素のみを厳選して使用

画素選択型テンプレートマッチングの概要

テンプレート画像
（LSIチップ）

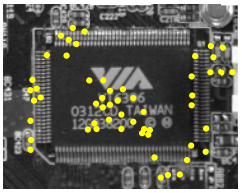
テンプレート画像中の
全ての画素を使用

（従来のマッチング）



入力画像（プリント基板）

戦略に基づいて
少数画素を選択



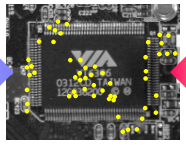
特殊な性質を持つ画素だけを使用
（画素選択型マッチング）

基本型：　独自性の高い画素を選択

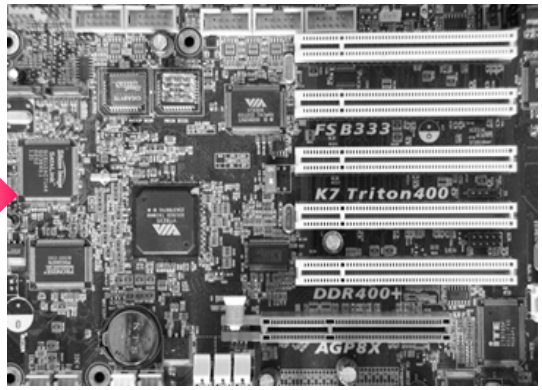
- テンプレート画像内の全ての点を使わない.
- 厳選された画素だけを使用する.

- テンプレート画像を**代表できる画素**を選択する.
  ＝　独自性の高い画素

画素選択



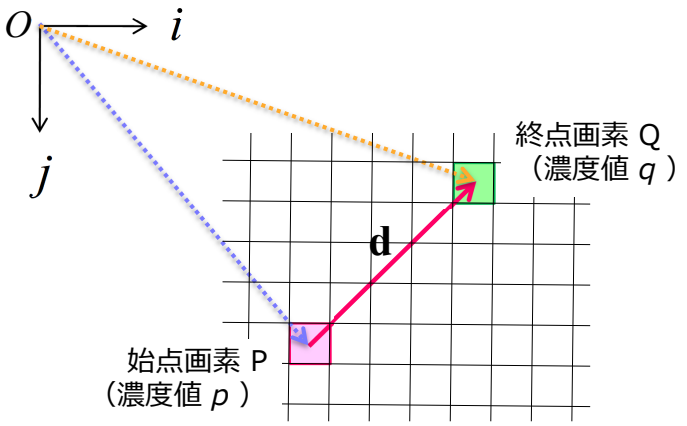
照合

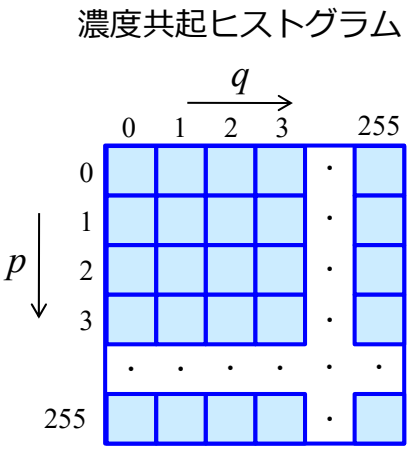


- **アイデア**
  濃度発生確率に基づき，画素の独自性を評価

濃度共起ヒストグラムによる発生確率の表現

- CPTM＝Co-occurrence Probability Template Matching
  - 始点と終点の２画素パターン（濃度ペア p, q ）を考える.
  - 濃度ペアの発生頻度を，濃度共起ヒストグラムとして表現する.
  - 共起ヒストグラムは，発生確率を表現している.



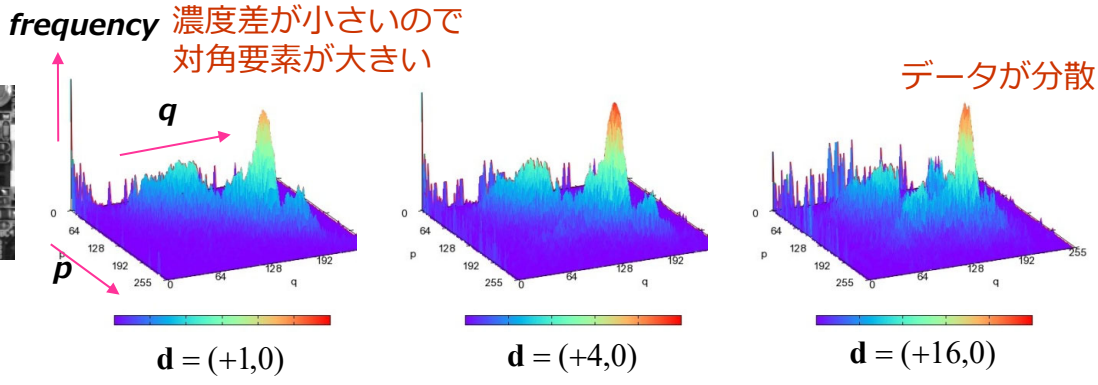


変位ベクトル d ：
共起性を測る方向と距離を指定するパラメータ

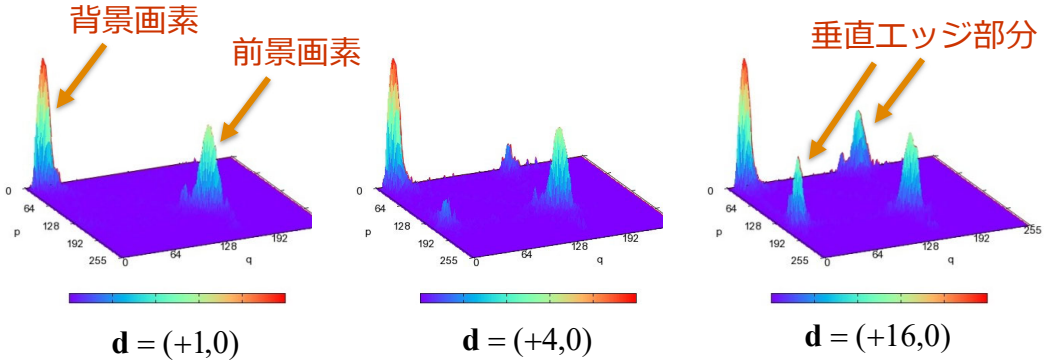
濃度共起ヒストグラムの例　　（発生確率分布）

- 水平2画素の共起ヒストグラムの場合

LSI

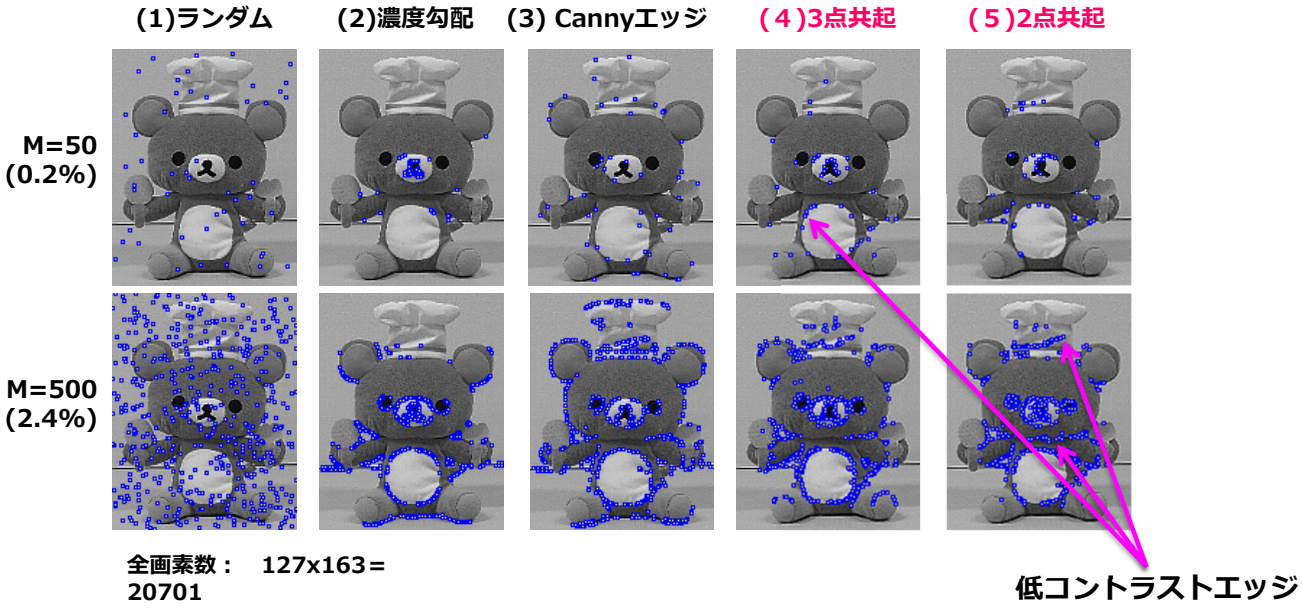


Parts



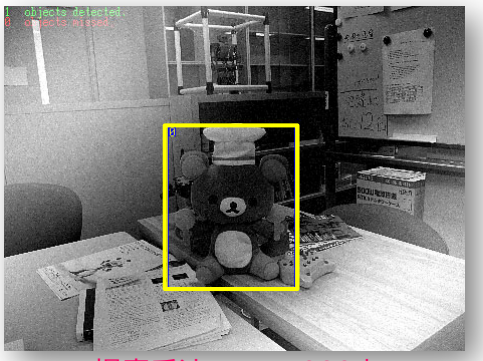
選択された画素の例

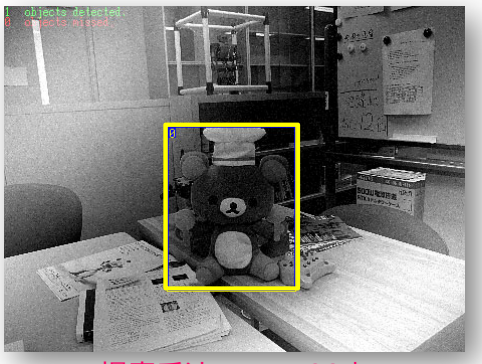
■ **濃度共起ヒストグラムにおいて，発生確率の低い画素から順に画素を選択する．**

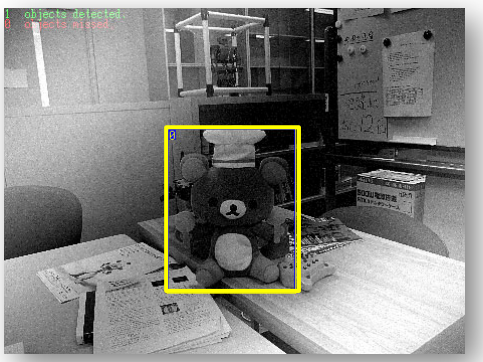


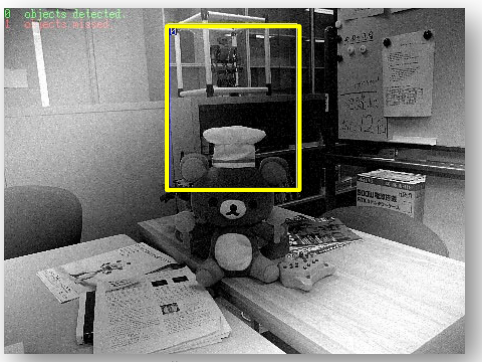
マッチング性能

マッチング結果の例 （画像： Noise=45%， Shading=80）


提案手法，M=1000点


提案手法，M=100点


提案手法，M=50点


ランダム法，M=50点

さまざまな戦略に基づく画素選択

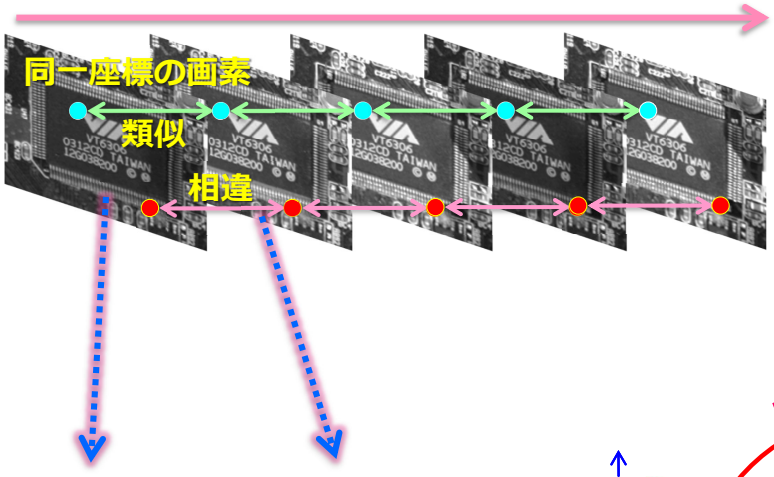
独自性の高い画素を選択（CPTM）

- 照明変動の影響を受けにくい画素を選択（SCPTM）
- 周辺対象物との識別に寄与する画素を選択（DPTM）
- 複数クラスへの分類性能が高い画素を選択（MCTM）
- １スキャンで多クラス分類できる画素を選択
- 部分的な隠蔽に頑健な画素を選択
- 回転粗精サーチに適した画素を選択（Manifold-CPTM）
- 任意の回転許容力を持つ画素を選択
- ユニークな色相を持つ画素を選択（CoP-TM）
- DNNの中間層分析に基づく画素を選択
- 中長期の画像変動に追従しうる画素を選択
- ・・・

照明変動の影響を受けにくい画素を選択
SCPTM

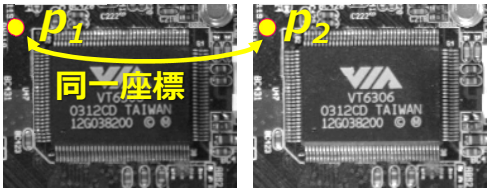
画素安定性推定のアイデア

一定の時間間隔で切り出された複数のテンプレート画像

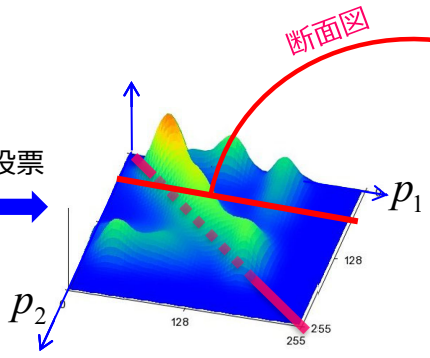


類似している -> 安定

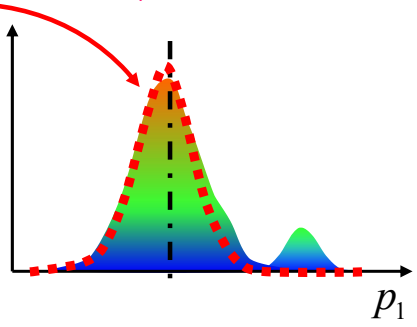
類似していない -> 不安定



投票

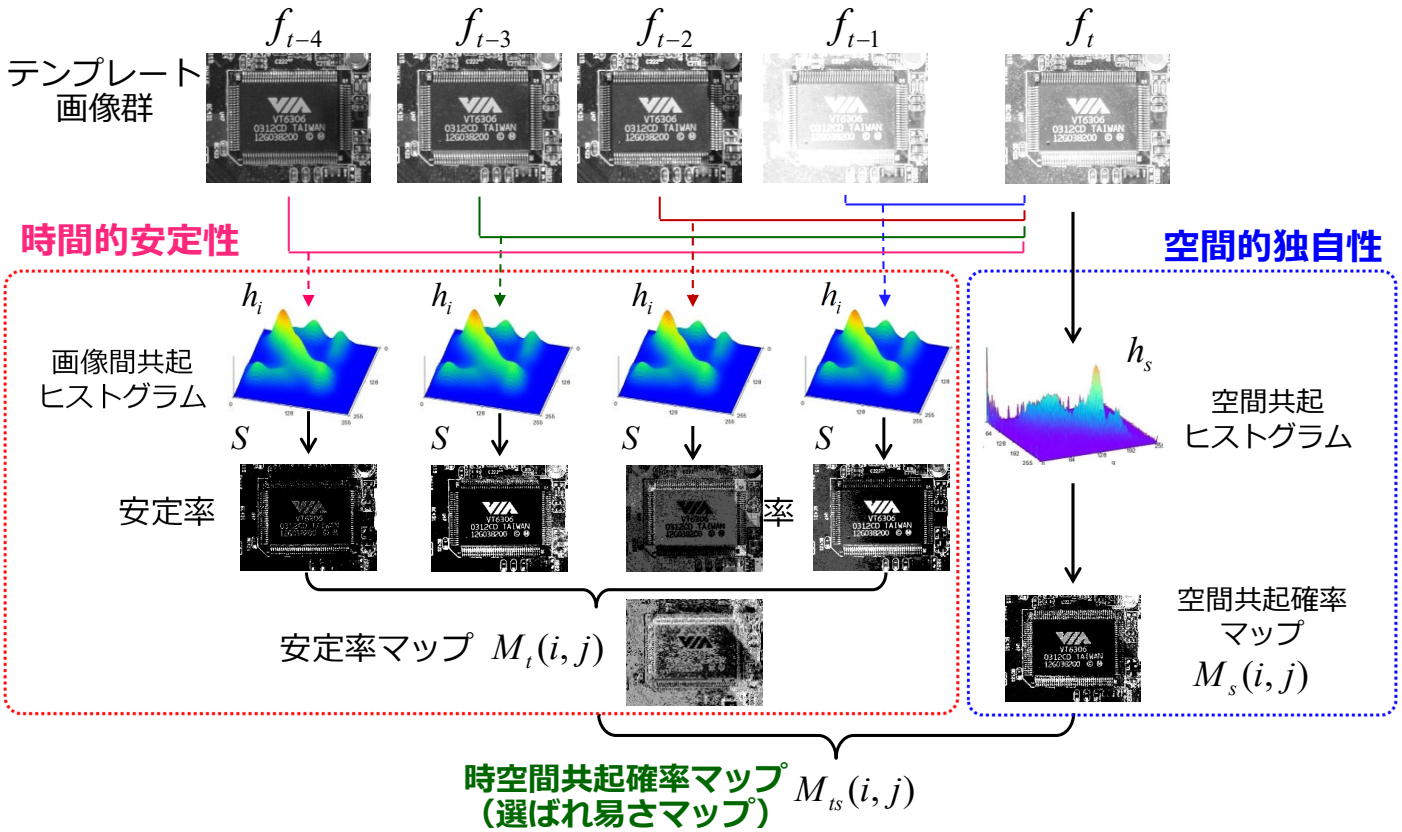


正規分布への近似性能をもとに，安定度を定量化



画像間共起ヒストグラム

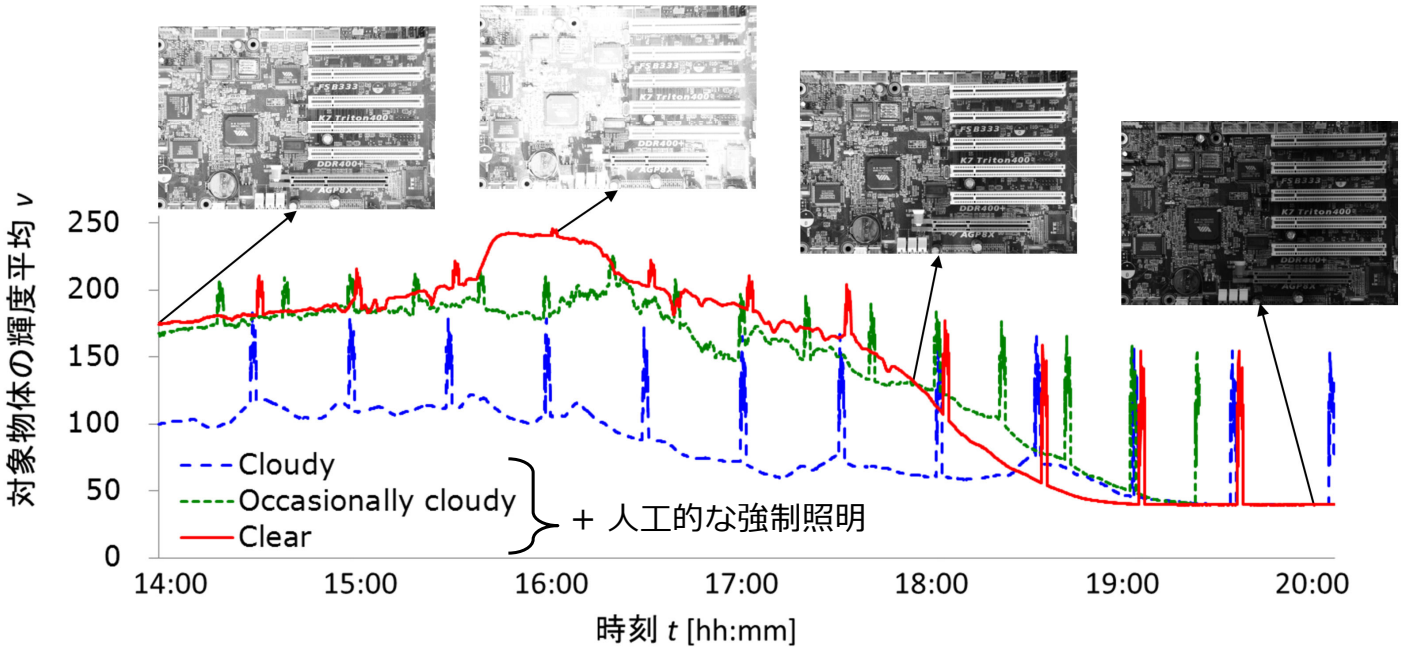
安定かつ独自性の高い画素の選択



空間的に独自性が高くかつ時間的に安定した参照画素を選択

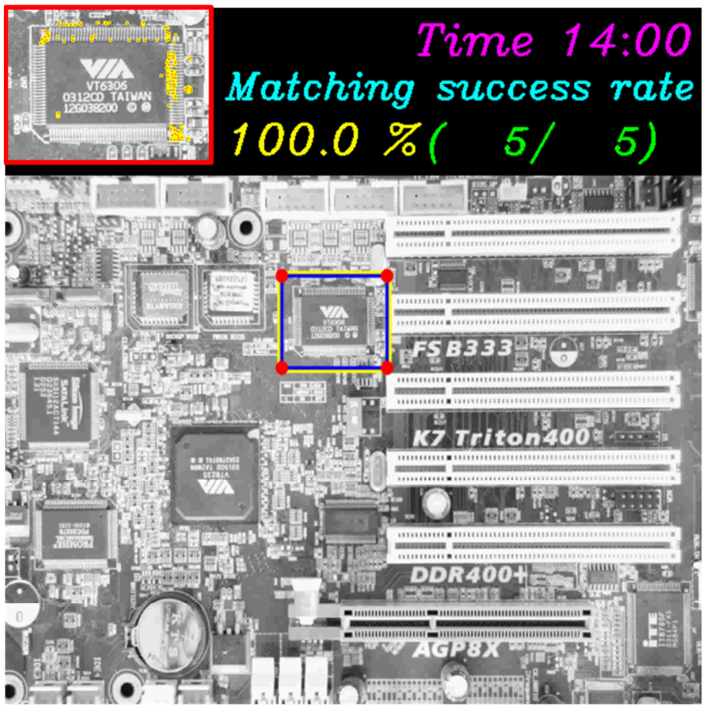
検証実験に用いた連続写真（プリント基板）

3日間×6時間＝計18時間にわたって，3秒おきに画像を自動取得（計21600枚）

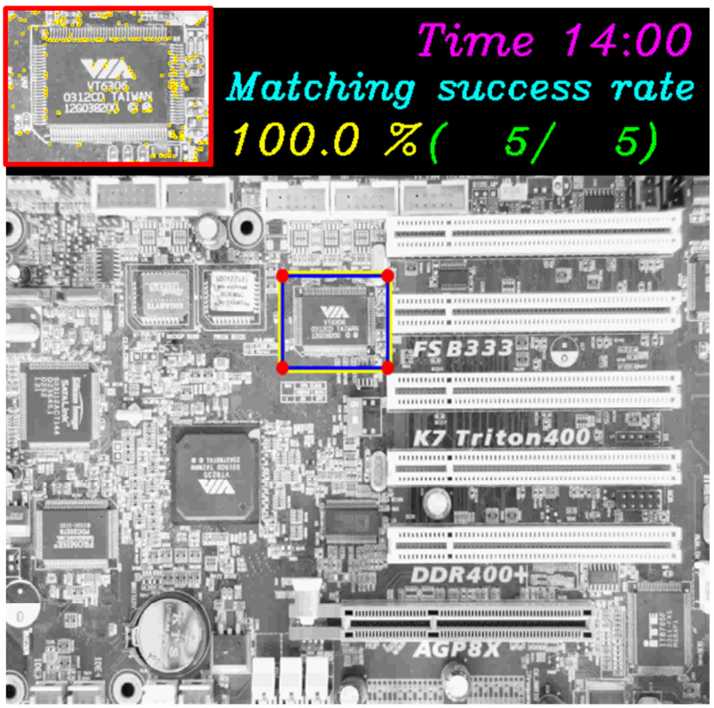


実験結果

従来法（CPTM）



提案手法

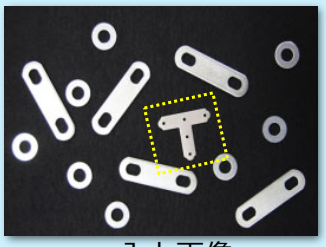


周辺対象物との識別に寄与する画素を選択
DPTM

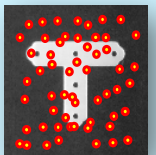
基本アイデア

対象物を（周辺物に対して）特徴づけている画素群を選択

➡ 識別性能向上



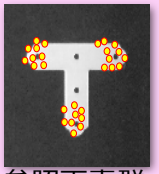
入力画像



参照画素群

周辺物が対象と似ていない場合

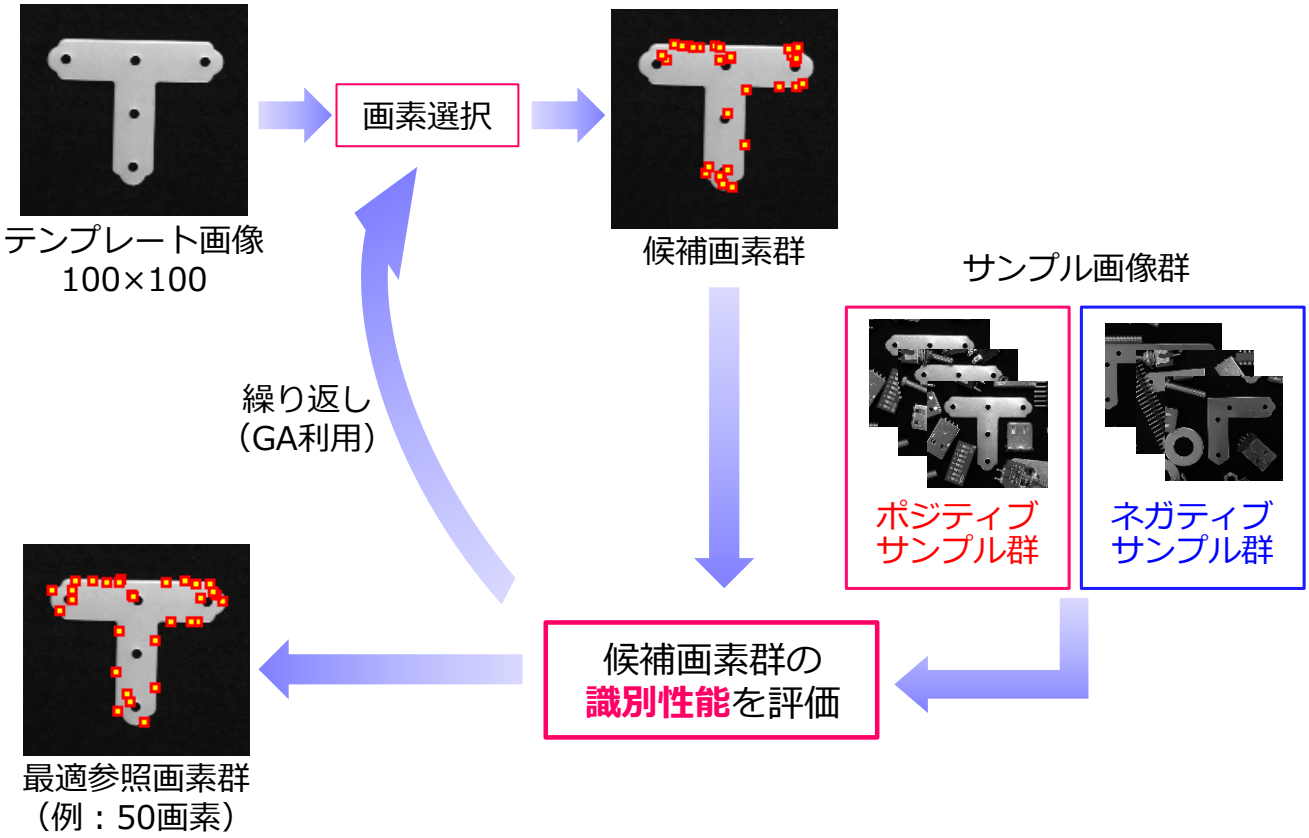
入力画像



参照画素群

周辺物が対象と似ている場合

識別に有効な画素選択アルゴリズム

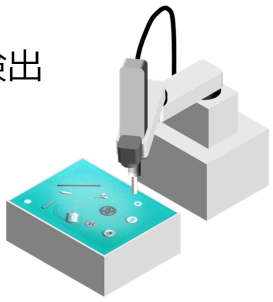


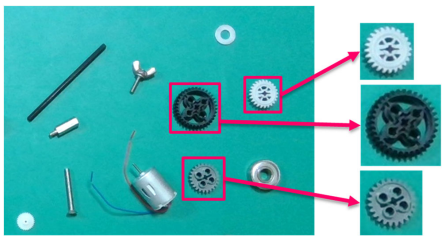
実験結果

１スキャンで多クラス分類できる画素を選択

多クラス物体検出における課題と本研究の目的

- **多クラス物体検出に求められるもの（セル生産ロボを想定）**
  - 多クラス識別
  - 各対象物の正確な位置検出





- **テンプレートマッチング　v.s. 機械学習**

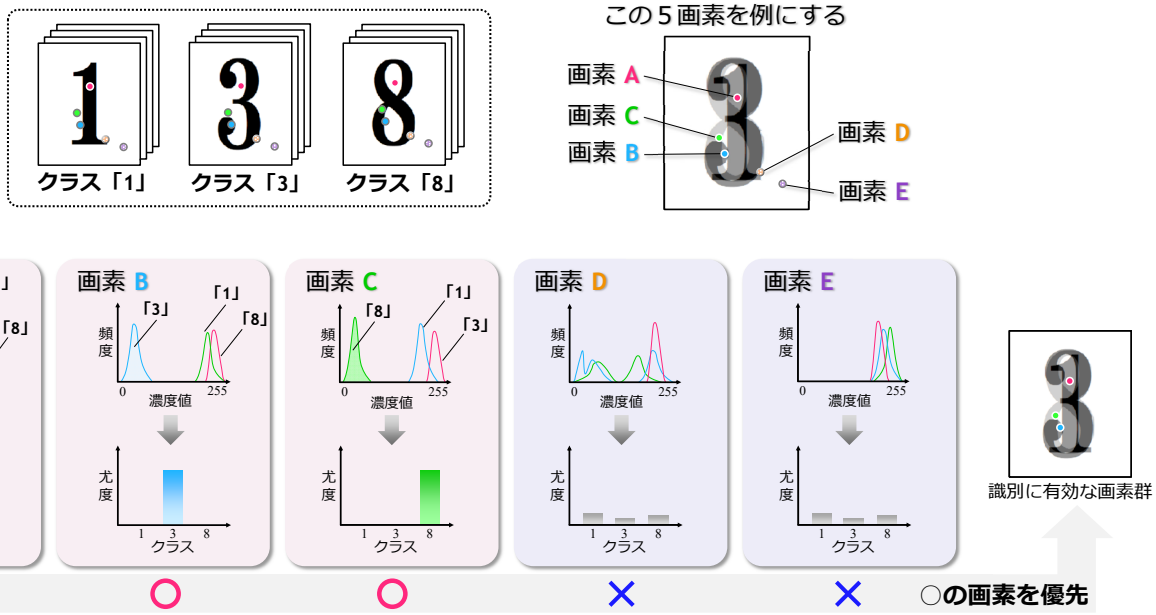
	位置検出性能	識別性能	処理時間
テンプレートマッチング	○	×	×
機械学習（深層学習）	×	○	△

- **本研究の目的**

高い位置検出性能と識別性能の両立＋高速化を目指した手法の開発

基本アイデア

テンプレート画像中のひとつひとつの画素がもつクラス識別能力に着目



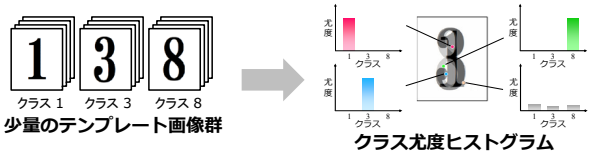
クラス識別に有効な複数の画素を選択することにより，1 回の探索で複数の対象物を検出

提案手法の概要

事前処理（オフライン）

1. 各画素の識別能力を指標化

画素の濃度値ヒストグラム分布に着目



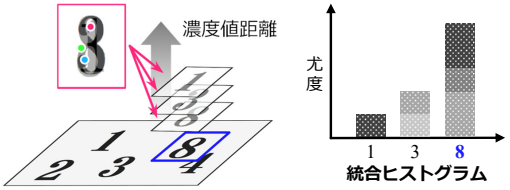
2. 多クラス分類できる画素群を選択

- 指標を利用して，各クラスの識別能力が均一になるように画素を選択
- 画素の組み合わせ最適化（GAで求解）

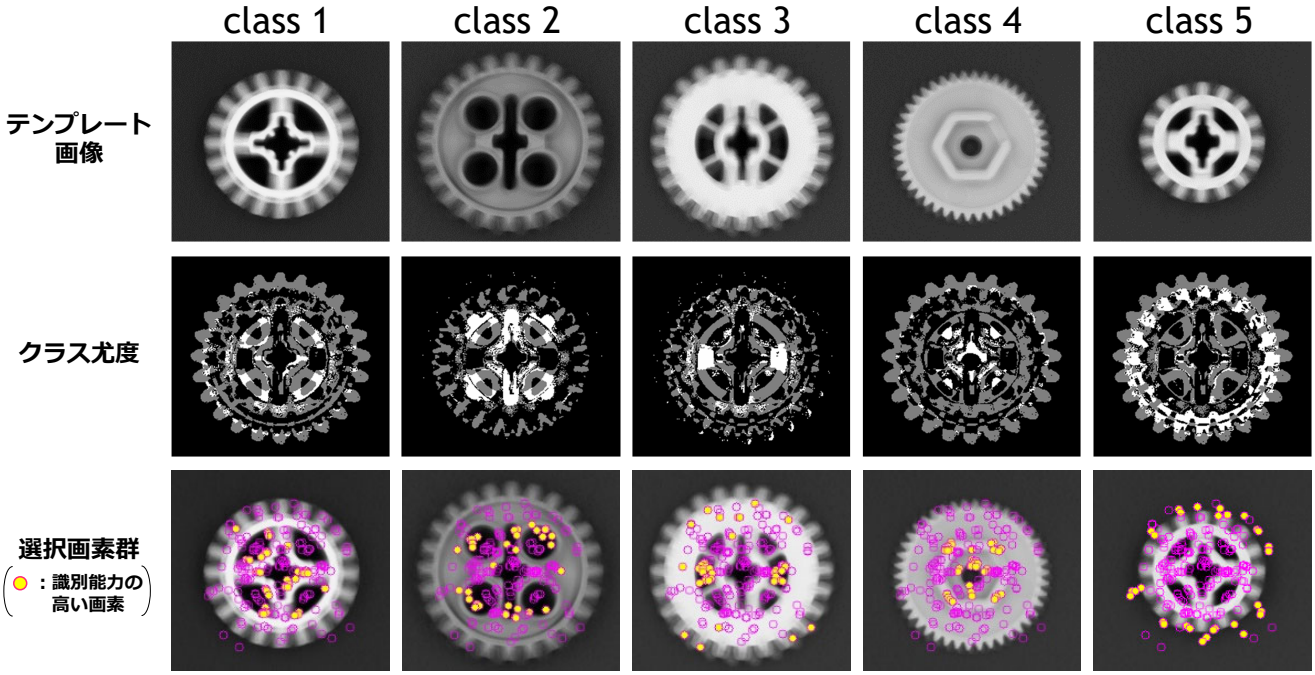
マッチング（オンライン）

3. ヒストグラム投票型マッチング

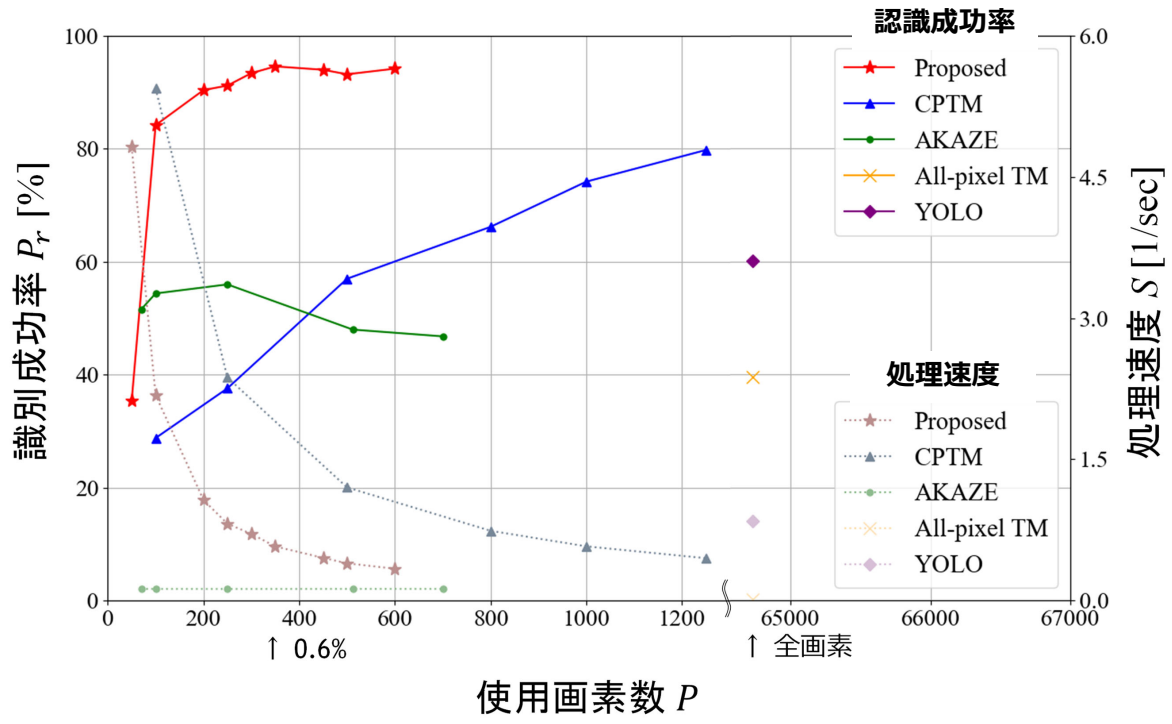
テンプレート画素と入力画素の濃度値距離をもとに，一つ一つの画素によるエビデンスを，クラス尤度ヒストグラムに投票



テンプレート画像と識別能力・選択された画素群



評価実験 ー識別性能と処理速度ー



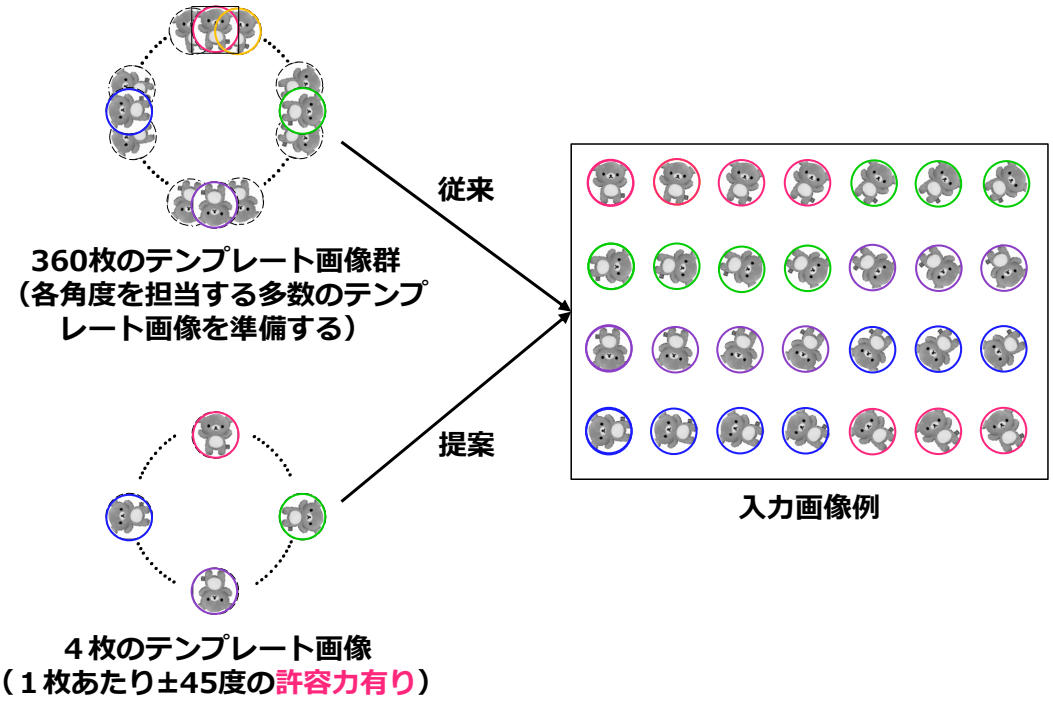
- 識別成功率： 他手法と比較して，提案手法は少画素数で高性能
  例：テンプレート全画素の **0.6%** を使用したとき，**95.4%**
- 処理時間： 識別成功率 80% のとき，提案手法が最も高速

任意の回転許容力を持つ画素を選択

本研究の目的

- 回転対応テンプレートマッチング開発の目的（想定用途）
  - 目的１：多少の回転ずれがあっても，それを許容して検出する.
  - 目的２：どんなに回転していても，検出する.
- これら２つの目的を達成できる方式を提案する

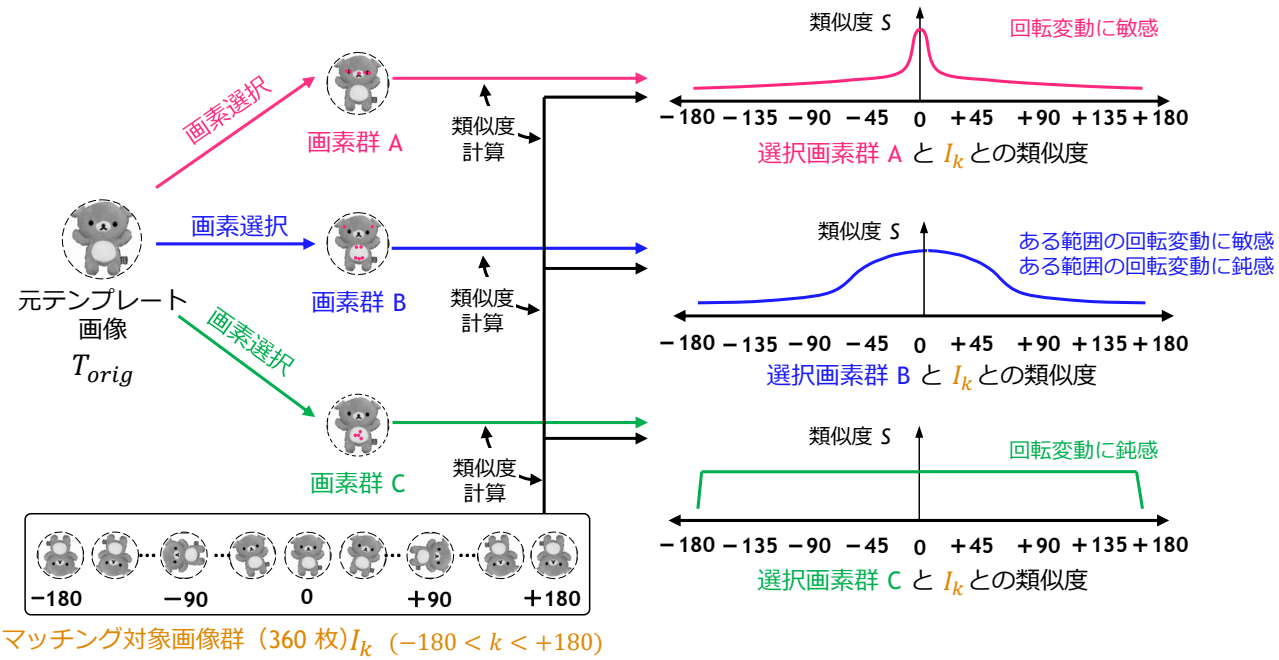
個々のテンプレートが持つ回転許容力とその利用法



もし，テンプレート１枚あたり ±45 度の回転許容力を持つならば，
４枚のテンプレートのみを使うだけで，回転に依らないパターン検出が可能

画素選択による回転許容力のコントロール

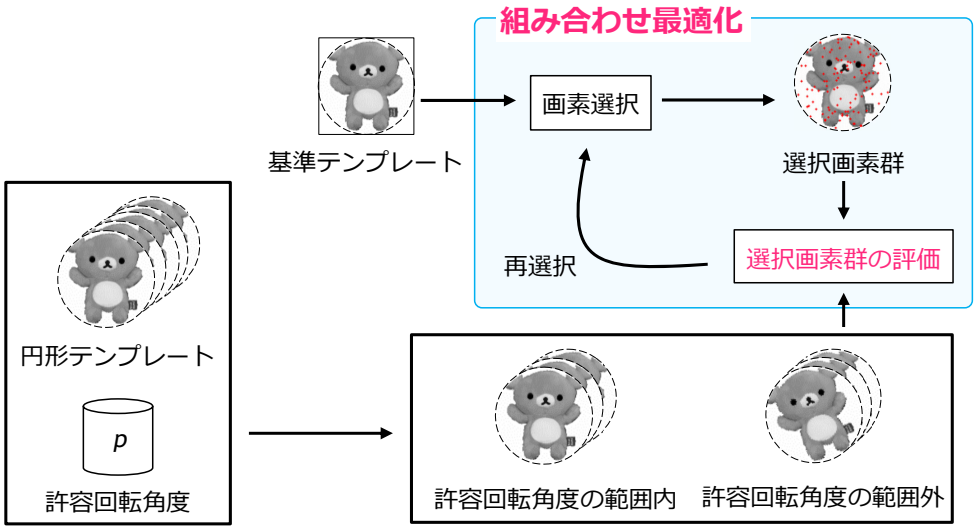
- **マッチングに使用される画素群によって回転許容力（どのくらい回転していても検出できるか）が異なる.**



すなわち，適切な画素を選択すれば，任意の回転許容力を持たせることができる.

最適な画素の探索方法

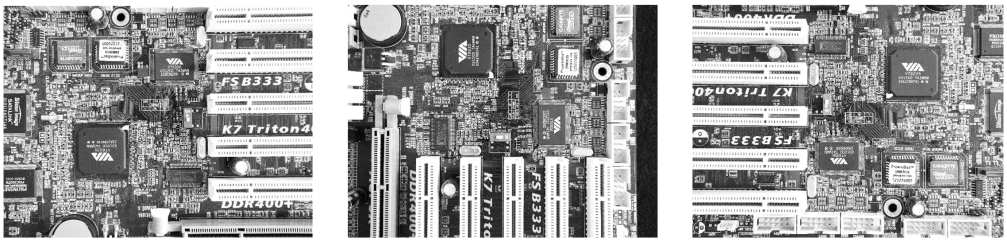
- 最適な選択画素群の探索を，組み合わせ最適化問題として扱う
  - 許容回転角度の基準となるテンプレートから選択された画素群の回転変動許容力を評価し，最適な選択画素群を探索



実験条件

- 実験画像
  - 対象物を回転させた 360 枚の入力画像
  - マッチング位置が正解位置から ±5pixel 以内の場合は検出成功

テンプレート画像
（128x128 pix）

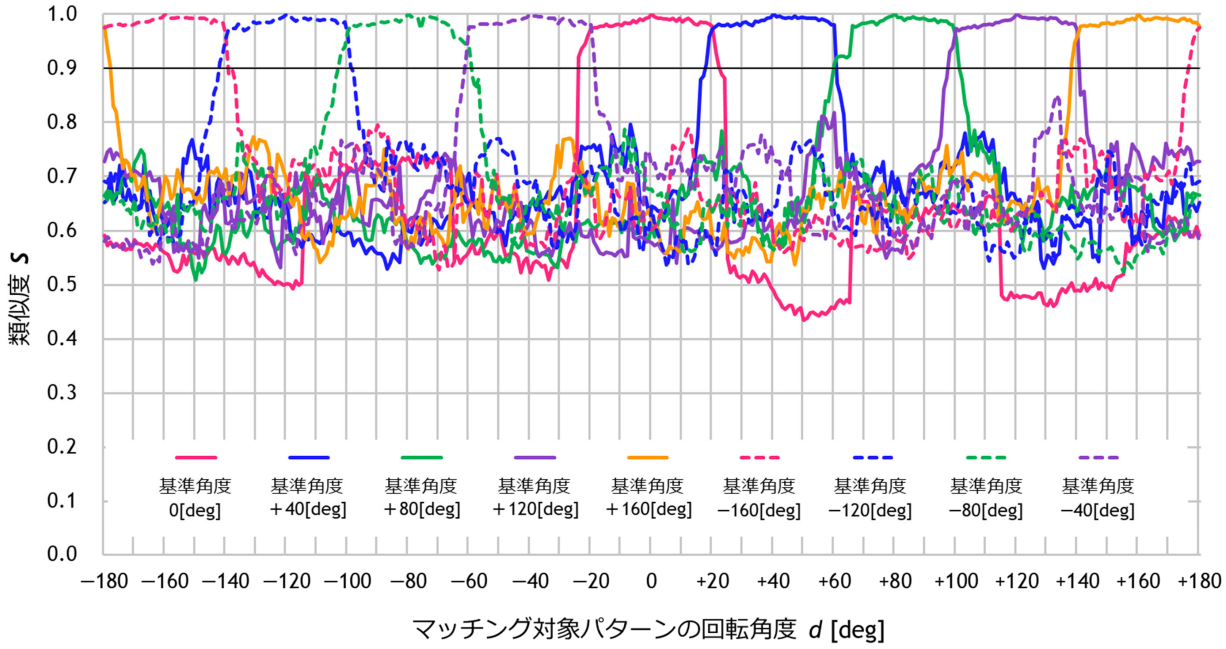


入力画像の例
（640x480 pix）

- 最適化（GA）のパラメータ
  - 世代数：500,000，個体数：5,000
  - 選択画素数の割合：5%
  - 許容回転角度：±20 度

類似度グラフ（許容範囲：±20 度）

選択画素数：5%



実験結果

検出率と処理時間

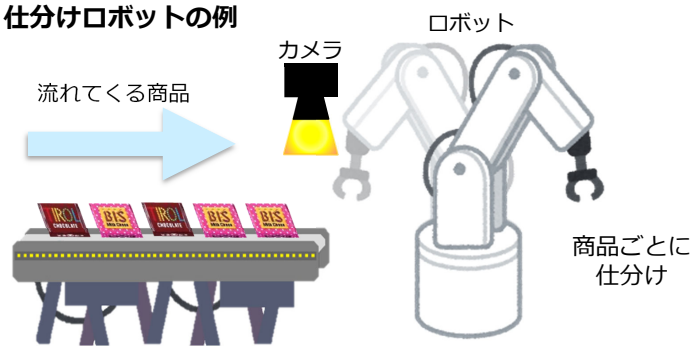
	画素数 [pixel]	テンプレート枚数 [枚]	検出率 [%]	処理時間 [msec]
All pixels	16,384	360	100.0	621,180
CPTM	18	360	0.01	330
AKAZE	–	1	6.1	339
Random	819	9	12.7	342
Proposed	819	9	**86.7**	**342**

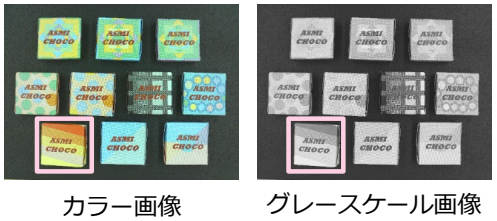
- All pixels：全画素選択し，360 枚のすべてのテンプレートを使用
- CPTM：ごく少数の CPTM の画素を選択し，360 枚すべてのテンプレートを使用
- AKAZE： 1 枚のテンプレートを使用し，AKAZE を使用
- Random：ランダムな画素を選択し，少数のテンプレートを使用
- Proposed：回転変動許容力を持つ最適な少数画素を選択し，少数のテンプレートを使用

ユニークな色相を持つ画素を選択
CoP-TM

背景と目的

- 背景
  - 高速処理の目的で，グレースケール画像が利用されることが多い．
  - 商品パッケージなど，色情報を使ったほうが高効率な場合がある．





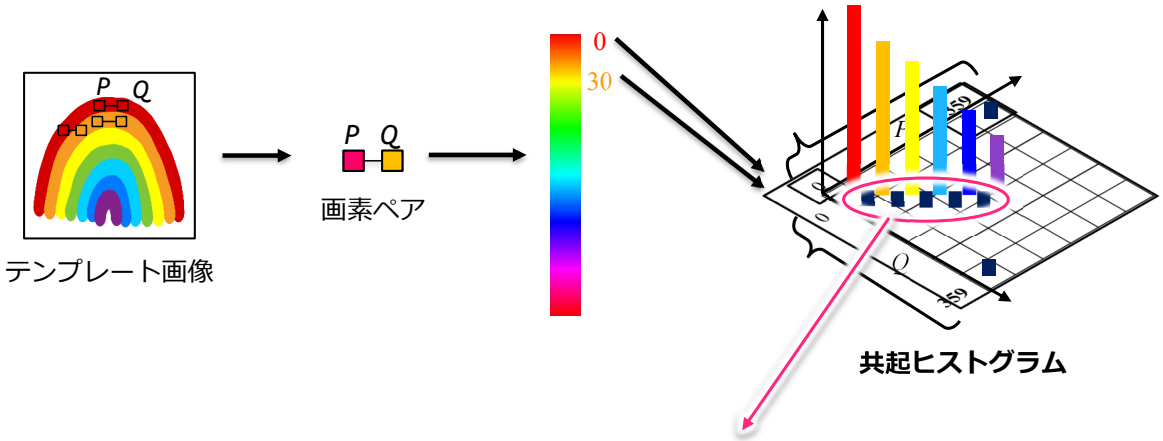
カラー画像　　グレースケール画像

どちらの画像の方が対象物を探しやすい？

- 研究の目的
  - 色情報の独自性に基づくTMの提案

基本アイデア：発生頻度をもとにして独自性を評価

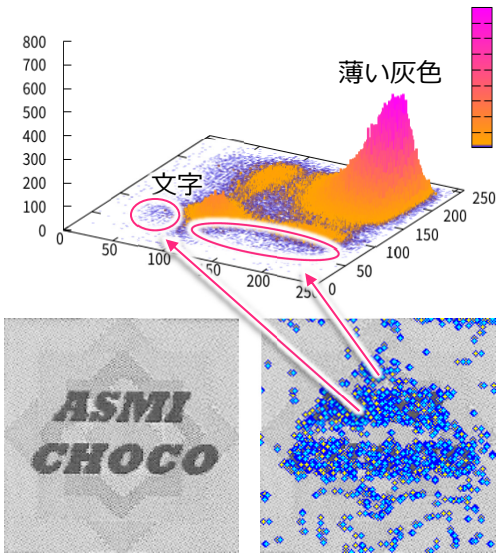
- 画素の色情報をもとに共起ヒストグラムを作成
  発生頻度　→　発生確率　→　独自性



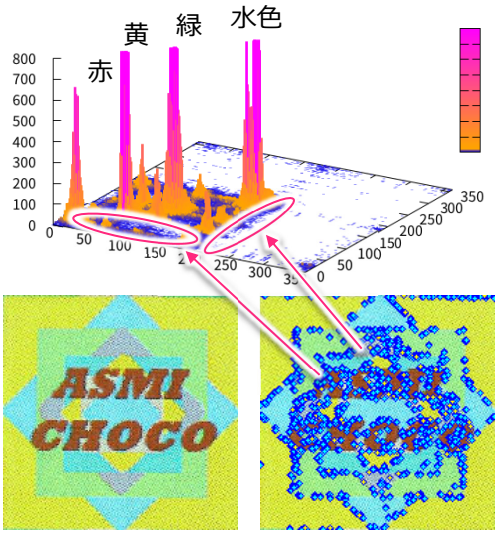
色情報からみて独自性の高い画素として選択する.

COlor co-occurrence Probability Template Matching（CoP-TM 法）

選択された画素群



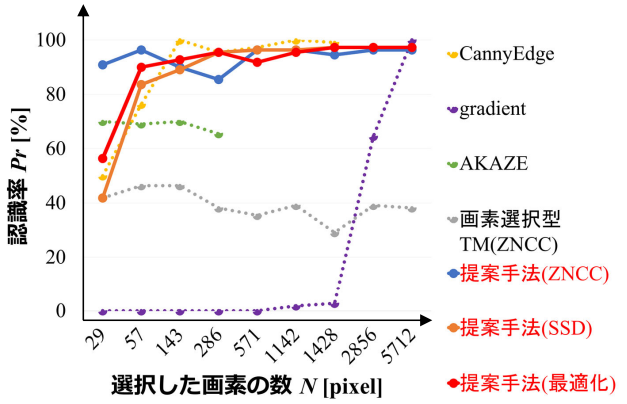
テンプレート画像　選択画素群
従来手法

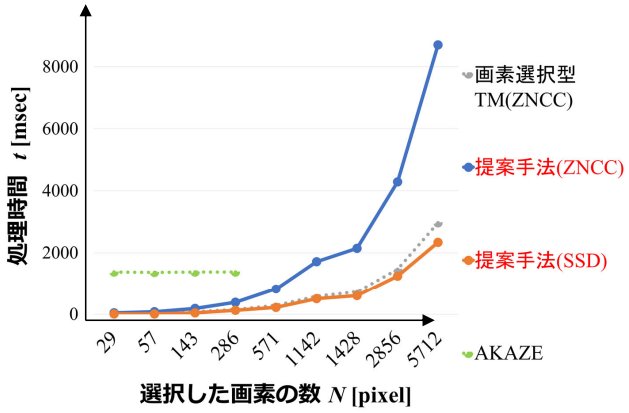


テンプレート画像　選択画素群
提案手法

実験

- テンプレート画像の全画素数：28561 [pixel]
- 用いた入力画像：110 枚
- 用いた CPU：i7-8700





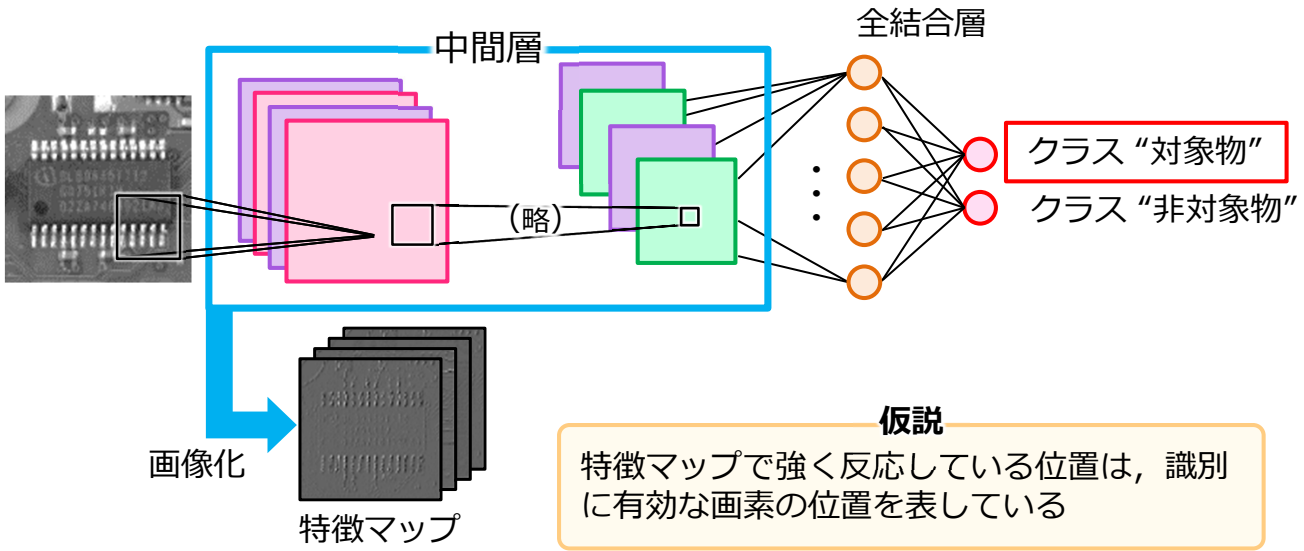
実験結果から言えること

- 提案手法は，他の画素選択やKPM（AKAZE）より認識成功率が高い
- 類似度尺度をSSDにすれば，ZNCCより高速

DNNの中間層分析に基づく画素を選択

DNN の中間層を利用したテンプレート画素の決定

- TMをクラス識別問題と捉えて学習させる.
- 学習後のネットワークを分析し，重要な画素を推定.
- それを画素選択型テンプレートマッチングとして利用



仮説

特徴マップで強く反応している位置は，識別に有効な画素の位置を表している

この仮説に基づいて，使用する画素を決定する.

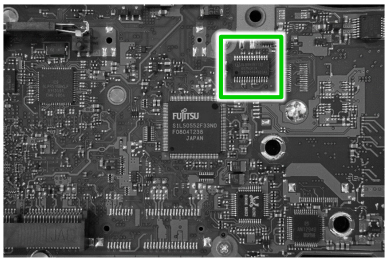
実験結果

各手法の認識性能

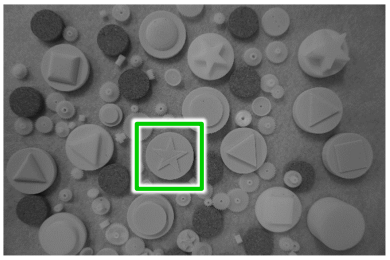
手法	参照画素数	認識成功率 [%]		処理時間 [msec]
		IC チップ	プラスチック部品	
全画素法	10000	96.6	96.4	6816
ピラミッド探索*	10000	96.6	96.4	276
ランダム	50	77.3	53.2	47
CPTM [Hashimoto 2011]	50	78.7	62.8	47
提案手法	**50**	**95.4**	**100.0**	**47**

*探索のしきい値：0.7

画素選択手法

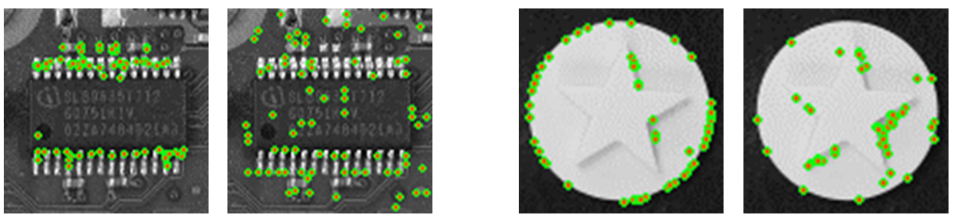


IC チップ



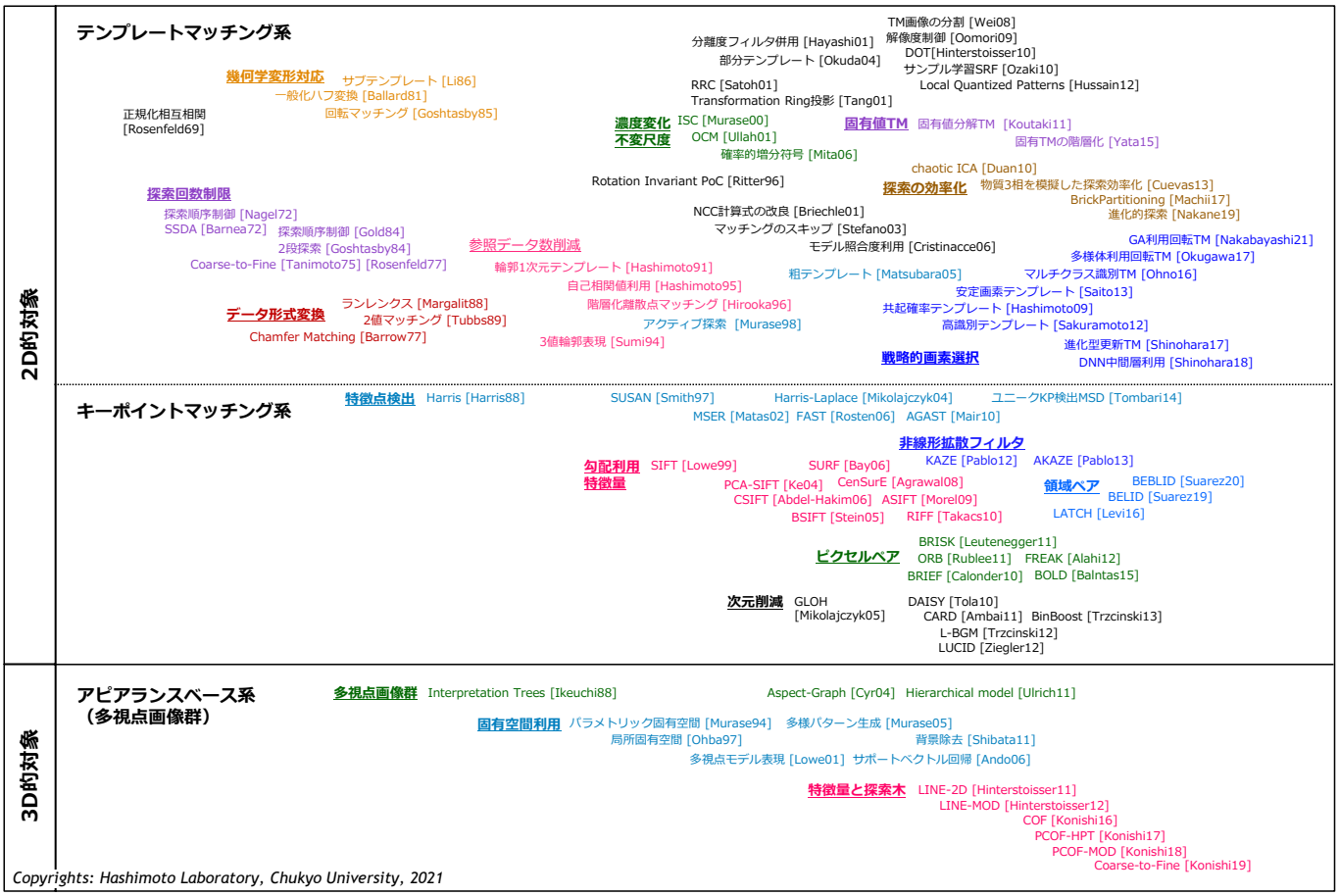
プラスチック部品

自動決定した参照画素



CPTM　提案手法　CPTM　提案手法

２Ｄ的アプローチの変遷（タイプA，タイプB）



Copyrights: Hashimoto Laboratory, Chukyo University, 2021

本日の話題

１．イントロダクション：　課題整理とロボットビジョンの概要

２．モノを視る‥‥２次元的アプローチ

戦略的画素選択に基づくテンプレートマッチングの高度化

（照明変動ロバスト化，識別性能向上，回転変動対応‥‥）

３．モノを視る‥‥３次元的アプローチ

ロボット動作生成にかかわる諸技術

（機能認識，動作教示簡素化，言語指示‥‥）

異なるセンサ間のギャップの低減

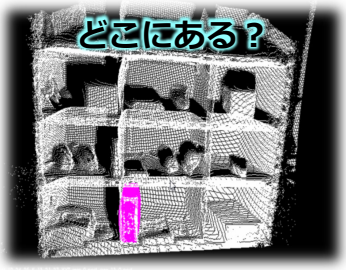
４．ヒトを観る

匠の技の分析

５．まとめ

ロボットビジョンの共通課題（再掲）

Pick → Operation

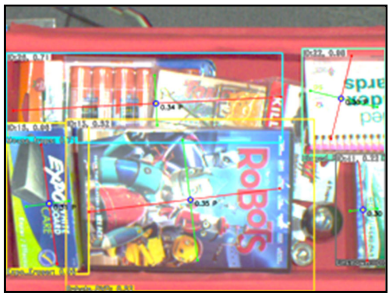


（本日はSKIP）　（本日はSKIP）　**本日の話題**

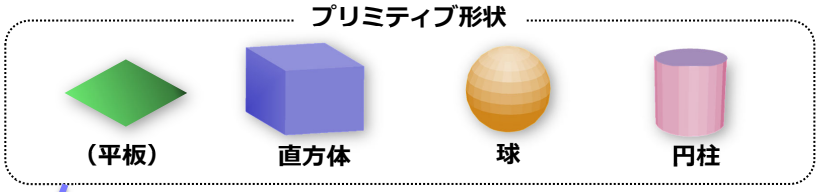
Intelligent Sensing Laboratory
Hashimoto Lab., Chukyo University

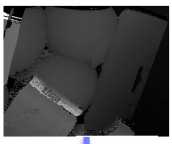
プリミティブ形状近似による
モデルレス把持位置決定

アイデア

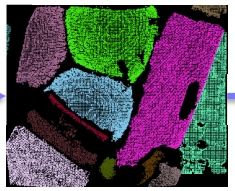


乱雑に積まれた物品





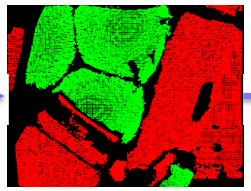
入力データ



領域セグメンテーション

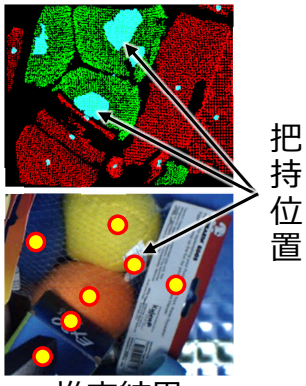
プリミティブ形状のあてはめ

あてはめ結果



緑：球　赤：平面

把持位置推定



把持位置

推定結果

プリミティブ形状近似に基づく把持位置推定

近似・把持位置推定モジュール

入力データ

点群のセグメンテーション

セグメントの抽出

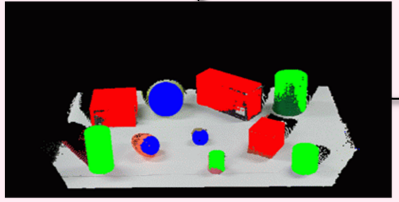
セグメントの統合

点群を分割した入力データ

面プリミティブの尤度推定

尤度

尤度統合による立体プリミティブ近似



立体プリミティブで近似した結果

学習モジュール

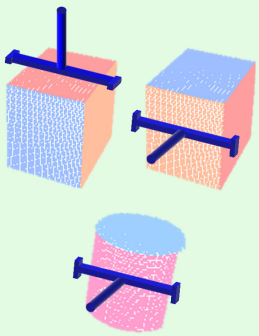
学習データ群

点群のセグメンテーション

点群を分割した学習データ

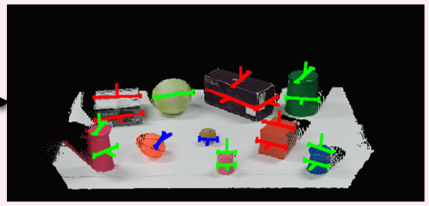
3D-DNNの学習

学習済みネットワーク



把持位置推定ルール

把持位置推定



把持位置推定の結果

評価実験結果

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

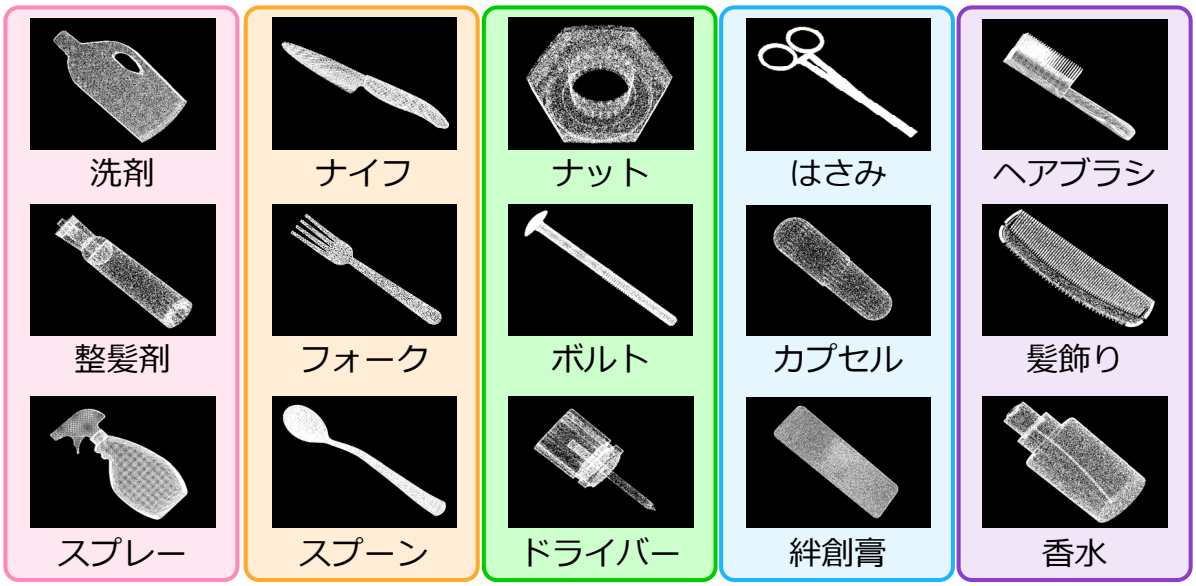
立体プリミティブの近似成功率 [%]

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

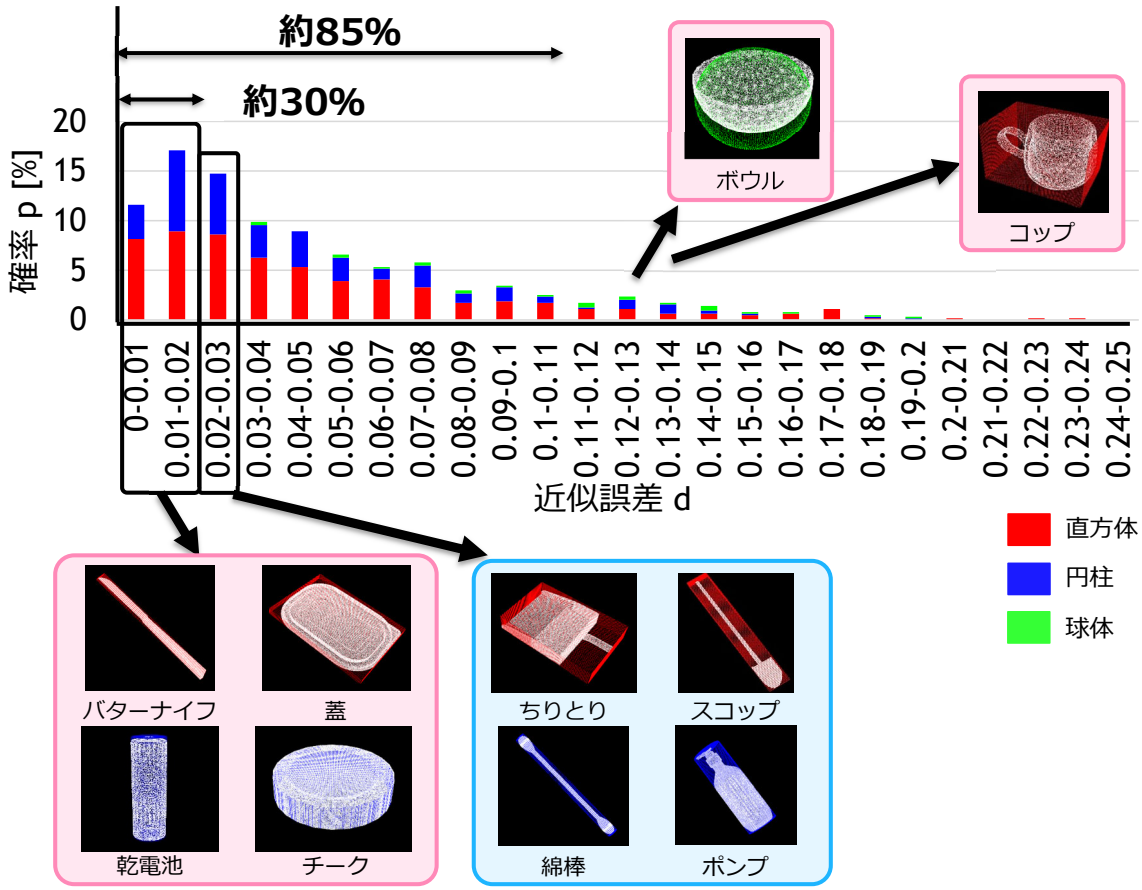
＊：3D ShapeNets 法

3種のプリミティブはどのくらい有効か？

- JICFS/IFDB （商品情報データベース）が定義する5カテゴリ
- ShapeNet, ObjectNet, ModelNet, 3Dwarehouse, Thingiverse, および独自に作成したモデル 638 個を用いた.



実験結果（速報）



Intelligent Sensing Laboratory
Hashimoto Lab., Chukyo University

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キング

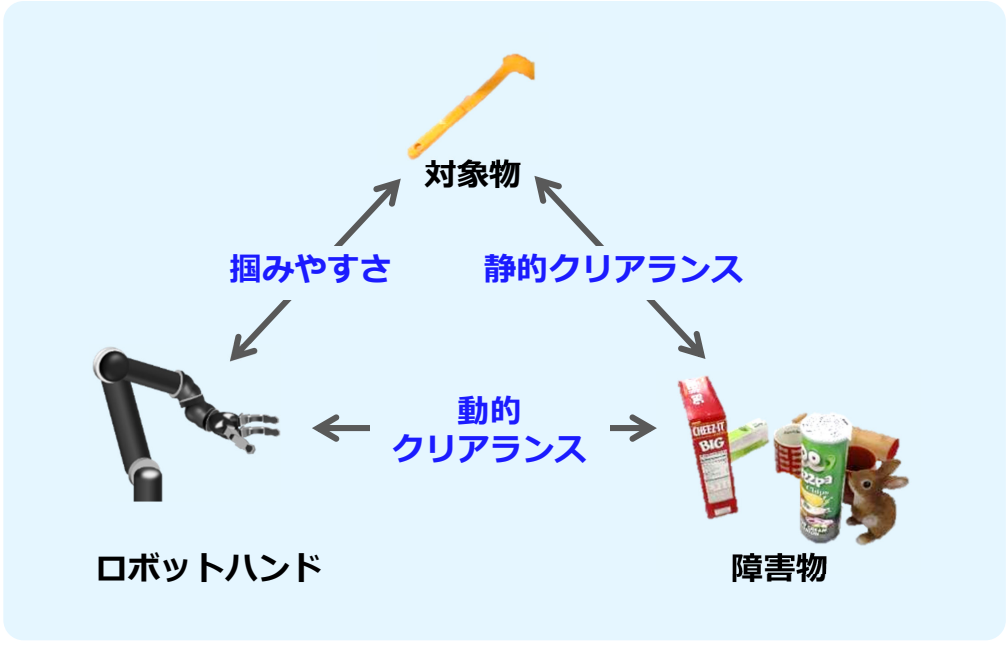
成功例

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

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

- $P_S(\mathbf{p})$：把持余裕度
- $P_{H\text{-}O}(\mathbf{p})$：①ハンドと障害物の動的クリアランス
- $P_{H\text{-}T}(\mathbf{p})$：②対象物の掴みやすさ
- $P_{T\text{-}O}(\mathbf{p})$：③対象物と障害物の静的クリアランス

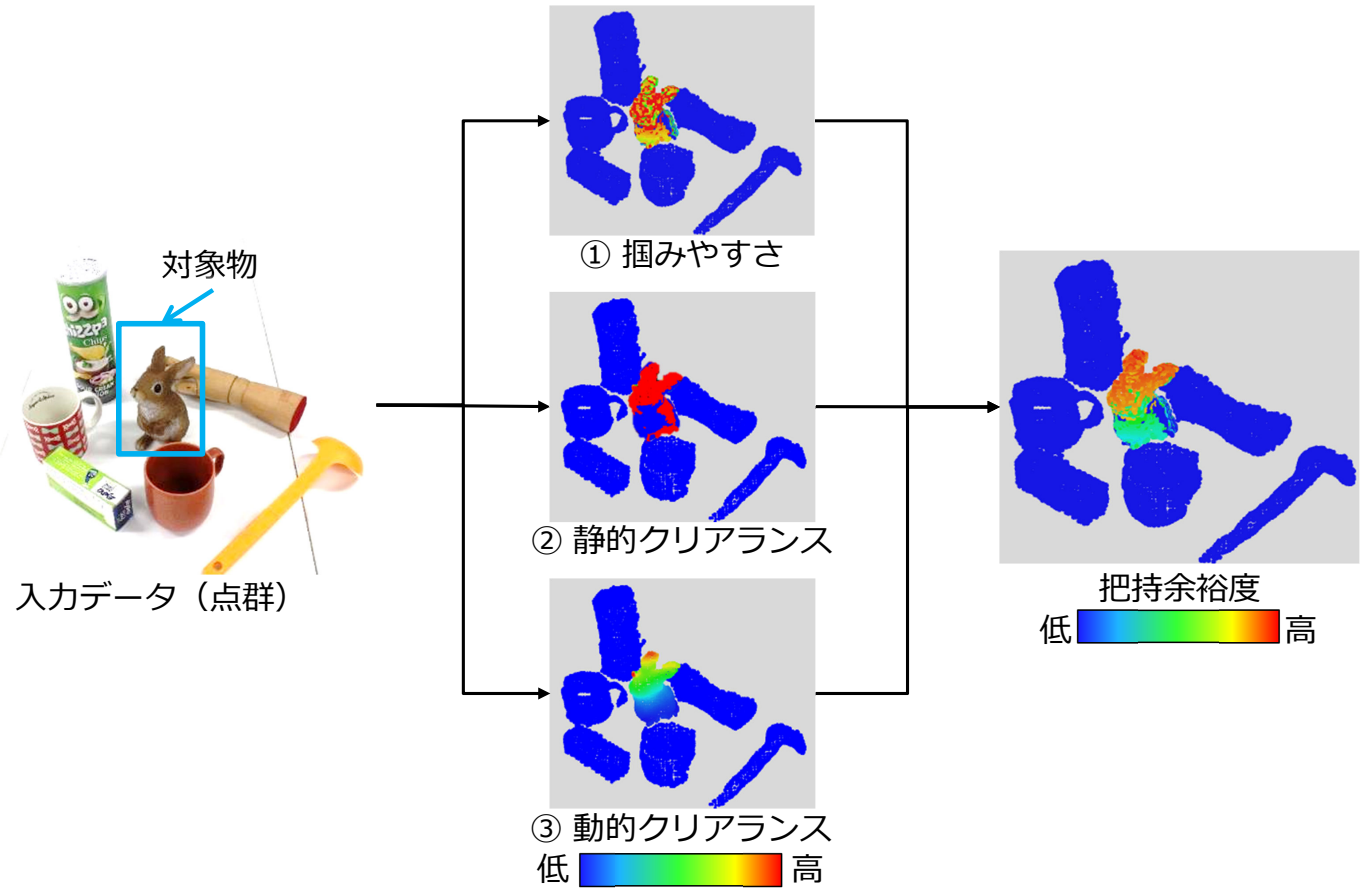
※ $\mathbf{p}$：把持パラメータ（把持位置，姿勢，開口幅）

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

$$\tilde{\mathbf{p}} = \underset{\mathbf{p}\in\mathbb{S}}{\operatorname{argmax}}\, P_S(\mathbf{p})$$

$\mathbb{S}$：$\mathbf{p}$ の候補集合

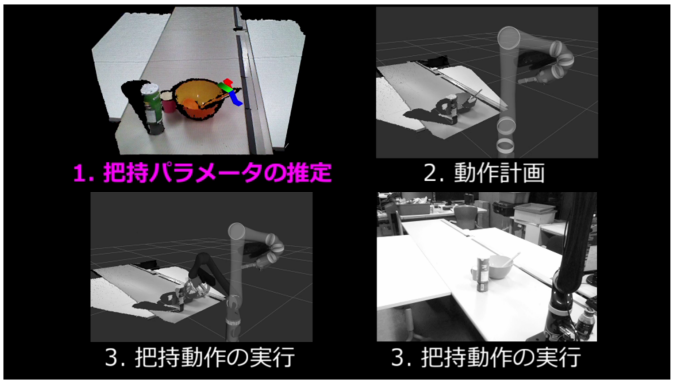
算出された余裕度指標



把持パラメータの推定実験の結果

15 シーン中の 47 個の対象物について把持パラメータを推定し，その非干渉率，把持成功率，処理時間を比較した．

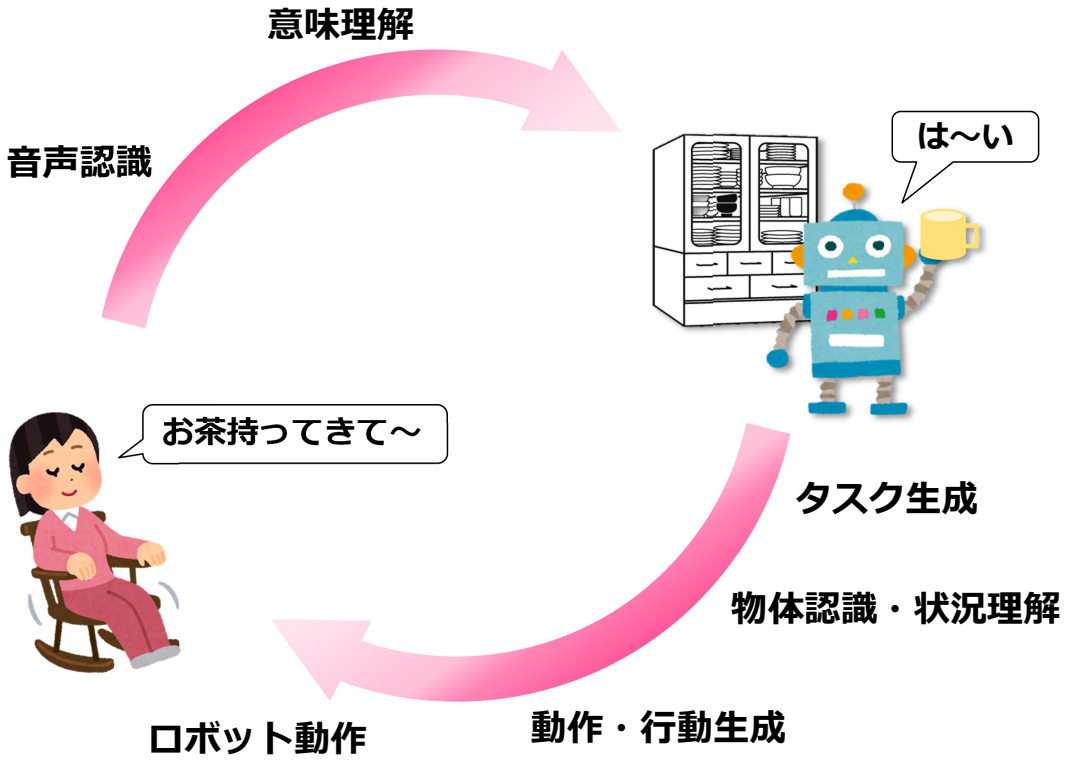
手法	非干渉率[%]	把持成功率[%]	処理時間[s]
Agile_Grasp [Pas 2015]	55	48	5.46
Grasp-Detector [Pint 2017]	64	45	15.0
GPD [Gualtieri 2017]	73	64	3.20
提案手法	**91**	**82**	**0.42**



ハンド右指　ハンド左指　把持姿勢

モノの “機能” 認識とその応用

お茶を持ってきてくれるロボットに必要な主要技術



お茶をたてる動作（人間の場合）

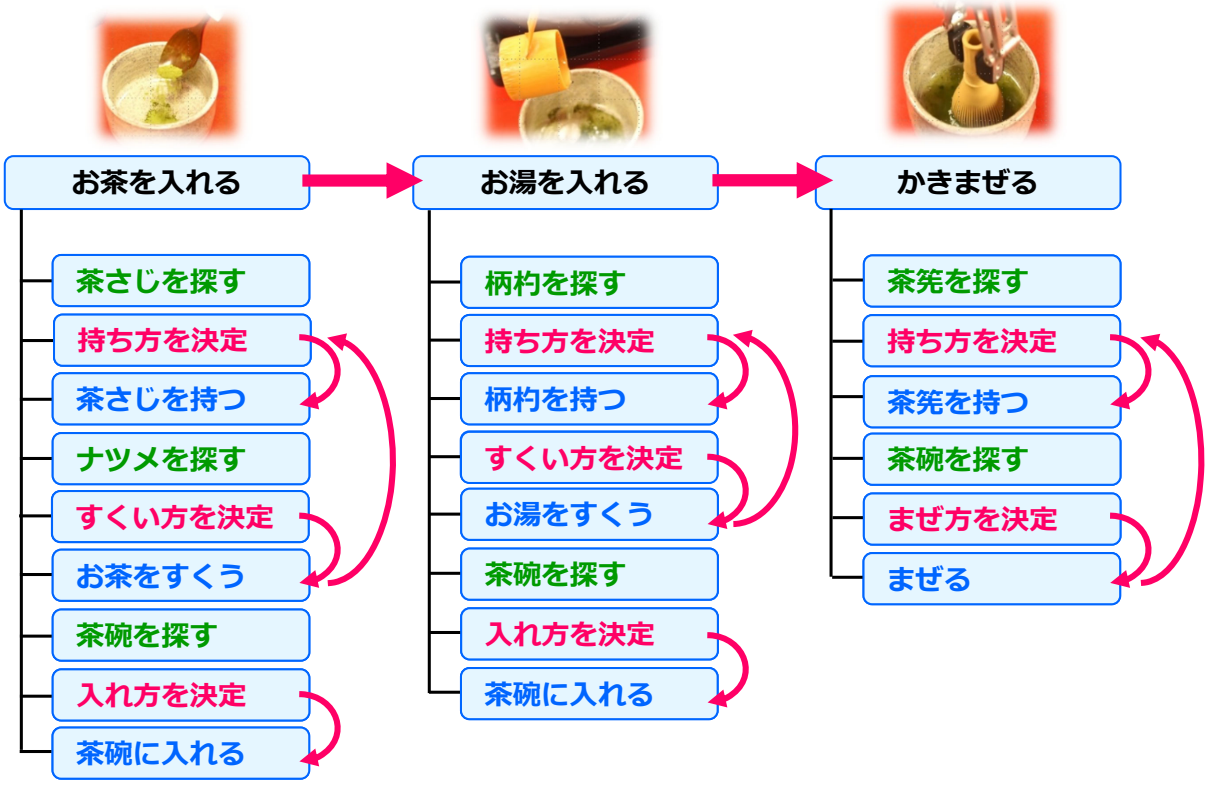
① お茶を茶碗に入れる

② お湯を茶碗に入れる

③ お茶とお湯をかきまぜる

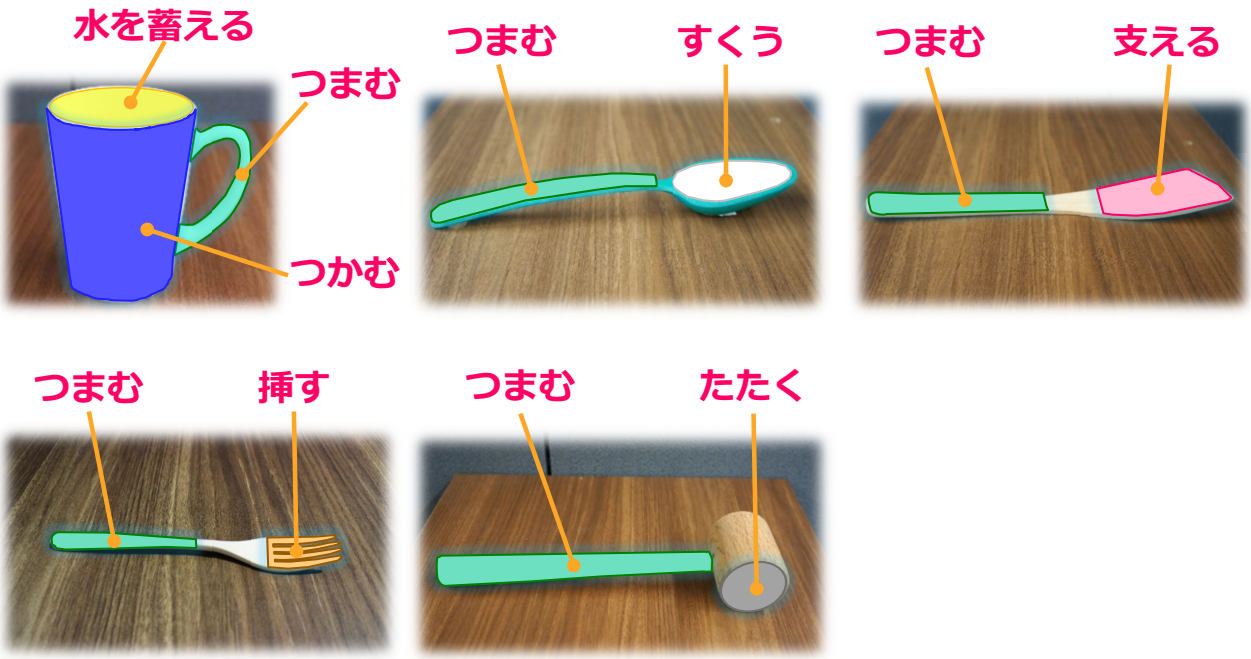
お茶をたてる動作（ロボットの場合）

「お茶をたてる」動作の内訳

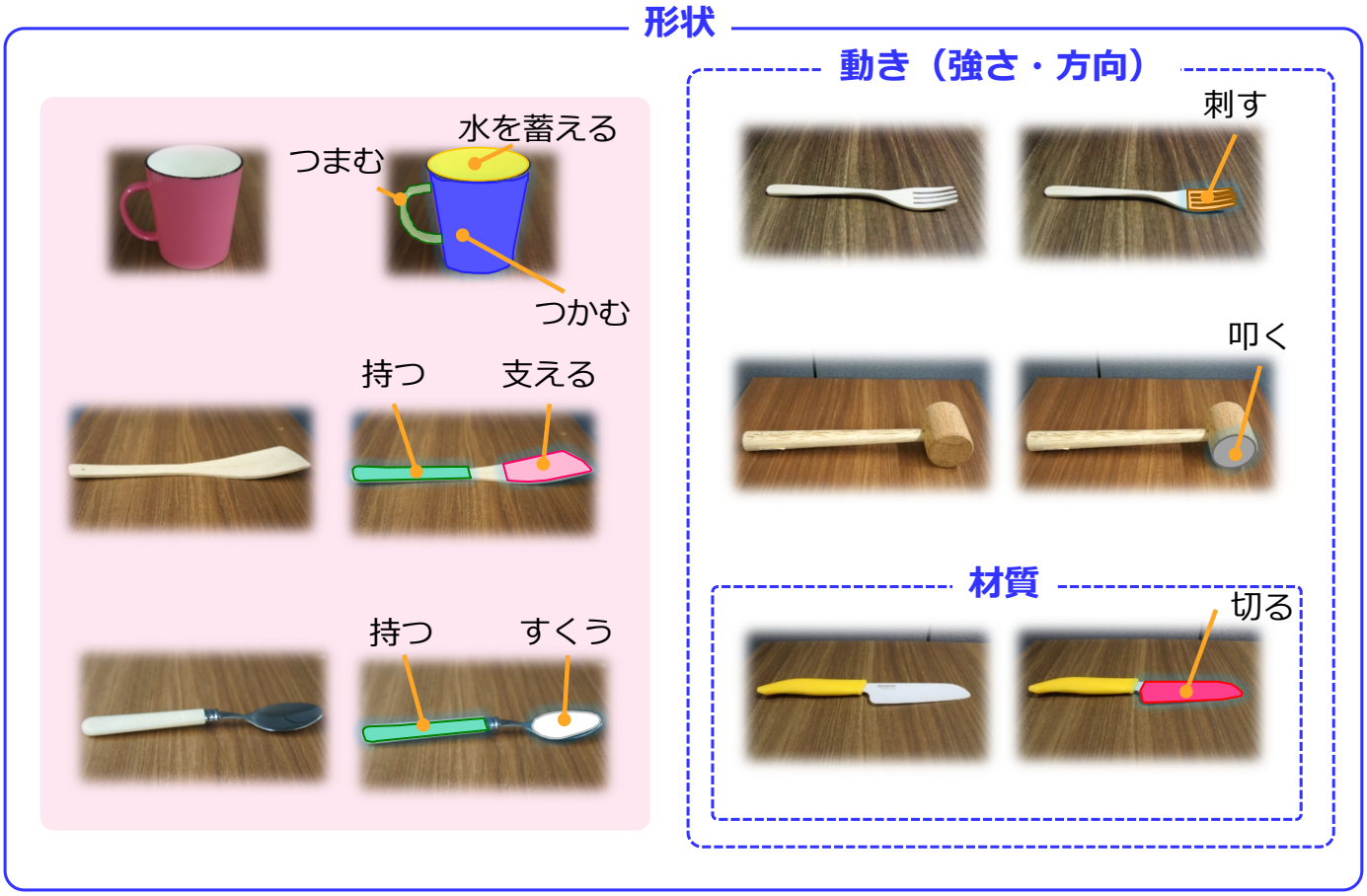


人間なら，道具の使用目的を事前に想定して持つことができる．

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さまざまな要素



主な Affordance（≒“機能属性”）認識手法

ハンドクラフト特徴量＋識別器

Hermans2011：RGB画像からSIFT特徴などのハンドクラフト特徴を抽出し，SVMによって graspable など8種の affordance を識別

Myers15：RGB-D画像をもとに，DのHoG特徴やShape-indexなどのハンドクラフトな局所特徴を抽出し，SRF（Structured Random Forests）で affordance を認識

深層学習ベース

Nguyen16：RGB-D画像をもとに，Encoder-Decoder型CNNを適用することによって affordance を認識

Roy16：RGB画像をもとに，Multi scale CNN型の segmentation によって affordance を認識

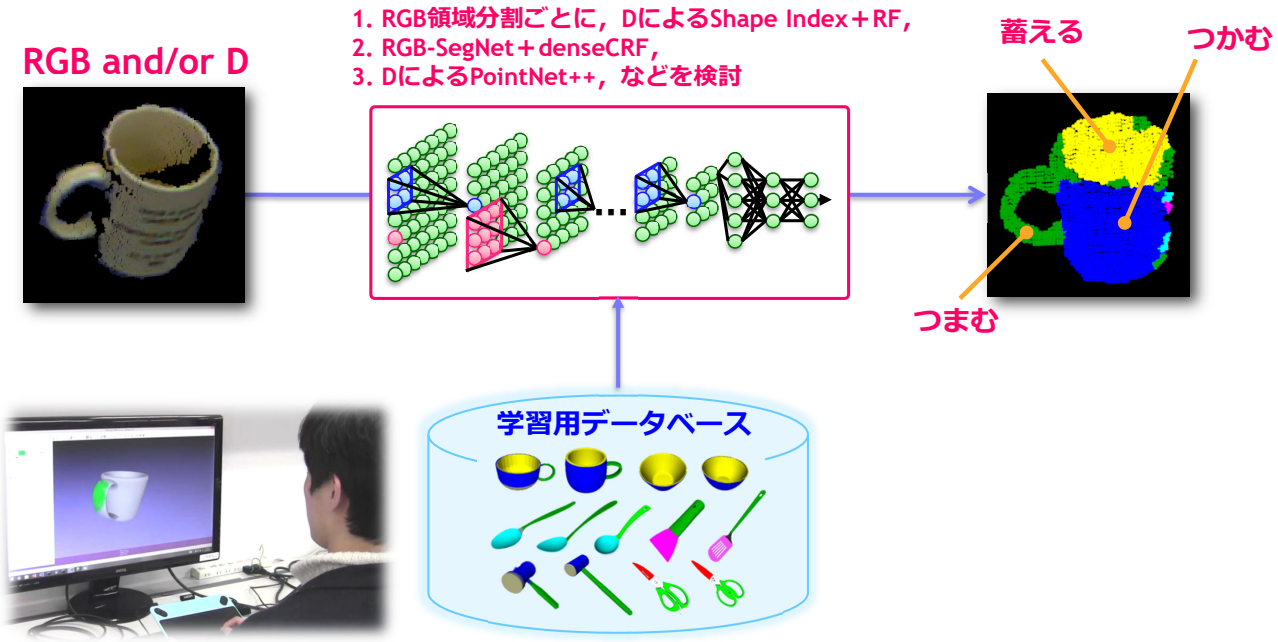
深層学習＋最適化

Nguyen17：前段としてRGB画像から対象物を検出し，出力されたBBをROIとしてCNN（VGG-16）によって affordance に関する特徴量を得，さらに後段にdense CRF（Fully Connected Conditional Random Field）を加えることによって画素間のつながりを最適化した出力を得る．

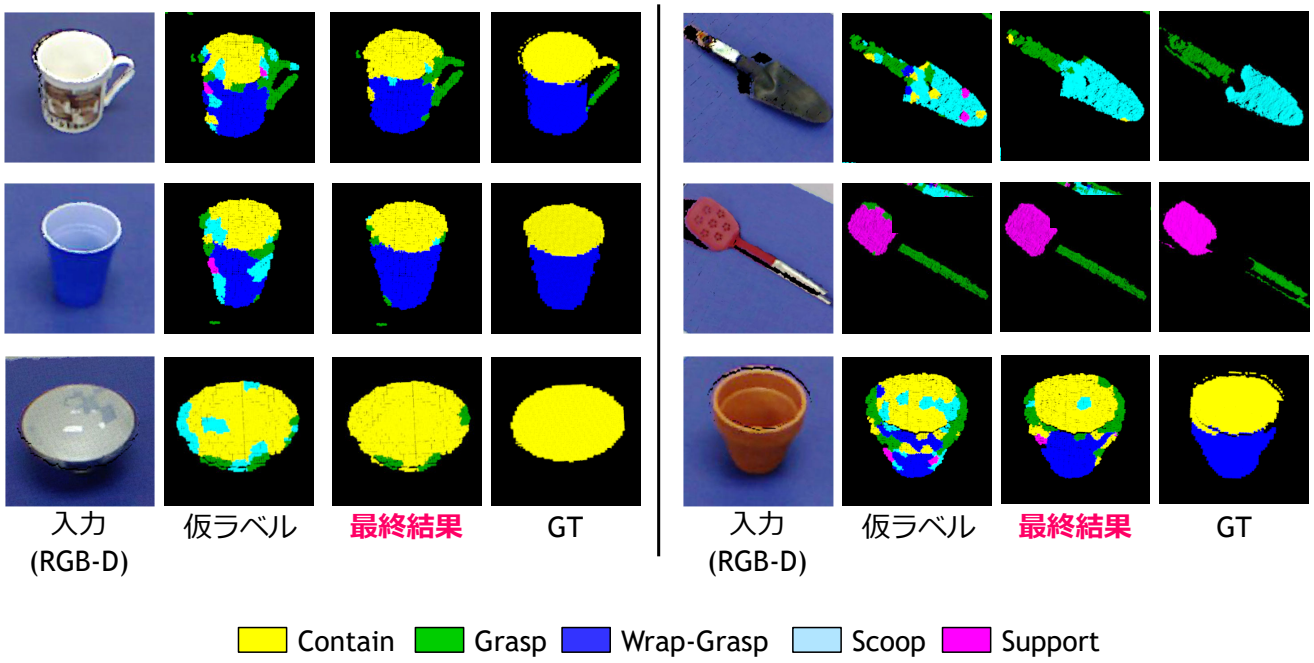
Iizuka18：RGB-D画像を利用する．RGBをもとに supervoxel 化して Random Forests により affordance を認識した後，depth を用いた denseCRF によって形状的に滑らかな出力になるように最適化．

[1] T. Hermans, J.M. Rehg and A.F. Bobick: “Affordance Prediction via Learned Object Attributes,” IEEE Int. Conf. on Robotics and Automation, pp.1–8, 2011.
[2] A.Myers, C.L.Teo, C.Fermuller, and Y.Aloimonos: ``Affordance Detection of Tool Parts from Geometric Features“, Proc. IEEE Conf. on ICRA, pp.1374-1381, Washington, USA (2015-5)
[3] A.Nguyen, D.Kanoulas, D.G.Caldwell, and N.G.Tsagarakis: ``Detecting object affordances with Convolutional Neural Networks",Proc. IEEE/RSJ Conf. on IROS, pp.2765-2770, Daejeon, Korea (2016-10)
[4] A. Roy and S. Todorovic: ``A Multi-Scale CNN for Affordance Segmentation in RGB Images"Proc. IEEE Conf. on ECCV, pp. 186-201(2016-9)
[5] A. Nguyen, D. Kanoulas, D. G. Caldwell, and N. G. Tsagarakis: ``Object-Based Affordances Detection with Convolutional Neural Networks and Dense Conditional Random Fields" Proc. IEEE/RSJ Conf. on IROS(2017-9)
[6] 飯塚正樹，秋月秀一，橋本学，物体形状を考慮したdenseCRFによる機能属性認識の高精度化，電気学会論文誌C，Vol.138，No.9，pp.1088-1093，2018.

機械学習による“機能”認識



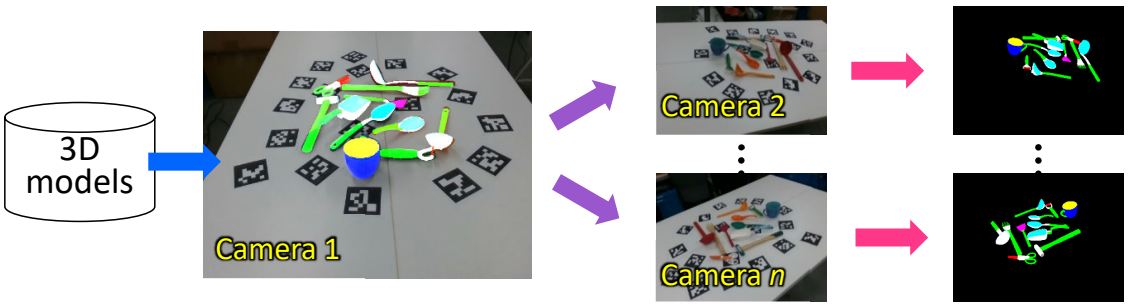
機能認識の結果



機能認識のための学習データ自動生成

課題：機能認識のためのアノテーションのコストが高い.

解決：物体の姿勢推定を利用し，3Dモデルに割り当てたアノテーションデータを実画像に自動マッピング.



1. **物体の姿勢を推定（ここだけ手動）**　2. **カメラの姿勢を推定（自動）**　3. **GT画像レンダリング（自動）**

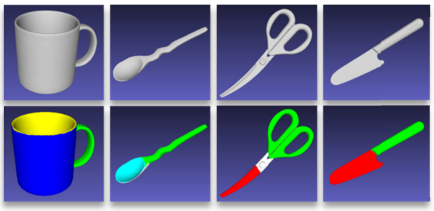
Shuichi Akizuki, Manabu Hashimoto,
Semi-automatic training data generation for semantic segmentation using 6DoF pose estimation, proceedings of VISAPP, pp.607-613, 2019.

機能属性ラベル付きデータセットの公開　（2019年5月24日公開開始）

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]

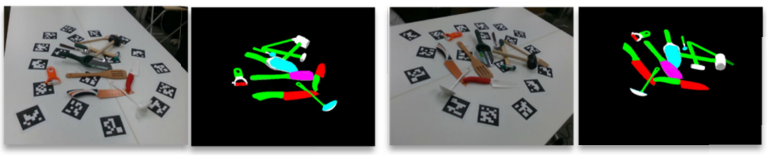
日用品の例



機能属性ラベル付き 3D モデルの例

データ形式		使用例
3Dモデル	ply 形式 obj 形式 stl 形式	全周囲データを用いたセグメンテーション【PointNet など】
機能属性ラベル	ply 形式	

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

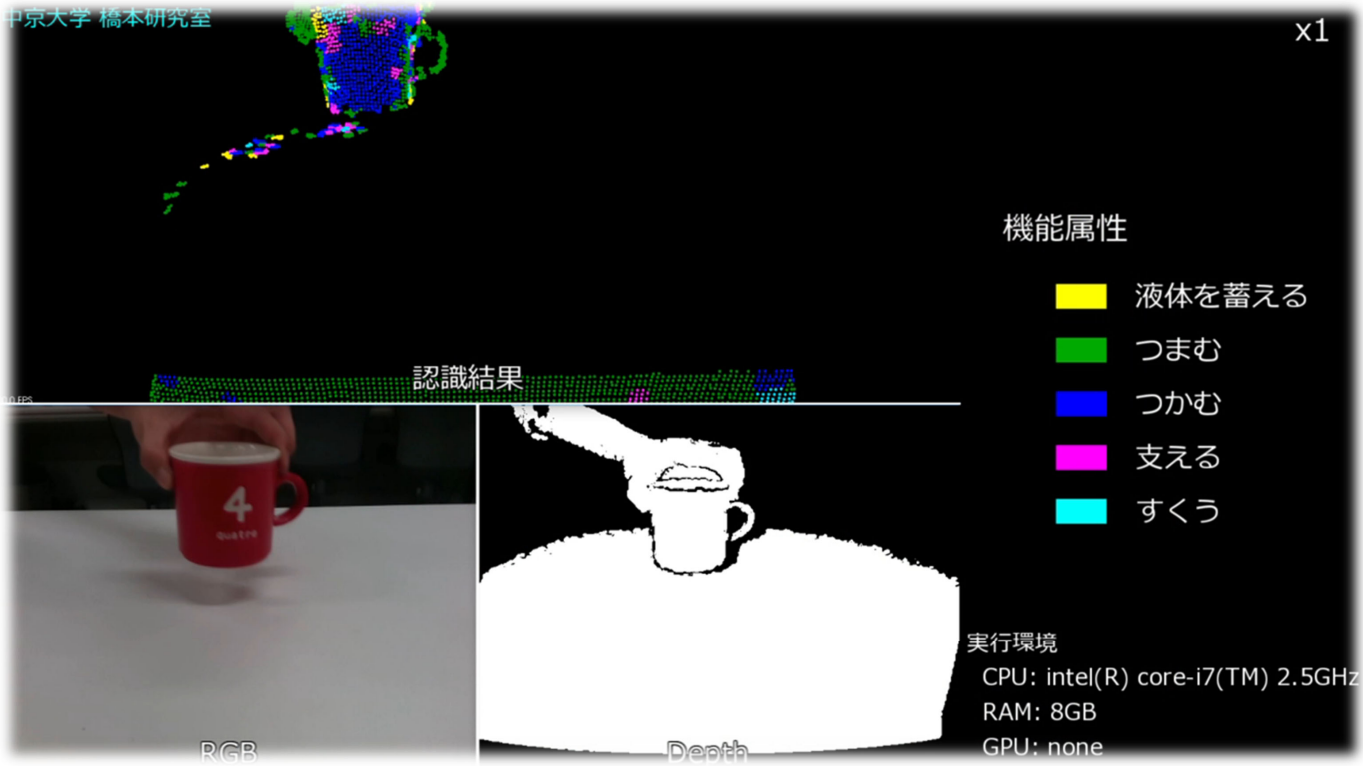
c. 機能属性セグメンテーション用シーン（2D）[モデル6品種2799枚，機能6クラス]

茶道具シーンと機能属性ラベルの例

データ形式		使用例
RGB画像	jpg 形式	画像ベースの機能認識（セグメンテーション）【SegNet，U-Net など】
機能属性ラベル	png 形式	

リアルタイム機能認識（10fps）

（44秒）



“機能” から対象物の名前（種類）を知ることもできる

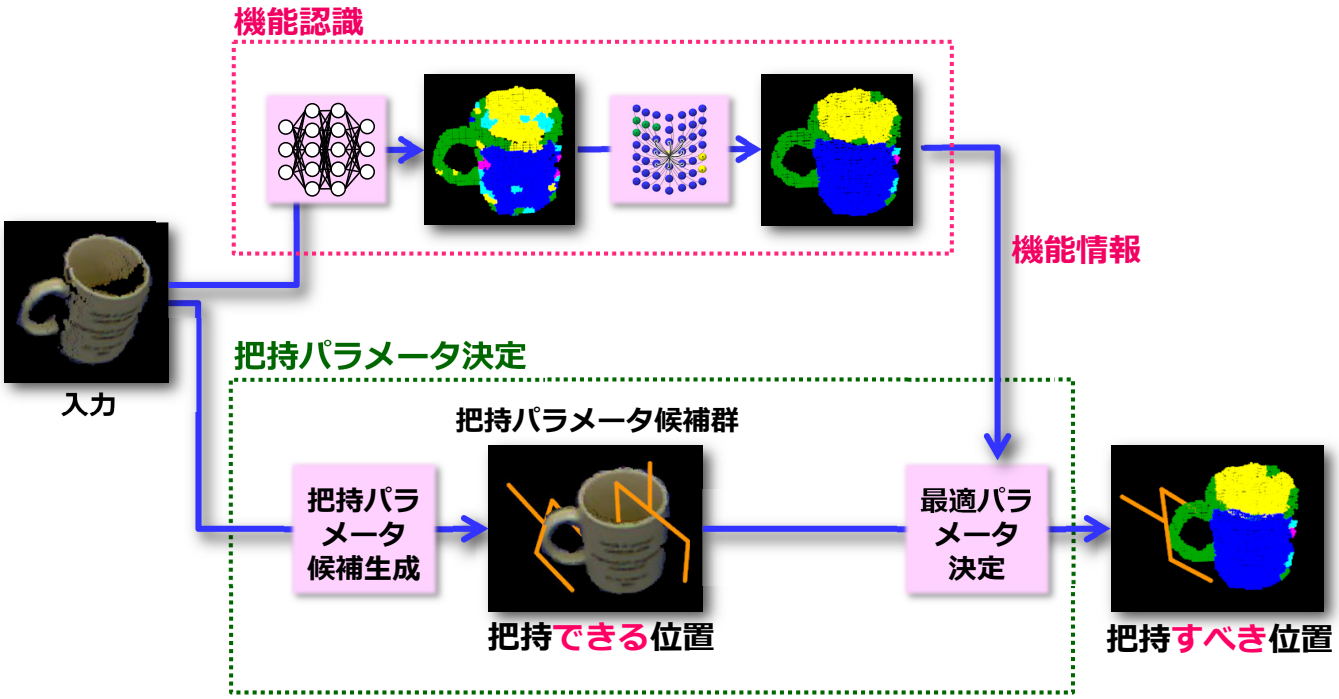
さまざまな機能属性の含有比率を利用してカテゴリ名を推定

（34秒）

Masaki Iizuka, Shuichi Akizuki, Manabu Hashimoto, Affordance-based 3D Feature for Generic Object Recognition, Proceedings of 13th International Conference on Quality Control by Artificial Vision (QCAV), Vol.10338, 103380W-1-6, DOI: 10.1117/12.2266917, 2017.

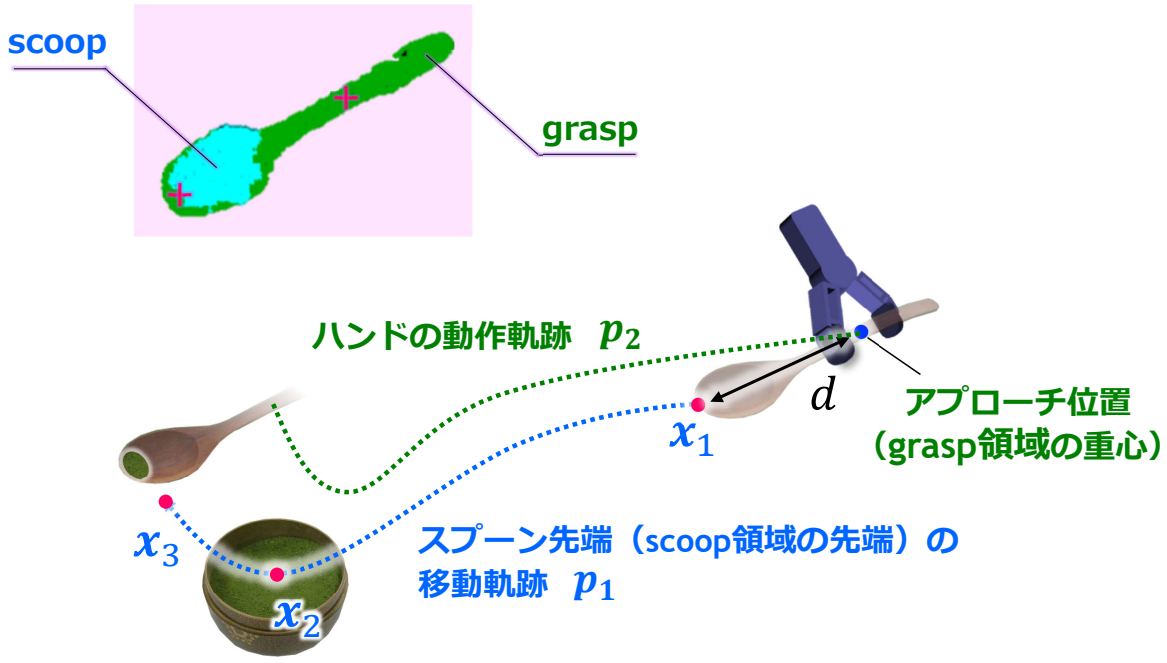
“機能” がわかると何がうれしいか？①

適切な把持位置の決定



“機能” がわかると何がうれしいか？②

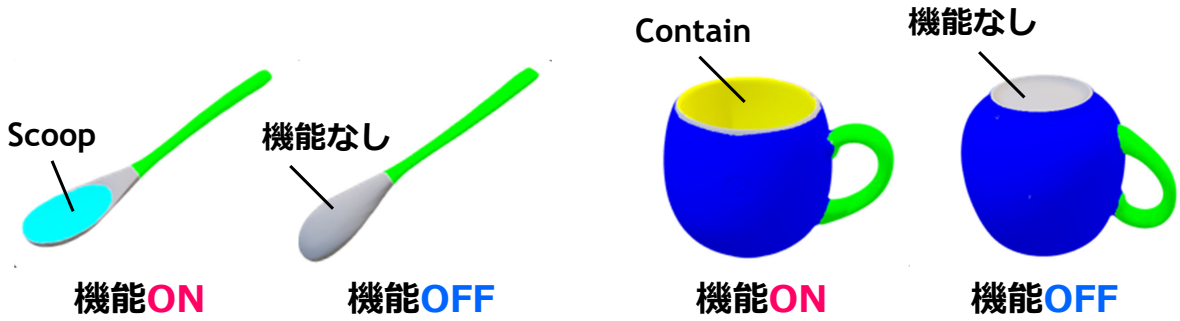
動作軌跡の決定



“機能” がわかると何がうれしいか？③

機能を満たすための（追加的な）動作の生成

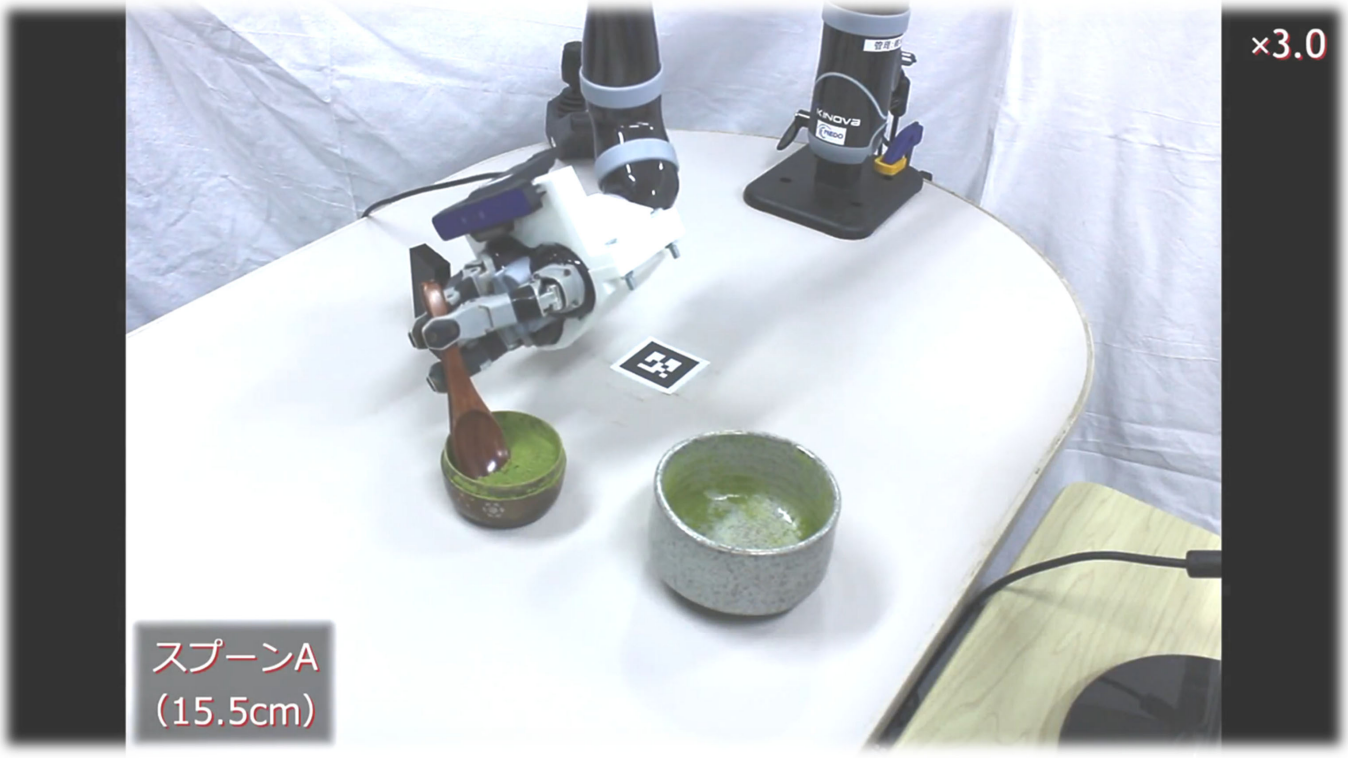
機能が　発現する姿勢（ON） / しない姿勢（OFF）



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

機能属性情報を用いた動作生成の例

（22秒）



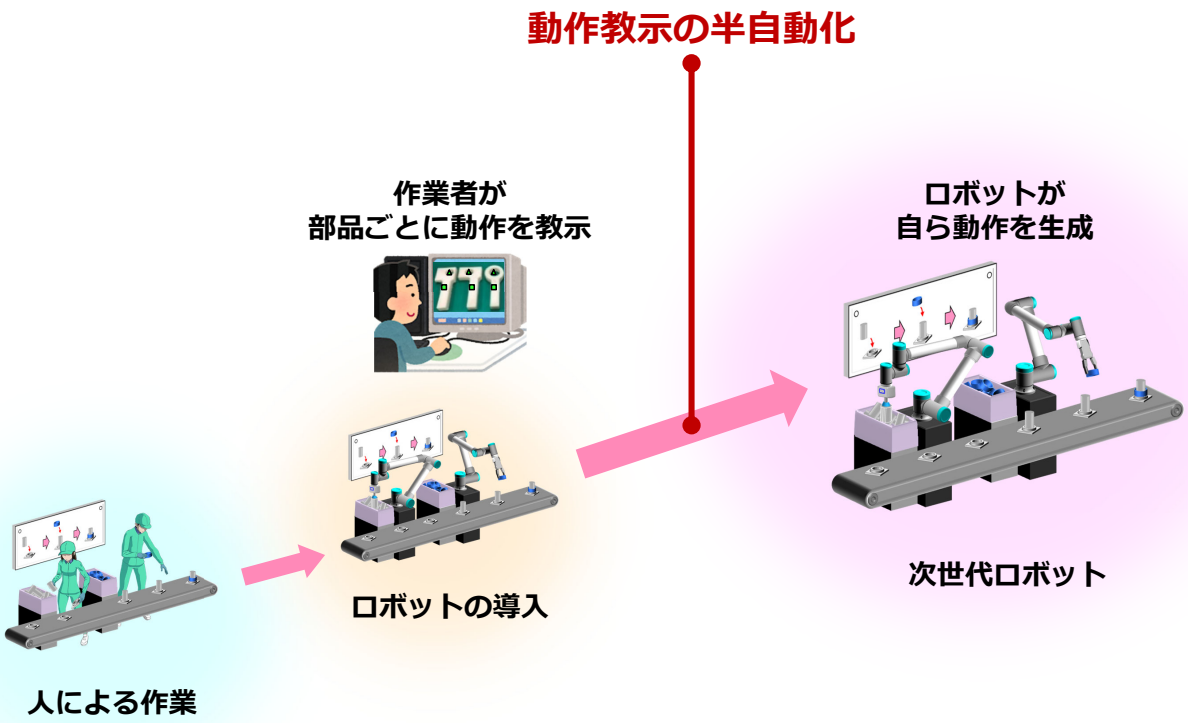
お茶会ロボット２号

（2020年1月16日完成，YouTube公開中）

（3分50秒）

ロボット動作教示の簡素化

研究の狙い



カテゴリ内での動作パラメータの転移

■ **組み立て作業で扱われる部品群（カテゴリ）の例**

接続ロッド

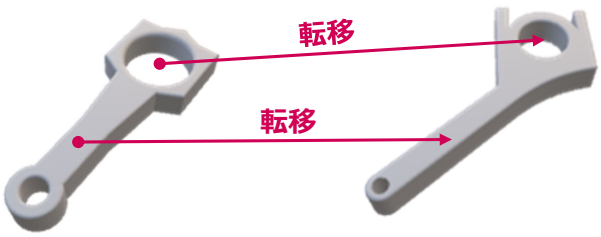
コネクタ

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/

■ **基本的な考え方**

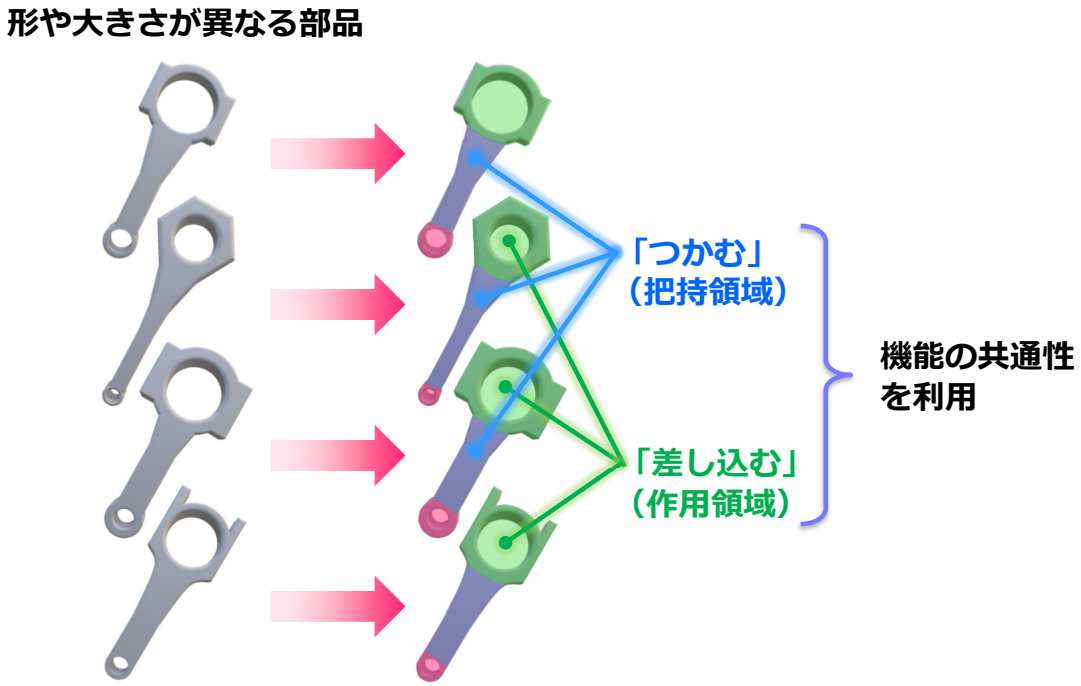
人による教示結果を参考に，別の部品に転移



人による教示点　別の部品

異形部品を同一視するための共通性

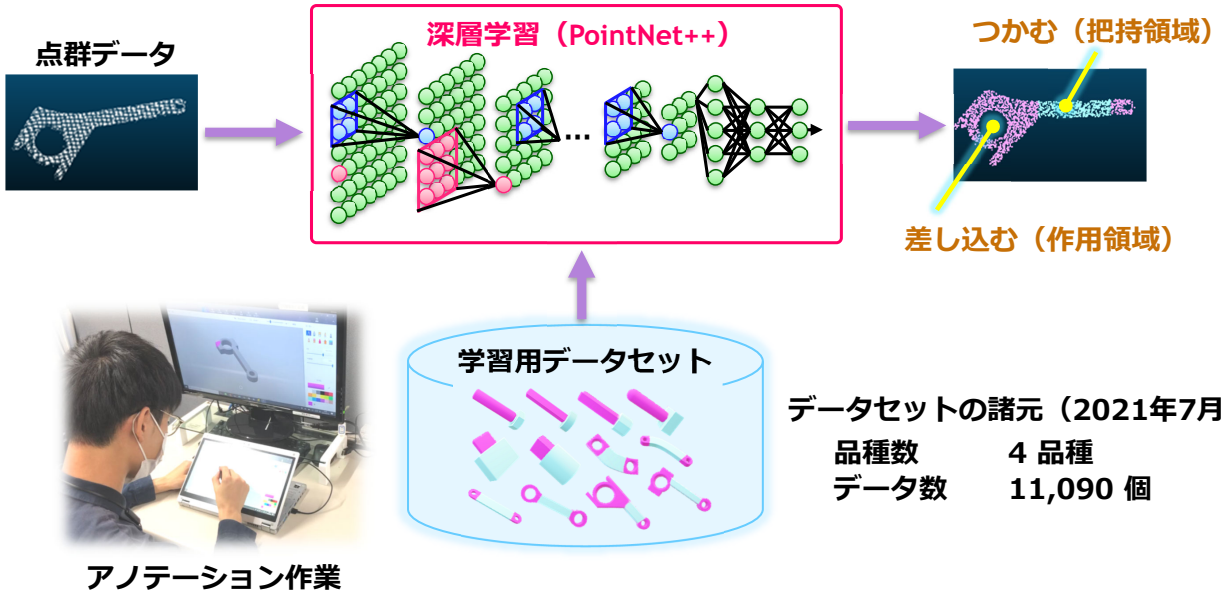
■ **同一カテゴリ内の部品が有している機能の共通性を利用**



深層学習を用いた部品の機能属性認識

■ カテゴリごとに，把持領域と作用領域を学習し，認識する

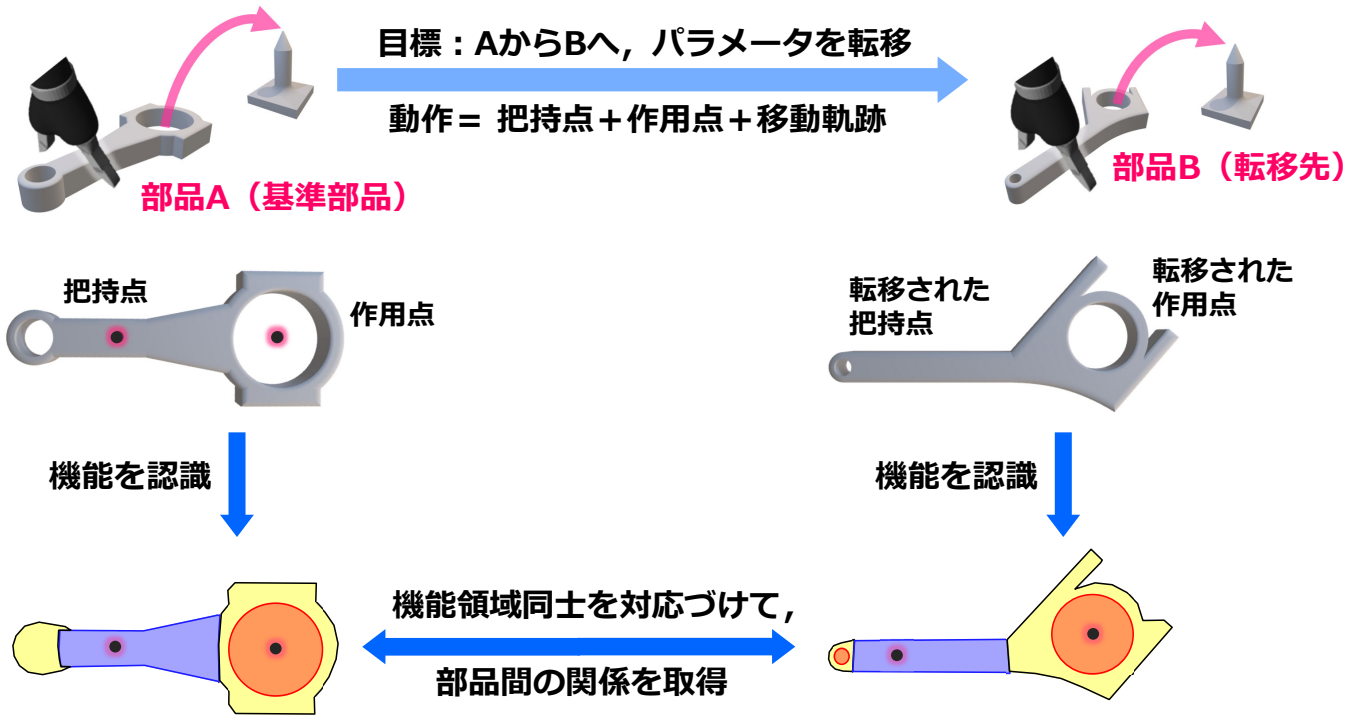
（一種の“機能”）



データセットの諸元（2021年7月現在）

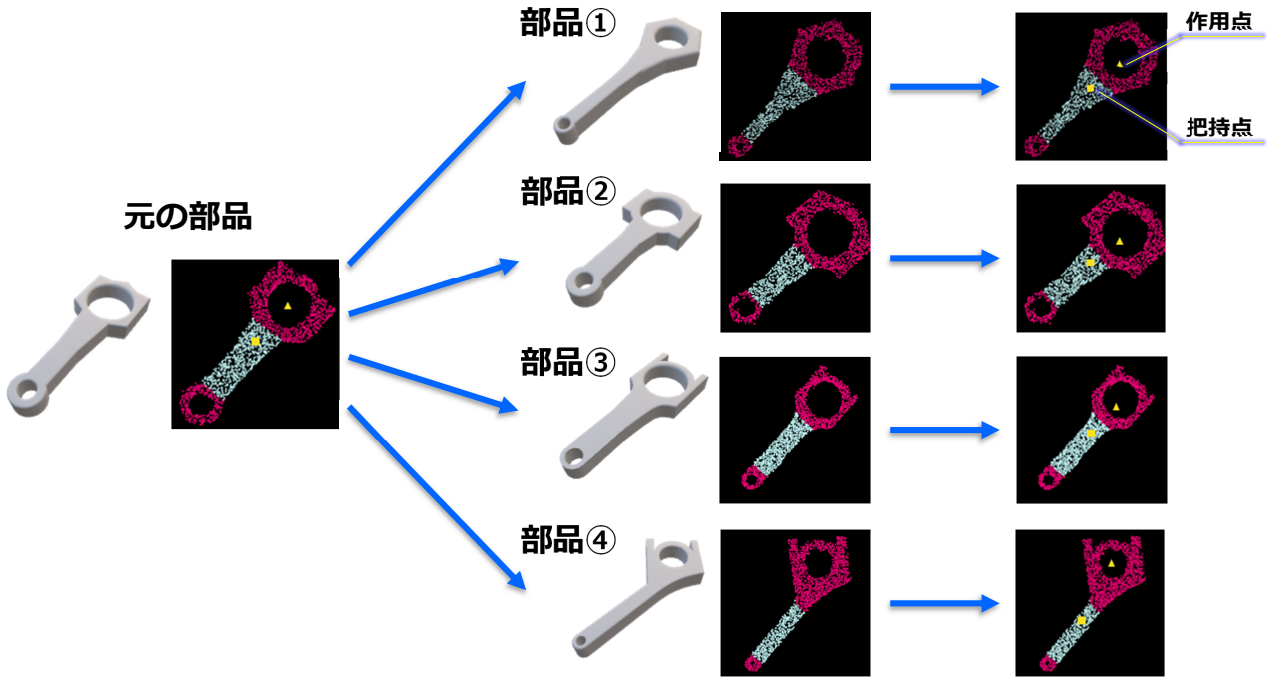
品種数	4 品種
データ数	11,090 個

“機能” を手がかりにした把持点と作用点の転移



把持点と作用点の転移結果

元の部品の把持点と作用点を，
大きさや形が異なる４種の部品に転移させた．

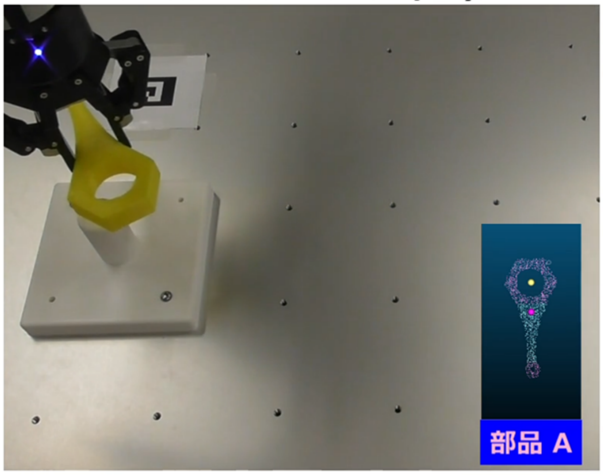


提案手法によって生成されたロボット動作

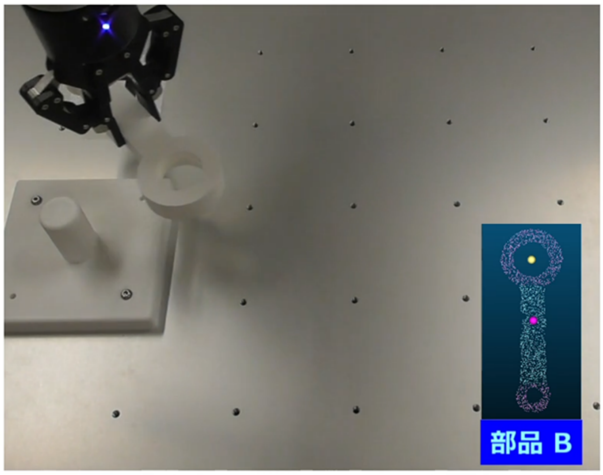
（1分10秒）

部品 A から部品 B への動作パラメータ転移

部品 A
人が教示した動作（Replay）



部品 B
提案手法により転移された動作

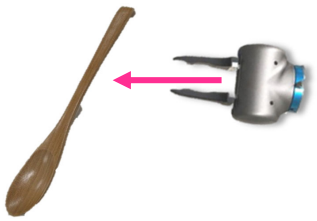


（サイズが異なる部品）

部品 B の挿入動作は，提案手法によって自動的に生成されています．

自然な言語からのロボット動作生成

本研究で想定する動作の種類



(a) Pick 動作

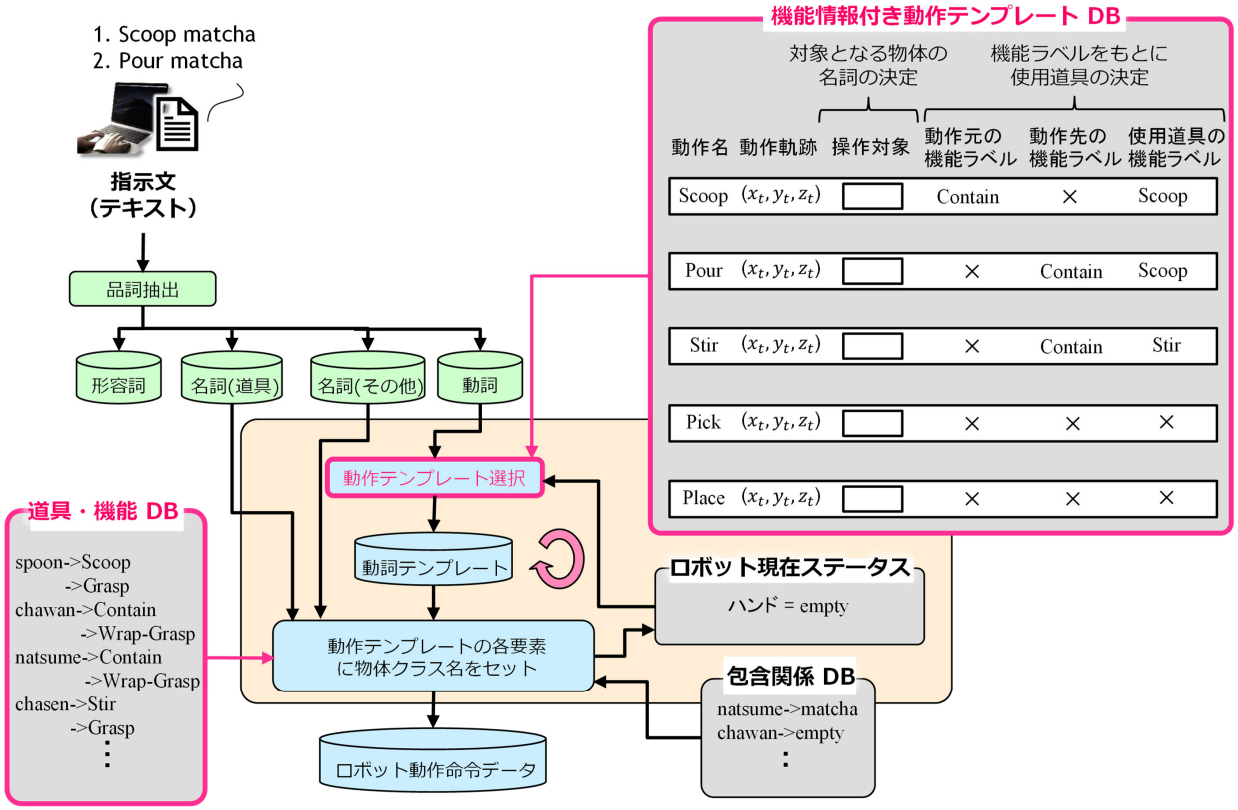
(b) Place 動作

(c) Scoop 動作

(d) Pour 動作

(e) Stir 動作

機能情報付き動作テンプレートを利用した提案手法



ロボット動作の生成例（お茶たてタスク）

入力したテキスト (English sentence)

L1: Put matcha.
L2: Put hot-water.
L3: Stir process2.

提案システムによって生成されたロボット動作
（動作を表す内部データを文章化したもの）

```
01      Pick spoon
02      Scoop matcha from natsume with spoon
03      Pour matcha to chawan with spoon
04      Place spoon
05      Pick hishaku
06      Scoop hot-water from chagama with hishaku
07      Pour hot-water to chawan with hishaku
08      Place hishaku
09      Pick chasen
10      Stir chawan with chasen
11      Place chasen
```


ロボット動作実験の結果

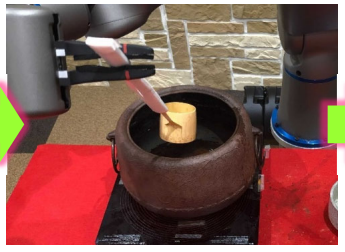
01 Pick spoon

02 Scoop matcha from natsume with spoon

03 Pour matcha to chawan with spoon

04 Place spoon

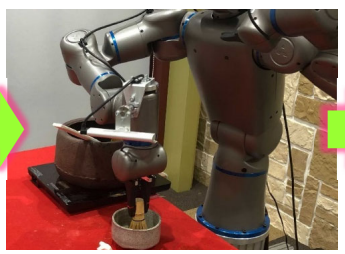
05 Pick hishaku


06 Scoop hot-water from chagama with hishaku

07 Pour hot-water to chawan with hishaku

08 Place hishaku

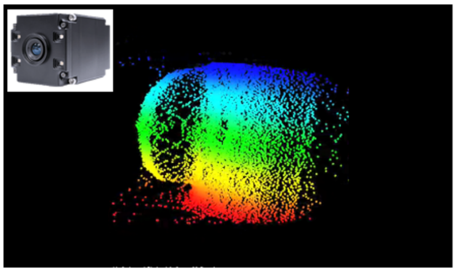
09 Pick chasen


10 Stir chawan with chasen

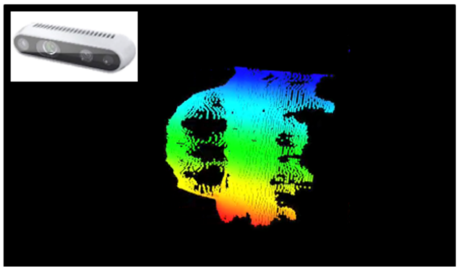
11 Place chasen

異なるセンサ間のギャップの低減

使用したセンサによって得られる点群が異なる



センサ A



センサ B

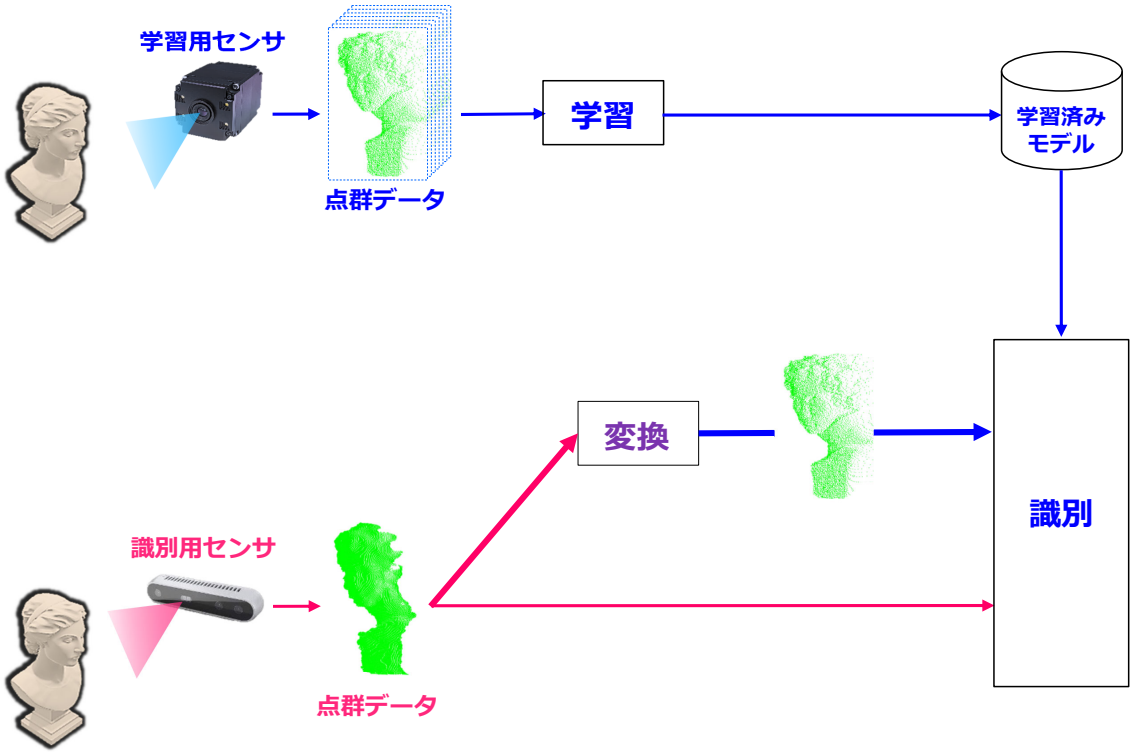
センサの差異が物体の識別性能に影響を及ぼす

内容：　　PointNet++を用いた４クラス識別実験
学習画像：　6600枚
テスト画像：　600枚

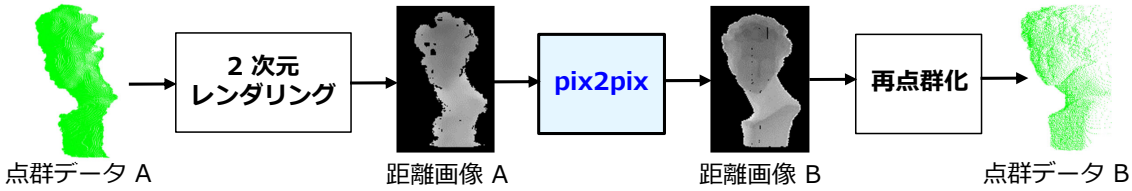
学習＼識別	D415 (Active Stereo)	D435 (Active Stereo)	Helios (TOF)
D415	100.0	78.7	50.0
D435	98.9	99.4	77.8
Helios	71.2	58.9	99.8

学習時と識別時で，センサが異なると性能が低下する．

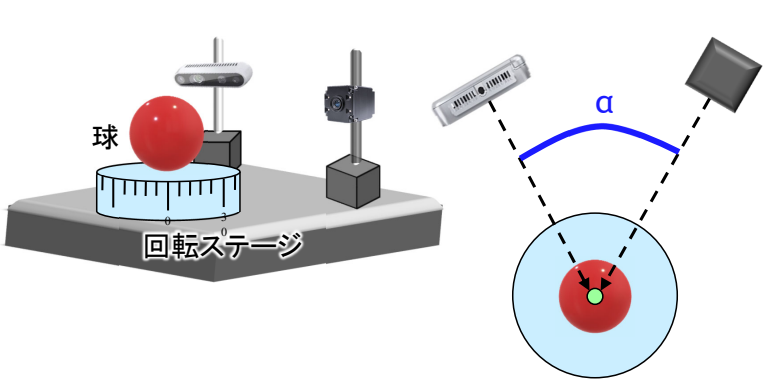
解決のためのアイデア



機械学習を利用した点群データ変換

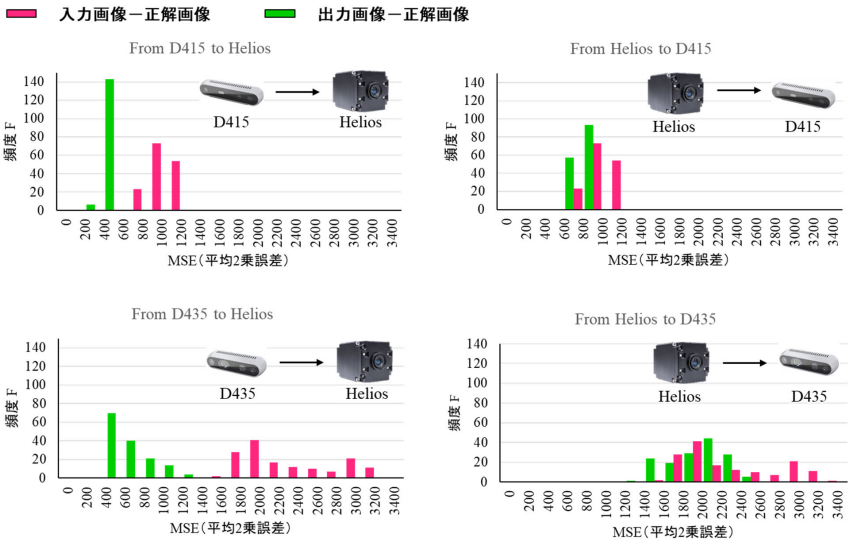


Pix2Pixをうまく学習させるためには,
同一視点から異なるセンサで撮影したデータ群が必要



実験結果

距離画像としての変換性能：出力画像のほうが正解画像に近くなった.



PointNet++による識別結果：RealSense D系とHeliosとの変換は効果あり

学習＼識別	D415	D435	Helios
D415		78.7 → 52.6	50.0 → 74.5
D435	98.9 → 92.5		77.8 → 90.6
Helios	71.2 → 73.7	58.9 → 87.5	

本日の話題

1．イントロダクション：　課題整理とロボットビジョンの概要

2．モノを視る‥‥２次元的アプローチ

戦略的画素選択に基づくテンプレートマッチングの高度化

（照明変動ロバスト化，識別性能向上，回転変動対応‥‥）

3．モノを視る‥‥３次元的アプローチ

ロボット動作生成にかかわる諸技術

（機能認識，動作教示簡素化，言語指示‥‥）

異なるセンサ間のギャップの低減

4．ヒトを観る

匠の技の分析

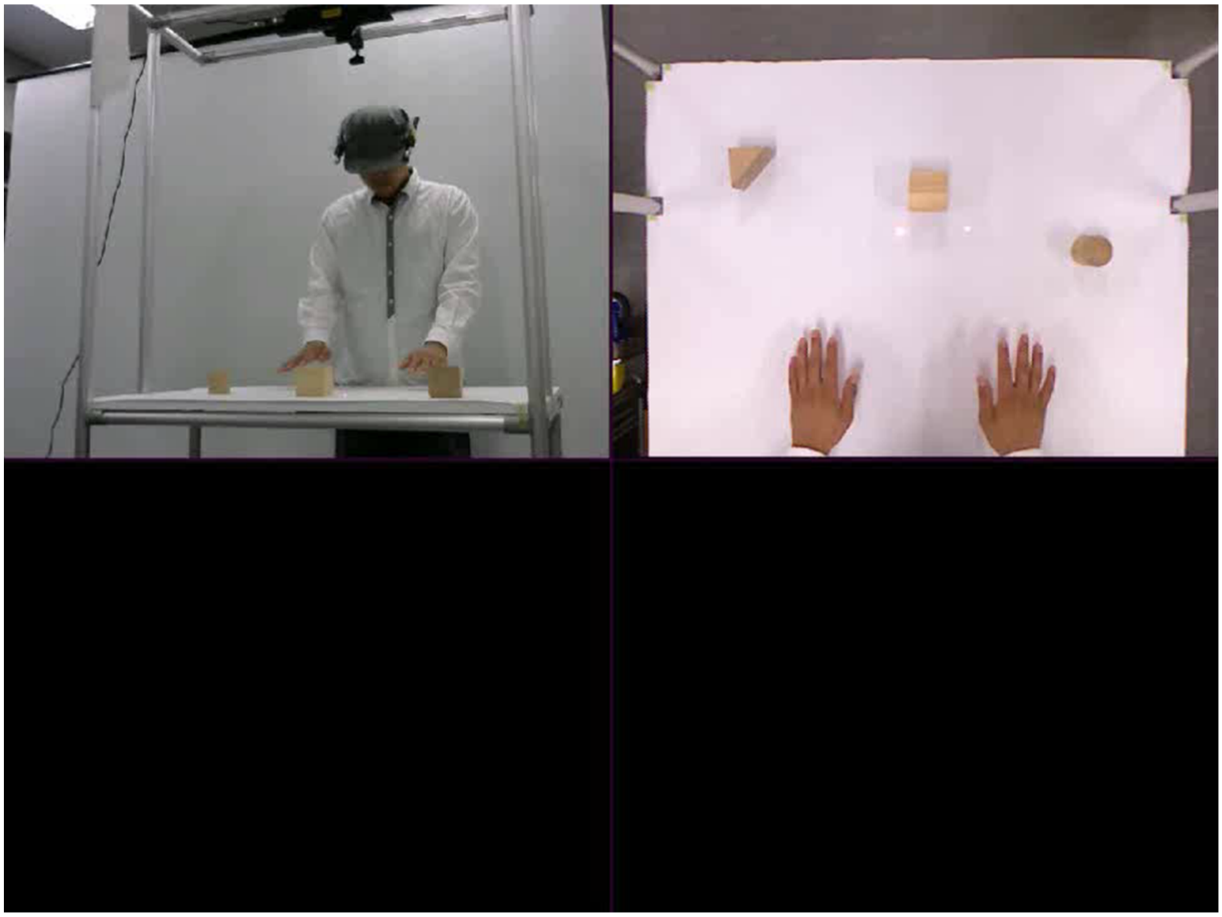
5．まとめ

「匠の技」の分析

- 本研究の目的
  - 技能レベルごとの動きの違いを分析するための特徴量の提案
- 基本アイデア
  1. 外部からセンシングした情報に基づいて，特徴量を抽出

  2. 技能レベルごとに視線と手の動きの違いを分析

動作のコード化と記述



コード化特徴量に基づく動作分析

特徴量に基づいて，スキルレベルごとの視線と手の動きを調査

特徴量抽出モジュール

作業状況の
動画像

A

視線の位置情報
のコード化

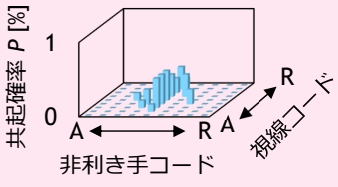
A C C A D
視線のコード列

手の位置情報
のコード化

右：A C C A D
左：A C O B D
両手のコード列

B

視線・手の
統合特徴量の抽出

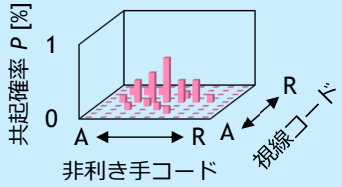


非利き手コード

視線・動作統合特徴量

スキル獲得プロセス分析モジュール

各スキルレベルの
平均モデル生成



非利き手コード

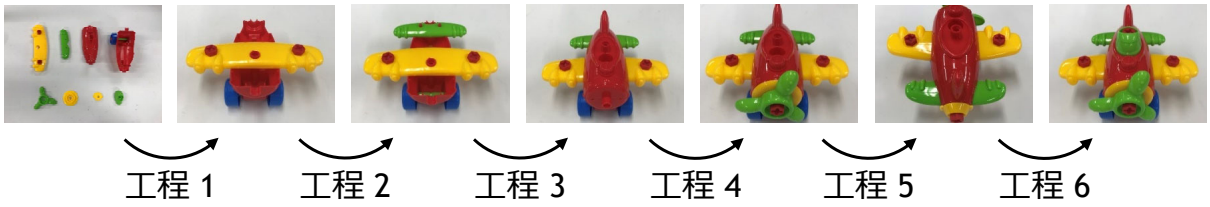
スキルレベルごとの平均モデル

スキルレベルごとの
平均モデルの比較

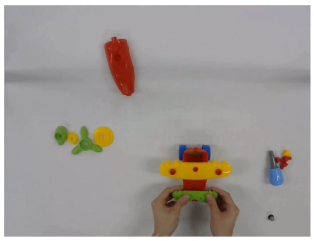
スキルレベルごと
の動きの特徴

対象動作

- 対象動作
  飛行機型のおもちゃを組み立てる作業



- 基本動作
  作業動作を 3 種類の基本動作に分類


Picking

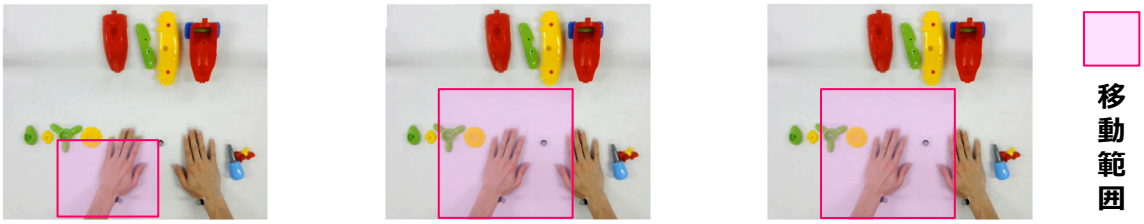
Assembling

Screwing

分析結果

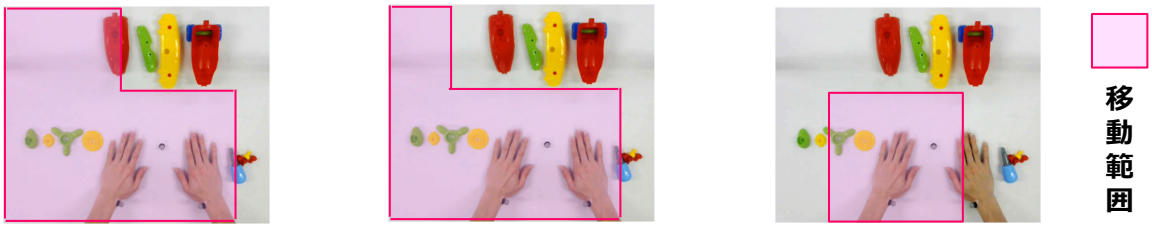
- 技能レベルごとに視線と手の移動範囲を調査

左手-(非利き手)-の移動範囲



移動範囲が増加　変化なし

視線の移動範囲

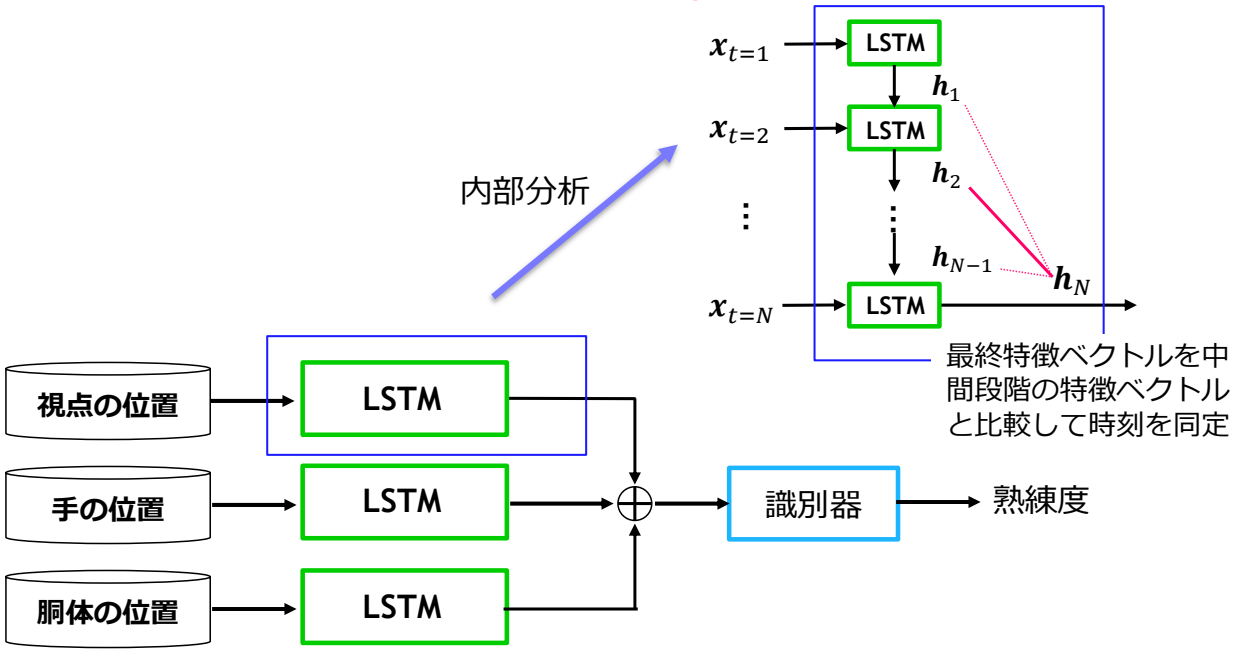


変化なし　移動範囲が減少

習熟の段階で移動範囲の変化する要素が異なる.

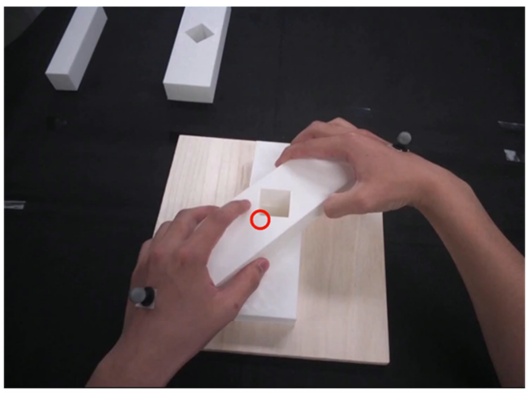
（最新）Multi-Stream LSTM を用いた分析

- 入力データ：モーキャプで撮影した時系列の視点，手，ボディ位置
- LSTMで特徴を抽出し，統合して識別．熟練度合いを推定させる．
- 学習後のLSTM内部を分析し，熟練/初心者の差が顕著な時刻を同定
- その時刻の画像を分析（現時点では手動）

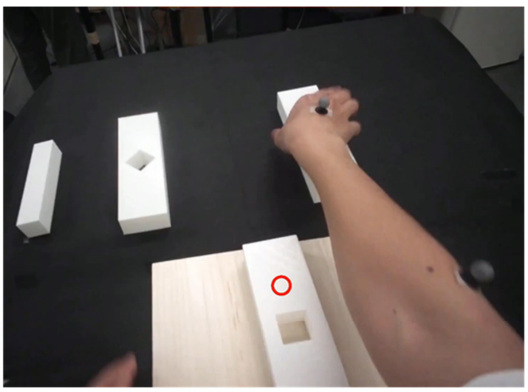


プロセス分析の結果例

初心者**物体を目で確認して把持**，熟練者は**物体を目で確認することなく把持**



Pick 時の**初心者**の視線の動き



Pick 時の**熟練者**の視線の動き

本日の話題

1．イントロダクション：　課題整理とロボットビジョンの概要

2．モノを視る‥‥２次元的アプローチ

戦略的画素選択に基づくテンプレートマッチングの高度化

（照明変動ロバスト化，識別性能向上，回転変動対応‥‥）

3．モノを視る‥‥３次元的アプローチ

ロボット動作生成にかかわる諸技術

（機能認識，動作教示簡素化，言語指示‥‥）

異なるセンサ間のギャップの低減

4．ヒトを観る

匠の技の分析

5．まとめ

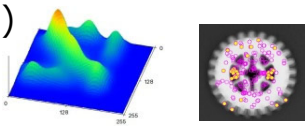
まとめ

1．イントロダクション：　課題整理とロボットビジョンの概要

2．モノを視る‥‥２次元的アプローチ

戦略的画素選択に基づくテンプレートマッチング

（照明変動ロバスト化，識別性能向上，回転変動対応‥‥）

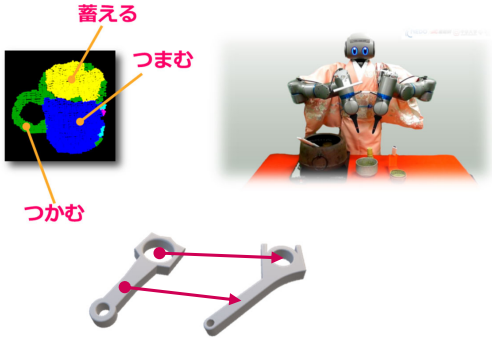


3．モノを視る‥‥３次元的アプローチ

ロボット動作生成にかかわる諸技術

（機能認識，動作教示簡素化，言語指示‥‥）

異なるセンサ間のギャップの低減



4．ヒトを観る

匠の技の分析

技術紹介のページ
http://isl.sist.chukyo-u.ac.jp/archives3/

3次元物体認識に関するチュートリアル資料
http://isl.sist.chukyo-u.ac.jp/wp-content/uploads/2021/05/ObjectRecognition-byHashimotoLab-3D.pdf