

## 3次元物体認識の現状技術と将来技術

中京大学 工学部 橋本 学

[mana@isl.sist.chukyo-u.ac.jp](mailto:mana@isl.sist.chukyo-u.ac.jp)

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

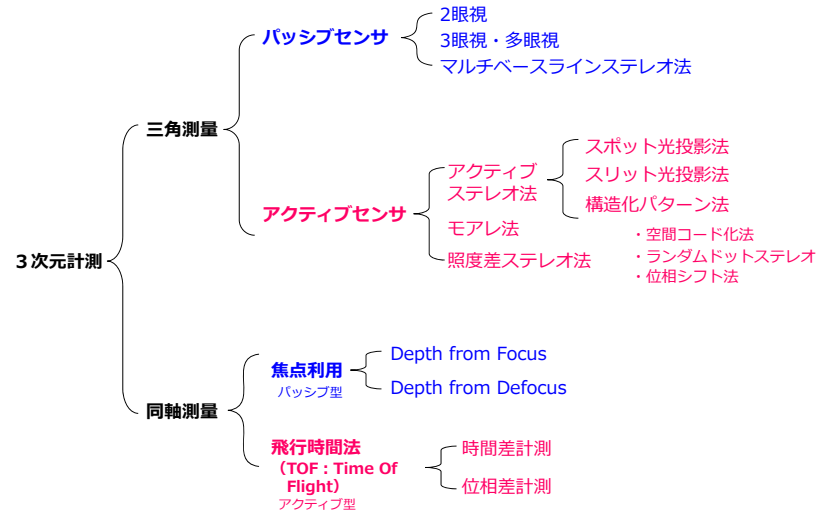
## 本日の話題

1. 3次元物体認識の基礎
  - (1)3次元計測手法の概要
  - (2)センサ主要技術と実センサ事例
  - (3)物体認識アルゴリズムの概要
  - (4)アピランスベースとモデルベース
  - (5)物体認識の周辺技術
2. 3次元局所特徴量を用いた物体認識技術
  - (1)キーポイントベース物体認識の概要
  - (2)3次元特徴量の主要技術
  - (3)局所参照座標系の主要技術
  - (4)3次元特徴量に関する最新技術
3. アマゾンピッキングチャレンジ2016に関連するロボットビジョン技術
  - (1)背景
  - (2)課題分析
  - (3)アプローチ 物体識別と姿勢認識
  - (4)大会のようすと世界的な技術トレンド
4. 物体認識技術の将来展望
  - (1)工場から家庭へ
  - (2)人間の内面を探る

## 1. 3次元物体認識の基礎

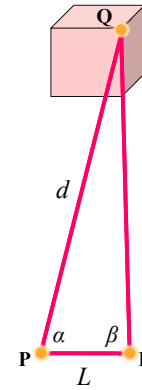
## 3次元計測手法の概要

## 3次元計測手法の分類



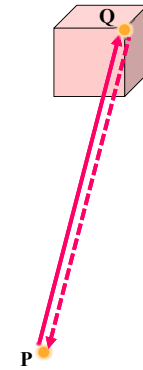
## 三角測量 v.s. 同軸測量

三角測量



異なる視点  $P, R$  をとり ( $L$  が既知),  $\alpha, \beta$  を計算することによって, 距離  $d$  を計測

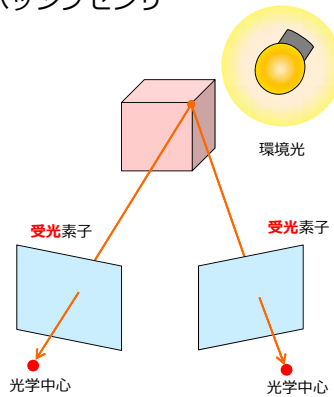
同軸測量



光波の性質 (時空間位相) や集光 (合焦; ぼけ) などを利用して,  $P$  と同じ軸上の  $Q$  までの距離  $d$  を計測

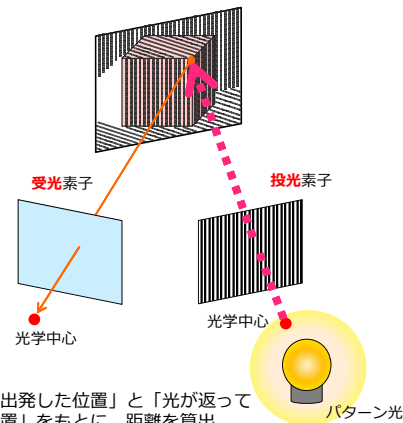
## パッシブセンサ v.s. アクティブセンサ

パッシブセンサ



それぞれの受光部が「光を受けた位置」をもとに, 距離を算出

アクティブセンサ



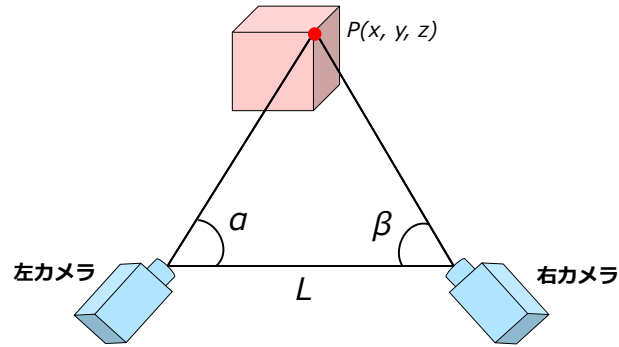
「光が出発した位置」と「光が返ってきた位置」をもとに, 距離を算出

## パッシブ法の概要と主要技術

- ステレオ視覚の考え方
- 3次元座標と観測座標の関係
- 画像マッチングを用いた対応点探索
- 対応点探索の高速化と高信頼化
- 実際のステレオカメラの例

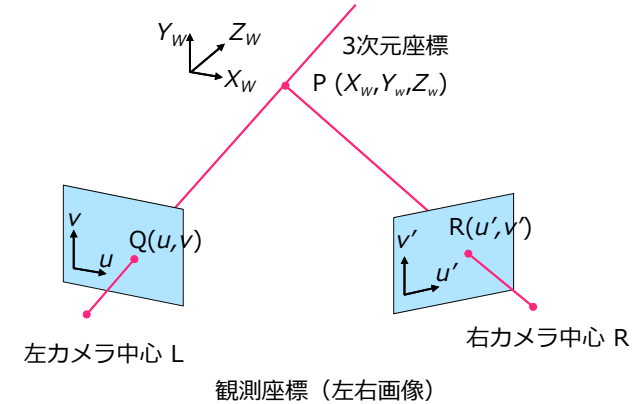
## ステレオ視覚の考え方

- $L$  が既知の条件下で,  $\alpha, \beta$  を計測することによって, 対象物の3次元位置を知る.



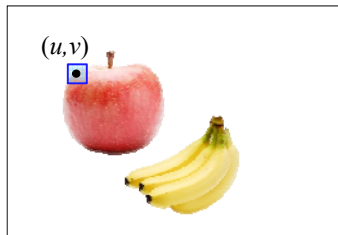
## 3次元座標と観測座標の関係

1. 空間中の  $P$  から発した光が, 左右2つのカメラ画像上に映る.
2. それぞれの画像上の点  $Q(u, v)$  と  $R(u', v')$  の位置を, 画像処理 (マッチング技術) によって特定する (= 対応点探索).
3. これらの位置から  $P$  の3次元位置を算出する.

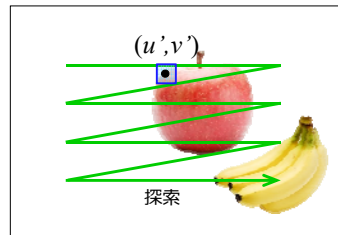


## 画像マッチングを用いた対応点探索

- 領域ベースマッチング (area-based matching)  
 $(u, v)$  周辺の領域内の画素を **テンプレート** とし, テンプレートマッチングを利用して  $(u, v)$  の対応点  $(u', v')$  を探索



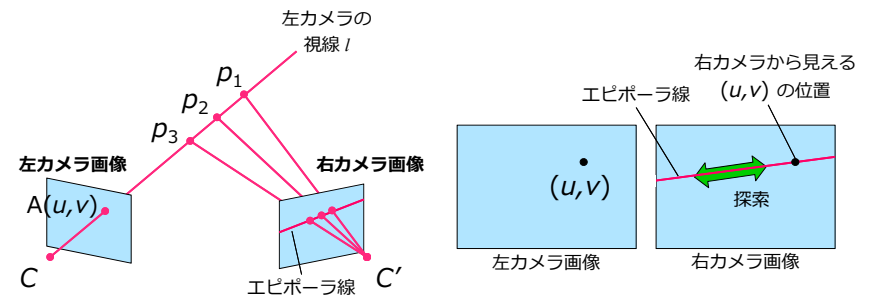
左カメラ画像



右カメラ画像

## 対応点探索の高速化と高信頼化

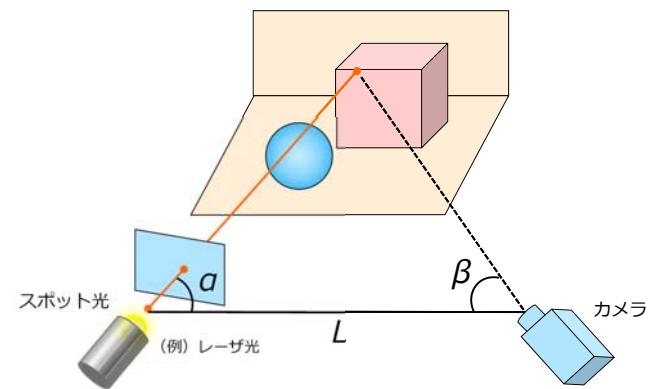
- エピポーラ線 (epipolar line)  
 左画像上のある位置  $A(u, v)$  に対する視線  $l$  を, 右画像に投影したときにできる画像中の直線のこと.
- エピポーラ拘束の利用  
 右画像において,  $A$  の対応点は必ず **エピポーラ線上に存在する** という制約を利用すれば, 対応点を, 高速かつ効率よく探索できる.



## アクティブ法の概要と主要技術

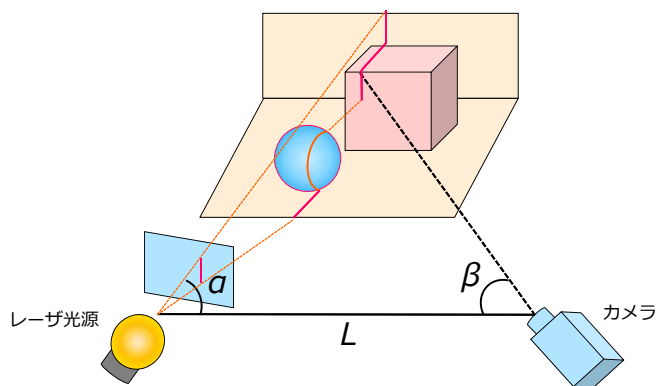
## スポット光投影法

- ステレオ視覚において、片方のカメラをスポット光源に置き換えた方法
- 対象物上に特徴点（テクスチャ）がなくても計測可能
- 1回につき、1点だけしか計測できない。



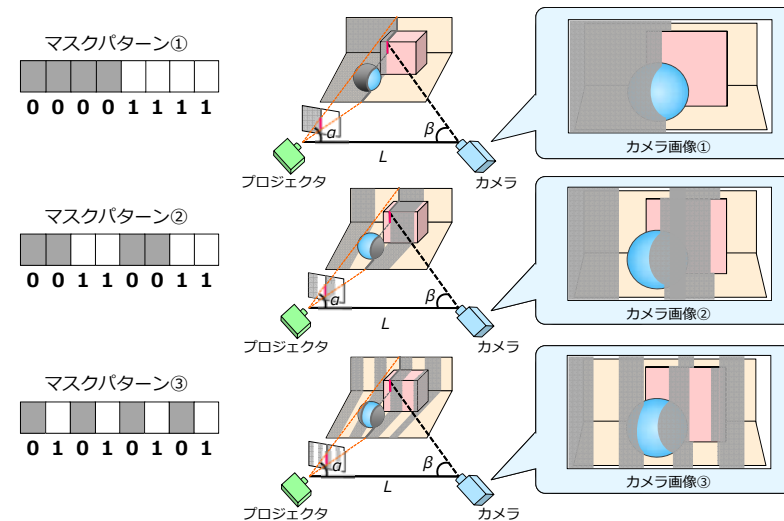
## スリット光投影法

- 対象物に、スポット光ではなく、スリット光を投影する。
- スポット光と比べて、スリット上の多数点を同時計測可能。
- 1回につき、1ラインだけしか計測できない。



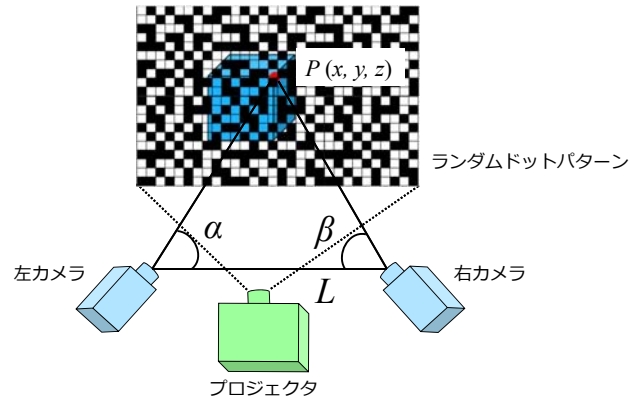
## 空間コード化法

- 2進コードに対応した光パターン（ストライプ）を投光して対象空間を分割
- 三角測量の原理を用いて距離計測



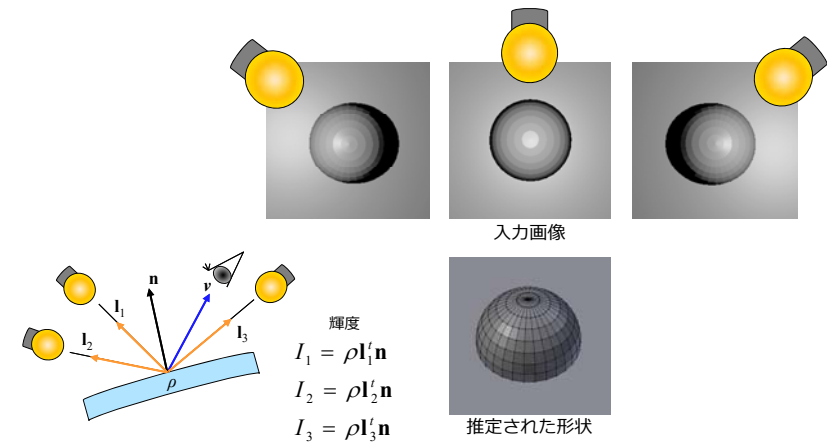
## ランダムドットステレオ

- 対象物にランダムドットパターンを投影。
- 対象物に模様をつけ、ステレオ対応付けの信頼性を向上。
- 図柄のない物体でも、距離計測が可能



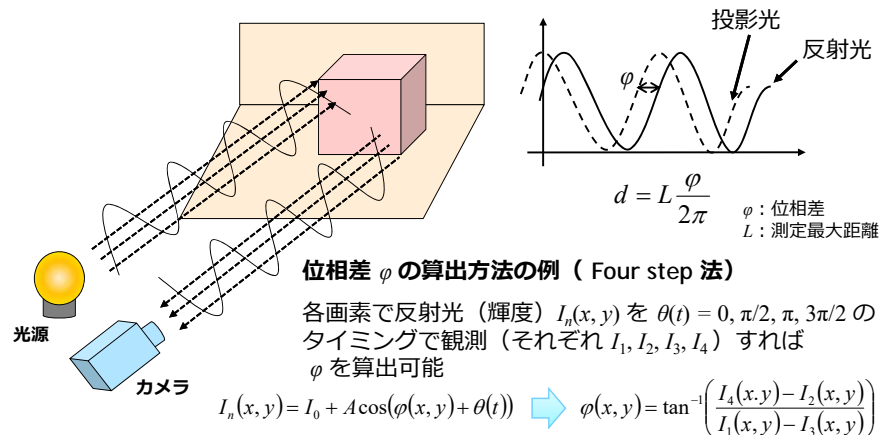
## 照度差ステレオ法

- 3つ以上の光源を一つずつ点灯し、同一視点から画像を撮影する。
- これらの画像の陰影の違いから3次元形状を復元する。
- 1つの固定したカメラを用いるので、対応点探索が不要



## 飛行時間法 (TOF : Time Of Flight)

- 光源から出た光が対象物で反射し、センサに帰還するまでの飛行時間 (位相) と光の速度から被写体までの距離を計測
  - 実際には、投影光と反射光の位相差から時間差を推定する。
  - 死角が生じないなどの利点があり、近年注目されている。



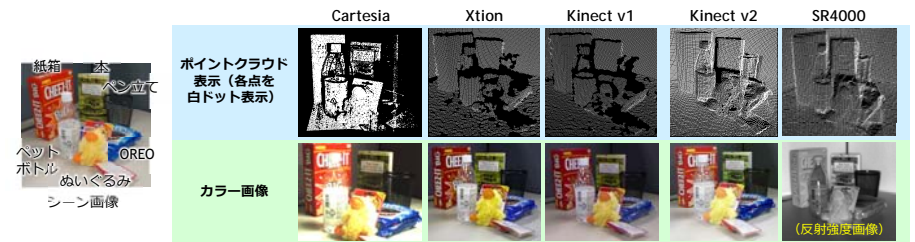
## センサ主要技術と実センサ事例

# 3D センサの分類 (光学センサ)

	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 / X3 (Point Grey 2006 / 2007) TVS (三次元メディア 2011) SV-M-S1 (リコー) R200 (Intel 2016)	光切断法 (スポット光) TDS-A (ノリステック 1997) 光切断法 (スリット光) VIVID9i (KONICA MINOLTA 2004)
	Bumblebee2 TVS SV-M-S1 R200	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) VF0810 (Intel 2016)
同軸測量	マルチベースラインステレオ法 [Okutomi1993] 視体積交差法 [Matsuyama2002]	ランダムドットパターン投光法 [Hashimoto1999] Kinect v1 (Microsoft 2010) Xtion PRO LIVE (ASUS 2011) Leap Motion (2013) 位相シフト法 [Halioua1989], [Zhao1994]
	Shape(depth) from (De) Focus [Hiura1999] Shape from Motion	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) ) 照度差ステレオ [Woodham1980] Kinect v2 SR4000 DepthSense325

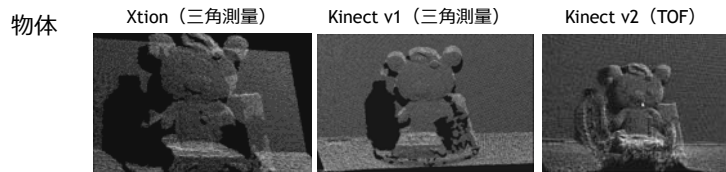
[Woodham1980]: R. J. Woodham, "Photometric method for determining surface orientation", Optical Engineering, Vol. 19, No. 1, pp.139-144, 1980.  
[Posdamer1982]: J. L. Posdamer, et al., "Surface measurement by space-encoded projected beam systems", Computer Graphics and Image Processing, Vol. 18, No. 1, pp.1-17, 1982.  
[Sato1985]: 佐藤ら, "空間コード化による非接触測厚システム", 電子情報通信学会論文誌, Vol. J88-D, No. 3, pp.349-375, 1985.  
[Halioua1989]: M. Halioua, et al., "Optical Three-dimensional Sensing By Phase measuring Profilometry", Optics and Lasers in Engineering, Vol. 11, No. 3, pp.185-215, 1989.  
[Okutomi1993]: M. Okutomi, et al., "A multiple baseline stereo", Trans. IEEE Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, 1993.  
[Zhao1994]: H. Zhao, et al., "Phase unwrapping algorithm for the measurement of three-dimensional objects", Applied Optics, Vol. 33, No. 20, pp.4697-4700, 1994.  
[Hashimoto1999]: 橋本ら, "ランダムドットパターン投光ステレオを用いた物体3次元形状計測システムの開発", 日本ロボット学会誌, Vol. 17, No. 1, pp.48-49, 1999.  
[Hiura1999]: 日高ら, "傾斜面計測用マルチベースラインステレオ画像処理システム", 電子情報通信学会論文誌, Vol. J82-C(1), No. 11, pp.1912-1929, 1999.  
[Matsuyama2002]: 松山ら, "3次元ビデオ映像の撮影・編集・表示", 日本バーチャリアリティ学会論文誌, Vol. 7, No. 4, pp.521-532, 2002.

# 代表的な 3D センサによるデータ取得例と比較

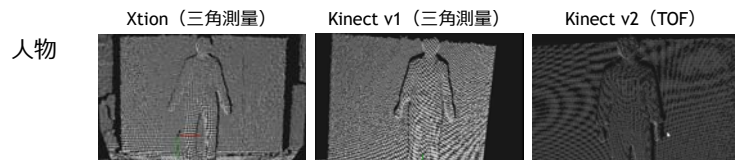


	三角測量			TOF	
	Cartesia 640 x 480 (307200 points) 2 fps	Xtion 640 x 480 (307200 points) 30 fps	Kinect v1 640 x 480 (307200 points) 30 fps	Kinect v2 512 x 424 (217088 points) 30 fps	SR4000 176 x 144 (25344 points) 50 fps
紙箱 (紙)	良好に形状取得が可能				
ペットボトル (透明)	透明部分については輝散光が乱れる (あるいは輝散しない) ため, 形状取得が困難				
本 (光沢紙)	黒色部分は輝散光の強度が不足し, 形状取得が困難	形状取得が可能			
へん立て (金属メッシュ)	メッシュ部分は, 輝散光が乱れるため, 形状取得が困難				
OREO (光沢プラスチック)	光沢面では輝散光が乱れるため, 形状取得が困難	鏡面反射が大きい部分は, 形状取得が困難	形状取得が可能		
ぬいぐるみ (布)	形状取得が可能	箇所によっては輝散光の強度が小さくなり, 形状取得が困難			形状取得が可能
まとめ	5つのセンサの中で精度が最もよいが, 面角が狭く, 黒色部分や透明部分の形状取得が困難。三角測量なので, 対象物の境界面等でデータ欠落がある。		透明部分やメッシュ部分の形状取得ができず, 精度は一着悪かった。三角測量なので, 対象物の境界面等でデータ欠落がある。		Xtion, Kinect v1, SR4000 よりも精度がよい。透明部分, メッシュ部分の形状取得が困難。同軸測量なのでデータ欠落が少ない。

# 代表的な 3D センサによるデータ取得例



物体の左側に計測欠落領域がある。 計測欠落は少ない。



左側に計測欠落領域があるが, 距離が遠いのであまり目立たない。

# ポイントクラウドデータ (3次元点群)

## xyz 3次元座標系におけるデータ点の集合

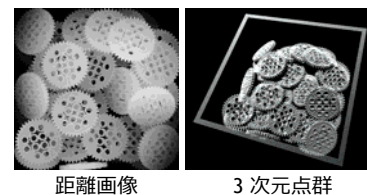
代表的な点群のデータ構造 (.pcd) PCL で定義されたフォーマット

```

# .PCD v0.7 - Point Cloud Data file format+
2 VERSION 0.7+ // pcd ファイルのバージョン
3 FIELDS x y z+ // 次元 (x y z)
4 SIZE 4 4 4+ // 点群のサイズ (4 バイト)
5 TYPE F F F+ // 点群の型 (float)
6 COUNT 1 1 1+ // 要素数
7 WIDTH 307200+ // データの幅
8 HEIGHT 1+ // データの高さ
9 VIEWPOINT 0 0 0 1 0 0 0+ // 視点方向
10 POINTS 307200+ // 点群数 (WIDTH x HEIGHT)
11 DATA ascii+ // データタイプ
12 -0.600881 0.449636 1.000000+
13 -0.598983 0.449636 1.000000+
14 : : :
15 : : :
    
```

307200 行 Kinect v1, Kinect v2, Xtion は, PCL を用いることによって .pcd ファイル形式で保存可。

## 距離画像と点群データ



距離画像 3次元点群

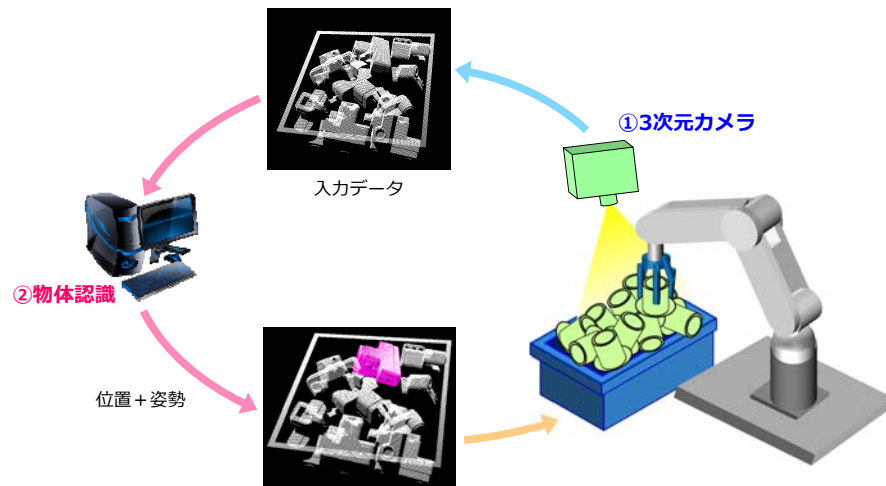
	距離画像	3次元点群
隣接点へのアクセス方法	データが整列されているので隣接データへの連続アクセスが容易。	データが整列されていないので最近傍探索が必要。
デプス解像度	値が濃度値分解能によって離散化される。	一般には離散化されない。実空間データが利用可能。
主な表示ツール	画像 Viewer (Irfan View, Windows フォトビューアーなど)	OpenGL, Mesh Lab, PCL の Viewer など
利用可能なオープンソースライブラリ	OpenCV, OpenGL	Point Cloud Library (PCL), OpenGL

## 物体認識アルゴリズムの概要

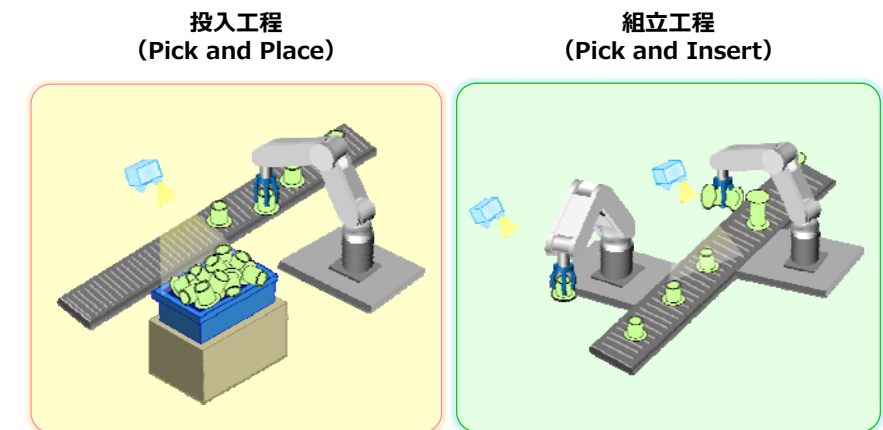
## ロボットピッキング



## ピッキングロボットにおける 3次元物体認識の課題



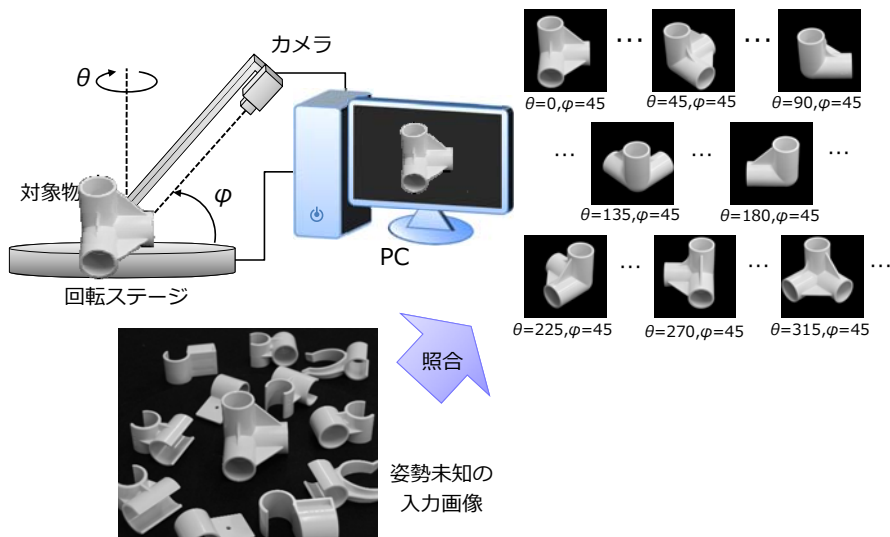
## FA分野における物体認識



## アピアランスベースとモデルベース

## アピアランスベース物体認識の基礎

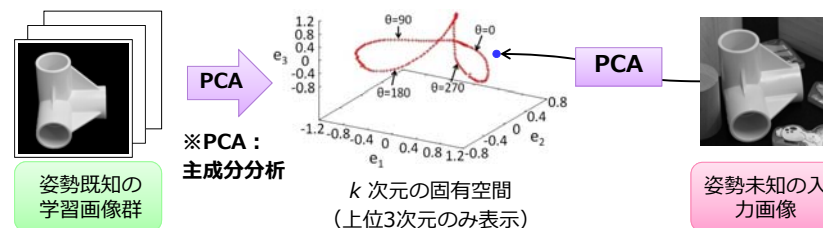
### アピアランスベース物体認識の原理



### パラメトリック固有空間法 (PES法)

[村瀬, 信学論1995]

1. 姿勢既知の画像を, それぞれPCAにより次元圧縮
2. 元の次元数より小さい空間に投影
3. 姿勢の連続性をもとに画像群を (空間内で) つなげて表現  
→多様体 (Manifold)
4. 未知画像を同様に圧縮, 投影して, 最近傍既知画像を同定

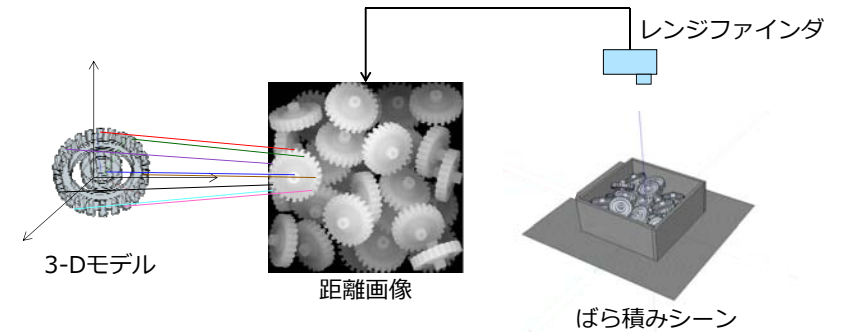




## モデルベース物体認識の基礎

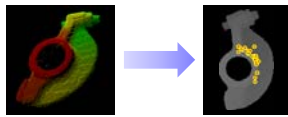
## モデルベース物体認識

モデル： 3次元CADモデルなど  
 入力データ： センサで取得されたポイントクラウドデータなど



## モデルベース物体認識の分類

### 1. キーポイント利用

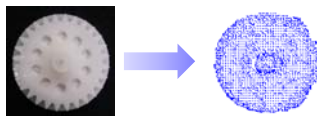


モデル上のキーポイント単位で照合

Spin Image法  
 DAI法  
 Local Surface Patches  
 Point Feature Histogram  
 Point Pair Feature  
 SHOT

[Johnson1997]  
 [Takeguchi2001]  
 [Chen2007]  
 [Rusu2008]  
 [Drost2010]  
 [Tombari2010]

### 2. 非キーポイント利用



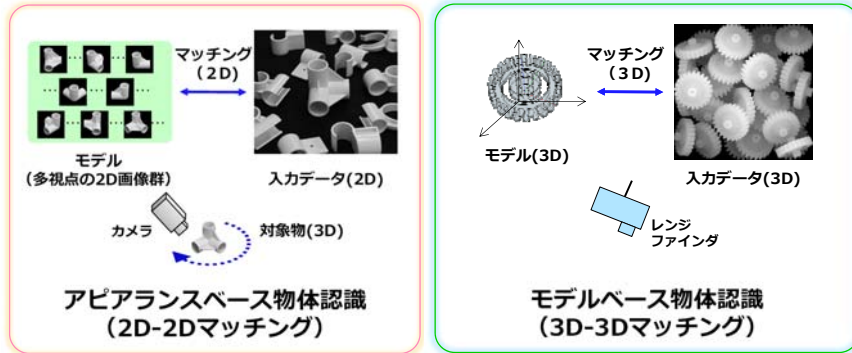
全点、あるいはエッジなど単位で照合

EGI (拡張ガウス像)  
 3DPO  
 3Dエッジ  
 VVV  
 DGI

[Horn1983]  
 [Bolls1986]  
 [Lowe1991]  
 [Sumit1997]  
 [Adan2011]

- かつては距離画像そのものを使う手法が主流.
- 最近は高速化のためにキーポイントベースが増えている.
- 特徴記述法が重要.

## まとめ アピランスベース v.s. モデルベース



物体認識の周辺技術

高精度位置合わせ ICP アルゴリズム

	アピランスベース	モデルベース
手法	入力データを多視点画像群と照合	入力データを3次元モデルと照合
メリット	単眼カメラ利用できる	3次元モデルが入手容易に
課題	多量の辞書データが必要	3次元センサはまだ高価・大型

Iterative Closest Point (ICP)

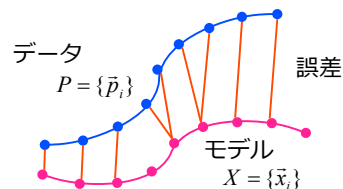
■ データ点群とモデル点群の誤差最小化手法

[Besl1992]  
[Kaneko2003]

■ 方法

- 最近傍点の対応付け
- 対応点からの幾何変換の推定
- 上記2ステップの繰り返しによって、位置合わせ誤差が最小となる剛体変換パラメータを出力

$$\bar{q} = [\bar{q}_R \mid \bar{q}_T]^T$$



$$f(\bar{q}) = \frac{1}{N_p} \sum_{i=1}^{N_p} \|\bar{x}_i - \mathbf{R}(\bar{q}_R)\bar{p}_i - \bar{q}_T\|^2$$

■ 派生手法として、M推定により外れ値にロバストなM-ICP等がある。

2. 3次元局所特徴量を用いた物体認識技術

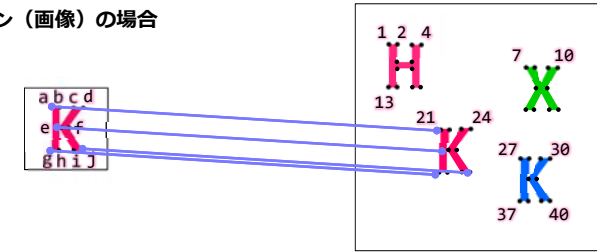
参考文献: P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes", IEEE Trans. PAMI, vol.14, no.2, 1992.  
S.Kaneko, T.Kondo, A.Miyamoto, Robust matching of 3D contours using iterative closest point algorithm improved by M-estimation, PR, vol.36, pp.2041-2047, 2003.



## キーポイントベース物体認識の概要

## キーポイントマッチング (特徴量マッチング)

2次元パターン (画像) の場合

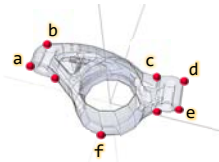


	座標 $i, j$	角度 (始・終)	色 (RGB)		座標 $i, j$	角度 (始・終)	色 (RGB)
a	32, 25	0, 90	255, 0, 0	→	1	55, 71	0, 90 255, 0, 0
b	37, 25	90, 180	255, 0, 0	→	2	60, 71	90, 180 255, 0, 0
:	:	:	:		:	:	:
d	80, 25	135, 180	255, 0, 0	→	4	103, 71	90, 180 255, 0, 0
:	:	:	:		:	:	:
g	32, 85	270, 0	255, 0, 0	→	21	125, 220	0, 90 255, 0, 0
j	80, 85	180, 225	255, 0, 0	→	24	173, 220	135, 180 255, 0, 0
:	:	:	:		:	:	:
				↳	27	251, 265	0, 0, 255
				↳	30	299, 265	135, 180 0, 0, 255
				↳	:	:	:

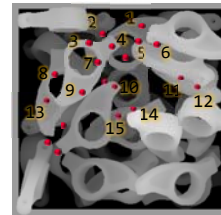
画素を特徴付ける数値  
= 特徴量 (ベクトル)

## キーポイントマッチングの基本 (特徴量マッチング)

### 3次元データ同士のマッチング



3次元モデル



入力シーン

座標 (x,y,z)	特徴量 (ベクトル)		座標 (x,y,z)	特徴量 (ベクトル)
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, ...	→	10 101.6, 35.3, 71.8	5, 2, 6, ..., 9, 20, ...
:	:		:	:
g 11.5, 9.5, 234.1	50, 32, 4, ..., 3, 13, ...	→	21 25.1, 220.9, 5.1	99, 49, 3, ..., 20, 4, ...
j 8.9, 794.4, 212.3	1, 95, 34, ..., 13, 5, ...	→	24 734.7, 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

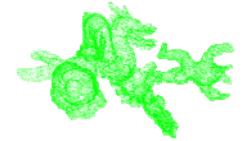
## 実験① キーポイント数と認識性能の関係

### 【実験方法】

実験データ: Stanford models の "Dragon" で構成

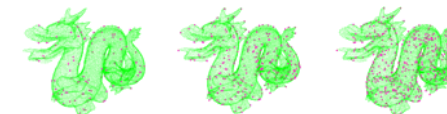
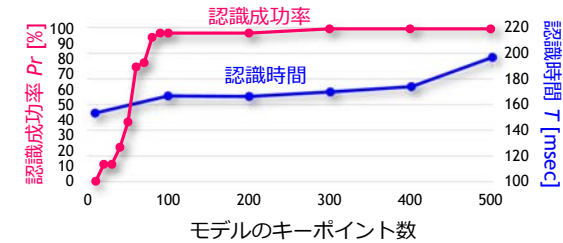
キーポイント: SHORT 法によって検出された点からランダムに選択

S. Takei, S. Akizuki, M. Hashimoto, "SHORT: A Fast 3D Feature Description based on Estimating Occupancy in Spherical Shell Regions". IVCNZ, 2015.



入力データ例

### 【実験結果】



キーポイント数が性能を左右する。

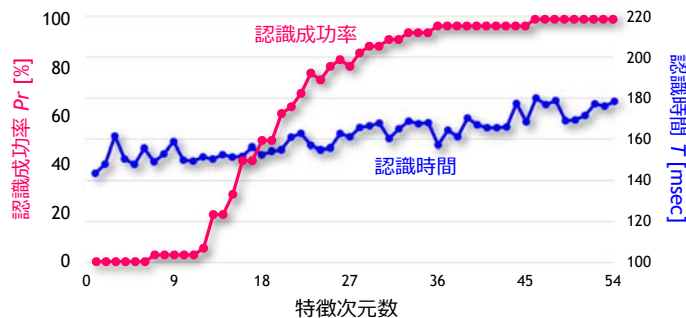
## 実験② 特徴量の次元数と認識性能

### 【実験方法】

実験データ：Stanford models の“Dragon”で構成

特徴量：SHORT 特徴量 54 次元（6 シェル x 9 ビンのヒストグラム）からランダムに要素を抽出  
 （特徴量の次元数 = ランダム抽出数）

### 【実験結果】



特徴量の次元数が性能を左右する。

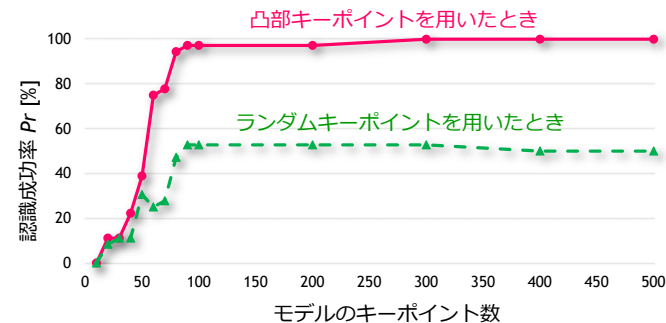
## 実験③ キーポイント検出方法と認識成功率の関係

### 【実験方法】

SHORT 法：凸形状の点だけをキーポイントとして検出

ランダム法：ランダムにキーポイントを検出

### 【実験結果】

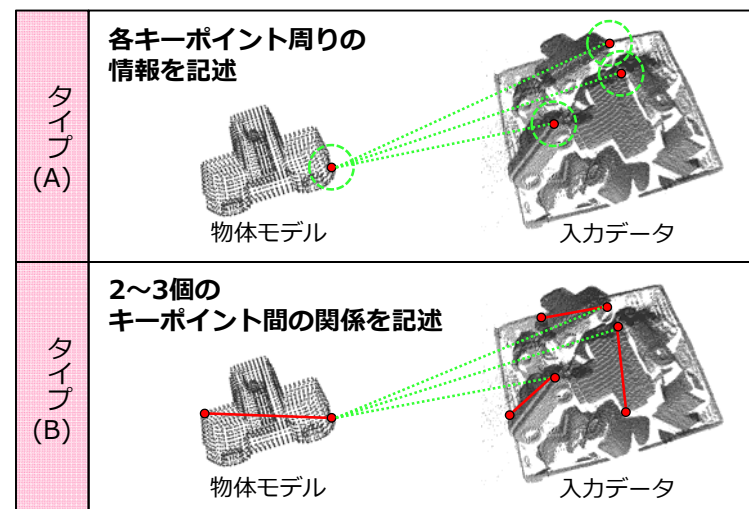


キーポイントの選び方が性能を左右する。

## 3次元特徴量の主要技術

## 2つのタイプの3次元特徴量

キーポイントにどのような情報（特徴量）を付加するか？



特徴量の記述方法			
(A) キーポイント周りの情報記述		(B) 複数点間の関係記述	
座標データ利用	法線ベクトル利用	その他の情報を利用	2点間 or 3点間
SHOT, PFH, CSHOT, B-SHOT, HONV		LSP, CCDoN, <u>NARF</u> , MHOG, <u>RoPS</u>	PPF, ER, VC-PPF, MPPF, PPF B2B or S2B or L2L, VPM
<u>SI</u> , SSI, DAI, <u>3DSC</u> , <u>SDLSD</u> , <u>USC</u>	<u>FPFH</u> , <u>DoN</u>		

PFH : Point Feature Histogram  
 SHOT : Signature of Histograms of Orientations  
 CSHOT : Color SHOT  
 B-SHOT: Binary SHOT  
 HONV : Histogram of Oriented Normal Vector  
 SI : Spin Image  
 SSI : Spherical Spin Image  
 DAI : Depth Aspect Image  
 3DSC : 3D Shape Contexts  
 USC : Unique Shape Context

SDLSD : Scale-Dependent Local Shape Descriptor  
 FPFH : Fast PFH  
 DoN : Difference of Normals  
 LSP : Local Surface Patches  
 CCDoN : Combination of Curvatures and Difference of Normals  
 NARF : Normal Aligned Radial Feature  
 MHOG : Mesh HOG  
 RoPS : Rotational Projection Statistics

PPF : Point Pair Feature  
 ER : Efficient Ransac  
 VC-PPF : Visibility Context PPF  
 MPPF : Multimodal PPF  
 PPF B2B or S2B or L2L : PPF Boundary-to-Boundary or Surface to Boundary or Line to Line  
 VPM : Vector Pair Matching

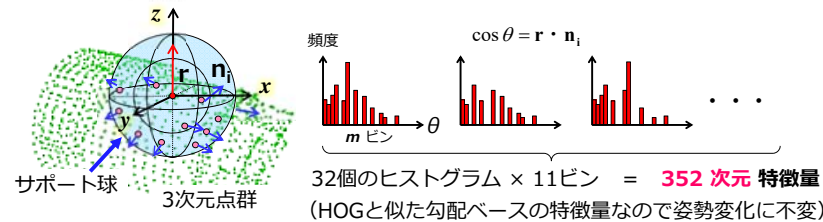
※下線は PCL に登録されている手法

## タイプ (A) キーポイント周りの情報記述

## SHOT 特徴量 (Signature of Histograms of Orientations)

### 特徴記述

- キーポイント周辺 (サポート球内) を分割する (局所座標系を利用) .  
 xy 平面で 2 分割, 球内を中心部と周辺部に 2 分割, さらに,  
 z 軸まわり 8 分割. (→  $2 \times 2 \times 8 = 32$  分割)
- 基準点の法線  $r$  と, 分割された 32 個のスペース内の法線ベクトル群  $n_i$   
 との内積値を計算し,  $m$  ビンのヒストグラムを作成 ( $\cos\theta$ 値). ( $m=11$ )



### 利点

- 次元数が高いので, 特徴量のアイデンティティは比較的高い.  
 (→誤照合は少なめ)

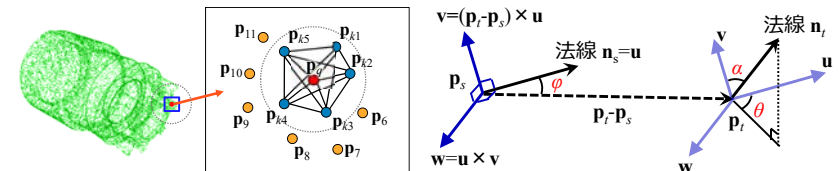
参考文献 : F. Tombari, S.Satti, L.D.Stefano, "Unique Signatures of Histograms for Local Surface Description", ECCV, pp.356-369, 2010.

## PFH 特徴量 (Point Feature Histograms)

### 特徴記述

- キーポイント周りの球領域内の  $k$  近傍点から**多数の 2点組**を選択.
- 2点から  $\alpha, \theta, \varphi, p_i-p_j$  を計算し, **ヒストグラム化 (=特徴量)**

各パラメータを離散化し, 125次元で表現 (PCL)



### キーポイントの選択方法

- ある特徴量が, モデル内の平均特徴量より大きければ候補として保存.
- 球の半径を変えながら同様の計算をおこない, 多くの半径から支持された候補点を, 最終的な **キーポイント** として選択する.

### 利点

- ヒストグラム特徴なので, 多少の外乱には頑健.
- 次元数がある程度高いので, モデルの表現力も高い.

参考文献 : R. B. Rusu, et al., "Aligning Point Cloud Views using Persistent Feature Histograms", IEEE Proc. IROS, pp.3384-3391, 2008.

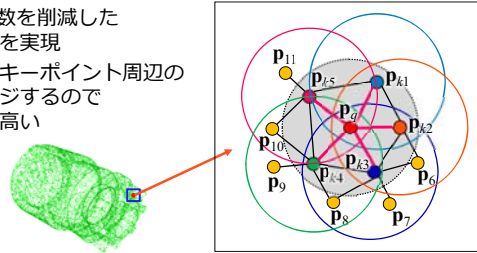
## タイプ (B) 複数点間の関係記述

## FPFH 特徴量 (Fast Point Feature Histograms)

- PFH の改良版 (特徴記述の組み合わせ数を削減)
- 2点組の選択を **キーポイント** と **近傍点群** の組み合わせに変更
  1. キーポイントと、球領域内の近傍点から PFH を作成  
各点が PFH を保持
  2. キーポイントと近傍点を持つ PFH をマージ (ヒストグラムの同一ビンを合成) することによって、最終的なヒストグラム (FPFH) を作成  
図では、 $p_q$  と灰色の領域内の近傍点  $p_{ki}$  の PFH をマージ  
**33次元で表現 (PCL)**

### 利点

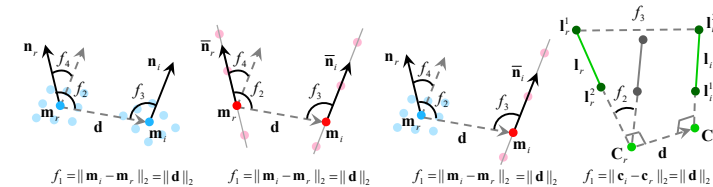
- 2点組の組み合わせ数を削減したことによって高速化を実現
- 次元数が低めだが、キーポイント周辺のヒストグラムをマージするので局所形状の表現力が高い



参考文献: R. B. Rusu, et al., "Fast Point Feature Histograms (FPFH) for 3D Registration", IEEE Proc. ICRA, pp.3212-3217, 2009.

## PPF 特徴量の改良 B2B, S2B, L2L

- PPF においてペアリングする点を限定することによって高速化
    - B2B: Boundary 上の点同士の組合せ
    - S2B: Surface と Boundary 上の点の組合せ
    - L2L: Line Segment 同士の組合せ
- ※論文においては B2B が最も高性能



S2S [Drost2010]

B2B

S2B

L2L



青点: Surface

赤点: Boundary

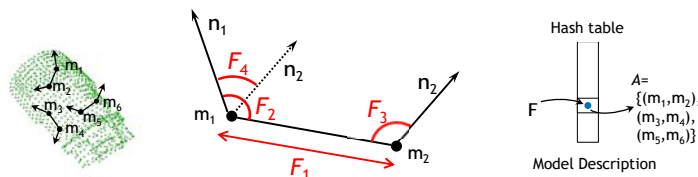
緑線: Line Segment

出典: C. Choi, et al., "Voting-Based Pose Estimation for Robotic Assembly Using a 3D Sensor", IEEE Proc. ICRA, pp.1724-1731, 2012.

## PPF 特徴量 (Point Pair Feature)

### 特徴記述

- 2点対から算出された4次元特徴量 ( $F_1 \sim F_4$  の4つの値)
- 物体上の **全ての点から** 構成される2点対を利用する.



### マッチング

- 4次元特徴量が類似する PPF を探索。  
(ハッシュテーブルを活用して効率化)
- それらの幾何変換パラメータを計算する.

### 利点

- 部分的なマッチング結果を集積し、多数決処理を適用するので、部分的な隠れに頑健.

参考文献: B. Drost, et al., "Model Globally, Match Locally: Efficient and Robust 3D Object Recognition", CVPR, pp.998-1005, 2010.

- (A) キーポイント周りの情報記述
- (B) 複数点間の関係記述
- ……どちらのタイプがよいか？

## SHOT と PPF の比較

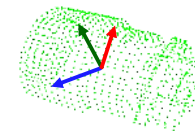
	タイプ (A) SHOT (キーポイント周辺の情報記述)	タイプ (B) PPF (複数キーポイント間の関係)
特徴量の次元数	352	4
特徴量のアイデンティティ	特徴ベクトルの次数が高い (アイデンティティが高い) ので、誤照合は少ない。	単体ではSHOTよりもアイデンティティが低い、投票処理 (統計処理) によって解決している例が多い。
外乱への耐性 = 再現性 (Repeatability)	ヒストグラムベースの特徴量なのでノイズには強い。	物体全面にばらまかれた多数の点対 (全点からの2点組み合わせ) を用いるので、オクルージョンには強い傾向あり。
照合に使用する特徴点数	特徴量の次元数が高いので、一般には少数で足りる。 →少数しか使わないので、オクルージョンには弱い。	多数のPPFを使う必要があるので、高速化のためには、マッチング時の工夫が重要 (ハッシュテーブル利用等)。
処理時間	特徴量次元と特徴点数の両方が関連しているので、一概に言えない。ばら積み物体認識の場合は、特徴量単体の ID が低い PPF は多数のマッチング候補が発生するので、時間がかかることがある。	

## 局所参照座標系の主要技術

Local Reference Frame

## 局所参照座標系 (LRF : Local Reference Frame)

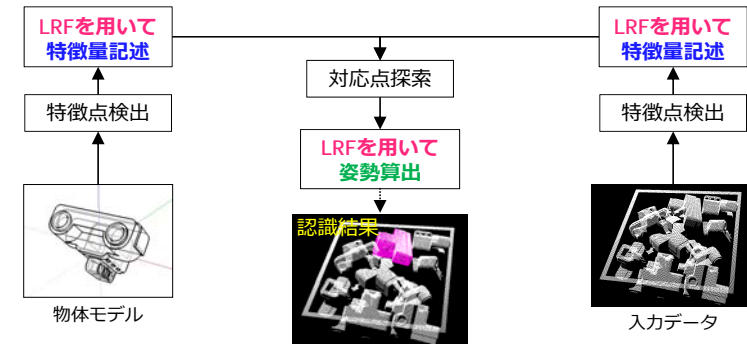
- LRF とは？ 特徴点ごとに設定される 3 次元 (直交) 座標系



LRFの決め方 (例)

- 第1軸： キーポイントを代表する法線ベクトル
- 第2軸： 第1軸と直交するベクトル
- 第3軸： 第1軸、第2軸の外積ベクトル

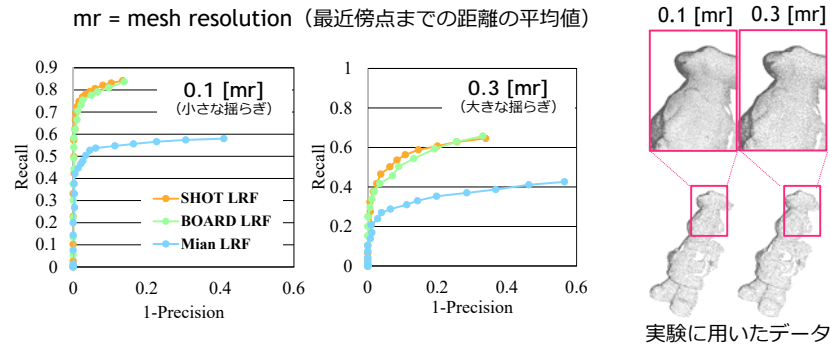
- 物体認識における LRF の必要性



## LRF が認識性能に与える影響 (実験例)

- 同一の特徴量に、さまざまな LRF を組み合わせて性能を評価。

使用特徴量: SHOT  
 マッチング対象: ガウシアンノイズを付加した点群  
 ノイズ付加方式: 元の点の xyz 座標位置に揺らぎを与える (点数は不変)  
 揺らぎの大きさ: 0.1 [mr] (揺らぎ小), 0.3 [mr] (揺らぎ大) の2種



使用する特徴量が同じでも、LRF によって性能が左右される。

## 代表的な LRF

一括算出型	個別算出型	
キーポイント周りの点群の座標から共分散行列を計算し、その固有ベクトル群を LRF とする。	3軸をそれぞれ独立に算出する。 z 軸 = キーポイント周りの法線ベクトル x 軸 = さまざまな工夫	
Mian, SHOT RoPS, EM, TriSI	y 軸の決定方法	
	最大値選択	方向分布分析
	BOARD, PS Petrelli-LRF	MH DPN, GRF

Mian : Mian の LRF  
 SHOT : Signature of Histograms of Orientations  
 RoPS : Rotational Projection Statistics  
 EM : Exponential Map  
 TriSI : Tri-Spin-Image

BOARD : BOrder Aware Repeatable Directions  
 PS : Point Signature  
 Petrelli-LRF : Petrelli 氏による LRF (2012年)  
 MH : Mesh HoG  
 DPN : Dominant Projected Normal  
 GRF : Global Reference Frame

※下線は PCL に実装されている手法

## 一括算出型 LRF

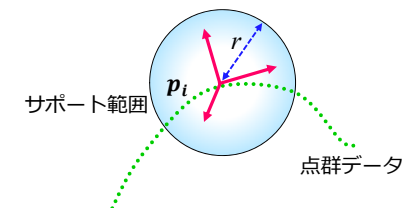
## Mian LRF

[Mian2010]

半径  $r$  の球領域内の点群の共分散行列の固有ベクトル群 (3次元)

$$\text{共分散行列 } \Sigma_{\hat{p}} = \frac{1}{k} \sum_{i=0}^k (p_i - \hat{p})(p_i - \hat{p})^T \quad (1)$$

$$\text{平均ベクトル } \hat{p} = \frac{1}{k} \sum_{i=0}^k p_i \quad (2)$$

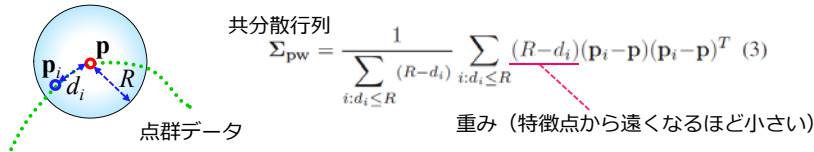


参考文献: A.Mian, et al., "On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes", International Journal of Computer Vision (IJCV), Vol.89, Issue 2-3, pp.348-361, 2010.

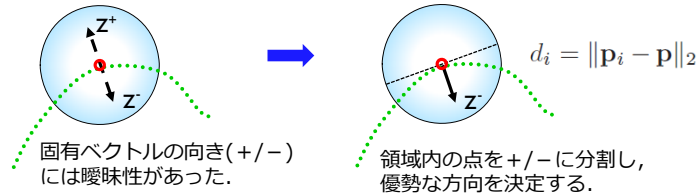


- Mian LRFの改良版.
- 注目点からの距離に応じた**重みつき共分散行列**の固有ベクトル

1. 周辺部外乱の影響を抑制



2. 固有ベクトルの向きの曖昧性を解決

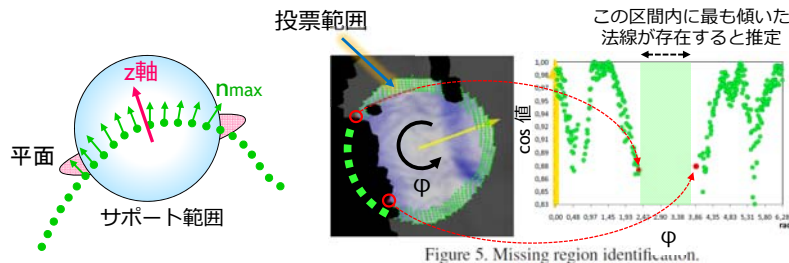


参考文献: F.Tombari, S.Satti, L.D.Stefano, Unique Signatures of Histograms for Local Surface Description, ECCV, pp.356-369, 2010.

個別算出型 LRF

BOARD (BOOrder AwaRe Repeatable Directions) LRF

- z 軸の決め方
  - サポート範囲内の点群を**平面フィッティング**したときの法線が z 軸.
- x 軸の決め方
  - サポート球内の各点の法線ベクトル群の中から, z 軸との内積 (cosθ) が最も小さい (=大きく傾いた=直角に近い) もの (nmax) を決定し, これを x 軸とする.
  - このとき, 欠損領域の有無を判定し, cosθの最小値がその領域内に存在する場合は, 両脇のデータをもとに, 最小位置を推定する.



参考文献: A. Petrelli, L. D. Stefano, On the Repeatability of the Local Reference Frame for Partial Shape Matching, ICCV, pp.2244-2251, 2011.

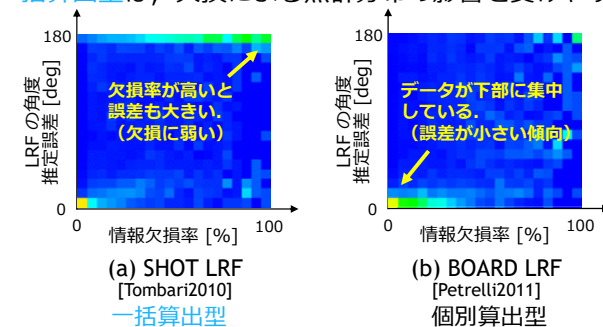
SHOT-LRF と BOARD-LRF の比較 (実験例)

- 点群データに欠損を混入 (オクルージョンを模擬)
- 欠損率とLRF の角度推定誤差との関係を調査

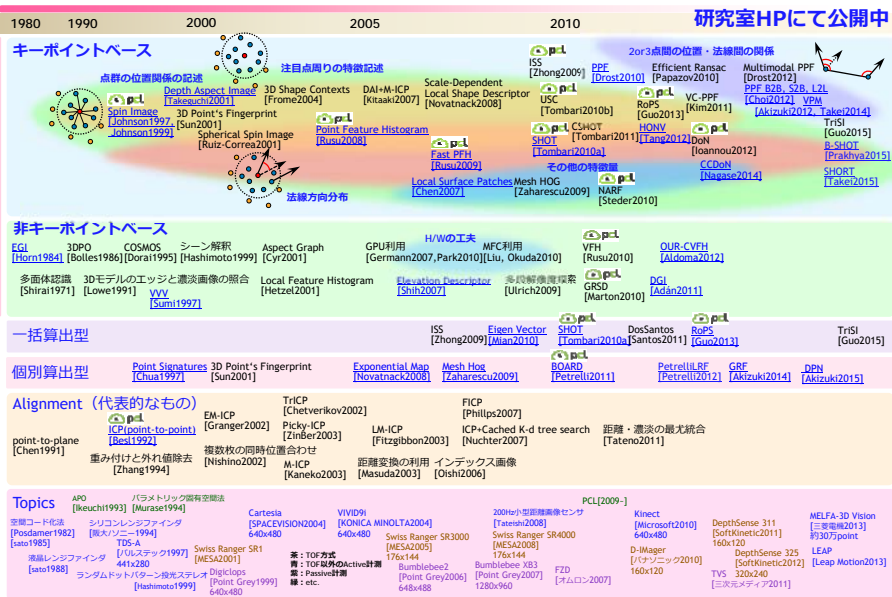
■ カラーマップの見かた

- 縦軸: 対応点間の LRF の角度誤差 [deg]
- 横軸: 対応点間の LRF 算出に用いた点群の欠損率 欠損率:  $\frac{|N_M - N_S|}{\max(N_M, N_S)}$
- 色は頻度を表す

■ 一括算出型は, 欠損による点群分布の影響を受けやすい.



# 3次元特徴量とその周辺技術に関する研究マップ



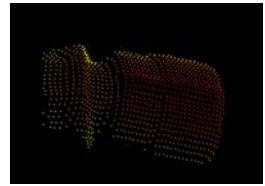
## 3次元特徴量に関する最新技術

## ベクトルペア特徴量

## 従来手法の問題点

- Johnson, et al., "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes", PAMI, 1999.
    - モデル全点から作成した視点変化に不変な照合データを利用
- 問題点** 照合に用いるデータ量が多いので、処理時間が膨大
- 
- Chen, et al., "3D Free-form Object Recognition in Range Images using Local Surface Patches", PRL, 2007.
    - 曲率の大きな微小形状をもとに、視点変化に不変な特徴データを作成
- 問題点** 曲率の大きな点が照合に有効とは限らず、類似した曲率値をもつ点と誤照合

## 基本アイデア

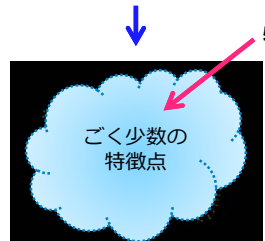


物体モデル  
(3-D point representation)

1. 少数の特徴点を用いる。(例: 全点の1%)  
高速化に寄与する。

重要なこと...

どのような3次元特徴点を選択されるべきか?



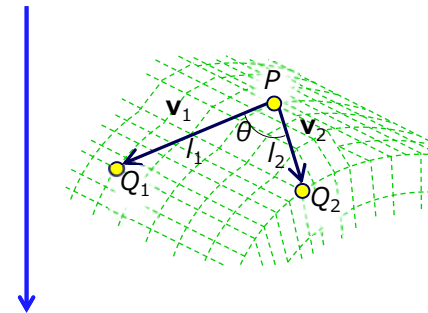
マッチングデータ

特徴点

2. 本研究では、共起発生確率が低い点を選択する。周囲には、これと似た形状が存在しないので、誤認識のリスクが低い。

## 3次元ベクトルペア

- 3つの3次元点
  - 3次元姿勢パラメータを決定するためには、3つの3次元点があればよい。

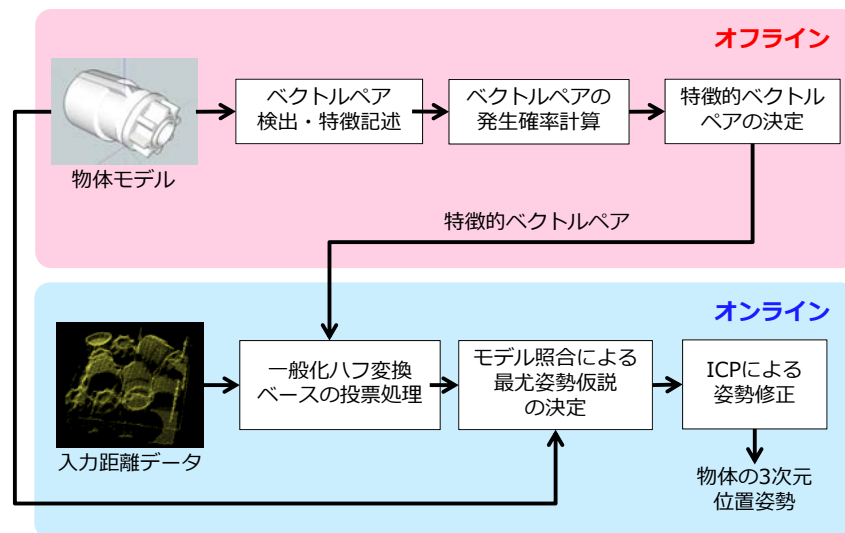


幾何学的パラメータ  
 $l_1 = |\mathbf{v}_1| = |PQ_1|$   
 $l_2 = |\mathbf{v}_2| = |PQ_2|$   
 $\theta$

Shape description value  
 $S_p, S_{q1}, S_{q2}$

- 視点を共有する2本のベクトル = 3次元ベクトルペア
  - この表現により、幾何学的な表現がシンプルになる。

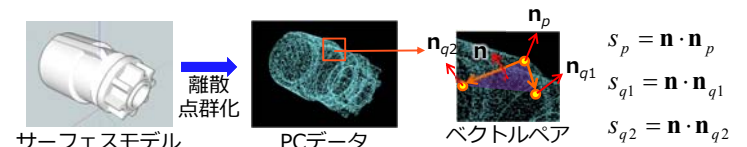
## ベクトルペアマッチング法：処理の流れ



## 特徴的3-Dベクトルペアの抽出

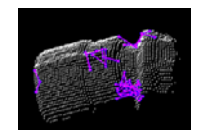
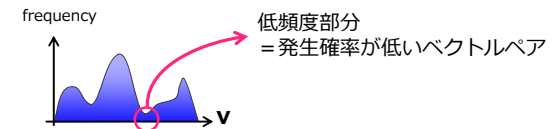
- 物体モデルにおいて「発生確率が低いベクトルペア」→特徴的
- ベクトルペアの特徴量 $\mathbf{v}=(s_p, s_{q1}, s_{q2})$ の発生確率を計算

1. 法線分布を用いた特徴量の計算
  - 三角パッチの法線と、各端点の内積値



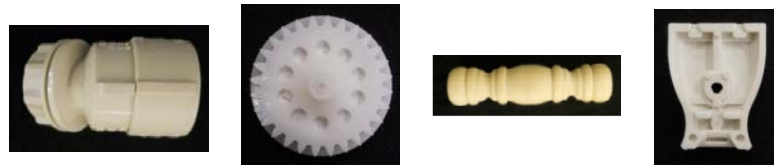
2. 特徴量の発生確率分布としての物体モデル表現

- 物体内全てのベクトルペアを $\mathbf{v}$ の各値を軸とする頻度分布化

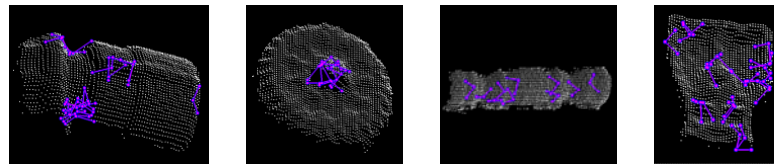


## 選択された特徴的ベクトルペア

- 4種類の物体からベクトルペアを選択

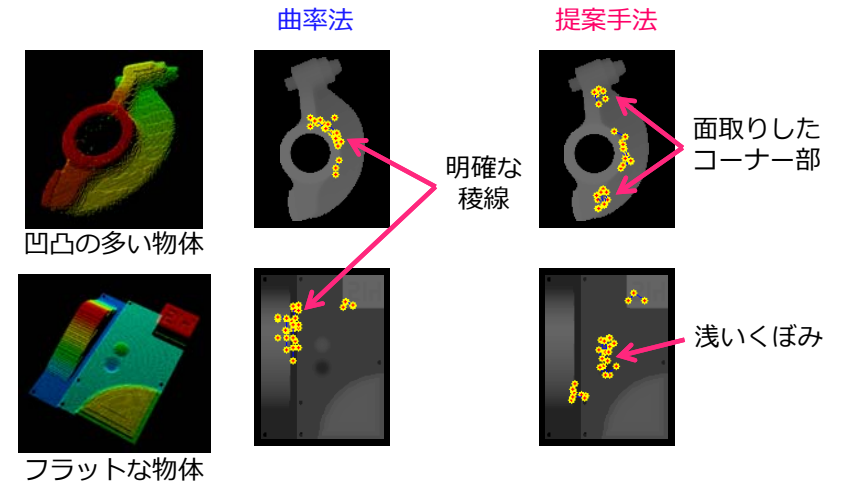


物体モデル外観



選択された特徴的ベクトルペア

## 特徴抽出結果の例： 曲率法 v.s. 提案手法

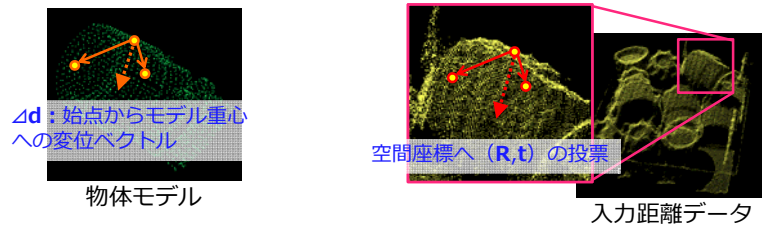


3次元ベクトルペアは独自性の高い形状から抽出される。

**この3次元ベクトルペアを照合に用いる。**

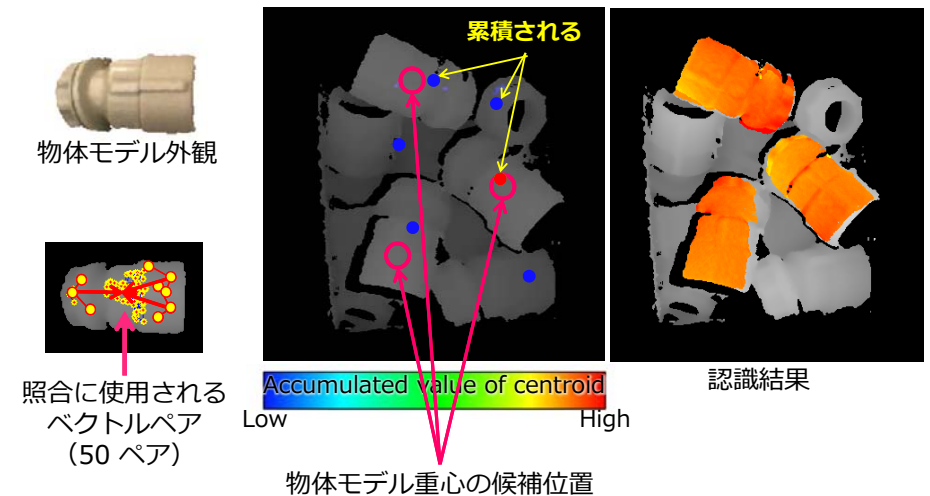
## マッチングモジュール

- 一般化Hough変換ベースの投票型手法



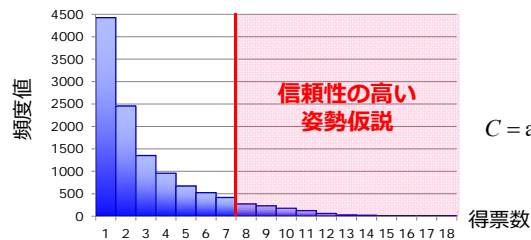
- モデル・シーン間のベクトルペアの類似度を算出
  - 形状評価値( $s_p, s_{q1}, s_{q2}$ )のノルム計算
- 姿勢仮説( $R, t$ )を算出
- 入力距離データ空間へ投票

## 3次元ベクトルペアマッチング



## 姿勢仮説の検証

- 投票数が上位の姿勢仮説について、モデル照合によって検証
  - 投票数に関する頻度分布を作成、上位  $m\%$  の得票数の姿勢仮説と入力距離データの整合性をチェック
    - モデル・シーン間の誤差の絶対値和を最小化する  $(R, t)$  を算出



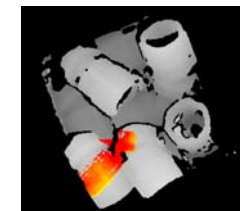
$$C = \arg \min_{(R, t)} \left\{ \frac{1}{M} \sum_{m=0}^M \|Rp_m + t - q\| \right\}$$

- 最も整合性の高い姿勢仮説にICPアルゴリズムを適用
  - ※ICPアルゴリズム・・・3次元データ点群同士を高精度に位置合わせする手法

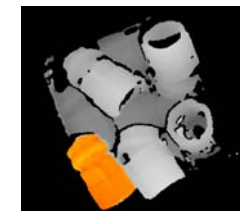
## 実シーンに対する認識性能

N: 抽出されたベクトルペア数  
Pr: 認識成功率  
T: 処理時間

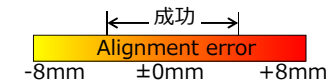
手法	特徴量	N	Pr	T(sec)
Spin Image法	Spin maps	-	93.7%	20.68
ランダム法	Vector pairs	10	47.6%	8.66
曲率法	Vector pairs	10	77.0%	0.74
提案手法	Vector pairs	10	81.7%	0.21
		40	92.9%	1.19



Spin Image法による結果 (失敗)

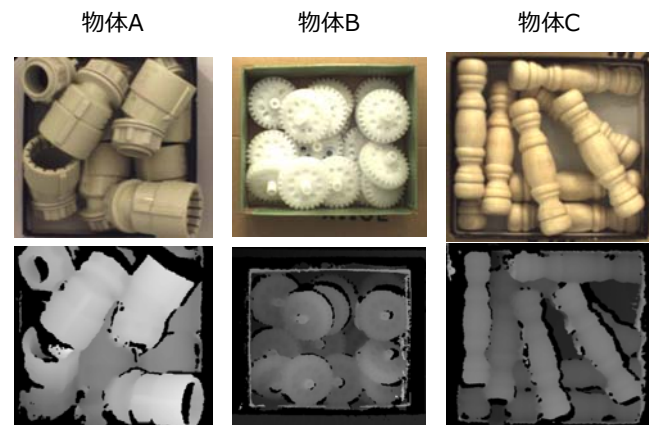


提案手法による結果 (成功)



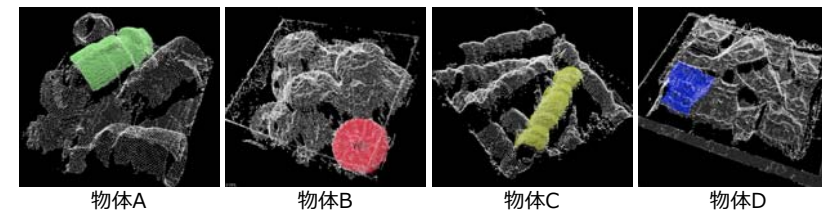
## 多様な形状の物体に対する認識結果

- 実験に用いた画像



## 多様な形状の物体に対する認識結果

- 認識結果例と位置合わせ誤差



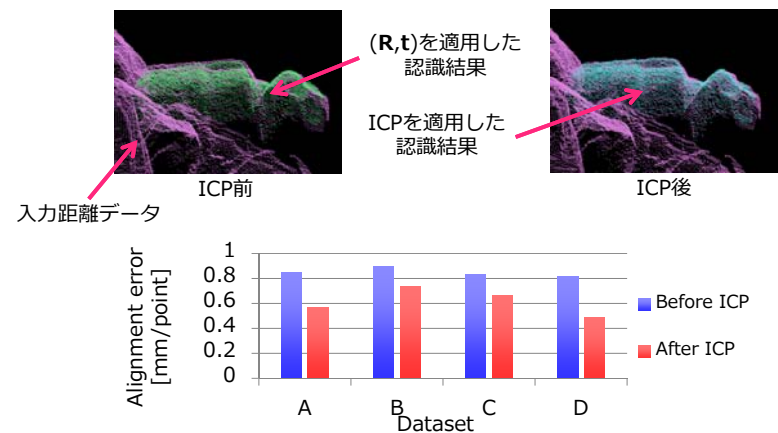
		A	B	C	D
Spin Image	Pr(%)	70.6	50.8	35.2	49.6
	T(sec)	20.68	55.34	31.21	24.95
Correspondence Grouping法	Pr(%)	84.9	96.2	65.6	74.4
	T(sec)	26.10	52.85	29.26	45.41
提案手法	Pr(%)	95.2	100.0	100.0	95.3
	T(sec)	0.95	3.17	5.12	2.64

Pr: 認識率, T: 処理時間

入力距離データは各物体約140例  
CPU: Intel® CORE™ i7, RAM: 4GB

提案手法は他手法と比べて、認識率、処理時間の両面において優位

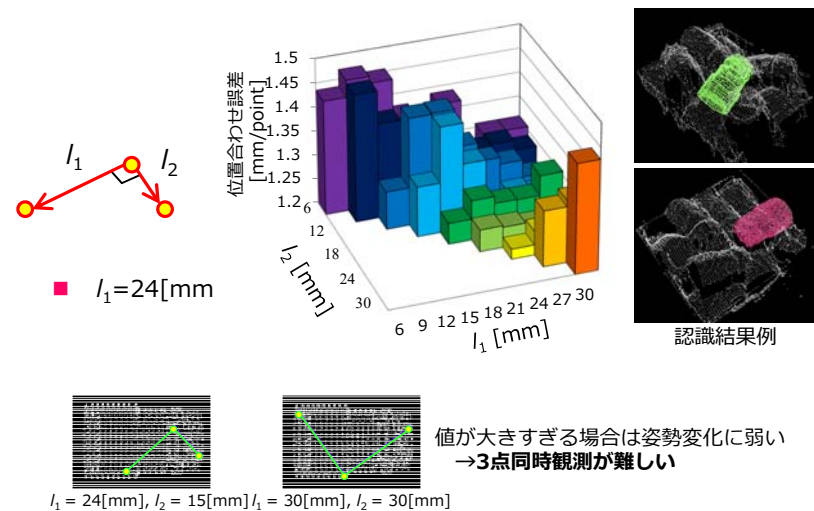
## 位置合わせ精度の評価



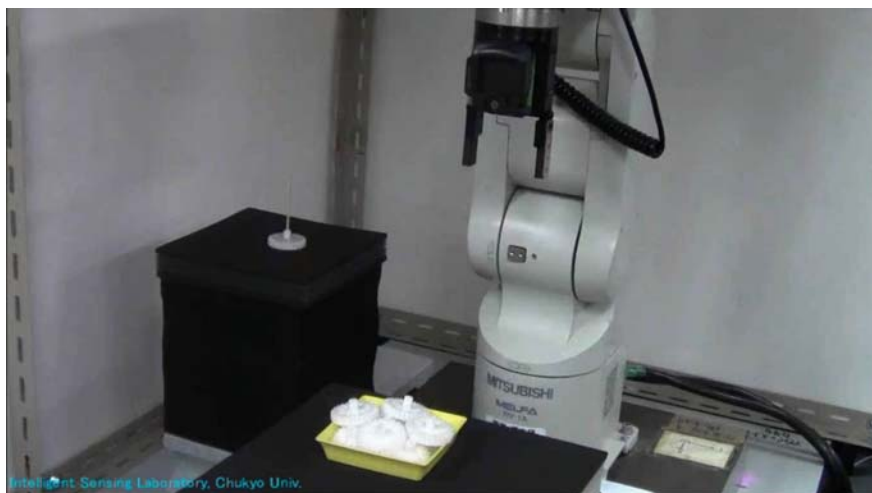
- レンジファインダの空間分解能 = 0.4[mm]
- 精度の高い位置合わせが実現

## 各ベクトルの長さとの認識精度の関係

- $l_1, l_2$ を変化させながら位置合わせ誤差を計測



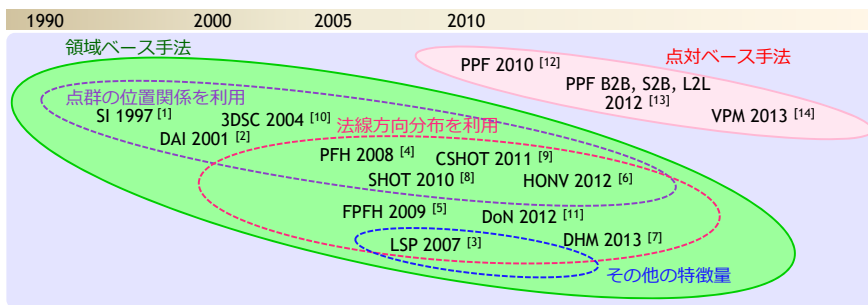
## VPMを用いたピンピックシステム



## CCDoN特徴量

～ 低次元・高アイデンティティ特徴量 ～

## 特徴量計算手法の長所・短所



	領域ベース手法	点对ベース手法
手法	SI <sup>[1]</sup> , DAI <sup>[2]</sup> , LSP <sup>[3]</sup> , PFH <sup>[4]</sup> , FPFH <sup>[5]</sup> , HONV <sup>[6]</sup> , DHM <sup>[7]</sup> , SHOT <sup>[8]</sup> , CSHOT <sup>[9]</sup> , 3DSC <sup>[10]</sup> , DoN <sup>[11]</sup>	PPF <sup>[12]</sup> , PPF 派生 <sup>[13]</sup> (B2B, S2B, L2L), VPM <sup>[14]</sup>
処理速度	遅い	速い
信頼性	高い	低い

[1] Johnson et al. 1999. [2] 武口 et al. 2004. [3] Chen et al. 2007. [4] Rusu et al. 2008. [5] Rusu et al. 2009. [6] Tang et al. 2012. [7] 永瀬 et al. [8] Tombari et al. 2010. [9] Tombari et al. 2011. [10] Frome et al. 2004. [11] Ioannou et al. 2012. [12] Drost et al. 2010. [13] Choi et al. 2012. [14] Akizuki et al. 2013.

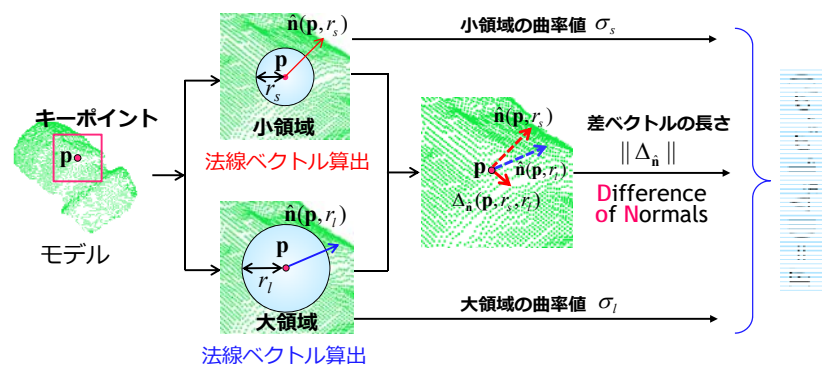
## 本研究の目的と基本アイデア

- 目的
  - 高速性と高信頼性を両立する新たな特徴量を提案
- 基本アイデア
  - 特徴点周りの領域内に含まれるデータを利用して低次元の特徴量を記述
    - 低次元特徴の利用 -> 処理の高速化
    - 特徴点周りのデータを利用 -> 認識の高信頼化
- Difference of Normals (DoN) 特徴量
  - スカラー値 -> 高速な処理が可能
  - 相対的な値 -> 誤照合の増加

DoN 特徴量に少数の特徴量を組み合わせることによって、特徴量の信頼性を高め、誤照合の低減を図る

## 基本アイデア

### Combination of Curvatures and Difference of Normals



$$F = (\|\Delta_{\hat{n}}\|, \sigma_s, \sigma_l)$$

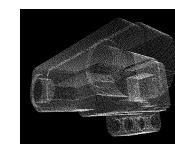
次元数は 3

局所形状を表現 (小領域)  
大域形状を表現 (大領域)  
法線間の相対値

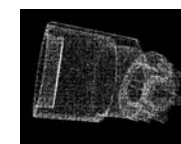
➢ Y. Ioannou, B. Taati, R. Harrap, and M. Greenspan, "Difference of Normals as a Multi-Scale Operator in Unorganized Point Clouds," International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, pp.501-508, 2012.  
➢ 永瀬誠信, 秋月秀一, 橋本学, "CCDoN: ばら積み部品の高速・高信頼な6自由度位置姿勢推定のための局所特徴量", 精密工学会誌, Vol.80, No.12, pp.1138-1143, 2014.

## 実験に用いたシミュレーションデータ

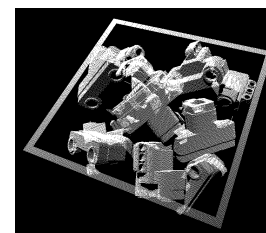
- 各物体それぞれ約 115 シーンのデータを用いた
- 点群データは 1 点あたり 0.40 mm ピッチの離散化点群



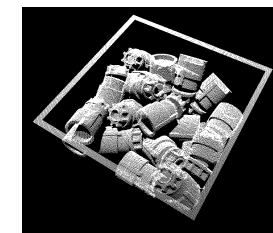
物体 A



物体 B

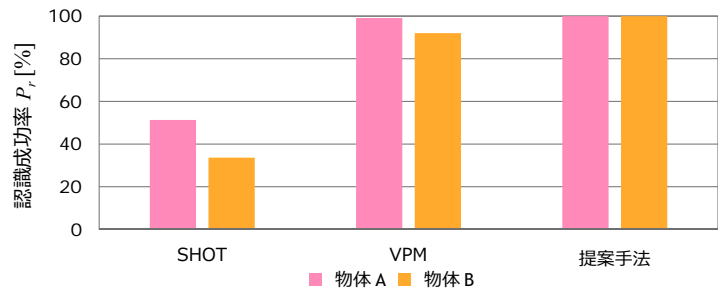
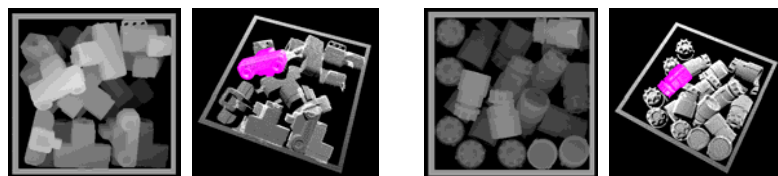


物体 A のシミュレーションシーンデータ



物体 B のシミュレーションシーンデータ

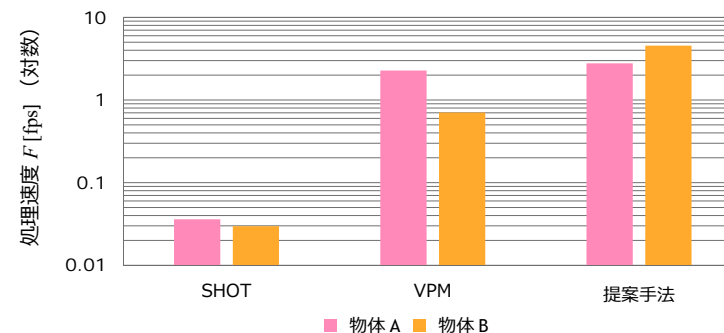
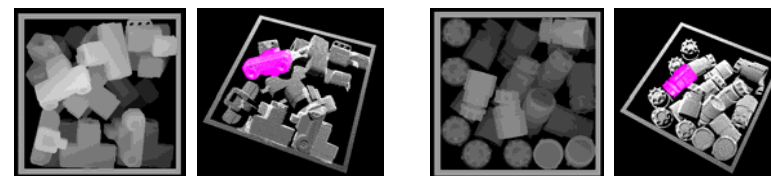
## シミュレーションデータに対する認識成功率 [%]



各物体それぞれ約 115 シーンのデータを用いた

CPU: Intel®CORE™i7-3.4GHz, System memory: 8GB

## シミュレーションデータに対する処理速度 [fps]



CPU: Intel®CORE™i7-3.4GHz, System memory: 8GB

## 高性能 LRF の提案 (DPN)

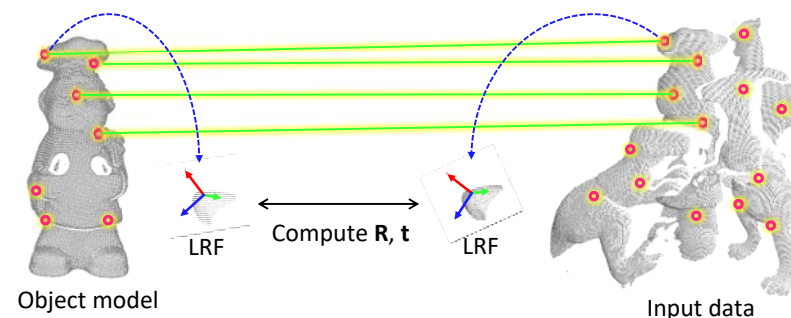
## 3次元物体認識の流れ

### ■ キーポイントマッチングの処理

Step 1. キーポイント検出

Step 2. 特徴量記述, 対応点探索

Step 3. 対応点群を利用した姿勢推定



特徴記述, 姿勢推定では LRF が必須  
LRF : Local Reference Frame

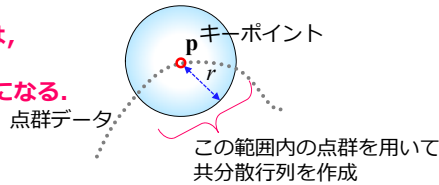


# 従来の LRF の問題点

## 1. 一括算出型 LRF [1~3]

- 3つの軸を同時に算出する手法
- 3次元点群の分布そのものの共分散行列から得られる固有ベクトルを LRF とする

点群が部分的に欠落した場合には、分布の広がり方そのものが変化  
→ LRF の方向ベクトルが不安定になる。



## 2. 個別算出型 [4, 5]

- LRF を構成する各軸をそれぞれ独立に算出
- z 軸に面法線を割り当てることが特徴
- x 軸の算出方法に特色がある

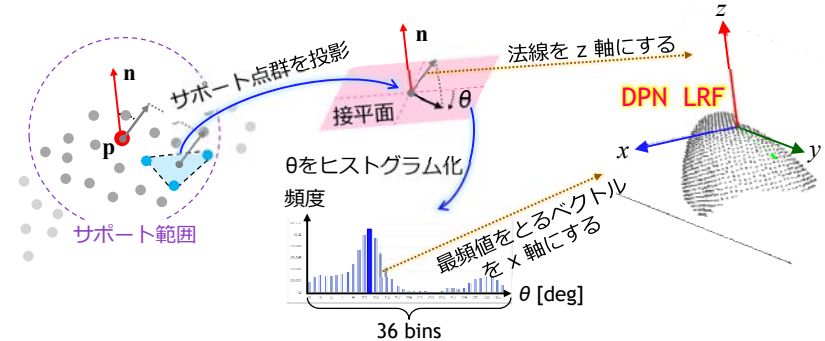
点群の密度が異なる場合には再現性が低下 [3]  
→ 密度変化を扱える手法が無い

[1] A.Mian, et al., "On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes", IJCV, Vol.89, Issue 2-3, pp348-361, 2010.  
 [2] F. Tombari, et al., "Unique Signatures of Histograms for Local Surface Description", Proc. ECCV, pp.356-369, 2010.  
 [3] Y.Guo, et al., "Rotational Projection Statistics for 3D Local Surface Description and Object Recognition", IJCV Vol.105, Issue 1, pp.63-86, 2013.  
 [4] A.Petrelli, et al., "On the Repeatability of the Local Reference Frame for Partial Shape Matching", ICV, pp.2244-2251, 2011.  
 [5] A.Zaharescu, et al., "Surface Feature Detection and Description with Applications to Mesh Matching", CVPR, pp.373-380, 2009.

# 提案手法 DPN (Dominant Projected Normals) LRF

## ■ アイデア: x 軸の算出方法に工夫 (z 軸は注目点の法線)

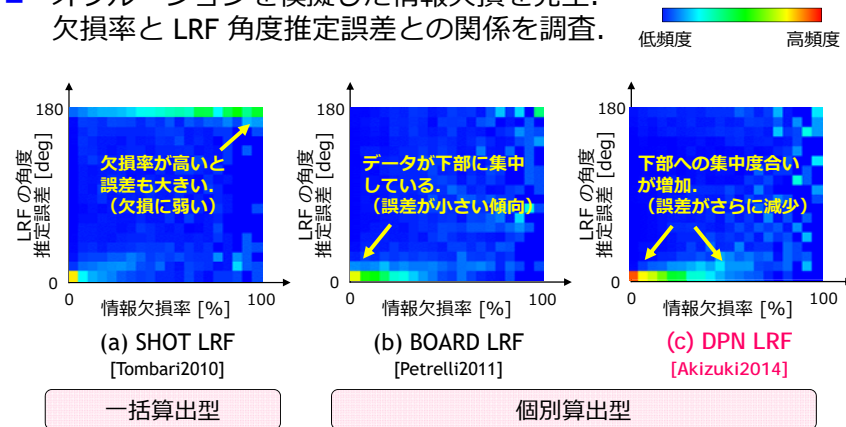
- 方向ヒストグラム of ピーク位置を利用 → 情報欠落の影響を改善
- 近傍点で作る三角パッチの面積で正規化 → 点群密度の違いを許容
  - WDが短い (センサに近い) 場合でも、面が横を向いていると点数が少なくなり「粗」であるように誤解する。
  - 三角が大きい場合は、大きな重みでヒストグラムに投票する。
  - これにより、「密」と同じ効果が得られる。



# DPN LRF の性能 情報欠損への耐性

- オクルージョンを模擬した情報欠損を発生。欠損率と LRF 角度推定誤差との関係を調査。

[Akizuki,VIEW2014]

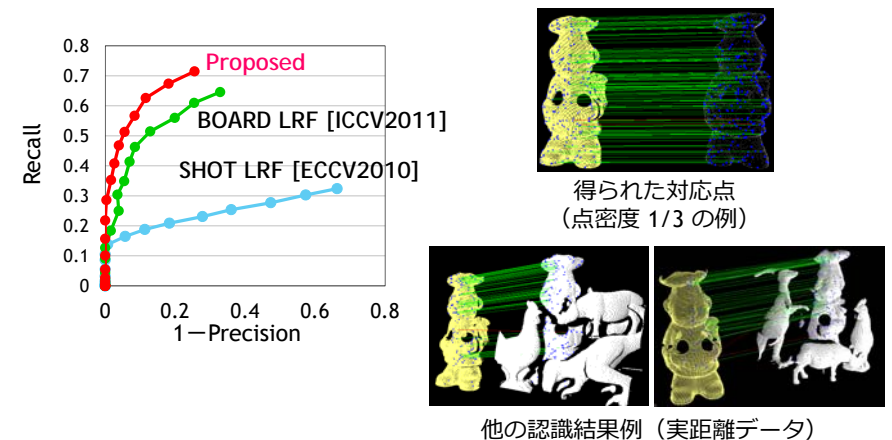


DPN LRF は、情報欠損の影響を受けにくい。

参考文献: 秋月秀一, 橋本学: "安定的な3D特徴量記述のための点群の密度変化にロバストな Local Reference Frame", VIEW, 2014.

# DPN LRF の性能 特徴量の記述性能の評価

- SHOT 特徴量を用いたキーポイントマッチングの性能

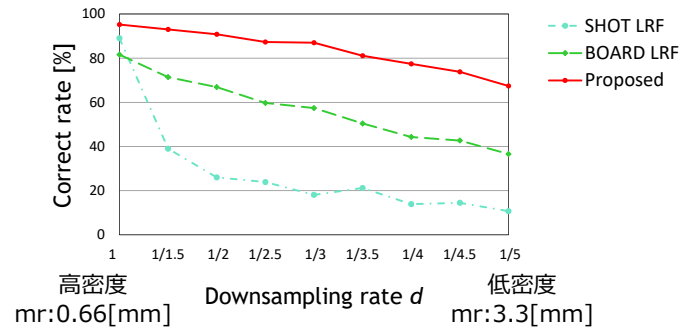


LRF を提案手法に差し替えるだけで認識性能が向上

## DPN LRF の性能 点群密度変化への耐性

[Akizuki, ViEW2014]

- ノイズ+点群密度 1~1/5 の外乱付きデータに対してDPN LRF を算出.
- 角度誤差 10 deg 以内のデータ数の割合をプロット.



DPN LRF は, 点群密度の違いによる影響を受けにくい.

参考文献: 秋月秀一, 橋本学: "安定的な3D特徴量記述のための点群の密度変化にロバストな Local Reference Frame", ViEW, 2014.  
Hashimoto Laboratory, Chukyo University Feb. 2017

101

## その他

## OUR-CVFH (Oriented, Unique and Repeatable CVFH)

CVFH: Clustered Viewpoint Feature Histogram

[Aldoma2012]

- 領域ベース特徴量 (領域ごとに特徴量を設定)

step1: RFの生成

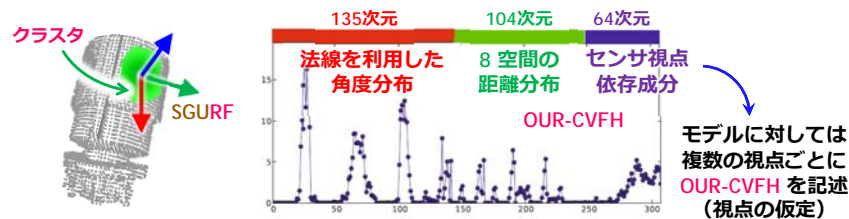
点群から滑らかな連結領域 (クラスタ) を抽出し, 共分散行列によってRFを生成  
 ... SGURF (Semi-Global Unique Reference Frames)

step2: 特徴量の記述

領域内の点群から次の3つの成分からなる計303次元の特徴ベクトルを記述

... OUR-CVFH (Oriented, Unique and Repeatable CVFH)

1. 法線を利用した (FPFH的な) 角度分布ヒストグラム (135ビン)
2. 3次元 LRFの8つの象限ごとに注目点と他点群の距離を13ビンで表現 (104ビン)
3. センサ視点からみたクラスタ内の法線ベクトル群の分布 (64ビン)



出典: A. Aldoma, F. Tombari, R. B. Rusu, M. Vincze, "OUR-CVFH - Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram for Object Recognition and 6DOF Pose Estimation", DAGM-OAGMPPRS, pp.113-122, 2012.

## 3. アマゾンピッキングチャレンジ2016に関連するロボットビジョン技術

## 課題分析

## APC2015@Seattle

### APC2015 ルール概要

- 目標
  - 非構造的環境における自動ピッキング装置の開発
    - 商用可能な物品
- 1m×1mの範囲の計12Binからピッキング
  - 身長170cmの人による作業を想定



- 時間内にどれだけアイテムをピッキングできたかを競う
  - 商品の誤選択や落下による破損は減点

### Binの構成と得点

- Single-item Bins:(10points)
  - 最低でも2つのBinに1種類のアイテムが存在
- Double-item Bins:(15points)
  - 最低でも2つのBinに2種類のアイテムが存在し  
 その中から指定されたアイテムをピッキング
- Multi-item Bins:(20points)
  - 最低でも2つのBinに3種類以上のアイテムが存在し  
 その中から指定されたアイテムをピッキング
- Bin内のアイテムの置かれ方
  - アイテムの姿勢はランダムに決定
  - Binを正面から見たときオクルージョンは発生しない



## ピッキング対象

- 全25アイテム (アイテムの種類によっては加点あり)

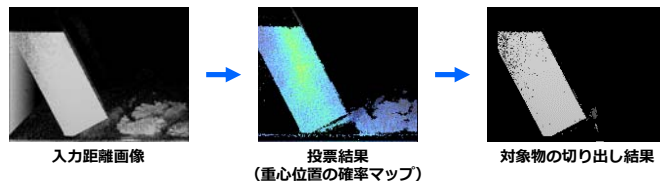
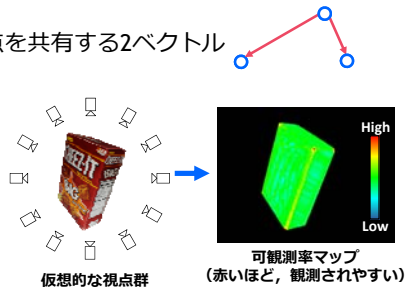


## 参加チーム (世界中から28チーム)

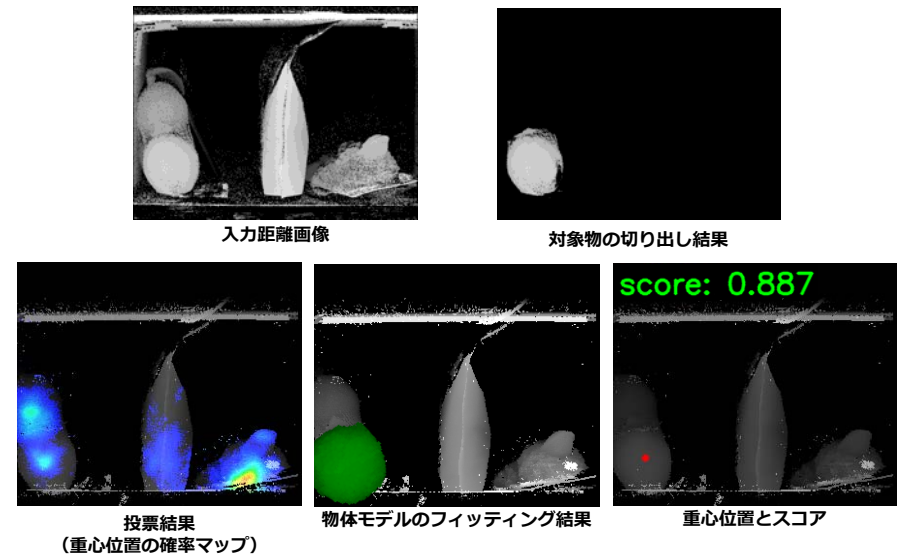


## VPMアルゴリズム (中京大学)

- 特徴量 (ベクトルペア)
  - 3点の3次元点で構成された、視点を共有する2ベクトル
- ベクトルペア選択
  - 物体モデルのまわりに設置した仮想的な視点群から、安定的に観測可能 (高可観測率) なベクトルペアのみを照合に利用
- マッチング
  - 投票型マッチングによって、物体の重心位置を推定



## VPMアルゴリズム (中京大学) 認識結果例

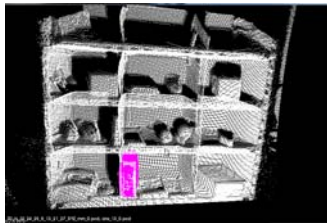


# VPMアルゴリズム (中京大学) 認識結果例

箱状物体

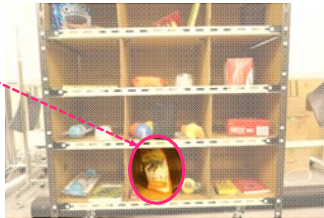


入力シーン外観

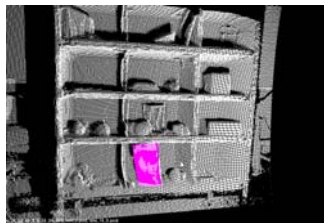


認識結果

形が変形しやすい袋状物体



入力シーン外観



認識結果

対象物

# チャレンジ本番のようす



ダイジェスト動画は→[こちら](#)



# APC2016@Leipzig

# APC2016における対象アイテム (39種類)

dove_beauty_bar  1	rawlings_baseball  +1 2	clorox_utility_brush  +2 3	dr_browns_bottle_brush  +2 4	dasani_water_bottle  2x 5	easter_turtle_sippy_cup  +2 6
cherokee_easy_tee_shirt  +1 7	folgers_classic_roast_coffee  8	crayola_24_ct  9	peva_shower_curtain_liner  10	barkely_hida_bones  11	kyjen_squeakin_eggs_plush_puppies  +1 12
expo_dry_erase_board_eraser  2x 13	scotch_duct_tape  +1 2x 14	jane_eyre_dvd  15	scotch_bubble_mailer  2x 16	woods_extension_cord  +1 17	womens_knit_gloves  +1 18
cool_shot_glue_sticks  19	elmers_washable_no_run_school_glue  2x 20	staples_index_cards  2x 21	laugh_out_loud_joke_book  22	i_am_a_bunny_book  23	kleenex_tissue_box  24

## APC2016における対象アイテム (39種類)



- 形状や重量, 表面の硬さが異なる多様なアイテム
- 認識や把持が困難なアイテムには加点あり

## 配点

Picking または Stowing した Bin の中のアイテム数により得点を決定

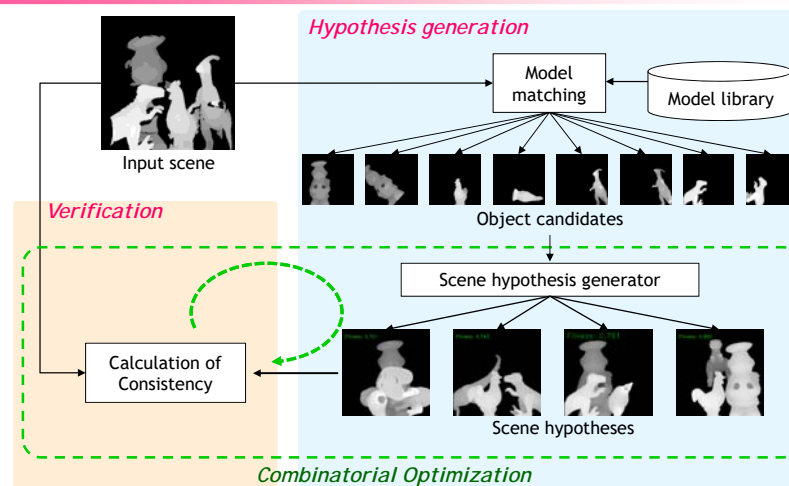
アイテム数 1~2 個の Bin	+10 point
アイテム数 3~4 個の Bin	+15 point
アイテム数 5 個以上の Bin	+20 point

### ペナルティ

誤った JSON ファイルの記述 (アイテムの誤認識)	-10 point
30cm 以上の高さから落下	-5 point
アイテムまたは shelf を傷つける	- 5 point
アイテムまたは shelf を破損	-20 point
0.5cm 以上 shelf からはみ出す	- 5 point
アイテムが床に落下	-10 point

## アプローチ 物体識別と姿勢認識

## 仮説検証型アルゴリズム



- シーン全体 (大局的) の整合性を計算するため, 誤認識が発生しにくい
- × 過検出が発生しやすい.

## 従来の仮説検証手法と提案アプローチ

### 従来の仮説検証



入力シーン

実シーン・シーン仮説  
間類似度



シーン仮説

画像類似度 [Hashimoto1999]  
点群間距離・法線間類似度 [Aldoma2012]  
RGB 利用 [Aldoma2013]

### 物体配置の実現可能性の導入



仮説シーンの物理的整合性を検証  
(物体候補同士に空間的な重なりを確認)



シーン仮説の実現可能性も考慮することによって、  
過検出を積極的に抑制する

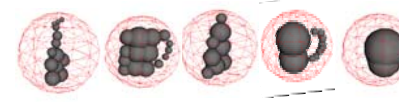
## 干渉モデル

### ■ 高速な干渉判定が可能なモデルを提案

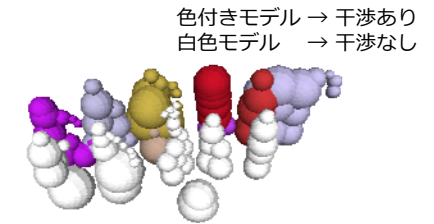
- 物体の外形を単純な形状によって近似表現
  - 球体が最も単純な基本形状 → 高速な干渉判定が可能
- 外接球 (赤) と内接球群 (グレー) の 2 段構成
  - 外接球の干渉判定によって、干渉の疎精探索が可能



物体モデル



干渉モデル

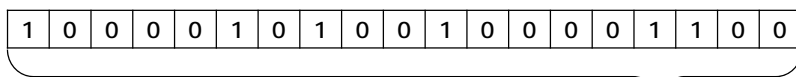


干渉モデルとして表現されたシーン仮説

色付きモデル → 干渉あり  
白色モデル → 干渉なし

## コスト関数の設計

### ■ 配置の実現性と形状的整合性を利用したコスト関数を定義



長さ  $n$  (=物体候補数) のビットストリング  
1 に対応する物体候補がシーン仮説を構成

$$\tilde{X} = \arg \max \{F(X) = w f_p(X) + (1-w) f_s(X)\}$$

配置実現性 (物体仮説間の干渉の少なさ)      形状的整合性 (画像類似度)

スコアを最大にするビットストリングを  
遺伝的アルゴリズムによって探索





## APC2016 : Pick task における認識結果

Amazon Picking Challenge 2016  
Pick task 認識率 : 92.5% (= 50/54)

## 4. 物体認識技術の将来展望

## 工場から家庭へ



## 生活支援ロボットのための認識技術

“アフォーダンス”概念を利用した一般物体認識

## 家庭内ロボットののための認識システム

誰もが簡単に利用できる生活支援ロボット  
～自然な命令を理解する～

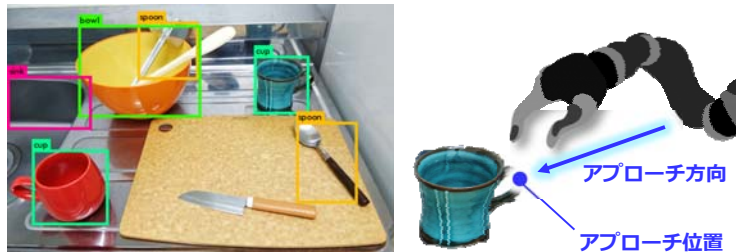


NEDO 人工知能プロジェクト (2015～2019)  
産総研 人工知能研究センターからの再委託研究として遂行中

## 研究の背景と必要なタスク

### ■ 背景

生活支援ロボットには、多品種の日用品を扱うことが求められる。



Deep Learning[1] を用いた識別結果

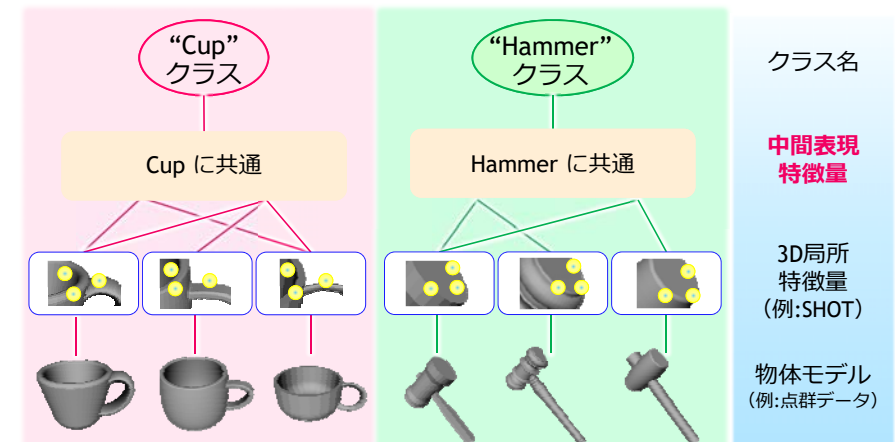
[1] J. Redmon, et al., "You Only Look Once: Unified, Real-Time Object Detection", CVPR, (2016).

### ■ ビジョンに求められるタスク

……生活支援ロボットのための物体認識である点を考慮……

- ① モデルレスでの品種識別 (クラス識別)
- ② マニピュレーションのためのアプローチパラメータ推定

## 一般物体認識に必要なクラス内共通の特徴量



ロボットビジョンのための中間表現特徴量に求められる条件

1. クラス内で共通していること, なおかつ,
2. アプローチパラメータ推定に利用できること.

# アイデア

## 日用品の特性

1. なんらかの**目的**のために製造されている。
2. その目的を達成するために、クラス内共通の**機能**がある。
3. 機能は**物体の形や大きさ**に作り込まれている。

"Cup"		
1. 目的	飲み物を入れて運ぶ	
2. 機能	水を蓄える	つかまれる
3. 形状	くぼみ形状	取っ手
AFラベル	Contain	Grasp

∴ **共通形状**を利用してクラス内の物体群を同一視する。

アフォーダンスラベル "Contain" ラベル



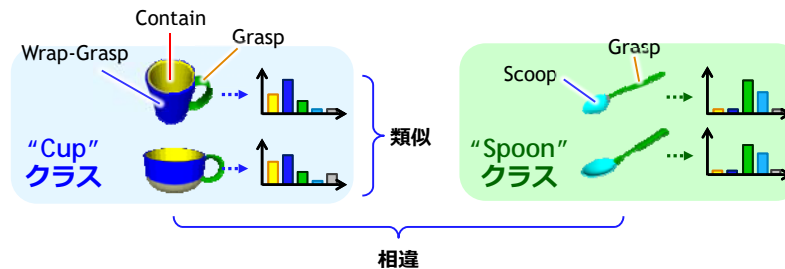
"Grasp" ラベル

# アフォーダンス特徴量

## アフォーダンス特徴量 = アフォーダンスラベルの含有率分布

同一クラス内なら  
異なるクラス間なら

→アフォーダンス特徴は類似  
→アフォーダンス特徴は似ていない。



AF ラベルと色の対応・・・ ■ : Contain ■ : Wrap-Grasp ■ : Grasp ■ : Scoop

# 実験に使用したシーン

4クラスに属する日用品のうち、3つが並べられたテーブルトップシーンを使用。

- シーンの数 : 12 種類
- 使用したクラス : Cup, Hammer, Spatula, Spoon

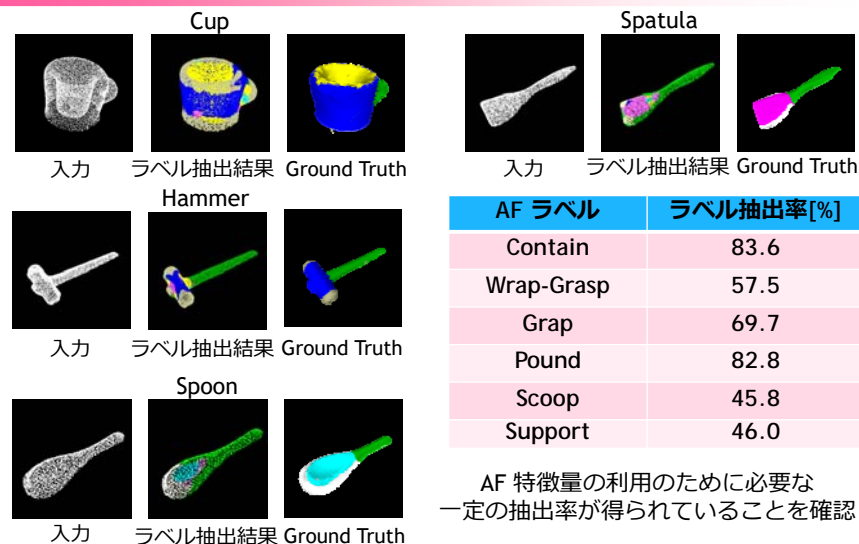


実験に使用した日用品の例



実験に使用したシーンの例

# アフォーダンスラベル抽出結果



AF 特徴量の利用のために必要な一定の抽出率が得られていることを確認

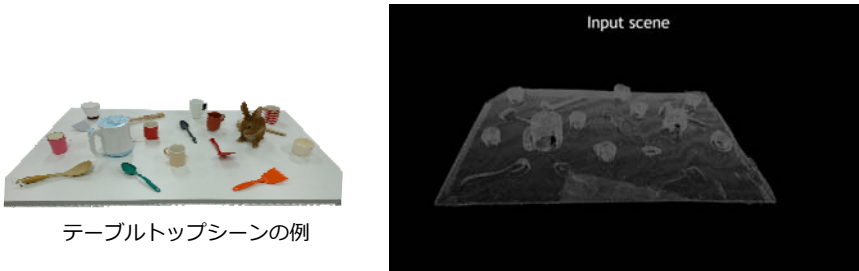
■ : Contain ■ : Wrap-Grasp ■ : Grasp ■ : Pound ■ : Scoop ■ : Support

## 認識結果

1 物体あたりの認識処理時間と認識成功率

手法	処理時間 [s]	認識成功率 [%]
比較手法 : CNN (GoogLeNet)	0.28	81
提案手法	0.47	92

(CPU: Intel® CORE™ i7-4790, メモリ: 8GB)



テーブルトップシーンの例

複数の日用品が並べられたテーブルトップシーンでの物体認識結果の例

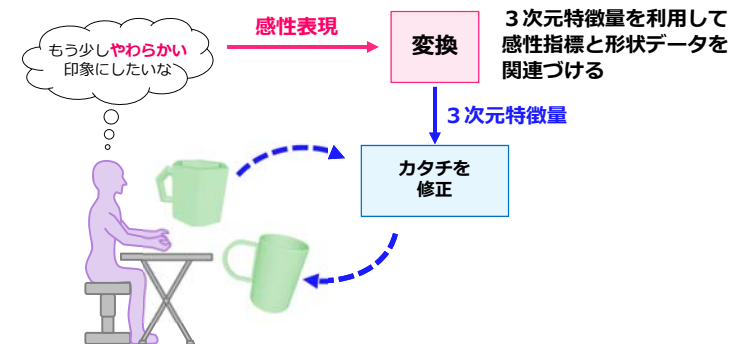
## 人間の内面を探る

## 人間の「感性」と3次元形状とのマッピング

## 人間の「感性」と3次元形状とのマッピング

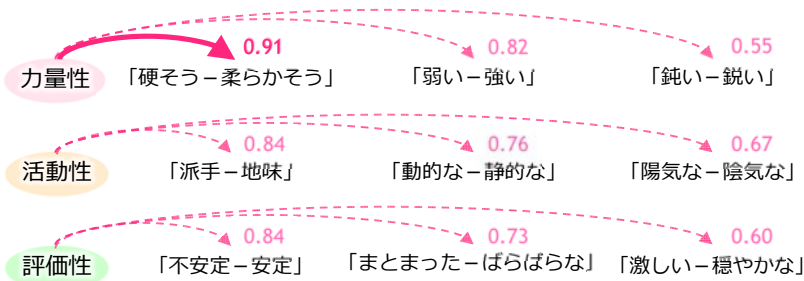
パーソナルファブ리케이션（個人によるものづくり）時代の到来

「こんなものを作りたい」という自然な感情を、ものづくりに反映させる。  
→感性をカタチに結びつける



## 人間の「感性」と3次元形状とのマッピング

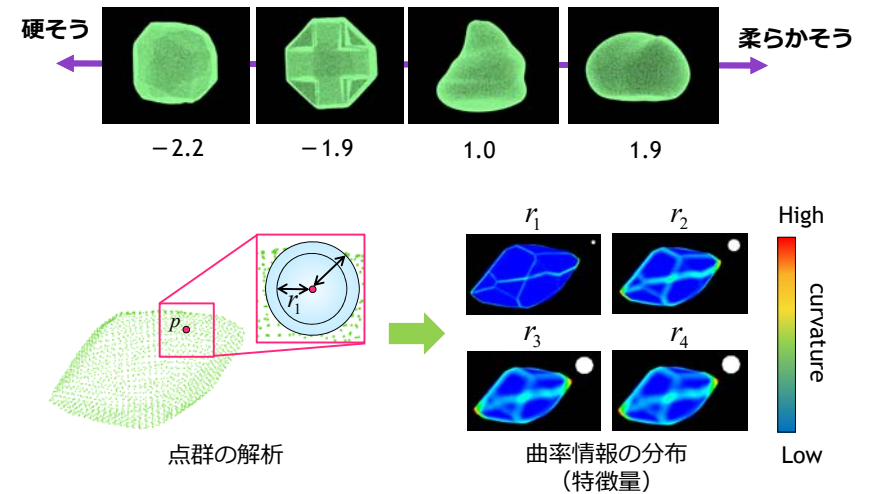
- 目的  
3次元特徴量を用いて、人の「感性（印象）」を自動推定すること
- 3次元形状から受ける印象に関する考察 [飛谷 2014]  
人間が受ける印象は **3つの因子** で表現できる。



「硬そう-柔らかそう」という印象について取り組む

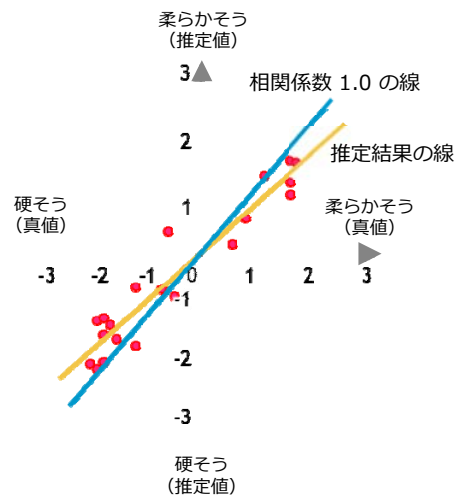
## 人間の「感性」と3次元形状とのマッピング

ユーザによって付与された感性指標（硬そう～柔らかそう）



## 人間の「感性」と3次元形状とのマッピング

真値（SD法による人間の感性）と提案手法による推定結果の比較



相関係数	相関の度合い
0.0 ~ 0.2	ほとんど相関関係がない
0.2 ~ 0.3	弱い相関関係がある
0.3 ~ 0.4	ある程度の相関関係がある
0.4 ~ 0.5	中程度の相関関係がある
0.5 ~ 0.7	かなり高い相関関係がある
0.7 ~ 1.0	強い相関関係がある

相関係数と相関の度合：  
(放送大学「社会調査の基礎」)

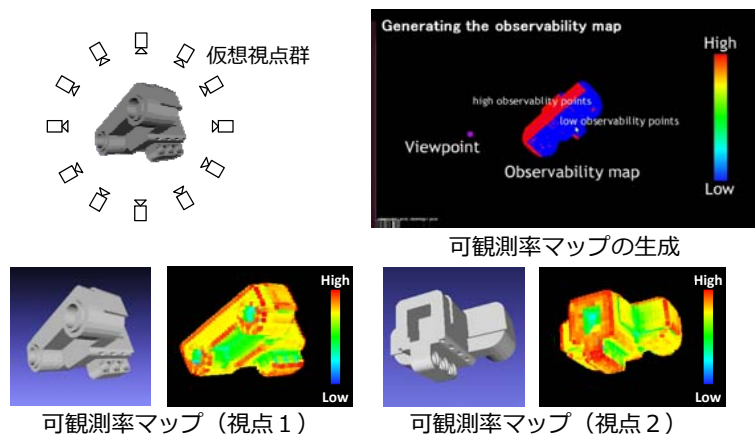
## 可観測性を考慮した物体認識

見えない点は使えない

## 可観測性を考慮した 3D特徴点選択

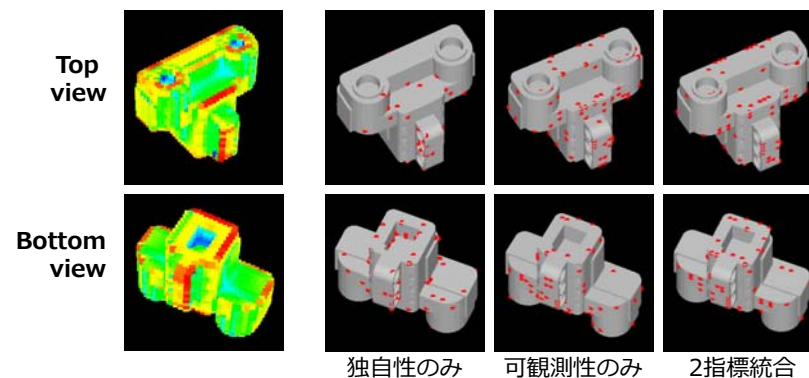
[Akizuki2014]

- 入力シーンにおける観測可能性の高い点を自動選択
  - 物体周りに複数の視点を設定し、各点の観測可能性をモニタリング
  - 観測可能性の高い点を特徴点とする。



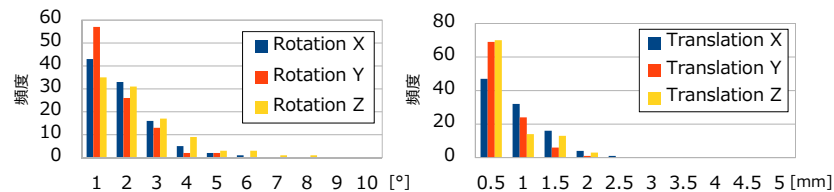
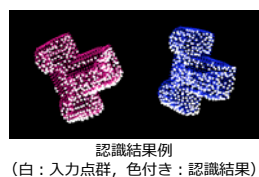
S. Akizuki, M. Hashimoto: "Position and Pose Recognition of Randomly Stacked Objects using Highly Observable 3D Vector Pairs", Proc. IECON, pp.5266-5271, Oct, 2014.

## 選択された特徴点



## 位置姿勢の推定精度

- 単一の物体モデルを姿勢変化 (超音波センサ)
  - 回転: XYZそれぞれ $\pm 30$ [°]
  - 平行移動:  $\pm 300$ [mm]
  - データ数: 100例
  - データ密度: 3.0[mm/point]



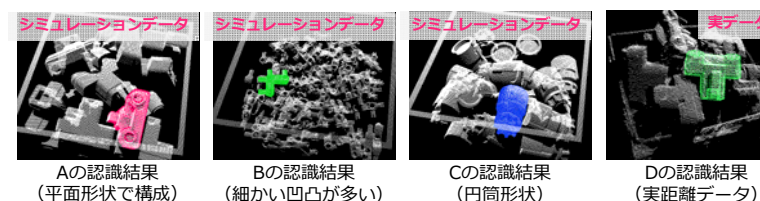
- 回転成分: 物体の端で約2.9mmずれる (1point cloud以内)
- 平行移動成分: 元データの解像度以内に収束

ピッキングに十分な精度で認識可能

## さまざまな形状の物体に対する認識結果

- 4種類×約113シーンで実験 認識率Pr[%]と処理時間T[sec]の関係

		A	B	C	D
SHOT+Correspondence Grouping [Tombari,PSIVT2011]	Pr	51.3	77.9	33.6	20.3
	T	27.78	19.46	33.69	22.48
FPFH+RANSACベース照合 [Buch,ICRA2013]	Pr	54.9	11.5	73.5	31.0
	T	0.67	3.97	1.07	0.91
VPM(独自性のみ)	Pr	40.7	31.0	47.8	33.6
	T	0.27	1.41	0.39	0.72
VPM(可観測性のみ)	Pr	97.3	100.0	94.7	85.0
	T	0.47	1.76	1.76	1.99
<b>VPM(独自性+可観測性)</b>	<b>Pr</b>	<b>99.1</b>	<b>95.6</b>	<b>92.0</b>	<b>85.8</b>
	<b>T</b>	<b>0.44</b>	<b>1.76</b>	<b>1.42</b>	<b>1.81</b>

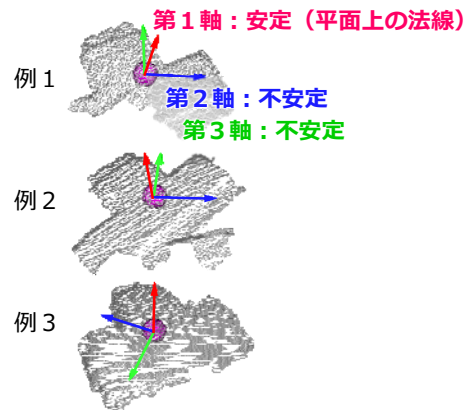


## 3次元的な起伏が少ない物体への対応 (GRF)

工業部品は平面部が多い

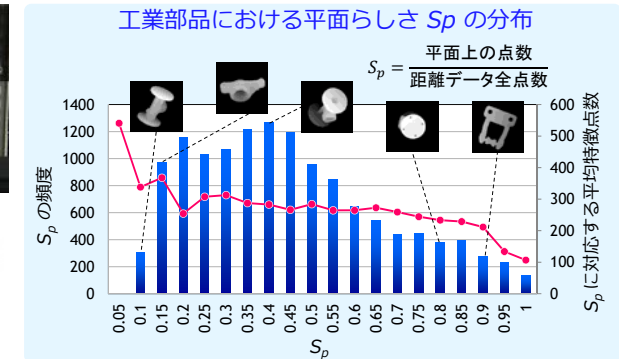
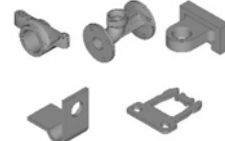
## 従来の LRF の問題点

- 平面的な形状の場合に、軸方向が不安定



## 背景

- 工業部品には平面的な見え方のものが多い (約32%)



工業部品の例 [1]

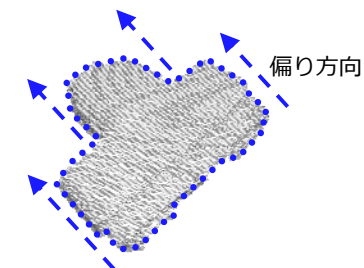
- 見え方が平面的 → 特徴点数が少なく **認識困難**.

**認識困難な理由:** ①特徴点が少ない, ②LRFが不安定.

[1] S. Jayanti, et al., "Developing an Engineering Shape Benchmark for CAD Models", Computer-Aided Design, Vol. 38, Issue 9, pp.939-953, 2006.

## 基本アイデア

- 物体の外形を構成する輪郭には特定方向の偏りが存在
  - シルエットの偏りを使えば安定的な方向ベクトルの算出が可能

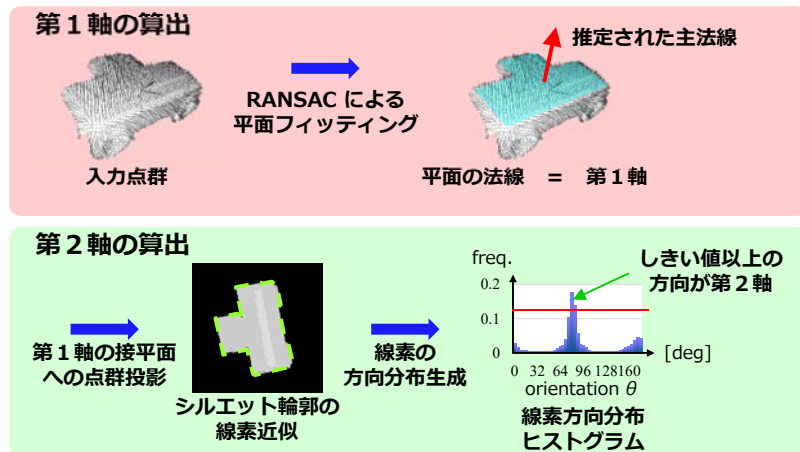


**外形 = 大域的 な情報を使った  
新たな Reference Frame を提案**

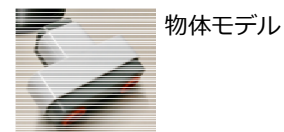
## 提案する Reference Frame

### ■ Global Reference Frame (GRF)

- 主法線方向と外輪郭の偏り方向によって構成



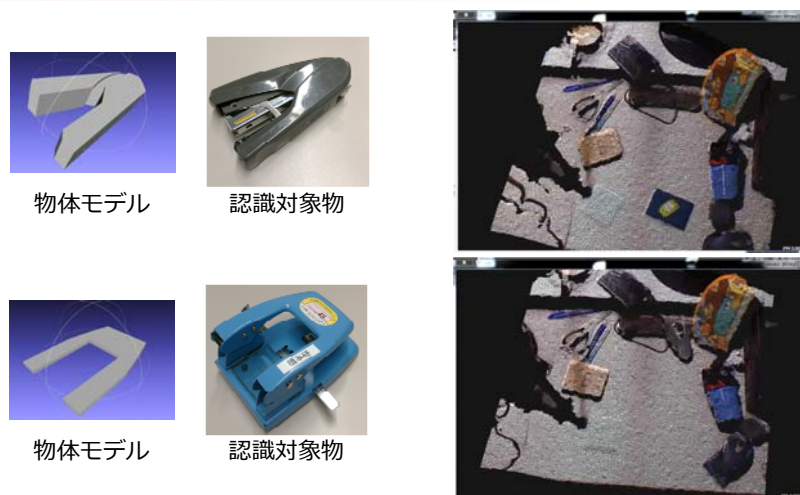
## 従来の LRF と GRF の比較



	Pose A	Pose B	Pose C	Pose D
LRF [Tombari2010]				
GRF (提案手法)				

GRFの方向ベクトルは姿勢変化・形状の欠けに対して頑健

## GRFを用いた平面的な物体の認識



GRFを用いれば、平面的形状に対しても頑健に認識可能

## 照合戦略切替による位置姿勢認識

- 処理対象を事前に領域分割し、「平面」と「非平面」に分類



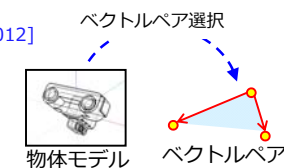
- 平面領域 → GRF 照合

- 非平面領域 → 特徴点ベース照合

- ベクトルペアマッチング (VPM) [秋月2012]

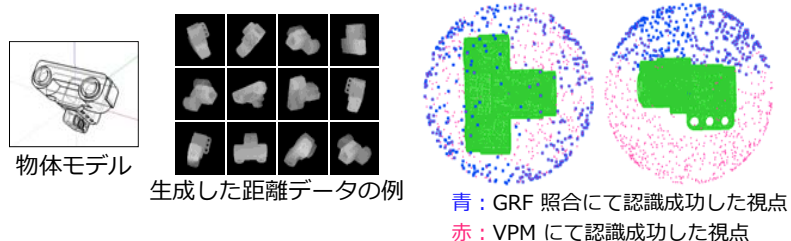
- ✓ 2つの性質を持つ特徴点を認識に利用

1. 高独自性の特徴ベクトルを記述可能
2. 観測可能性が高い



## 照合戦略の相補性能評価

- ランダムな1000視点から生成した形状データに対して照合
  - 平面 : 389視点, 凹凸 : 611視点



照合戦略	平面	凹凸
GRF照合	94.6[%]	2.8[%]
VPM	68.9[%]	94.8[%]

GRF 照合と VPM を組み合わせることによって位置姿勢認識の信頼性を向上可能

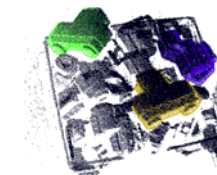
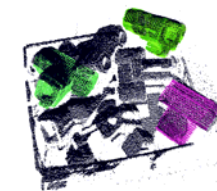
2手法は相補性が高い

## 認識性能の比較

CPU : Intel@Core™ i7 860 2.80GHz, RAM : 12GB

Feature	Pr [%]	T [sec]
SHOT [Tombari2010]	75.2	5.39
FPFH [Rusu2009]	31.0	0.91
Vector pair [Akizuki2012]	80.5	2.05
GRF	70.8	0.23
GRF+Vector pair (提案手法)	93.8	2.31

- 提案手法は処理領域の形状特徴に合わせて、用いる特徴量を自動選択する
  - 平面的な領域 : GRF
  - 非平面的な領域 : Vector pair



認識結果例

## 点群密度変化への対応 (RPD特徴量)

センサからの距離に依存しない特徴量

## 移動ロボットでの3次元物体認識に必要なこと

- 計測点群の密度が異なる場合における認識の頑健性
  - 点群の密度の変化を引き起こす原因
    1. センサと対象物間の距離の変化
    2. 使用するセンサの種類の違い





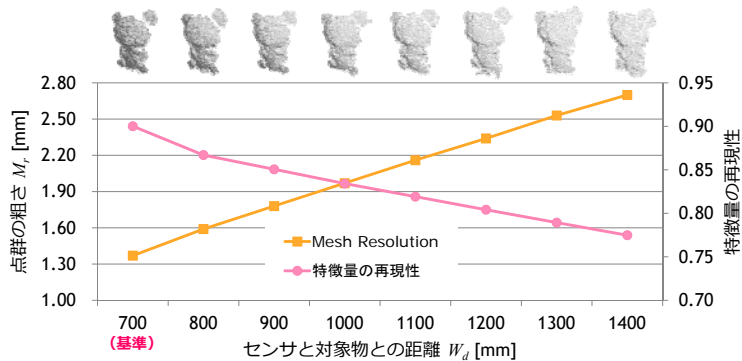
## 点群密度の差に起因する特徴点の再現性の低下

### ■ 予備実験

計測距離に起因する点群の密度変化 v.s. 特徴量の再現性

※再現性：2つの特徴ベクトル間の内積値

※Wd=700mm時をモデル（基準）として、遠方から撮影したデータと照合（使用特徴量：SHOT）



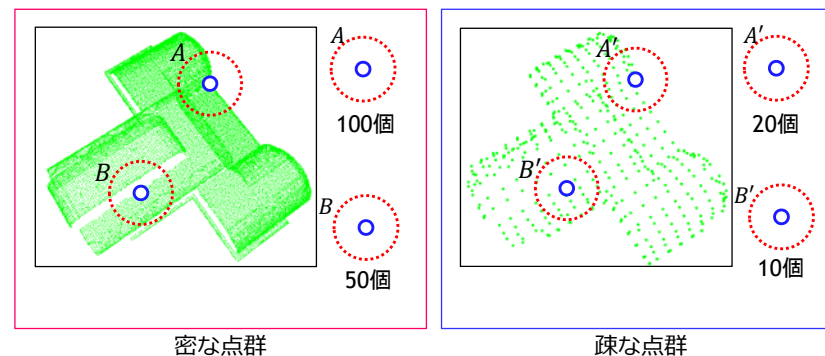
計測距離の変化により、特徴量の再現性が低下する。

このほか、ノイズの影響もある（∴Wd=700mmでも1.0にならない）

## 基本アイデア

■ センサからの距離等によって点群の密度は変化するが・・・  
**局所領域間の相対点数比は不変**

■ 複数の局所領域内に存在する点数の相対関係を利用



## RPD (Relative Point Density) 特徴量

### 局所領域間の点数比から構成される特徴量

■ 3つの3次元点を基準として特徴量を記述

■ 3つの3次元点は位置姿勢変換パラメータを算出するための最小単位

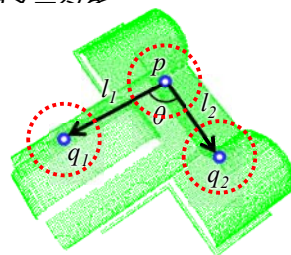
→ Vector Pair [Akizuki2014]

■ 特徴量の記述方法

$C(x)$ : 点  $x$  周辺に存在する点の個数

$$\mathbf{f} = \left( \frac{C(p)}{S}, \frac{C(q_1)}{S}, \frac{C(q_2)}{S} \right)$$

$$S = C(p) + C(q_1) + C(q_2)$$



ベクトルペア幾何パラメータ

$l_1$ :  $p$ - $q_1$  間の距離

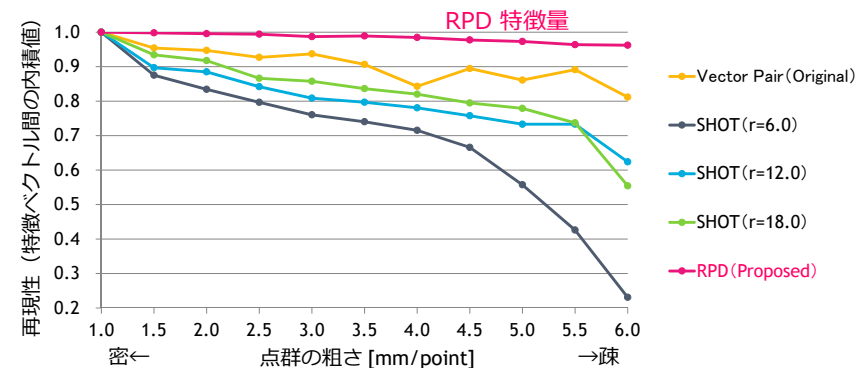
$l_2$ :  $p$ - $q_2$  間の距離

$\theta$ :  $p$ - $q_1$  と  $p$ - $q_2$  の間の角

## RPD 特徴量の性能① (再現性)

比較特徴量：①オリジナルの Vector Pair 特徴量 [Akizuki2014]

②SHOT 特徴量 [Tombari2010]

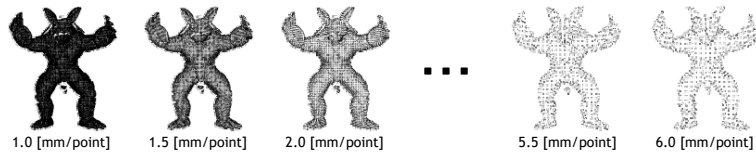


RPD 特徴量は粗い点群に対しても再現性が高い

## RPD 特徴量の性能② (認識率)

### ■ 実験に用いたデータ

11 種類の粗さのデータセット × 100シーン



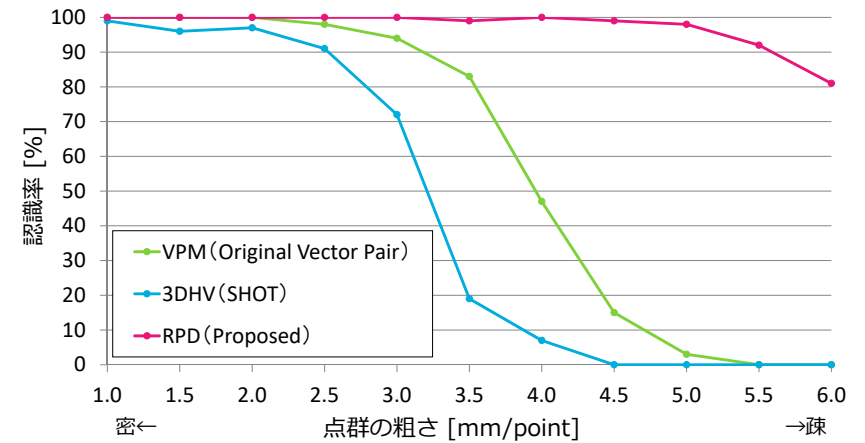
### ■ 実験に用いたパラメータ

		SHOT+3DHV	Vector Pair+VPM	RPD+VPM
ベクトルペア 幾何パラメータ	$l_1$	-	48.0 [mm]	48.0 [mm]
	$l_2$	-	36.0 [mm]	36.0 [mm]
	$\theta$	-	90.0 [deg]	90.0 [deg]
特徴量の 記述半径	$r$	6.0 [mm]	-	6.0 [mm]

3DHV : 3D Hough Voting [Tombari2010]

VPM : Vector Pair Matching [Akizuki2014]

## RPD 特徴量の性能② (認識率)



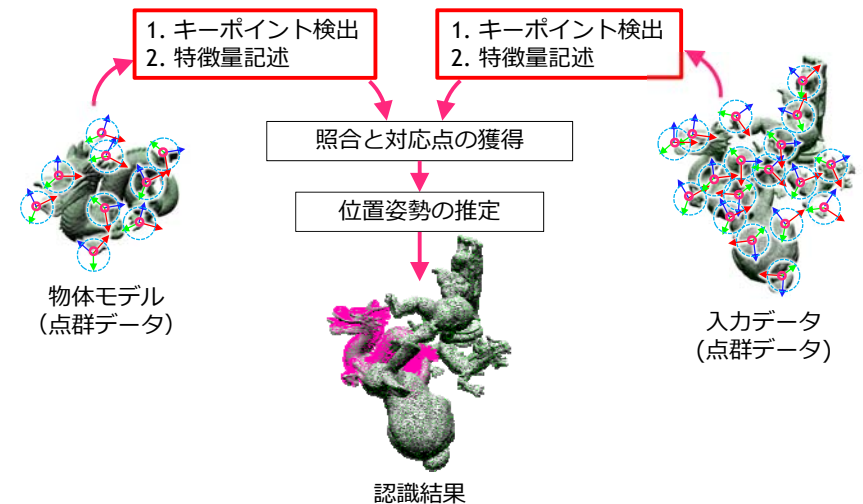
RPD 特徴量は点群が粗くても認識率が高い

## SHORT: 点群の空間占有率とマルチスケール シェルを用いた高速 3D 物体認識

Shell Histograms and Occupancy from Radial Transform

キーポイント検出と特徴量記述の高速化

## モデルベース 3次元物体認識の性能を決める要因



高速なキーポイント検出・特徴量記述が重要

## 従来研究の問題点と本研究の目的・基本アイデア

### ■ キーポイント検出の従来研究

局所領域内の点群から共分散行列などの統計量を算出し、形状を評価 (ISS法 [Zhong2009], LSP法 [Chen2007])

問題: 形状評価のための処理コストが高い

### ■ 特徴量記述の従来研究

局所領域内の点群から、法線ベクトルの方向分布などを記述 (SHOT法 [Salti2014], FPFH法 [Rusu2009])

記述領域を分割 → **アイデンティティを強化**

問題: 領域内の全点を用いるため、処理コストが高い

法線ベクトルの算出コストも高い

方位角と仰角の分割は基準座標系に影響されやすい



### ■ 本研究の目的

従来法の認識性能を維持しつつ、特徴算出を高速化すること

### ■ 基本アイデア

キーポイント検出: **統計量の算出を必要としない形状評価**

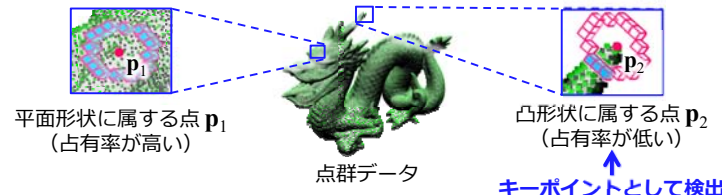
特徴量記述: **半径方向の限定領域のみを使用** (法線は不使用)

## 占有率に基づくキーポイント検出

### 1. 占有率の算出

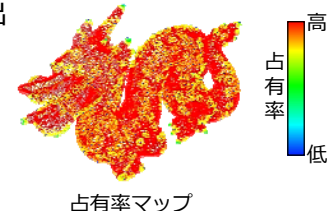
局所領域内に対する点数の割合 (占有率) を用いて形状を評価

- 領域内の点数をカウントするだけ
- 設定した領域を球の最外殻 (シェル) のみに限定



### 2. 占有率マップから極小値を検出

注目点とその周辺点の占有率を比較して、**占有率が低い点**をキーポイントとして検出



## マルチスケールシェルを用いた特徴量の記述

マルチスケールのシェル領域に存在する点群の分布を記述

### 1. 主方向ベクトル $d_k$ の算出 → **姿勢不変な特徴量記述**

キーポイント  $p_k$  を原点とした重心ベクトルを主方向とする

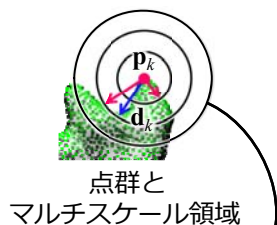
$$g_k = \frac{1}{L} \sum_{i=1}^L o_i (p_i - p_k)$$

$p_i$ : 領域内の点

$o_i$ : 占有率

$L$ : 領域内の点数

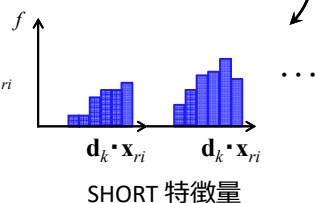
$$d_k = \frac{g_k}{\|g_k\|}$$



### 2. 特徴量 (ヒストグラム) の算出

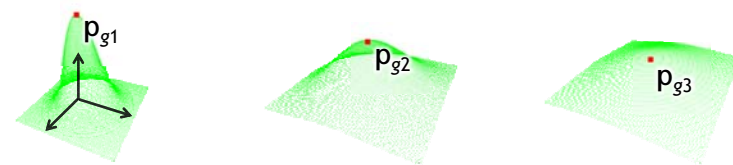
$d_k$  と、領域内の点への方向ベクトル  $x_{ri}$  との内積ヒストグラムを算出

→ **点群の分布を高速に記述**

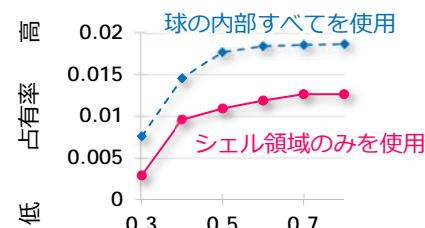


## 点群の空間占有率と形状の関係

$\sigma$  が異なるガウス分布の点群を生成し、 $p_{gk}$  に対して占有率を計測



ガウス分布 ( $\sigma = 0.3$ )    ガウス分布 ( $\sigma = 0.5$ )    ガウス分布 ( $\sigma = 0.7$ )



**凸は占有率が低く、平面は占有率が高いことを確認**

↓  
**占有率を形状評価値として、キーポイント検出と特徴量記述に利用**

小 (凸)    ガウス分布の  $\sigma$     大 (平面)

## 認識性能の評価実験

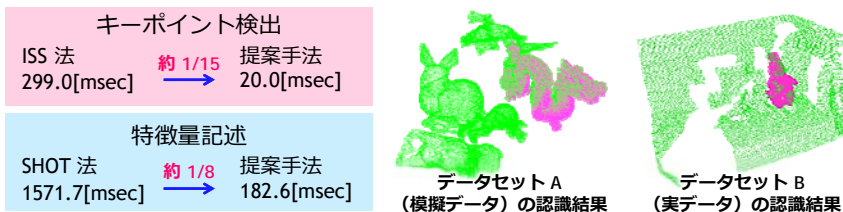
公開データセット [1] に対して物体認識

処理時間  $T$  [msec] と認識率  $Pr$  [%] の関係

キーポイント検出 + 特徴量記述	認識アルゴリズム	データセット	処理時間 $T$ [msec]	認識率 $Pr$ [%]
ISS 法 [Zhong2009] + SHOT 法 [Salti2014]	3D Hough Voting 法 [Tombari2010]	A	840.2	98.9
		B	4146.9	85.7
SHORT 法 (提案手法)		A	211.7	100.0
		B	460.2	91.8

データセット B に対する処理時間の比較

(CPU: Intel®CORE™i7-3770, メモリ: 16GB)



[1] <http://vision.deis.unibo.it/research/78-cvlab/80-shot>

## 3D-CG技術の物体認識への応用

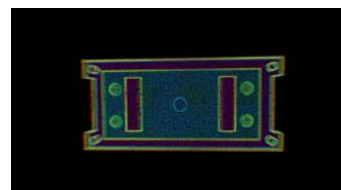
コンピュータグラフィックスを用いた  
3次元視覚の高度化

## 従来研究の問題点と本研究の目的

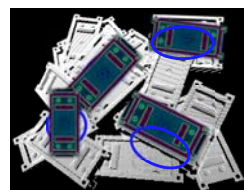
### ■ 従来研究の共通の問題

複数物体の接触に起因した認識エラー

原因: 認識には物体モデル単体のデータから得られる情報のみ利用



物体モデル (凹凸をカラー表示)



ばら積みシーン

### ■ 研究の目的

ばら積み状況下においても、高速かつ確実に物体の位置姿勢を認識する新しい手法の提案

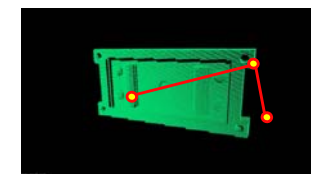
## 基本アイデア

### 1. 3点を用いた認識

膨大な3次元データから、3点のみを選択して認識に使用

→ 処理の高速化

3点の選び方が重要



3D-CADデータ (3D点の集合)

### 2. 3点の選択方法

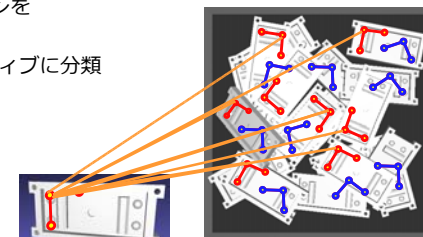
認識性能が高い3点を選択

- 3D-CGを用いて入力シーンを精密にシミュレート
- 3点をポジティブとネガティブに分類
- 3点の性能を数値化

→ 確実性の高い認識

ポジティブ

ネガティブ

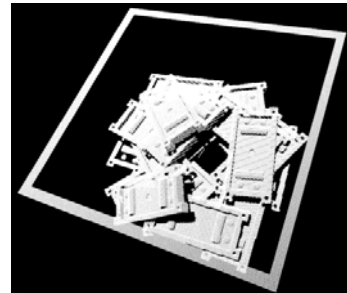
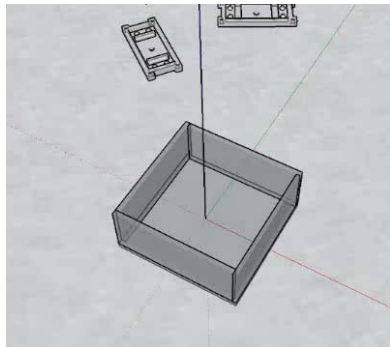


物体モデル

シミュレートした入力シーン

## アイデア：入力シーンを予測して特徴点の良否判定

シミュレータを用いてリアリティの高いばら積みシーンを自動生成  
物理エンジンには Sketchy Physics を使用



ばら積みシーン  
(センサ入力を模擬した3次元データ)

さまざまなパターンのばら積み状態をシミュレート

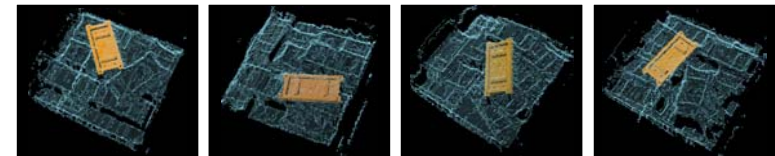
## 認識性能の評価実験②（本実験）

実際のばら積みシーン140枚に対する認識実験

### ライバル手法との性能比較

	認識成功率	処理時間
Spin Image法 [Johnson1999]	71.4%	57.99秒
Correspondence Grouping法 [Tombari2010]	62.1%	16.66秒
曲率法 [Takeguchi2004]+VPM法 [Akizuki2013]	91.4%	5.28秒
Vector Pair Matching(VPM)法 [Akizuki2013]	75.7%	0.61秒
<b>提案手法</b>	<b>94.3%</b>	<b>0.36秒</b>

CPU : Intel@CORE™i7-3770 システムメモリ : 16GB



提案手法はロボットの動作サイクル時間以内に高信頼な認識を実現