

# ばら積み物体の位置姿勢推定

○荒井翔悟 (東北大学), 青木吾郎 (東北大学), 小菅一弘 (東北大学)

## Pose Estimation for piled-up objects

○Shogo ARAI (Tohoku University), Goro AOKI (Tohoku University)

and Kazuhiro KOSUGE (Tohoku University)

Abstract : This paper proposes a method for pose estimation for piled-up objects. In the proposed method, a problem of pose estimation is converted into an optimization problem and the problem is solved by LASSO (Least Absolute Shrinkage and Selection Operator). Experimental results show that the proposed method can perform pose estimation and compute poses of multiple objects simultaneously.

### 1. はじめに

ばら積み物体の位置姿勢推定は、ロボット工学やコンピュータビジョン分野で古典的でありながら、現在も研究が続けられている最新のトピックでもある。三次元計測センサの低価格と普及、機械学習の進展などの影響を受けつつ、様々な位置姿勢推定方法が提案されてきた。本稿では、三次元計測センサでばら積みシーンを計測し、位置姿勢推定を行う方法に焦点を当てる。

三次元計測センサから得られた三次元点群を使った位置姿勢推定は、ICP[1][2]、局所特徴量を用いた手法 [5][6][7]、および深層学習 [3][4] に大別される。ICP は繰り返し計算による非線形最適化によって位置姿勢推定を行うので、初期推定位置と姿勢が必要とされ、この初期推定の精度に結果が大きく依存する。深層学習を使った方法では、ネットワークの学習のために大量の学習データを計算リソースが必要とされる。局所特徴量を用いた手法では、数多くのパラメータのチューニングが必要である。

こうした問題を解決するため、本稿では、スパース推定を用いたばら積み物体の位置姿勢推定を提案する。提案法では、三次元計測センサによって得られたばら積みシーンの奥行き方向画像から位置姿勢推定を行う。まず、物体の三次元形状モデルを使って、様々な位置姿勢の辞書画像を生成する。つぎに、ばら積みシーンの奥行き方向画像は、少数の辞書画像を使って構成できるという仮定を利用して、スパースモデリングを使った最適化計算を実施し、位置と姿勢を推定する。

本稿では、実数集合を  $\mathbb{R}$ 、0 以上の自然数の集合を  $\mathbb{N}_0$  で定義する。行列  $A \in \mathbb{R}^{m \times n}$  の列展開を  $\text{cs}(A)$  とかく。

### 2. 提案する位置姿勢推定手法の概要

本節では、位置姿勢推定問題を Least Absolute Shrinkage and Selection Operator(LASSO) と呼ばれる凸最適化問題 [9]

$$\min_{x \in \mathbb{R}^p} \|y - Ax\|_2 + \sum_{u=1}^p \lambda_u \|x_{(u)}\|_1 \quad (1)$$

に帰着させる。ここで、 $y \in \mathbb{R}^p$  は  $p$  次元のデータベクトル、 $A$  は係数行列、 $x \in \mathbb{R}^p$  は求める解、 $\|x\|_1$  は  $x$  の  $1$  ノルム、 $x_{(u)}$  は  $x$  の第  $u$  要素を表す。

まず、与えられた物体の三次元モデルを使って、様々な姿勢に対する物体の奥行き方向画像を生成し、これを辞書画像とする。いま、 $k \in \{1, 2, \dots, N\}$  枚目の辞書画像を  $X_k \in \mathbb{N}_0^{m \times n}$  で表す。後の理論展開の都合上、生成された辞書画像を使って構成される行列

$$D = [\text{cs}(X_1^\top) \quad \text{cs}(X_2^\top) \quad \cdots \quad \text{cs}(X_N^\top)] \quad (2)$$

を定義しておく。

つぎに、ばら積み物体を三次元計測センサで撮影して得られたシーンの奥行き方向画像  $S \in \mathbb{N}_0^{p \times q}$  に対して、 $S$  の画素  $(i, j)$  を左上頂点、画素  $(i + m - 1, j + n - 1)$  を右下頂点とする画像  $Y_{i*j+i} \in \mathbb{N}_0^{m \times n}$  から構成される行列

$$T = [\text{cs}(Y_1^\top) - z\mathbf{1}^\top \quad \text{cs}(Y_2^\top) - z\mathbf{1}^\top \quad \cdots \quad \text{cs}(Y_{pq}^\top) - z\mathbf{1}^\top] \quad (3)$$

を定義する。ここで、 $\mathbf{1}$  は適切なサイズのすべての要素が 1 の縦ベクトルである。

以上の設定の下で、ばら積み物体の位置姿勢問題は、

$$\min_{\xi \in \mathbb{R}^p} \left\| \begin{bmatrix} \mathbf{0} \\ \gamma_1 \\ \gamma_2 \end{bmatrix} - \begin{bmatrix} D & -T & \mathbf{1} \\ \gamma_1 \mathbf{1}^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \gamma_2 \mathbf{1}^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} \right\|_2 + \gamma_3 \|\xi_1\|_1 + \gamma_4 \|\xi_2\|_1 \quad (4)$$

を解くことに帰着される．ここで， $\gamma_\ell$  ( $\ell \in \{1, 2, 3, 4\}$ ) は重み， $\mathbf{0}$  は適切なサイズの零行列である．問題 4 は，シーン画像から構成される行列  $T$  が辞書画像集合  $D$  の少数の画像の足し合わせで表現できると仮定した上で， $\xi_1, \xi_2, \xi_3$  を求める問題になっている． $\xi_1$  は辞書画像を選択するベクトル<sup>\*1</sup>， $\xi_2$  はシーン画像中の物体の位置を表すベクトル， $\xi_3$  は推定された物体の奥行きを表す．

### 3. 検証と考察

提案法を検証するための実験環境を図 1 に示す．図 2 に示す物体をばら積みにし，これを三次元計測センサ Ensenso N35 を使って計測した．得られた奥行方向画像から計算機を使って，提案法による位置姿勢推定を行った．Ensenso N35 はステレオカメラとプロジェクタから構成される三次元計測センサで，その解像度は  $1280 \times 1024$  である．計算機として，動作周波数 2.8GHz の CPU (intel(R) Core(TM) i7-7700HQ)，DDR4: 16.0GB を備えた PC を利用した．辞書画像の作成には 3D CG ソフト blender[10] を用いた．計算のために，式 (4) において， $\gamma_1 = 1.0 \times 10^5$ ， $\gamma_2 = 1.0 \times 10^{11}$ ， $\gamma_3 = 1.0 \times 10^1$ ， $\gamma_4 = 3.0 \times 10^2$  と設定した．

Fig. 3 に生成した辞書画像の一例，Fig. 4 にシーン画像，Fig. 5 に位置姿勢推定結果を示す．Fig. 5 においては，物体の位置を表す  $\xi_2$  を矩形で表し，各矩形に対応す



Fig. 1: Experimental environment

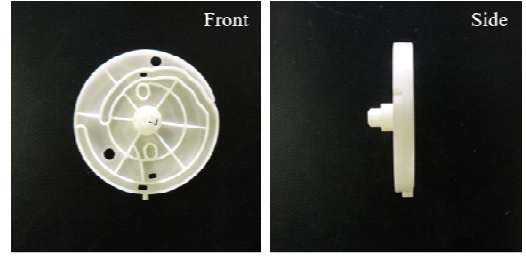


Fig. 2: Target parts

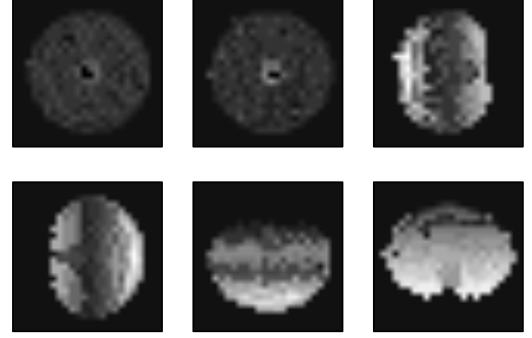


Fig. 3: Dictionary images

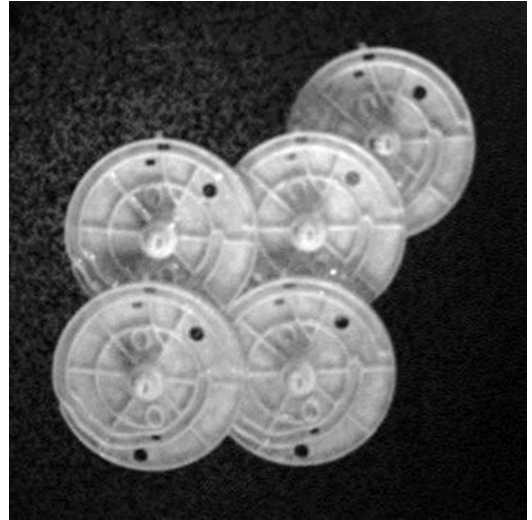


Fig. 4: Scene image

る辞書画像を右側に表示している．同図から提案法によって，位置姿勢推定が行えていることがわかる．結果として 4 つの物体の位置姿勢が提案法によって求められたが，これは最適化問題 (4) を複数回解くことによって求められた訳ではなく，1 回解いて求められることに注意する．

### 4. おわりに

本論文では，スパース推定を用いたばら積み物体の位置姿勢推定手法を提案した．位置姿勢推定問題を最適化問題

<sup>\*1</sup> 各辞書画像は物体の姿勢に依存するので， $\xi_1$  は姿勢を表すベクトルと考えることができる

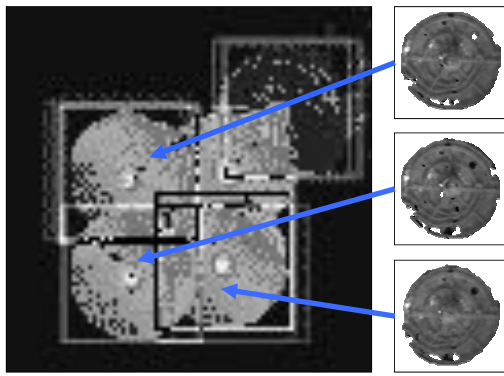


Fig. 5: Results of pose estimation

に帰着し，LASSO のよって解く方法を提案した．実験結果によって，複数の物体の位置姿勢推定が最適化問題を解くことによって同時に可能であることを示した．

## 参考文献

- [1] J.K. Oh, et al., “Development of a structured-light sensor based bin-picking system using ICP algorithm,” Proceeding of International Conference on Control, Automation and Systems, pp. 1673-1677, 2010.
- [2] H.Y. Kuo, et al., “3D object detection and pose estimation from depth image for robotic bin picking,” Proceeding of Conference on Automation Science and Engineering, pp. 1264-1269, 2014.
- [3] Y. Xiang, et al., “PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes,” arXiv:1711.00199, 2017.
- [4] A. Zeng, et al., “Multi-view self-supervised deep learning for 6D pose estimation in the Amazon Picking Challenge,” Proceeding of International Conference on Robotics and Automation, pp. 1386-1383, 2017.
- [5] F. Tombari, et al., “Unique Signatures of Histograms for Local Surface Description,” Proceeding of Conference on Image Processing, pp. 356-369, 2010.
- [6] B. Drost, et al., “Model globally, match locally: Efficient and robust 3D object recognition,” Proceeding of Computer Society Conference on Computer Vision and Pattern Recognition, pp. 998-1005, 2010.
- [7] D. Liu, et al., “Point Pair Feature-Based Pose Estimation with Multiple Edge Appearance Models (PPF-MEAM) for Robotic Bin Picking,” Sensors. 18. 2719. 10.3390/s18082719.
- [8] 福島 雅夫, “非線形最適化の基礎”, 朝倉書店, 2001.
- [9] 永原 正章, “スパースモデリング - 基礎から動的システムへの応用 -”, コロナ社, 2017
- [10] Blender, <https://www.blender.org/> (accessed Feb. 23th, 2019)