

3次元点群データの位置合わせのための深層学習を用いた部分的特徴抽出 Part-based Deep Feature Extraction of 3D Point Clouds for Registration

中央大学 ○顔 世荀, Pathak Sarthak, 梅田 和昇
Chuo University ○Shixun Yan, Sarthak Pathak, Kazunori Umeda

1. 序論

複数の距離センサを組み合わせることで、より高いロバスト性と汎用性なシステムを構築する研究はたくさん存在する。そのため、距離センサから得られた点群の位置合わせが重要な役割を担う、近年深層学習をベースにした点群を対象とした位置合わせ手法も提案されている[1][2]。しかし、点群を位置合わせする際に、範囲が異なる点群を対象とすると位置合わせするの精度が低下する傾向がある。点群同士のマッチング領域を計算する手法も提案されているが、ロバスト性などの点で不十分で、実用化がまだ遠い状況である[3]。

これらの問題を解決するために、本論文は深層学習を用いた位置合わせのため部分特徴抽出によるマッチング領域探索手法を提案する。マッチング領域探索によって位置合わせは簡単になる。

2. 関連研究

古くから沢山な位置合わせ手法が提案されている[4][5]。しかし、点群の範囲や姿勢など様々な要素で位置合わせの難易度が大きく変化する。近年、深層学習の普及によって点群ベースの深層学習モデルが沢山提案されている[6][7]。PointNet[6]は点群を対象としクラス分類とセマンティックセグメンテーションを目的とした最初の深層学習モデルである。簡単なニューラルネットワークを用いて、点群の順序不変性を実現し、高い性能の結果を得ている。PointNet[6]を用いた位置合わせ手法として PointNetLK[1]は一つである。PointNetLK[1]は PointNet[6]を用いて点群全体の特徴を抽出し、点群同士の特徴を合わせる剛体変換行列を求める。しかし、この手法は二つの点群がほとんど同じでなければならないという前提条件がある。点群が欠損しているあるいは片方がもう片方の一部の時、位置合わせ精度が大幅に落ちる。この問題を解決するために提案されたのが MaskNet[3]というモデルである。MaskNet[3]は片方の点群ともう片方の点群のマッチング領域を計算するモデルである。しかし、一つの点群と見なすため、範囲が大きく異なる点群同士或いは姿勢が大きく異なる点群同士に対応できず、マッチングが失敗することが多い。本研究で切り分けと部分的特徴抽出でロバストと高精度なマッチング領域探索可能な手法と提案する。

3. 提案モデル

本研究では、点群を切り分けることで独立した部分点群を作る。部分点群の特徴抽出することで、点群同士の精確なマッチング領域を探索する。以下のフローに従って、部分的特徴抽出及びマッチング探索を行う。

1. 点群 1 と点群 2 に対し、ダウンサンプリングを行うことで点群 1 と点群 2 を同様の密度にする。得られた点群を **part** と **full** と呼ぶ。 $\mathbf{part} = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$, $\mathbf{full} = \{\mathbf{f}_1, \dots, \mathbf{f}_g\}$ 。また、点 $\mathbf{p}_i = (x_i, y_i, z_i)$, $\mathbf{f}_j = (x_j, y_j, z_j)$ 。 n は **part** の点数であり、 g は **full** の点数である。

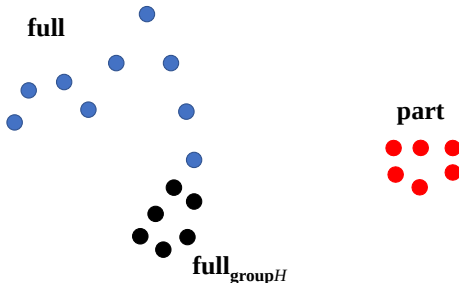


Fig. 1 Point cloud sample.

2. **full** を $\mathbf{full_group} = \{\mathbf{full_group}_x \mid x = 1, \dots, g\}$ のグルーピング処理を行う。グループ $\mathbf{full_group}_i$ は、点 $\mathbf{f}_i$ を中心とした最も近い n 点の部分点群である。
 3. **part** と $\mathbf{full_group}$ それぞれを PointNet[6] に入力してそれぞれの特徴 $\mathbf{Feature}_p = \{\mathbf{F}_p \mid \mathbf{F}_p \in \mathbb{R}^d\}$, $\mathbf{Feature}_f = \{\mathbf{F}_f \mid \mathbf{F}_f \in \mathbb{R}^d, i = 1, \dots, g\}$ を抽出する。
 4. それぞれの特徴を連結し、MLP (Multilayer Perceptron) に入力してスコアを計算する。MLP のサイズは (256, 128, 128, 1) とする。
 5. 一番高いスコア $\mathbf{Score}_H$ を選択し、それと対応する $\mathbf{full_group}_H$ を **part** とマッチングする領域と判断する。
- 点群のサンプルを Fig. 1 に示す。提案モデルを Fig. 2 に示す。

3.1. 部分的特徴抽出

PointNet[6]を用いてそれぞれの入力点群の特徴抽出を行う。本研究で用いる PointNet[6]の MLP サイズは (64, 128, 128, 512, 512) である。次に、それぞれの layer の結果を連結し、プーリングすることで、入力点群の大域的特徴を得る。部分点群 $\mathbf{full_group}$ を入力として、部分点群の大域的特徴を抽出することで **part** 点群とマッチングする領域が探索可能となる。

3.2. 損失関数

学習時には、クラス分類でよく使われる負の対数尤度損失を損失関数として使用する。

$$\text{loss}(\mathbf{Score}, \mathbf{target}) = -\log\left(\frac{\exp(\mathbf{Score}[\mathbf{target}])}{\sum_{i=1}^g \exp(\mathbf{Score}[i])}\right) \quad (1)$$

$\mathbf{target}$ は真値となる中心点 $\mathbf{f}_{g_i}$ の index である。

3.3. マッチング領域探索

PointNet[6]で出力された $\mathbf{Feature}_f$ から $\mathbf{Feature}_p$ と対応する領域を探索し、 $\mathbf{full_group}_H$ を導き出すのが本研究の目的である。マッチング領域探索の結果を評価するため、点群の位置推定を行う。

$$\mathbf{full_group}_H \approx \mathbf{part} + \mathbf{t} \quad (2)$$

$\mathbf{t}$ は並行移動行列である。

4. 実験

データセット ModelNet40[8]を使用し、学習を行った。ModelNet40[8]は 40 種類の CAD モデルからなる。データセットを半分にわけ、20 種類の学習セットで学習を行い、もう 20 種

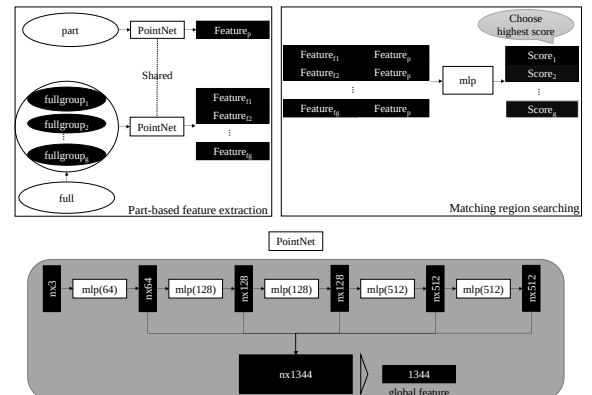
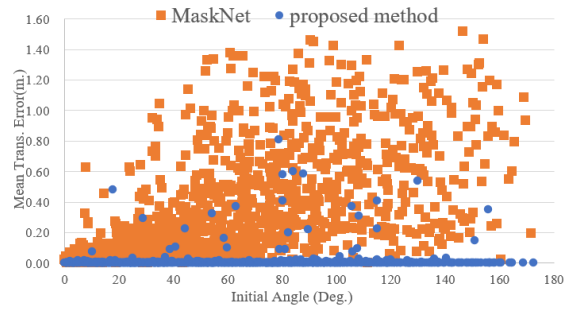
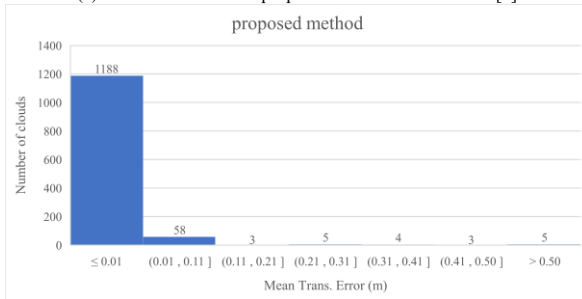


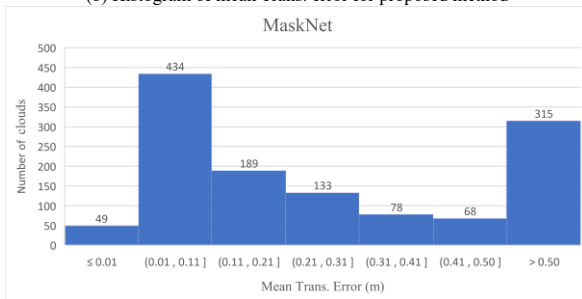
Fig. 2 Model architecture



(a) Mean Trans. error of proposed method and MaskNet[3]



(b) Histogram of mean Trans. error for proposed method



(c) Histogram of mean Trans. error for MaskNet[3]

Fig. 3 Experimental results comparing proposed method and MaskNet[3]

類のテストセットで評価を行う。点群 **full** には ModelNet40 のデータからリサンプリングした点群を用いた。点群 **part** は点群 **full** から 1/4 を切り取った後、平均 0 m、分散 0.05 m² の正規分布に従うノイズを加えた点群である。学習時の剛体変換行列の真値は、回転 [0, 90]°, 並行移動 [0, 1.57] m でランダムに生成した。テスト時の剛体変換行列の真値は、回転 [0, 180]°, 並行移動 [0, 3.14] m でランダムに生成した。Fig. 3(a)(b) に提案モデルの位置推定結果を示す。実験のサンプルを Fig. 4 に示す、また、MaskNet[3]の比較実験も行った。Fig. 3 と Table 1 に提案モデルと MaskNet[3]の比較実験結果を示す。

Fig. 3 が示されるように、MaskNet[3]は初期回転が 40°を超えた時位置推定誤差が急激に増大している。位置推定の成功率は 36.65%であり、全体の位置推定平均誤差が 0.326 m である。また、位置推定が成功した時の位置推定平均誤差は 0.043 m である。一方、モデルは全体的に安定的な結果を得られた。位置推定の成功率は 98.26%、全体の位置推定平均誤差は 0.008 m である。また、位置推定が成功した時の位置推定平均誤差は 0.002 m である。MaskNet[3]と比べ、提案モデルの位置推定精度は大幅に向上している。また、初期回転に対し、高いロバスト性を持っていることを示されている。

5. 結論

本論文は範囲や姿勢にロバスト性を持つ位置合わせのための深層学習に基づく部分的特徴抽出手法を提案した。このモデルでは、部分点群をグルーピングすることで、部分点群の大局的特徴を抽出可能にした。部分点群の大局的特徴からマッチング領域を探索

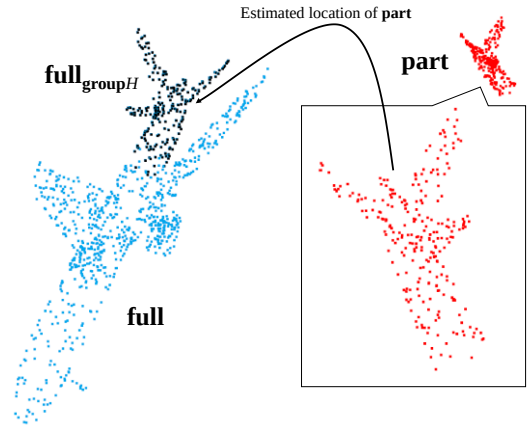


Fig. 4 Point cloud matching sample.

Table 1 Experimental results for each method

	success rate	Positional mean error (m)	Positional mean error on success (m)
MaskNet[3]	36.65%	0.326	0.043
Proposed model	98.26%	0.008	0.002

し、従来のモデルでは適用できなかった範囲の異なる点群のマッチングや位置推定に成功した。既存手法との比較実験により本手法の有用性を示した。今後は、実環境での有用性を検証し、データセットだけでなく実際距離画像センタから得られた点群を対象とした位置合わせを目標とする。

謝辞

本研究は JSPS 科研費 19H04191 の助成を受けたものである。

参考文献

- [1] Y. Aoki, H. Goforth, R. A. Srivatsan and S. Lucey, "PointNetLK: Robust & Efficient Point Cloud Registration Using PointNet," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7156-7165, 2019.
- [2] Y. He and C. Lee, "An Improved ICP Registration Algorithm by Combining PointNet++ and ICP Algorithm," 6th International Conference on Control, Automation and Robotics (ICCAR), pp. 741-745, 2020.
- [3] V. Sarode, A. Dhagat, R. A. Srivatsan, N. Zavallos, S. Lucey and H. Choset, "MaskNet: A Fully-Convolutional Network to Estimate Inlier Points," 2020 International Conference on 3D Vision (3DV), pp. 1029-1038, 2020.
- [4] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," Proceedings Third International Conference on 3-D Digital Imaging and Modeling, pp. 145-152, 2001.
- [5] P. Biber and W. Strasser, "The normal distributions transform: a new approach to laser scan matching," Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), vol.3, pp. 2743-2748, 2003.
- [6] C. R. Qi, H. Su, K. Mo and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation", Proc. Computer Vision and Pattern Recognition (CVPR), pp. 77-85, 2017.
- [7] Y. Li, R. Bu, M. Sun, W. Wu, X. Di and B. Chen, "Pointcnn: Convolution on x-transformed points", Advances in neural information processing systems, pp. 820-830, 2018.
- [8] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao: 3d shapenets, "A deep representation for volumetric shapes," In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1912-1920, 2015.