

# 深層学習を用いた多様体構築による 物体姿勢推定に関する検討

二宮 宏史<sup>1,a)</sup> 川西 康友<sup>1</sup> 出口 大輔<sup>1</sup> 井手 一郎<sup>1</sup>  
村瀬 洋<sup>1</sup> 小堀 訓成<sup>2</sup> 橋本 国松<sup>2</sup>

## 1. はじめに

ロボットの導入が産業分野だけでなく、介護・福祉、家事等の生活分野にも進みつつあり、ロボットが物体を掴むための技術が求められている。ロボットが物体を掴むには、対象物体を掴める方向を知るために、物体の姿勢推定を行う必要がある。

物体の姿勢推定手法として、あらかじめ対象物体を様々な角度から撮影した大量の画像とのテンプレートマッチングを行ない、最も類似した画像が対応する姿勢を、推定結果として出力する手法がある。しかし、多種多様な姿勢変化に対応するために、膨大な数のテンプレートを記憶しておく必要がある。

この問題に対し、Murase らは、姿勢変化による画像上での見えの変化を、主成分分析によって導出した低次元空間における多様体で表現するパラメトリック固有空間法を提案した [1]。多様体構築に用いたテンプレートの系列をキュービックスプライン等で補間することで未知の姿勢にも対応できるため、記憶しておくべきテンプレートを減らすことができる。

しかし、主成分分析による多様体構築では画像全体の見えに着目するため、図 1 のように姿勢変化量が同じだとしても見えが似ている場合に、低次元空間上でほぼ同一の点に写像される場合がある。その様子を図 2 に示す。このような多様体では近い点に写像される姿勢を分離することが困難なため、姿勢推定精度が低下してしまう。これは、主成分分析は画像全体の見えを考慮した教師なし学習であり、姿勢の分離性について考慮していないことが原因である。

そこで我々は、姿勢の分離性に着目した特徴量による多様体構築手法を提案する。姿勢を教師信号として学習した Deep Convolutional Neural Network (DCNN) [2] の中間層から、姿勢の分離性が高い特徴を抽出する。この特徴量を用いて多様体を構築する。その様子を図 3 に示す。

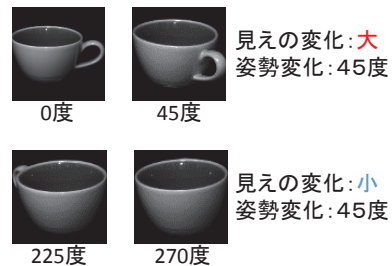


図 1 見えの変化と姿勢変化

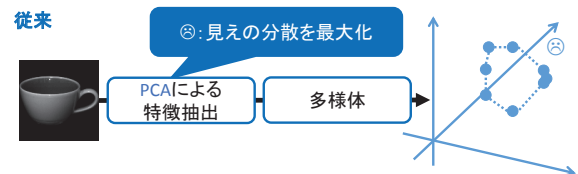


図 2 PCA による多様体構築

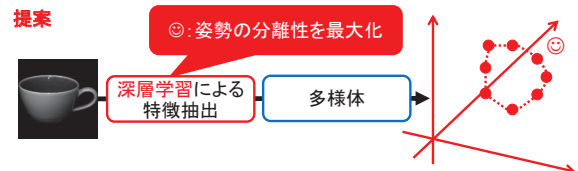


図 3 深層学習による多様体構築

DCNN は学習過程において識別に適した特徴量を自動的に獲得することができ、一般物体認識やシーン認識など様々なベンチマークで高い性能を示している [3]。そのため、図 1 に示すように、画像全体の見えの分散を最大化する教師なし学習である主成分分析では区別できない見えの変化が小さい姿勢の違いであっても、姿勢を教師信号として教師あり学習を行った DCNN を用いて抽出した姿勢の分離性の高い特徴量を用いることで区別することができる。

## 2. 深層学習を用いた多様体構築による 姿勢推定

姿勢を教師信号として学習した DCNN から抽出した特徴量により多様体を構築し、入力画像と最も近い多様体上の 1 点を求めることによって、物体姿勢推定を行う。以上

<sup>1</sup> 名古屋大学, 〒 464-8601 愛知県名古屋市千種区不老町

<sup>2</sup> トヨタ自動車株式会社 〒 471-8571 愛知県豊田市トヨタ町 1

<sup>a)</sup> ninomiyah@murase.m.is.nagoya-u.ac.jp

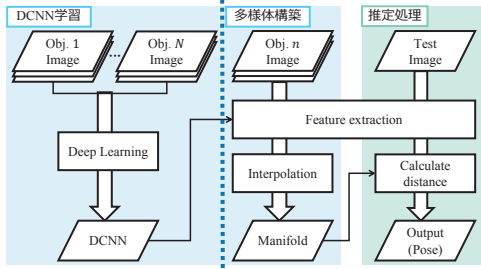


図 4 提案手法の処理手順

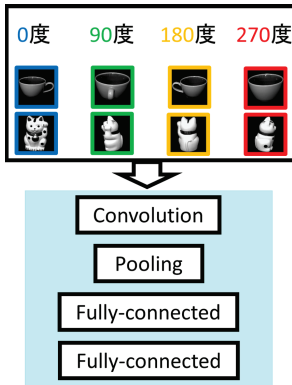


図 5 姿勢を教師信号とした DCNN の学習手順

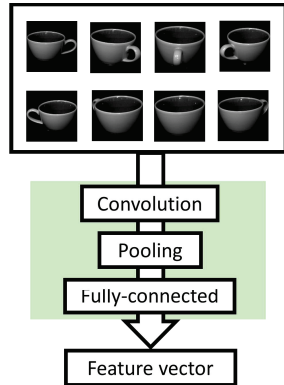


図 6 学習済み DCNN からの 特徴抽出手順

の流れを図 4 に示す。

まず特徴抽出のために、深層学習モデルの 1 つである DCNN の学習を行う。学習サンプルには、推定対象物体を任意の回転軸に従って、一定の角度ごとに回転させた画像を用いる。ここで、図 5 に示すように、教師信号として各物体の姿勢を与える。これにより、姿勢の分離性に着目した学習が行われる。そして、学習した DCNN に再び学習サンプルを入力し、中間層の出力を特徴量として抽出する。その流れを図 6 に示す。このように抽出した特徴量に対し、パラメトリック固有空間法と同様に補間処理を行い、多様体を構築する。

姿勢推定を行うには、入力画像を学習した DCNN に入力し、特徴抽出を行う。そして、多様体上の各点に対して距離計算を行い、最も近い 1 点が示す姿勢を推定結果として出力する。

### 3. 実験

提案手法の有効性を確認するため、公開データセットを用いた物体の姿勢推定実験を行なった。なお、本報告では予備実験として鉛直軸周りの姿勢推定を行なった。

公開データセットには Columbia Object Image Library (COIL-20) [4] を用いた。推定対象物体は 20 種類で、それぞれについて鉛直軸に沿って 5 度刻みで姿勢を変化した画像があり、画像の総数は 1,440 枚である。

評価は学習サンプルを以下のように分割し、2 分割交差

表 1 多様体による姿勢推定評価値

	平均絶対誤差
Pixel	1.16 度
PCA (パラメトリック固有空間法)	1.39 度
DCNN (提案手法)	1.09 度

検定を行った。

- セット 1 : 0 度, 10 度, 20 度, ..., 350 度
- セット 2 : 5 度, 15 度, 25 度, ..., 355 度

また、比較手法として、画素値をそのまま並べて特徴ベクトルとしたものと、そのベクトルの集合に対して主成分分析を行ない導出した固有ベクトルを用いて構築した多様体による姿勢推定を行った。

実験結果を表 1 に示す。提案手法が最も高精度であることがわかる。Pixel 特徴や PCA 特徴は画像全体の見えを考慮した特徴であり、姿勢の分離性を考慮していないため、見えの変化が小さい姿勢の違いを区別することが難しく、精度が低かった。それに対し、DCNN 特徴は姿勢を教師信号とした教師あり学習により抽出した特徴であるため、見えの変化が小さい姿勢の違いも区別でき、高精度となった。

以上の結果より、提案手法の有効性を確認した。

### 4. むすび

本報告では、深層学習を用いた多様体構築により物体姿勢を高精度に推定する手法について検討した。姿勢を教師信号として深層学習モデルを学習することで姿勢の分離性の高い特徴量を抽出し、その特徴量を用いて多様体を構築することで、見えの変化が小さい姿勢の違いも区別できる手法を提案した。評価実験より、多様体構築に主成分分析を用いる従来手法等と比較して、高精度に姿勢を推定できることを確認した。

今後の課題として、学習パラメータについて調査を進めることや、複雑背景・照明変動に対する頑健性の調査等が挙げられる。

### 謝辞

本研究の一部は、科学技術研究費補助金による。

### 参考文献

- [1] H. Murase et al., "Visual learning and recognition of 3-D objects from appearance," Int. J. Computer Vision, vol.14, no.1, pp.5-24, Jan. 1995.
- [2] A. Krizhevsky et al., "ImageNet classification with deep convolutional neural networks," Proc. 26th Annual Conf. on Neural Information Processing Systems, pp.1097-1105, June 2012.
- [3] A.S. Razavian et al., "CNN features off-the-shelf: An astounding baseline for recognition," Proc. 2014 IEEE Computer Society Conf. of Computer Vision and Pattern Recognition, pp.806-813, June 2014.
- [4] S.A. Nene et al., "Columbia Object Image Library (COIL-20)," Technical Report, Columbia Univ. Dept. of Computer Science, CUCS-005-96, Feb. 1996.