一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
PRMU2019-62(2019-12)

距離画像からの物体姿勢推定のための遮蔽に頑健な姿勢表現の学習

立道　大樹†　川西　康友†　出口　大輔†　井手　一郎†　村瀬　洋†
安間　絢子††

† 名古屋大学 大学院 情報学研究科
〒464-8601 愛知県名古屋市千種区不老町
†† トヨタ自動車株式会社
〒471-8571 愛知県豊田市トヨタ町 1
E-mail: †tatemichih@murase.is.i.nagoya-u.ac.jp, ††{kawanishi, ide, murase}@i.nagoya-u.ac.jp,
†††ddeguchi@nagoya-u.jp, ††††ayako_amma@mail.toyota.co.jp

あらまし　ロボットが物体を持ち運ぶためには，物体の姿勢を推定する必要があるが，物体が他の物体に遮蔽される場合，観測値の欠損と，観測できた部分を囲む矩形で切り出した距離画像中での物体中心の位置ずれが生じ，姿勢推定が困難である．本発表では，欠損を復元できるよう学習した AutoEncoder により，位置ずれの補正と欠損の復元をすることで得られる姿勢表現を用いた，遮蔽に頑健な姿勢推定を提案する．

キーワード　距離画像，遮蔽，物体姿勢推定，AutoEncoder

Occlusion-Robust Pose-Feature Representation Learning for Object Pose Estimation from a Depth Image

Hiroki TATEMICHI†, Yasutomo KAWANISHI†, Daisuke DEGUCHI†, Ichiro IDE†,
Hiroshi MURASE†, and Ayako AMMA††

† Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
†† Toyota Motor Corporation
Toyota-cho, Toyota-shi, Aichi, 471–8571 Japan
E-mail: †tatemichih@murase.is.i.nagoya-u.ac.jp, ††{kawanishi, ide, murase}@i.nagoya-u.ac.jp,
†††ddeguchi@nagoya-u.jp, ††††ayako_amma@mail.toyota.co.jp

Abstract　In order for a robot to carry an object, it is necessary to estimate the pose of the object. However, when an object is occluded by other objects, the observation is partially lost and the object center is shifted in the cropped depth image. Therefore, the pose estimation of an occluded object is difficult. In the proposed method, we realize the pose estimation from the pose representation obtained by de-occluding and centerizing the object in the image via AutoEncoder that have been trained so as to de-occlude it.

Key words　Depth Image, Occlusion, Object Pose Estimation, AutoEncoder

1. はじめに

近年，産業用ロボットや日常生活支援のための生活支援ロボットの開発が進んでいる．生活支援ロボットには，マグカップなど人が指示した物体を把持して運ぶという基本機能を備える必要がある．このロボットによる物体把持は，ロボット分野における重要課題となっている．物体を把持するためには，対象物体の認識だけでなく，その姿勢を詳細に推定する必要がある．しかし，机上など物体が密集して置かれている状況では，対象物体が他の物体に遮蔽されることで姿勢推定が困難となる．本研究では，このような遮蔽された物体の姿勢推定問題に取り組む．

ロボットには周囲を観測するために，主に RGB 画像センサや距離画像センサが搭載される．特に距離画像センサは色や照

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.
Copyright ©201　by　IEICE

(a) 遮蔽されていないマグカップ　　(b) 遮蔽されたマグカップ

図 1　遮蔽による物体の見えの違い

(a) 欠損のない距離画像　　(b) 欠損のある距離画像

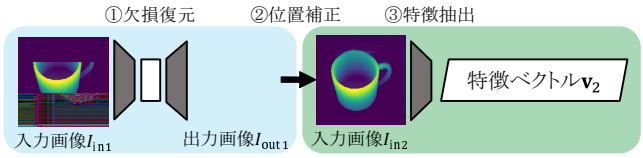
図 2　欠損による位置ずれ

明条件の変化に頑健であり，物体の切り出しも容易であるため，近年ロボットに搭載されることが多くなっている．本研究では，このセンサにより観測した距離画像からの物体姿勢推定を扱う．

図 1(b) に遮蔽されたマグカップの例を示す．本研究では，このような机上に置かれた物体を想定し，物体の上下左右いずれかが他の 1 つの物体に遮蔽され，最大で半分程度観測値に欠損が生じた物体画像を入力とする．また本研究では，特定の物体ではなくマグカップなど 1 つの物体クラスを対象とする．

画像から物体姿勢推定を行うには，観測画像全体のうち物体領域を抽出し，その領域が中心となるような矩形で切り出した画像を入力とすることが一般的である（図 2）．Ninomiya らは同一クラスに属する複数の物体の距離画像を用いて畳み込みニューラルネットワーク（Convolutional Neural Network; CNN）に基づく回帰モデルを学習し，そのネットワークを用いて特徴抽出することで，様々な形状のマグカップなど同一クラスに属する形状が未知の物体の未知姿勢に対しても高精度な姿勢推定手法を提案した [1]．しかし，この手法は遮蔽された物体のように，観測した距離画像中の物体に欠損が含まれる場合は想定していない．そのため，遮蔽による観測値の欠損が含まれる距離画像を入力とすると，特徴量が変化し精度良く姿勢推定できない．

このように，観測した距離画像中の物体に欠損が含まれていても，欠損していない元の形状が分かれば，本来の特徴量を用いて高精度な姿勢推定が可能と考えられる．この考えに基づき，Sundermeyer ら [2] は，AutoEncoder [3] を利用した Augmented AutoEncoder（AAE）を提案し，物体が一部欠損している，物体の背景に乱れがある，さらに照明条件が異なる物体画像から，対象物体だけを表現するような特徴量の抽出を実現した．この手法では，観測値に欠損が含まれるとき，欠損箇所を復元するよう AE を学習させることの有効性が示されている．しかし，遮蔽されることで，真の物体中心の位置が画像



(a) 2 段階 AE 法

(b) End-to-end 法

図 3　提案する 2 種類の特徴抽出法

図 4　提案手法の処理手順

中心からずれるという問題点には対処していないため，欠損が大きいほど姿勢推定精度が低下する．また，学習データにない形状の物体の姿勢推定には対応していない．

これらの問題点に対し，本研究では，欠損と，それによる位置ずれに対処できる欠損復元 AutoEncoder（AE）を提案する（図 3）．その実現方法として，一旦 AE を用いて欠損を復元したうえで位置ずれを修正し，再度 AE により特徴抽出する手法（2 段階 AE 法）を提案する．また，この処理を End-to-end に行う手法（End-to-end 法）も提案する．このとき，様々な形状のマグカップを学習データとすることで，学習データに存在しない形状の物体に対しても姿勢を推定できる手法を実現する．

以下，2 節で遮蔽に頑健な姿勢表現の学習と姿勢推定について述べる．続いて，3 節で CG データに対する評価について述べる．さらに，4 節で実画像に対する評価について述べる．最後に，5 節でまとめと今後の課題について述べる．

2.　遮蔽に頑健な姿勢表現の学習と姿勢推定

2.1　概　　要

他の物体に遮蔽された物体の姿勢推定を考えるとき，距離画像センサにより観測される物体画像には，遮蔽された部分の距離値が不明，物体の範囲が分からないために真の物体中心と入力画像の中心が一致しない，という問題点がある．これらの問題点により，対象物体の全体像を観測できる場合に対して特徴量にずれが生じるため，高精度な姿勢推定は難しい．本研究で

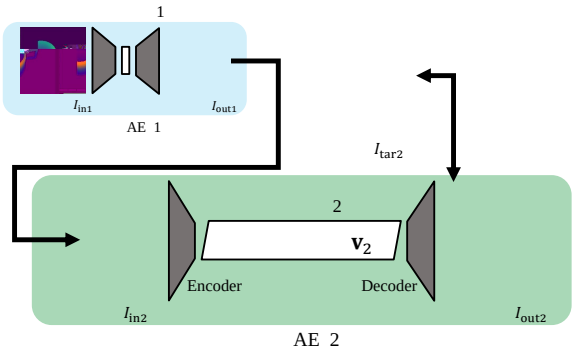


図 5　欠損復元 AutoEncoder（1 段目）の学習手順

図 6　欠損復元 AutoEncoder（2 段目）の学習手順

は，これらの問題点を同時に解決する姿勢表現の学習方法を提案する．

提案手法の処理手順を図 4 に示す．提案手法は，姿勢表現の学習及び，姿勢推定器の学習からなる学習段階と，推定対象の距離画像から姿勢推定を行うまでの推定段階からなる．

姿勢表現の学習では，欠損復元 AE を学習する．その手法として 2.2 にて図 3(a)(b) に示す 2 通りの手法を提案する．姿勢推定器の学習では，まず姿勢推定の学習用距離画像から欠損復元 AE を用いて特徴抽出を行う．このとき抽出される特徴ベクトルは，欠損がない物体が画像中心にあるような画像を表す特徴量となっていることが期待される．そして，パラメトリック固有空間法 [4] に倣って，連続的に姿勢を変化させた各画像に対し，抽出した特徴ベクトルを主成分分析（Principal Component Analysis; PCA）に基づき低次元特徴空間へと射影し，特徴空間中で姿勢変化に関する多様体を得る．

姿勢推定段階では，まず推定対象の距離画像から，学習用距離画像と同様に欠損復元 AE を用いて特徴抽出を行い，その特徴量を低次元特徴空間中へ射影する．そして，この特徴ベクトルと多様体との距離計算を行い，多様体上で最近傍のベクトルに対応する姿勢を推定結果として出力する．

以下，手法の詳細について述べる．

2.2　姿勢表現の学習

2.2.1　2 段階 AE 法による姿勢表現学習

2 段階 AE 法では，図 3(a) に示すようにまず 1 段目の AE を用いて，物体の位置は保ったまま，欠損を含む画像の欠損部分を復元する．そして復元した物体の中心位置を画像中心に補正し，2 段目の AE の Encoder 部に入力することで，出力として特徴量を得る．この 2 段階 AE 法における 1 段目と 2 段目の欠損復元 AutoEncoder の学習方法について述べる．

図 5 に，欠損復元 AutoEncoder（1 段目）の学習手順を示す．1 段目の AE ではまず，大きさ $H \times W$ 画素の元画像 I_{org} に対し，欠損させる方向を上下左右いずれかから無作為に選択する．続いて，最大欠損率 $N\%$ の範囲で無作為に欠損率を決め，物体を囲む矩形領域のうち，選択した欠損方向から，欠損率に応じた幅の距離値を背景と同じ距離値に変更することで遮蔽による欠損を模擬する．そして欠損していない部分を囲う縦横比 $H:W$ の最小矩形領域を大きさ $H \times W$ 画素に拡縮し，画像の中心にくるよう $(\Delta x, \Delta y)$ だけ平行移動させる．この画像について，復元後の画像に物体の大部分が含まれるよう，背景部分を拡張して，大きさ $H' \times W'$ 画素の入力画像 I_{in1} とする（図 5 ①）．大きさ $H' \times W'$ の背景のみからなる画像に対し，入力画像 I_{in1} における欠損していない部分が同じ位置にくるように，元画像 I_{org} を $(\Delta x, \Delta y)$ だけ平行移動して配置したものを目標画像 I_{tar1} とする（図 5 ②）．そして，画像 I_{in1} を入力，欠損部分を復元した画像 I_{tar1} を正解の出力とする AutoEncoder（欠損復元 AE（1 段目））を学習する（図 5 ③）．出力画像 I_{out1} と目標画像 I_{tar1} の誤差を評価（図 5 ④）し，この誤差を最小化するように，誤差逆伝播法により欠損復元 AE（1 段目）のパラメータを最適化する（図 5 ⑤）．

図 6 に，欠損復元 AutoEncoder（2 段目）の学習手順を示す．2 段目の AE では，前述の出力画像 I_{out1} を入力とする（図 6 ①）．しかしこの出力画像 I_{out1} では，物体は欠損の影響で中心位置からずれて復元される．そのため，まず物体位置の補正を行う．物体位置の補正では，出力画像 I_{out1} に対して距離値に関する 2 値化処理を施し，物体領域と背景領域に分ける．次に，この物体領域を囲む縦横比 $H:W$ の最小矩形を求め，その矩形の中心位置を，対象物体の中心位置とする．そして，求めた矩形領域を大きさ $H \times W$ 画素に拡縮し，対象物体の中心位置が画像の中心にくるように出力画像 I_{out1} を $(\Delta x', \Delta y')$ だけ平行移動し，入力画像 I_{in2} を得る（図 6 ②）．このとき，欠損復元 AutoEncoder（1 段目）と同様に入力画像 I_{in} における欠損していない部分が同じ位置にくるように，大きさ $H' \times W'$ の背景のみからなる画像に元画像 I_{org} を同じ倍率で拡縮し，$(\Delta x + \Delta x', \Delta y + \Delta y')$ だけ平行移動して配置したものを目標画像 I_{tar2} とする（図 6 ②）．そして，画像 I_{in2} を入力，欠損部分を復元した画像 I_{tar2} を正解の出力とする AutoEncoder（欠損復元 AE（2 段目））を学習する（図 6 ③）．出力画像 I_{out2} と目標画像 I_{tar2} の誤差を評価（図 6 ④）し，この誤差を最小化するように，誤差逆伝播法により欠損復元 AE（2 段目）のパラメータを最適化する（図 6 ⑤）．

2.2.2　End-to-end 法による姿勢表現学習

End-to-end 法では，図 3(b) に示すように，位置ずれの補正と欠損の復元を学習した AE の Encoder 部に，欠損を含む画像を入力すると，出力として特徴量を得る．この End-to-end 法の学習方法について述べる．

表 4　評価用データセットの概要

俯角 [°]	10	30	50
実モデル数	4 種類		
姿勢間隔	0 始点に 10 刻み		
姿勢パターン数	27	27	36
合計枚数 [枚]	108	108	144

表 5　実画像に対する欠損復元結果例の比較

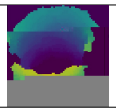
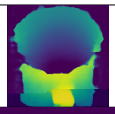
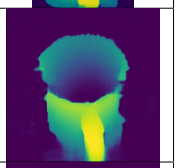
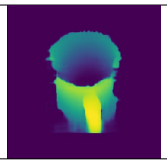
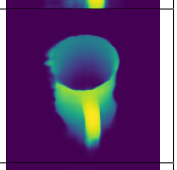
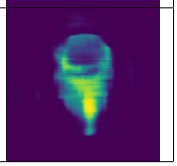
	入力画像 I_{in}	復元結果 I_{out}
比較手法 2		
提案手法 1（1 段目）		
提案手法 1（2 段目）		
提案手法 2		

表 6　実画像に対する姿勢推定結果

	平均絶対角度誤差 [°]			5° 以内割合 [%]		
俯角 [°]	10	30	50	10	30	50
比較手法 1	57.67	32.40	57.08	8.3	14.8	9.7
比較手法 2	**40.24**	29.08	34.19	14.8	**28.7**	19.4
提案手法 1	48.33	**28.75**	**31.51**	**16.7**	**28.7**	**25.7**
提案手法 2	47.31	44.85	43.92	12.0	13.0	16.0

他，この評価用データセットの詳細を表 4 に示す．

4.2　前　処　理

YOLOv3 [6] を用いてカラー画像からマグカップを検出し，その検出枠を得た．さらに，取得された距離画像に対し，ターンテーブルの領域からサンプリングした距離値を用いて最小 2 乗法により平面を算出し，ターンテーブルの領域の距離値を取り除いた．以上の処理により，マグカップ領域のみの距離画像を生成し，これを $H = 128$，$W = 128$ 画素の大きさに拡縮した．以後の処理，及び実験条件については 3 節と同様である．

4.3　実 験 結 果

まず，欠損復元 AE による俯角 50° の欠損復元結果例の比較を表 5 に示す．表 5 より，提案手法 1 では距離値が平滑化されたマグカップが復元されているが，提案手法 2 ではうまく復元できていないことが分かる．

次に，実画像に対する姿勢推定結果の比較を表 6 に示す．まず，各手法の精度を比較すると，提案手法 1 と比較手法 2 の精度がほぼ同程度で，わずかに提案手法 1 が優位な結果となった．一方で，提案手法 2 の精度はほとんどの項目で提案手法 1 の精度を下回った．これは，提案手法 2 では 1 段目の AE で距離値が平滑化された画像を用いて 2 段目の AE で特徴抽出するため，抽出された特徴量がより CG データからの特徴量に近づくからであると考えられる．

続いて，CG データによる姿勢推定結果（表 3）と比較すると，全ての手法で平均絶対角度誤差は大きく増加し，絶対角度誤差の 5° 以内割合は低下した．この要因として，撮影条件による観測値の欠損が挙げられる．距離画像センサは，光沢がある面や，センサから見た勾配が大きい面の距離値を観測することができない．それにより，遮蔽が要因ではない距離値の欠損が生じてしまう．提案手法を実用的な手法とするためには，このような実画像に特有の問題を解決するための改良が必要である．

5.　む　す　び

本研究では，遮蔽により観測値に欠損が生じた物体に対して欠損復元 AE を適用することで欠損と位置ずれの問題に対処し，高精度な姿勢推定を実現した．CG データに対する評価実験より，提案手法による姿勢推定精度の向上を確認した．さらに，提案手法について実画像に対する評価を行い，実用化に向けた問題点を明確化した．

今後は，この実画像に対する評価を元に，実用化に向けた手法の改良を進めるとともに，特徴抽出手法の改良，姿勢の回転軸数の拡張などについても検討する．

謝辞 本研究の一部は科学研究費 JP17H00745 の助成による．

文　　献

[1] H. Ninomiya, Y. Kawanishi, D. Deguchi, I. Ide, H. Murase, N. Kobori, and Y. Nakano, “Deep manifold embedding for 3D object pose estimation,” Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol.5, pp.173–178, March 2017.

[2] M. Sundermeyer, Z.-C. Marton, M. Durner, M. Brucker, and R. Triebel, “Implicit 3D orientation learning for 6D object detection from RGB images,” Proceedings of the 15th European Conference on Computer Vision, pp.699–715, Sept. 2018.

[3] G. Hinton and R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol.313, no.5786, pp.504–507, July 2006.

[4] H. Murase and S.K. Nayar, “Visual learning and recognition of 3-D objects from appearance,” International Journal of Computer Vision, vol.14, no.1, pp.5–24, Jan. 1995.

[5] A.X. Chang, T.A. Funkhouser, L.J. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, “ShapeNet: An information-rich 3D model repository,” Computing Research Repository, arXiv preprint, arXiv:1512.03012, Dec. 2015.

[6] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” Computing Research Repository, arXiv preprint, arXiv:1804.02767, April 2018.