欠損復元 AutoEncoder による 遮蔽に頑健な物体姿勢推定の検討

立道 大樹[1,a)]　川西 康友[1,b)]　出口 大輔[1,c)]　井手 一郎[1,d)]　村瀬 洋[1,e)]　安間 絢子[2,f)]

概要

ロボットが物体を把持するには，物体の姿勢推定が必要である．本研究では，遮蔽された物体の姿勢推定という特に困難な問題に取り組む．遮蔽された物体は，遮蔽された部分を観測できない上に，真の物体中心が分からないという問題点がある．これに対し，欠損を復元できるように学習した AutoEncoder を用いて，遮蔽による欠損に頑健な姿勢推定を実現するための特徴抽出法を提案する．また，同種の物体であっても，学習データにない形状や姿勢をした物体の姿勢推定は困難である．これに対し，パラメトリック固有空間法を拡張した姿勢補間手法を提案する．実験により，提案手法により観測値に欠損が生じた物体の高精度な姿勢推定ができることを示す．

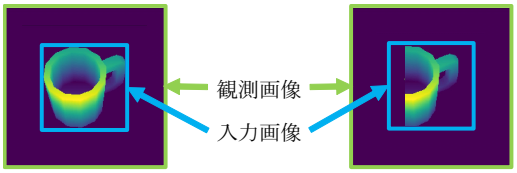
1. はじめに

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，受け渡すという基本機能を備える必要がある．このような，ロボットによる物体把持は，ロボット開発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，机上など物体が密に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ている．特に後者は色や照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，この距離センサにより観測した距離画像から

(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) 欠損がない距離画像　(b) 欠損がある距離画像

図 2　欠損による位置ずれ

の物体姿勢推定を検討する．

図 1(b) に遮蔽されたマグカップの例を示す．このような机上に置かれた物体を想定し，本研究では，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して想定する．また，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未知形状の物体の未知姿勢であっても高精度な姿勢推定手法を提案した [4]．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．遮蔽があると，観測値に欠損が生じ，またその結果，真の物体中心が画像中心からずれるという問題が生じる．そのため，特徴量が変化し精度良く姿勢推定できない．

1　名古屋大学
2　トヨタ自動車株式会社
a)　tatemichih@murase.is.i.nagoya-u.ac.jp
b)　kawanishi@i.nagoya-u.ac.jp
c)　ddeguchi@nagoya-u.jp
d)　ide@i.nagoya-u.ac.jp
e)　murase@i.nagoya-u.ac.jp
f)　ayako_amma@mail.toyota.co.jp

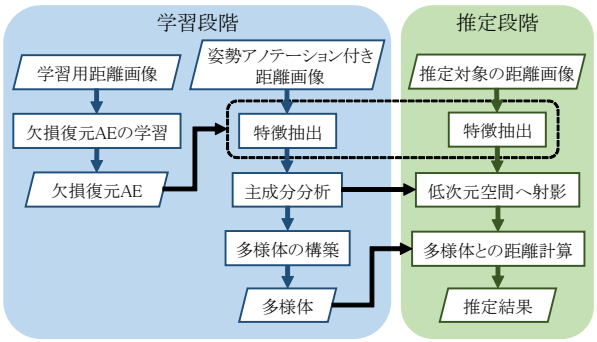


図 3 提案手法の処理手順

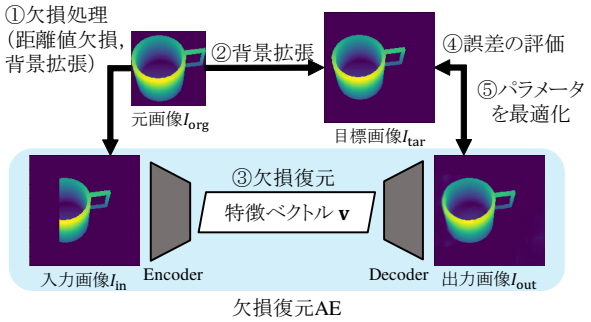


図 4 欠損復元 AutoEncoder の学習手順

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 [5] は，Augmented AutoEncoder（AAE）を提案し，物体に一部欠損が含まれる，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，真の物体中心が分からず，その位置が画像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，同一クラスに属する物体でも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．欠損復元 AE により欠損復元した画像を元に位置ずれを補正し，再度欠損復元 AE を適用することで欠損とそれによる位置ずれを補正した特徴抽出を実現し，高精度な姿勢推定を実現する．また，パラメトリック固有空間法 [3] を拡張した姿勢補間手法により，学習データに存在しない形状・姿勢の物体であっても姿勢推定ができる手法を提案する．

以下，2 節で欠損復元 AE による特徴抽出手法と，特徴空間中での姿勢補間と推定について述べる．続いて，3 節で提案手法の有効性を評価した実験について述べる．最後に，4 節でまとめと今後の課題について述べる．

2. 提案手法

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，遮蔽された部分の距離値が分からない，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，姿勢推定がより困難である．本研究では，これらの問題点を同時に解決する特徴抽出法を提案する．また，同一クラスに属するが学習データにない形状・姿勢をした物体の姿勢を推定できるようにするために，パラメトリック固有空間法を拡張した姿勢補間手法を提案する．提案手法の処理手順を図 3 に示す．

学習段階ではまず，欠損復元 AE の学習用距離画像から，距離画像の欠損部分を復元するよう欠損復元 AE を学習する．次に，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．特徴抽出には，前述の問題点を解決するために，欠損復元 AE により概形を推定し，それに基づいて物体中心のずれを推定する．そして，物体中心の位置を画像中心に合わせた画像を再び同じ欠損復元 AE の Encoder 部に入力する．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時の画像を表す特徴量となっていることが期待される．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．このとき，姿勢推定の学習用距離画像に含まれない姿勢を補間することで，未知の姿勢に対する姿勢推定をできるようにする．

姿勢推定段階ではまず，推定対象の距離画像から，姿勢推定の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 欠損復元 AutoEncoder の学習

図 4 に，欠損復元 AE の学習手順を示す．まず，大きさ $H \times W$ 画素の元画像 I_{org} 中の物体の上下左右いずれかを無作為に選択する．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，選択した欠損側から，欠損率分の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が画像の中心に配置されるよう $(\Delta x, \Delta y)$ 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が完全に含まれるよう，背景部分を拡張して，大きさ

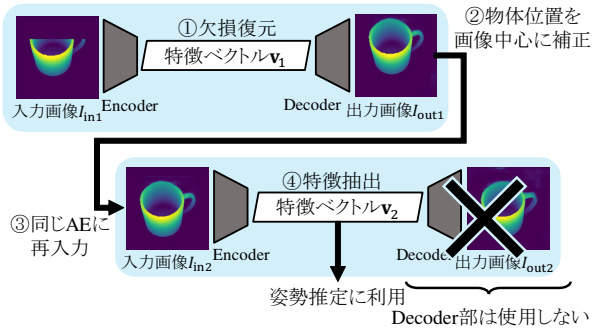


図 5　特徴抽出の処理手順

$H' \times W'$ 画素の入力画像 I_{in} とする（図 4 ①）．このとき，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配置されるように，大きさ $H' \times W'$ 画素の背景のみからなる画像に元画像 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_{tar} とする（図 4 ②）．

続いて，欠損復元 AE に画像 I_{in} を入力し，特徴量 $\mathbf{v}$ が計算され，Decoder 部から欠損部分を復元した画像 I_{out} を得る（図 4 ③）．そして，出力画像 I_{out} と目標画像 I_{tar} の誤差を評価する（図 4 ④）．最後に，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 のパラメータを最適化する（図 4 ⑤）．

2.3　欠損復元 AutoEncoder による特徴抽出

図 5 に，特徴抽出の処理手順を示す．まず，欠損していない部分を中心とした距離画像 I_{in1} を欠損復元 AE に入力し，Decoder 部分から欠損部分を復元した出力画像 I_{out1} を得る（図 5 ①）．しかし，出力画像 I_{out1} において物体は欠損の影響で中心位置からずれて復元される．そのため物体位置の補正を行う．物体位置の補正では，まず出力画像 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，対象物体の中心位置が画像の中心に配置されるように出力画像 I_{out1} を平行移動し，入力画像 I_{in2} を得る（図 5 ②）．これを同じ欠損復元 AE の Encoder 部に入力し，特徴ベクトル $\mathbf{v}_2$ を得る（図 5 ③④）．この特徴ベクトル $\mathbf{v}_2$ を姿勢推定に用いる．

2.4　特徴空間中での姿勢補間

図 6 に，多様体構築と姿勢推定の処理手順を示す．欠損復元 AE の中間層から抽出される特徴ベクトルは，元の画像を復元できるだけの情報を持つため，高次元である．その高次元特徴ベクトルから多様体を構築した場合，特徴空間中での姿勢補間や最近傍探索の計算量が大きい，また次元の呪いにより最近傍点と最遠隔点との差が小さくなり近傍点の探索が困難であるという問題点がある．これらの問題点に対処するため，PCA により特徴量を低次元空間へ

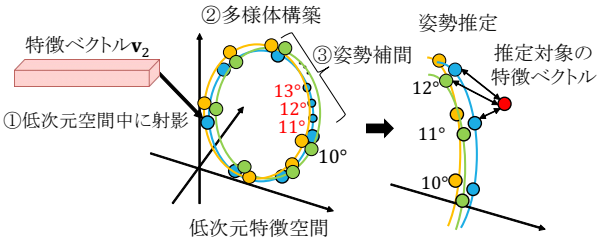


図 6　多様体構築と姿勢推定

射影する．PCA を施す前の特徴ベクトル $\mathbf{v}_2$ の集合を $\mathcal{V}$ で表現する．

$$\mathcal{V} = \{\mathbf{v}_2\} \tag{1}$$

PCA により，低次元空間への写像 U を得る．この U を用いて低次元化した特徴ベクトルの集合 $\mathcal{V}_{\text{d}}$ は以下の式で表現される．

$$\mathcal{V}_{\text{d}} = \{\mathbf{u} | \forall \mathbf{v}_2 \in \mathcal{V}, \mathbf{u} = U\mathbf{v}_2\} \tag{2}$$

次に，低次元特徴空間中で，ある物体の特徴ベクトルを回転順につなぐように 3 次スプライン曲線で補間し，1 つの多様体を構築する．同様の処理を全ての物体に対して施すことにより，複数の多様体が得られる．図 6 では，異なる物体の特徴ベクトルを異なる色で示している．

2.5　姿勢推定

まず，推定対象の距離画像から，図 5 に示す欠損復元 AE の 2 回適用により特徴ベクトル $\mathbf{v}'_2$ を抽出し，学習段階で生成した低次元空間に射影し，$\mathbf{u}' = U\mathbf{v}'_2$ を得る．そして，2.4 で得た多様体上の点で $\mathbf{u}'$ に最も近いベクトルを探索する．本研究では，近似的に全ての多様体上から姿勢に関して一定間隔にサンプリングしたベクトルのうち最近傍のベクトル $\hat{\mathbf{u}}$ を探索する．推定対象の特徴ベクトル $\mathbf{u}'$ から見て最近傍の特徴ベクトル $\hat{\mathbf{u}}$ を求め，それに対応する姿勢を推定結果として出力する．

$$\hat{\mathbf{u}} = \underset{\mathbf{u} \in \mathcal{V}_{\text{a}}}{\arg\min} \|\mathbf{u} - \mathbf{v}\| \tag{3}$$

ただし，$\mathcal{V}_{\text{a}}$ は全ての多様上からサンプリングしたベクトルの集合である．

3.　実験

3.1　データセット

本実験では，ShapeNet [1] に含まれるマグカップの 3D モデルを使用した．各マグカップの 3D モデルについて，ロール角を 0°，ピッチ角を 60° で一定とし，ヨー角のみを変化させながら仮想的に距離画像を生成した．欠損復元 AE の学習には，100 種類のマグカップについて，ヨー角を 1° 刻みに回転させながら生成した計 36,000 枚の画像を使用した．姿勢推定の学習には，同じ 100 種類のマグカッ

プについて，ヨー角を 0° 始点に 10° 刻みに回転させながら生成した計 3,600 枚の画像を使用した．姿勢推定精度の評価には，学習用とは異なる形状の 35 種類のマグカップについて，ヨー角を 5° 始点に 10° 刻みに回転させながら生成した計 1,260 枚の画像を使用した．

3.2 欠損復元 AutoEncoder の構成

欠損復元 AE は Convolution 層を Encoder 及び Decoder にそれぞれ 5 層ずつ持つ構成とした．Encoder 部から抽出される特徴ベクトルは 12×12×1,024 次元とした．活性化関数には Rectified Linear Units（ReLU）を用いた．Decoder 部からの出力画像の評価には平均 2 乗誤差を用い，Adam [2] により，ネットワークパラメータの最適化を行った．

3.3 欠損処理

元画像は生成した画像から切り出し，$H = 128$, $W = 128$ 画素の大きさに揃えた．欠損位置は上下左右いずれかを無作為に選択し，また欠損率は 0% から 50% の範囲で無作為に決定し，欠損処理を施した．背景部分を拡張した後の画像の大きさは，$H' = 192$，$W' = 192$ 画素とした．

姿勢推定の学習用距離画像は，各物体について，欠損位置，欠損率を無作為に変化させながら，1 回転分の入力画像を生成した．この処理を計 10 回転分行い，各元画像について計 10 枚の入力画像を作成した．これは，限られた枚数の学習用距離画像から，様々な欠損パターンに対応したいくつもの多様体を構築するためである．評価用距離画像については，各元画像について上記の欠損処理を施し，各 1 枚の入力画像を生成した．

3.4 姿勢補間

低次元空間の次元数は 512 とし，各物体の連続的な姿勢変化が多様体をなすことから，3 次スプライン曲線により，各物体の特徴ベクトル間を結び，10° 刻みの特徴ベクトルから 1° 刻みの特徴ベクトルを補間した．この処理により，36,000 個の姿勢アノテーション付き特徴ベクトルを 360,000 個に拡張し，1° 刻みでの姿勢分類を可能とした．

3.5 評価方法

姿勢推定結果と真値との絶対角度誤差を算出し，それらを平均した平均絶対角度誤差により評価を行った．また，姿勢推定の難易度は，物体の姿勢によって異なるため，マグカップの姿勢により，Easy，Difficult の 2 つの推定難易度に分けて評価を行った．具体的には，持ち手部分がカップ部分の後方にまわり，持ち手が見えづらくなる 45° から 135° を Difficult とし，残りの姿勢を Easy とした．これらの難易度別評価に対し，両方を含めた評価は All とした．

本実験では，欠損のある物体を学習しない CNN に基づく回帰モデルによるもの（Deep Feature [4]），欠損復元 AE を 1 回のみ適用するもの（[5] に相当），欠損復元 AE を 2 回適用するもの（提案手法）を比較した．なお，姿勢補間は 3 つの手法全てで実施した．

表 1 平均角度誤差の比較

特徴抽出手法	Easy	Difficult	All
Deep Feature [4]	47.64	73.13	54.72
AE1 回適用（ [5] に相当）	15.74	24.42	18.15
AE2 回適用（提案手法）	**10.40**	**17.64**	**12.41**

3.6 実験結果

姿勢推定結果の平均角度誤差の比較を表 1 に示す．表 1 より提案手法の精度が最良であった．特に難易度 Easy の姿勢では，提案手法が 10.40° と高精度であった．よって，欠損復元 AE を 2 回適用して特徴抽出する提案手法の有効性を確認した．

4. 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 により欠損復元した画像を元に位置ずれを補正し，再度欠損復元 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．また，パラメトリック固有空間法を拡張した姿勢補間により，同一クラスであるが学習データに含まれない形状・姿勢をした物体であっても精度良く姿勢推定ができる手法を提案した．評価実験により，提案手法による姿勢推定精度の向上を確認した．

今後は，特徴抽出手法の改良，実画像による評価実験，姿勢の回転軸数の拡張などについて検討する．

謝辞 本研究の一部は科学研究費補助金による．

参考文献

[1] Chang, A. X., Funkhouser, T. A., Guibas, L. J., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L. and Yu, F.: ShapeNet: An Information-Rich 3D Model Repository, arXiv preprint, arXiv:1512.03012 (2015).

[2] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv preprint, arXiv:1412.6980 (2014).

[3] Murase, H. and Nayar, S. K.: Visual Learning and Recognition of 3-D Objects from Appearance, *Int. J. of Computer Vision*, Vol. 14, No. 1, pp. 5–24 (1995).

[4] Ninomiya, H., Kawanishi, Y., Deguchi, D., Ide, I., Murase, H., Kobori, N. and Nakano, Y.: Deep Manifold Embedding for 3D Object Pose Estimation, *Proc. 12th Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Vol. 5, pp. 173–178 (2017).

[5] Sundermeyer, M., Marton, Z.-C., Durner, M., Brucker, M. and Triebel, R.: Implicit 3D Orientation Learning for 6D Object Detection from RGB Images, *Proc. 15th European Conf. on Computer Vision*, pp. 699–715 (2018).