

# 視覚・触覚センサを用いた透明物体の把持姿勢推定

○松岡尚吾 (大阪大) 万偉偉 (大阪大) 清川拓哉 (大阪大)  
小山佳祐 (大阪大) 原田研介 (大阪大)

## 1. はじめに

近年、医薬品の開発現場において試薬の取り出しや調合などの作業をヒトに代わってロボットで代替し、自動的に実験することの需要が高まっている。しかしながら、これらの現場での試験管の把持・移動などの作業では、透明物体の視覚認識の誤差や物体の操作時に伴う誤差などの様々な誤差要因が累積し、実際の把持姿勢が計画していたものと異なる。これにより、正確な操作が実現できないために、自動化は困難となる。そこで、本研究ではハンド内の物体の把持姿勢を推定することを目的とする。

把持姿勢推定に関して多くの研究が行われており、触覚センサを用いた手法・ハンド外部に取り付けられた RGB-D センサを用いたものなどがある [1–4]。触覚センサは形状の特徴的な小さな物体の推定は得意であるが、形状が均一で特徴の少ない物体の推定は不得意である。一方、RGB-D センサはサイズが大きい物体の推定を得意とするが、物体の性質による認識結果への影響が問題となり、透明物体などの推定が不得意である。そこで本研究では、2 台の RGB カメラ及び 3 次元形状復元の機能を有した GelSight [5] を用い、視覚と触覚センサを組み合わせることで物体の性質・形状に関わらない把持姿勢推定を行う。視覚センサで物体の上下方向への移動をとらえ、触覚センサで物体の把持の深さ・角度をとらえることが可能となる。例えば、細長い透明物体に対して、RGB カメラは認識を行うことが可能であり、触覚センサで物体の傾きなどの情報を同時に得ることができる。図 1 に示すように視覚・触覚センサを組み合わせ、シミュレータ上で学習した Autoencoder [6] を用いた特徴量抽出を行い、抽出したシミュレータと実世界の特徴量の比較し、把持姿勢推定を行う。

## 2. 提案手法

本研究では視覚センサと触覚センサを組み合わせる。図 2 に示すように、触覚センサは 3 次元形状復元の機能を有する GelSight [5] を用い、同時にロボット手首に取り付けられた 2 台のハンドカメラを使用する。また、本研究では複数センサの情報を統括して分析するために AutoEncoder [6] を用いた特徴量抽出器を用いて特徴量を抽出し、把持姿勢推定を実現する。

### 2.1 ハンドカメラ及び GelSight からのセンサ情報

本研究で使用するハンドカメラならびに GelSight から得られる情報について説明する。図 3 に示すように、ロボットハンド手首に取りつけられた 2 台のハンドカメラからはそれぞれ RGB 画像を取得する。GelSight [5] は接触部分に関しての 3 次元形状復元によって得られる点群を深度画像に変換して使用する。

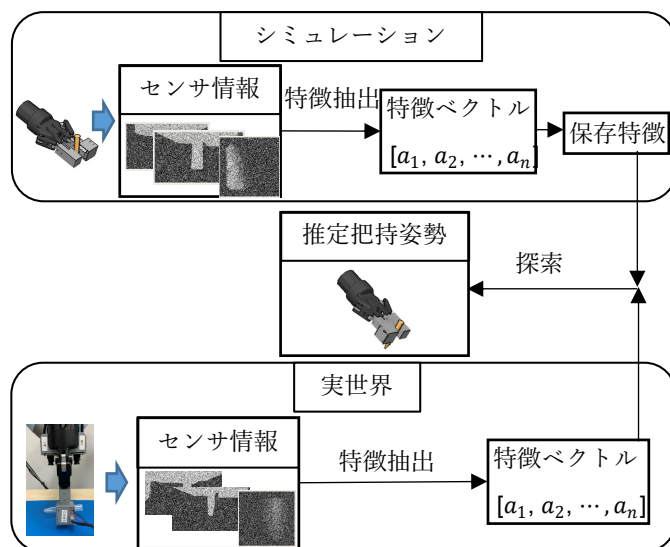


図 1 把持姿勢推定システムの概要

### 2.2 視覚及び触覚情報からの特徴量抽出

本研究では、物体認識を背景や物体のテクスチャに依存しないシステムとするために、ハンドカメラ画像から背景を除去したマスク画像を入力とする 2 つの特徴量抽出器ならびに GelSight から得られる深度画像を入力とする特徴量抽出器を組み合わせる。ノイズを除去する Autoencoder である Denoising Autoencoder (以下、DAE と呼ぶ) は効率的に特徴量を抽出できるという性質を用い、本研究では特徴量抽出器として DAE を用いる [6]。本研究では、図 4 のような構造の DAE を用い、3 つ目の Maxpooling2D 層からの出力を各ハンドカメラ画像から  $16 \times 12 \times 8$  の特徴と Gelsight から  $12 \times 12 \times 8$  の特徴としてそれぞれ取得する。

### 2.3 シミュレータを用いたネットワークの学習

特徴量抽出器を学習するためのデータを作成する。図 5 に示すように、学習データを多様な物体形状に対して対応可能にするために基本形状を組み合わせ、新たな形状を生成して学習データを作成する。生成した基本図形に対してシミュレーション上で把持可能なすべての把持姿勢を生成する。生成した新しい形状の立体に対する把持姿勢において図 6 のように 2 台のハンドカメラからの画像を保存する。

2 台のハンドカメラ画像を取得すると同時に、シミュレータ内で物体のモデル表面をサンプリングし、GelSight の接触可能な領域に対してクリッピングを行い、接触部分の変位を描画することで、図 6 のような GelSight からの深度画像を作成し、保存する。

得られた各把持姿勢に対応するセンサ画像を入力として Autoencoder を基とした特徴量抽出器を作成する。

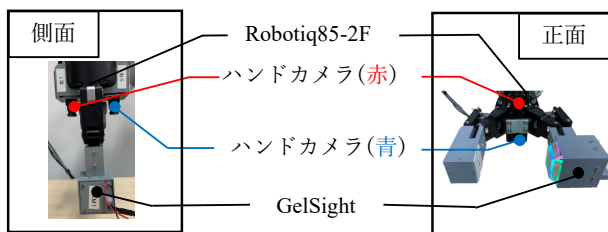


図2 本研究で使用するセンサ

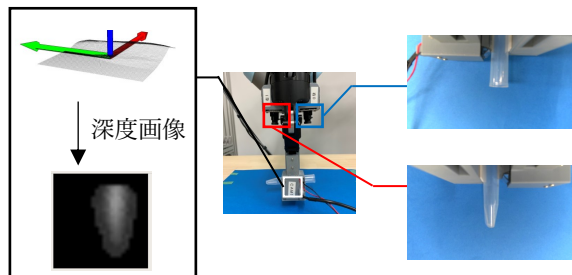


図3 実世界のセンサ情報の取得

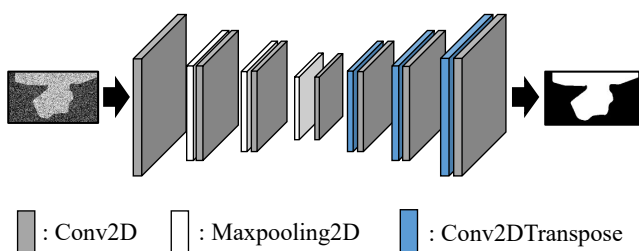


図4 特徴量抽出器として用いる AutoEncoder の構造

得られたシミュレーション上の各ハンドカメラ画像の背景を除去したマスク画像に対してガウシアンノイズを付与したものを入力としを図4のように Autoencoder の学習を行い、各ハンドカメラに対して特徴量抽出器を作成する。また、GelSight [5] から取得した深度画像に関してもガウシアンノイズを付与し、同様の構造の Autoencoder の学習を行う。各センサに対して3つ目の MaxPooling2D 層からの出力を特徴とする特徴量抽出器を作成する。

## 2.4 使用物体に対してのセンサ特徴の取得

次に把持姿勢推定を行う物体に対しての把持姿勢と特徴量を取得する。基本図形の把持姿勢を獲得する処理と同様に、使用する物体に対してシミュレーション上で把持可能なすべての把持姿勢を生成する。各把持姿勢に対応する2台のハンドカメラ画像と Gelsight から深度画像を獲得する。GelSight から得られる深度画像に関しては把持の強さを考慮する。把持の強さは弱いときは GelSight の変位は小さく、把持の強さが強いときは GelSight の変位が大きくなることを考慮し、1つの把持姿勢に対して複数の情報を取得する。把持の強さに応じてクリッピングする点群の深さを変化させることにより実現する。図7に示すように、得られた画像を2.3節で各センサごとの特徴量抽出器にガウシアンノイズを付与し、入力する。特徴量抽出器から得られた特徴量マップに対して1行のベクトルに変換する。また、各カメラに対して変換したベクトルを

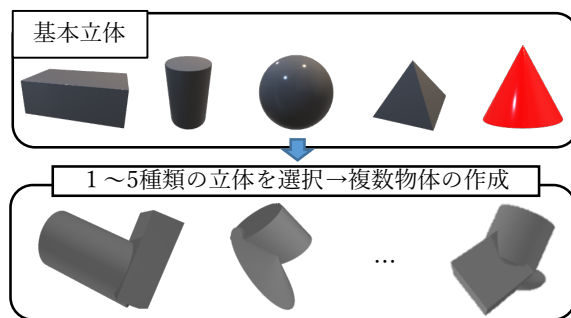


図5 基本立体の組み合わせによる新しい形状の立体の作成とその一例

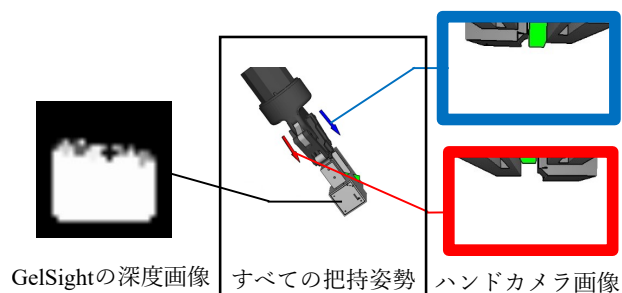


図6 基本図形の把持姿勢に対するシミュレータ上でのセンサ情報の獲得

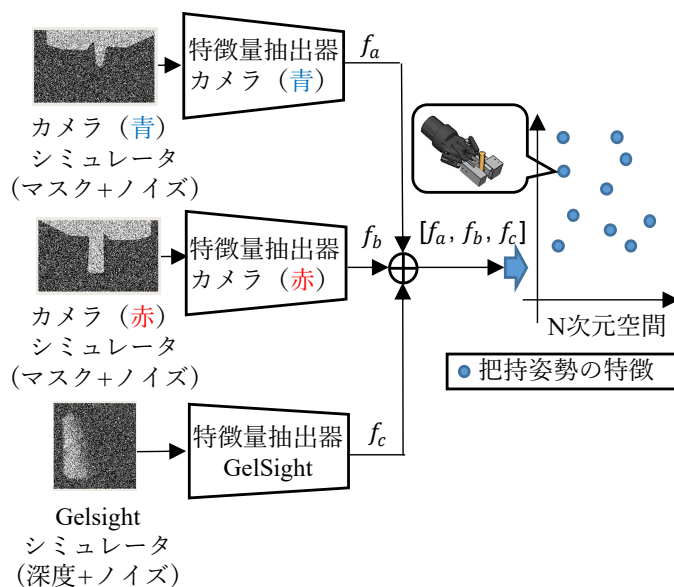


図7 対象物体に対する把持姿勢と各センサからの特徴の紐づけ

1行に再構成し、各把持姿勢に対応する特徴を1つのベクトルとしてシミュレータの特徴として保存する。

## 2.5 特徴量による把持姿勢推定

ロボットが物体を把持した状態における実世界の情報を取得する。図3のようにロボットハンド手首に取りつけられた2台のハンドカメラから画像を取得する。今回、実際の画像に関しては背景除去を容易とするために背景を青色に設定した。また、図3のように GelSight から変位部分に対して点群を取得し、深度画像に変換

する。

2.3 節で作成した各センサごとの特徴量抽出器に実世界において獲得した 2 台のハンドカメラのマスク画像と GelSight より得られた深度画像にガウシアンノイズを付加したものを入力し、これらの出力を実画像の特徴量データとして獲得する。2.4 節で保存した各把持姿勢に対してのシミュレーション画像の特徴量ベクトルと実画像の特徴量ベクトルのコサイン類似度を計算し、数値が最大の姿勢を大まかな推定把持姿勢とする。しかし、大まかな推定把持姿勢は離散化されたものであるため、離散化した把持姿勢の間のセンサからの特徴がより一致度が高い可能性が考えられる。そこで大まかな推定把持姿勢周辺において、離散化された把持姿勢を埋めるように、シミュレーション上での把持姿勢を再収集し、その時のハンドカメラ画像のマスク画像ならびに Gelsight から得られた深度画像にガウシアンノイズを付与したものを特徴量抽出器に入力することで特徴量ベクトルを再構成し、同様の比較を行うことによってさらに詳細な推定把持姿勢を獲得する。

### 3. 把持姿勢推定実験

#### 3.1 実験内容

図 2 のようなシステムを用いて実験を行う。物体をロボットハンドに把持させ提案システムを用いて把持姿勢推定の出力について検証する。本実験では、半透明の試験管の 2 種類について出力結果を確認した。また、各物体はシミュレータ上で、近似モデルを用いてセンサからのデータを取得した。

#### 3.2 結果と考察

実世界における把持姿勢を提案システムによって出力された把持姿勢の一例を図 8, 9 に示す。概ね実世界における把持姿勢を推定できているが、失敗例も存在する。半透明の物体に関しては近似モデルを使用しており、図 8 の 4 段目に示すように、GelSight のシミュレータと実機の接触情報の曲面の深度画像が異なってしまう、異なる把持姿勢を出力してしまうと考えられる。また、図 9 の 2 段目の例に示すように、透明な試験管に関しては上下が反転して推定されるものが存在する。背景除去を行う際に輪郭が透明部分に関して背景が余分に除去される箇所があり、これが影響し、上下が反転して推定されることが考えられる。また、実験に関してはセンサ情報に依存し、ハンドカメラにおける背景除去時や Gelsight の接触情報のノイズが推定結果に影響することがあることが考えられ、同一の把持姿勢において異なる推定結果を得ることがある。これに対して、ロボステ性向上のために手法の改善が必要と考えられる。

### 4. おわりに

本稿では、2 台のハンドカメラならびに GelSight [5] を用いたシミュレーションと実世界から特徴を抽出し、比較することによるインハンドの把持姿勢推定の手法を提案した。

現在は実世界のハンドカメラからの情報に対して確実に背景を除去できる設定にて実験を行っており、実際のタスクにおける煩雑な背景に関して考慮していな

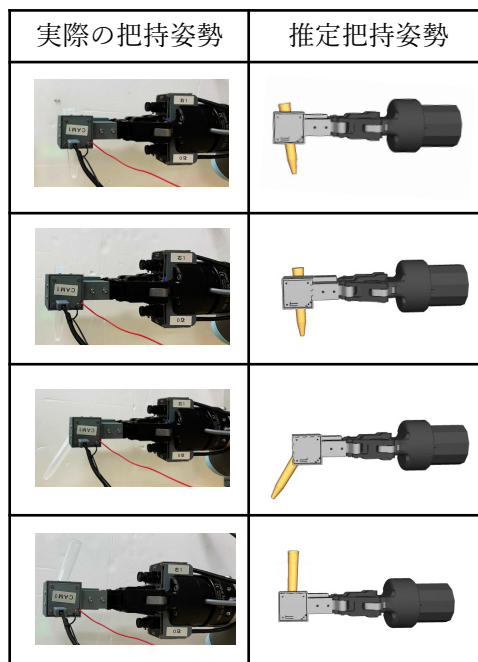


図 8 半透明な試験管に関する把持姿勢推定の結果の一例

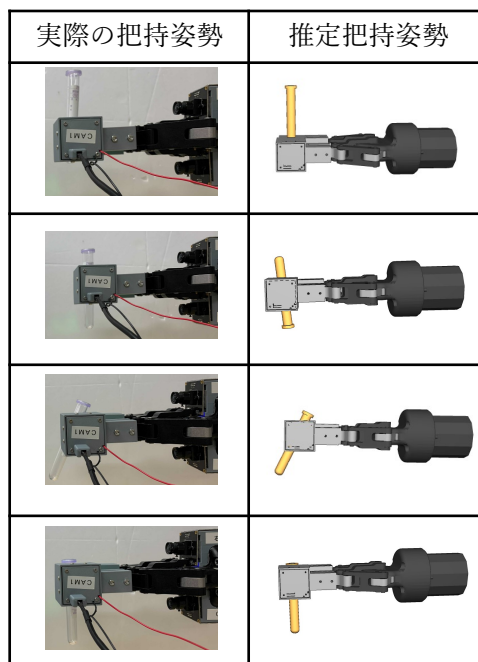


図 9 透明な試験管に関する把持姿勢推定の結果の一例

い。そこで今後の展望として以下のようなものが挙げられる。基本図形と深層学習を組み合わせた背景除去の試みと背景除去時に発生するノイズを AutoEncoder の中で除去するよう学習し、ノイズを考慮した特徴抽出器の作成を行うことで把持姿勢推定を行う。また、定量的な実験を行うことが求められる。

### 参 考 文 献

- [1] J. Bimbo, S. Luo, K. Althoefer, and H. Liu. “In-hand Object Pose Estimation using Covariance-based Tac-tile to Geometry Matching”. In *IEEE Robotics and Automation Letters*, Vol. 1, pp. 570–577. IEEE, 2016.
- [2] 花井亮, 高瀬竜一, 堂前幸康, 植芝俊夫. “視触覚によるネ

ジの把持姿勢推定”. 計測自動制御学会システムインテグレーション部門講演会 (SI), pp. 3012–3014, 2019.

- [3] M. Bauza, E. Valls, B. Lim, T. Sechopoulos, and A. Rodriguez. “Tactile object pose estimation from the first touch with geometric contact rendering”. *arXiv preprint arXiv:2012.05205*, 2020.
- [4] B. Wen, C. Mitash, S. Soorian, A. Kimmel, A. Sintov, and K. E. Bekris. “Robust, Occlusion-aware Pose Estimation for Objects Grasped by Adaptive Hands”. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2020.
- [5] W. Yuan, S. Dong, and E. H. Adelson. “Gelsight: High-resolution Robot Tactile Sensors for Estimating Geometry and Force”. In *Sensors*, Vol. 17, p. 2762. Multidisciplinary Digital Publishing Institute, 2017.
- [6] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. “Extracting and composing robust features with denoising autoencoders”. In *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.