

3-D シーン観察に基づく 手と物体の関係性の学習と把持パタンの想起

川上 拓也[†] 松尾 直志[†] 小川 陽子^{††} 島田 伸敬[†]

[†] 立命館大学情報理工学部 ^{††} 立命館大学大学院 情報理工学研究科

1. はじめに

人間が把持を行う物体というのは様々な機能を持っている。また、人間は物体を把持する際に、その物体の機能に応じて手の形を変えて把持する。本稿では機械学習を用いて、人間が物体にどういったアプローチを行うかという視点から物体の把持パターンを推定する。

2. 3-D シーン観察に基づく把持画像の生成

池上らの室内ロギングシステム[1]を仮定して、机などに置いてある物体を把持し、持ち上げるというシーンから点群の時系列情報を撮影する。初期フレームには物体のみが映っていると仮定し、その後物体を把持して持ち上げる。図1の『把持画像の生成』に示す通り、その点群から手と物体の点群のみを抽出するため、3次元トリミングと平面除去を行う。撮影した点群に対し、ICP アルゴリズムを用いて初期フレームの物体点群と重なるように位置合わせを行う。その後、位置合わせをした点群から初期フレームの点群と重なっている点を物体点群とし、それ以外の点を手の点群とする。その点群を用いて、深度画像、手のマスク画像、物体のマスク画像の3chからなる把持画像を作成する。

3. 機械学習を用いた把持パタンの想起

物体のみ画像から、把持パターンを表すパラメータである持ち方パラメータの想起を行う。把持パターンを想起するモデルは Matsuo らの AutoEncoder(AE)と Convolutional Neural Network(CNN)を組み合わせたモデルを使用する[2]。

図1の把持パタンの想起に示す通り、AE を用いて把持画像から 30 次元の持ち方パラメータが写像される空間を学習する。次に学習済み AE の学習結果である持ち方パラメータを教師とし、CNN で物体のみ画像との関係を学習させる。本稿では、この AE と CNN を組み合わせたモデルを把持パターン想起モデルと呼び、このモデルを用いて把持パターンを想起する。また、想起された持ち方パラメータを AE の Decoder に入力することにより、把持画像の想起も行う。

4. 把持パターン想起結果

使用した物体は、マグカップ、(取っ手無し)コップ、ボール、スプレーの4カテゴリの物体である。カテゴリごとに1種類の物体を用意し、把持画像をそれぞれ40枚作成し、把持パタンの想起を行った。図2に示す通り、各物体がパラメトリックに分類できている事が分かる。また、想起画像の手マスク画像を見ると、カテゴリごとに異なった手の形をしている事が分かる。

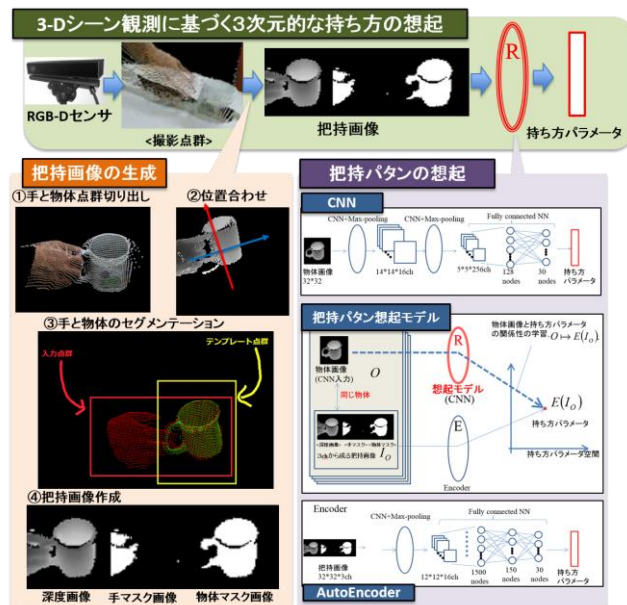


図1. 本手法の流れ

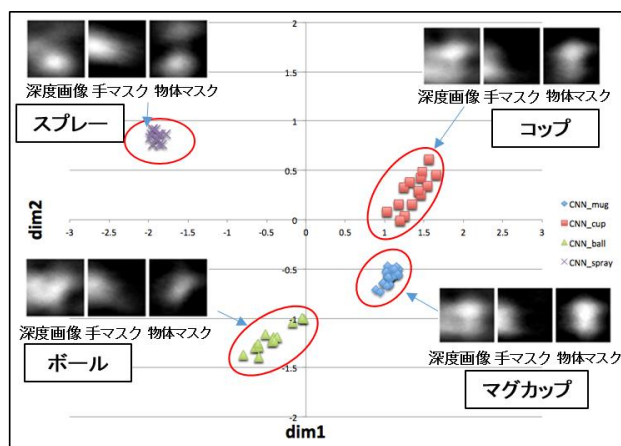


図2. 想起した把持パターンと把持画像

5. 今後の課題

学習に使う画像を日常的なシーンから自動で収集し、未知のカテゴリの物体に対する持ち方の想起を行う予定である。

参考文献

[1] 池上ほか “階層型イベント検知に基づく人と物の関わりのロギングシステム”, 第18回画像の認識・理解シンポジウム, SS5-37
 [2] T.Matsuo et.al “Extraction of Descriptor of Hand-Object interaction”, 第18回画像の認識・理解シンポジウム, OS1-4