

物体の面セグメンテーションに基づく 6D-PoseNet による位置姿勢推定

○大西剛史 平川翼 山下隆義 藤吉弘亘（中部大学）

1. はじめに

コンビニ等の陳列廃棄タスクにおいて、ロボットで物体を把持するには多品種の物体の位置姿勢を推定する必要がある。姿勢推定には距離情報を用いるのが一般的であるが、光の反射や表面の材質により計測の際に点群が欠損するという問題がある。これは姿勢推定の失敗を引き起こす大きな要因となる。そこで、本研究では距離情報を用いない姿勢推定を実現することを目的とする。また、同じ形状の物体でも商品によりテクスチャは異なるため、新たな商品に対して姿勢推定精度は低下する。そこで、物体のテクスチャに依存しない姿勢推定が求められる。本研究では、対象物体に対して面セグメンテーションを行うことで、テクスチャの差異を吸収し、セグメンテーション画像を用いた6D姿勢推定法を提案する。

2. 提案手法

本研究では、コンビニの陳列廃棄タスクを想定し、サンドイッチを対象とし、物体の面セグメンテーション画像を用いた物体の6D姿勢推定手法を提案する。提案手法の流れは図1のようにStep1で物体検出、Step2で検出領域に対する面セグメンテーション、Step3で6D姿勢推定である。

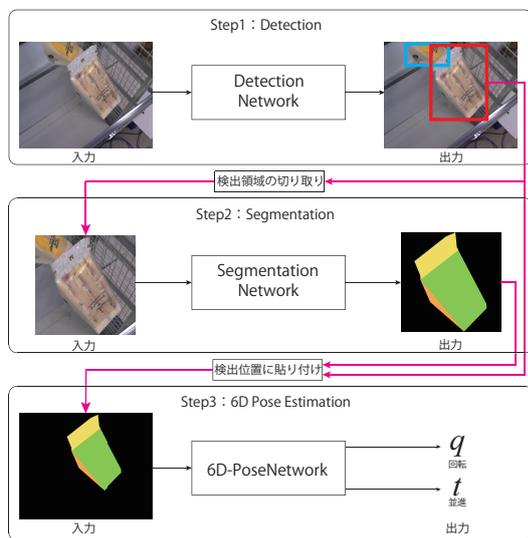


図1 本手法の流れ

2.1 物体の検出と面セグメンテーション

まず、物体検出手法である Single Shot MultiBox Detector(SSD)[1]を用いてサンドイッチを検出する。次に、SegNet[2]を用いて対象物体に面セグメンテーションを行う。把持対象は、画像中心に最も近い検出を対象とする。面ごとのクラスラベルを図2に示す。クラ

スは正面、底面、背面、左側面、右側面、耳部の6クラスとする。また、図3のように画像中心にある把持対象に隣接する物体は背景とする。これにより把持対象外の物体による姿勢推定への悪影響を低減させることができる。

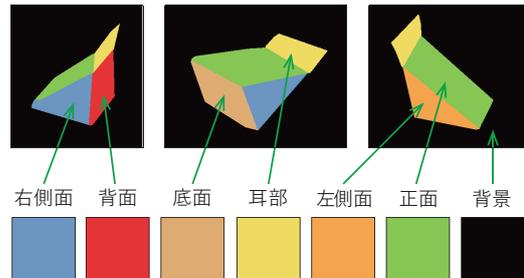


図2 面のクラスラベル

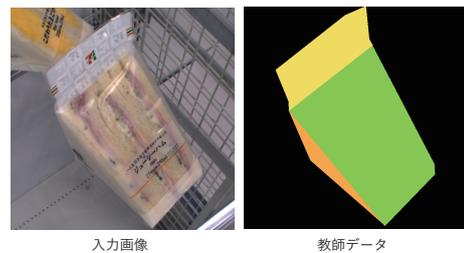


図3 面ラベルのアノテーション例

2.2 6D-PoseNet による位置姿勢推定

提案する6D-PoseNetは、物体の面セグメンテーション画像を入力とし、回転と並進のパラメータを出力する。ここで、回転パラメータはクォータニオンで表す。これにより、9つのパラメータを持つ回転行列と比べ4つのパラメータに削減できる。並進パラメータは、 x , y , z 軸方向に対するそれぞれの移動量を表す。学習には、セグメンテーションラベルをテクスチャとした3Dモデルを用いる。本研究では、サンドイッチの3Dモデルを用いて学習する。3Dモデルに対してランダムな回転と並進移動を加え、画像に投影したものを入力とする。ここで、回転行列をクォータニオンに変換した値と並進の値を教師データとする。

6D-PoseNetのネットワーク構造を図4に示す。Feature Extractorでは畳み込み処理とプーリング処理によりダウンサンプリングを繰り返すことで、物体の特徴を獲得する。Quaternion Branch, Translation Branchでは、畳み込み処理によりそれぞれ回転、並進に関する特徴を獲得し、回転パラメータ、並進パラメータを直接出力する。

学習に用いる誤差関数を式1に表す。

$$L = \alpha L_q + \beta L_t \quad (1)$$

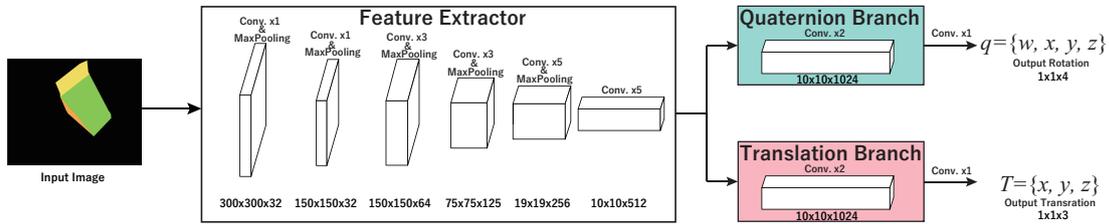


図4 6D-PoseNetのネットワーク構造

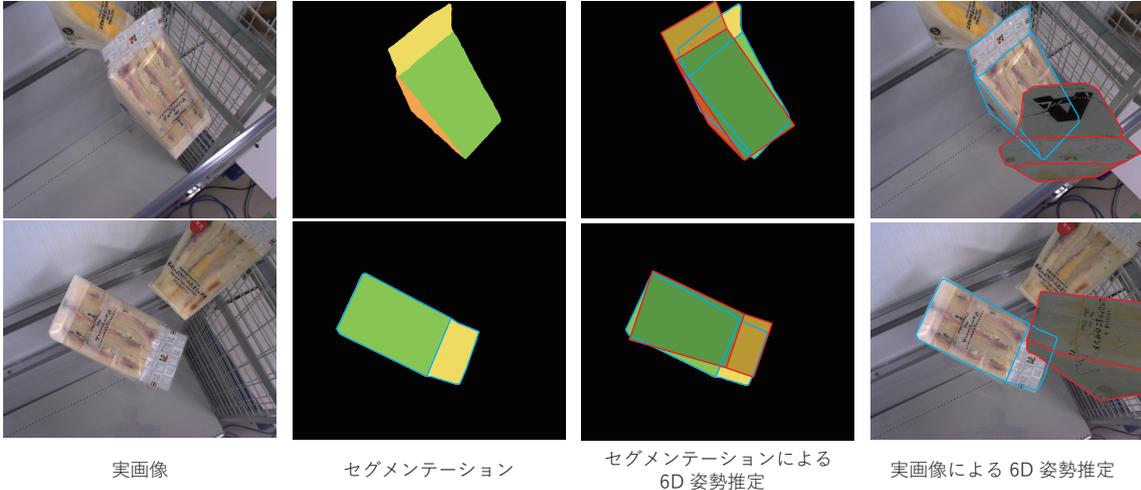


図5 実画像に対する6D姿勢推定例

ここで L_q は Quaternion loss, L_t は Translation loss であり, 平均二乗誤差関数を用いる. また, α と β は各誤差に対する重みであり, 本研究では $\alpha = 0.1$, $\beta = 1.0$ とする.

3. 評価実験

提案手法の有効性を示すために, セグメンテーション画像を用いて学習したモデルと実画像を用いて学習したモデルで比較をする.

3.1 実験概要

評価に使用するモデルは, セグメンテーション画像を入力としたモデルは 600,000 iteration, 実画像を入力としたモデルは 800,000 iteration 学習したものを用いる. また, 実画像の学習では, 多視点から撮影した商品棚の画像に 3D モデルを投影した画像を使用し, data augmentation として輝度変化を加える. 評価データには 3D モデルを投影した画像 1,000 枚を用いる. 評価には教師ラベルとの x , y , z 方向の並進誤差と回転誤差の平均と分散を用いる.

3.2 実験結果

実画像に対して推定した例を図5に示す. 対象の輪郭を青, 推定したモデルの輪郭を赤で表す. 提案手法は, 類似した物体が隣接する場合においても頑健な姿勢推定が可能であることがわかる.

3D モデルの投影画像に対するパラメータごとの平均誤差と分散を表1に示す. ここで, 分散は () に示す. 表1から, セグメンテーション画像は, すべてのパラメータにおいて誤差を小さくすることができた. また, z と比べ x , y の分散が小さいことがわかる. 正面から把持をする場合において, 奥行き方向の誤差は吸着判

定によって吸収できる. そのため, x , y 方向の推定が安定していることが重要である.

表1 各パラメータの誤差

| 入力 | x [mm] | y [mm] | z [mm] | rot [deg] |
|-----------------|-------------------|------------------|-------------------|------------------|
| セグメンテーション 画像 | 2.42 (4.35) | 2.69 (4.52) | 6.64 (31.16) | 7.76 (132.32) |
| 実画像 | 14.67 (110.12) | 13.59 (93.22) | 43.46 (940.51) | 77.30 (2271) |

4. おわりに

本研究では, 面セグメンテーションを用いた 6D 姿勢推定を行った. 面セグメンテーションを用いることで, 隣接する類似物体やテクスチャの変化に対して頑健な 6D 姿勢推定ができた. 今後の展望は側面からの把持も考慮し, z 軸方向に対する位置合わせ精度の向上が挙げられる.

謝辞

この成果は, 国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務とオムロン株式会社, 中京大学との共同研究の結果から得られたものである.

参考文献

- [1] W. Liu, *et al.* SSD: Single Shot MultiBox Detector, ECCV, 2016.
- [2] V. Badrinarayanan, *et al.* SegNet :A Deep Convolutional EncoderDecoder Architecture for Image Segmentation. , IEEE trans. on PAMI, Vol. 39, pp. 2481-2495, 2017.