PoseRBPF: A Rao-Blackwellized Particle Filter for 6D Object Pose Tracking

Yu Xiang, 9/27/2019
6D OBJECT POSE ESTIMATION

Pose information useful for:
- Object manipulation
- Semantic navigation
- Human robot interaction

3D Model

Input image

6D Object Pose

3D Translation

3D Orientation
TRADITIONALLY

• Feature matching
  
  Rothganger et al. IJCV, 2006
  
  MOPED, Collet, Martinez, Srinivasa, IJRR, 2011

• Template matching
  
  Hinterstoisser et al., ACCV, 2012
  
  Choi et al., IROS, 2012
CHALLENGES

• Model capability
  • Texture, texture-less objects
  • Symmetry objects
  • Clutter scenes

• Accuracy and Robustness
  • Lighting change
  • Different background
  • Uncertainty
  • Speed
DEEP LEARNING

• Better image features and stronger model capacity

- Rad & Lepetit, ICCV 2017
- Kehl et al. ICCV 2017
- Tremblay et al. CoRL 2018
- Tekin et al. CVPR 2018
- Xiang et al. RSS 2018
- Sundermeyer et al. ECCV 2018
- Li et al. ECCV 2018
- Wang et al. CVPR 2019

Our goal:
✓ Symmetry objects
✓ Pose Tracking
✓ Pose uncertainty
ORIENTATION UNCERTAINTY

Depends on context, shape, sensor

Observation

Orientation uncertainty

Shape symmetry

Texture breaks symmetry

View-based uncertainty
IMPLICIT ROTATION LEARNING

ROTATION ESTIMATION WITH CODEBOOK MATCHING
[Sundermeyer et al. ECCV 2018]

- Inherently handles symmetric views
- Only orientation, no translation
- Single image, single estimate

191,808 discrete rotations

Similarity scores

Encoder

Codebook

Input

Detection
PoseRBPF

Generic and Efficient Framework for 6D Object Pose Tracking

- **Main idea:** Instead of sampling all state dimensions, sample some of the dimensions and solve remaining ones analytically

- Successfully applied to SLAM, tracking, activity recognition, ...

- **Here:** Sample translation and estimate discrete orientation distribution over orientation

PoseRBPF: Particle Representation

\[ \mathcal{X}_i = \{ T_i, P(R_i|T_i, Z_{1:k}) \} \]

- 3D Translation \( T_i \)
- Orientation Distribution \( P(R_i|T_i, Z_{1:k}) \)

Discretized Rotations
Codebook
Particle Code
Rotation Likelihood

Encoder

191,808 bins
PoseRBPF: Observation Update

Compute posterior

\[ P(R_k | T_i^k, Z_{1:k}) \propto P(R_k | T_i^k, Z_k) P(R_k | R_{k-1}) \]

Weights

Observation likelihood

Normalizer

Orientation Distribution

Particle Codes

Encoder
Results: YCB Objects

Example: YCB mug (50 particles, ~20fps)

<table>
<thead>
<tr>
<th>YCB-Video RGB</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoseRBPF:</td>
</tr>
<tr>
<td>ADD: 62.1, ADD-S: 78.4</td>
</tr>
<tr>
<td>PoseCNN:</td>
</tr>
<tr>
<td>ADD: 53.7, ADD-S: 75.9</td>
</tr>
</tbody>
</table>
Results: TLess Objects

Example: TLess 01 (100 particles, ~11fps)

TLess RGB
Object recall for Err_vsd < 0.3:
- PoseRBPF: 41.47%
- Sundermeyer et al: 18.35%
ROBUSTNESS?

Self-supervised Learning

Lift-long Learning
SELF-SUPERVISED 6D POSE ESTIMATION

Interactive data collection (5x)  Generated pose annotations

The robot can automatically interact with the objects to create new scenes by grasping and pushing.
DATA COLLECTION

The system can automatically collect large scale datasets.
GRASPING RESULTS

Better pose estimates lead to higher grasp success.

Success Rate (30 grasps)
- Synthetic: 46.7%
- Fine-tuned: 86.7%

trained with only synthetic data

fine-tuned with self-annotated data
GRASPING RESULTS

Scene 1

Scene 2

Here, we show the performance of our system on pick-and-place tasks.
POSERBPF

• Estimates full 6D object pose distributions

• Combines Bayesian filtering with deep learning for embeddings

• Handles symmetric objects and pose uncertainty

• Fully trained in simulation, state-of-the-art results on RGB only datasets

• Enables us to build a self-supervised 6D pose estimation system for manipulation
Questions?