**Problem Statement**

- **Task**: Identify type and pose of every object in the scene (point cloud/depth image).
- **Given**: 6 DoF camera pose, 3D models of objects in the scene, camera intrinsics.

**Cost**

\[ J(O_1; K) = J_{\text{observed}}(O_1; K) + J_{\text{rendered}}(O_1; K) \]

- Feature and template-based methods are brittle (e.g., occlusion).
- Learning methods need training data to capture the combinatorics of inter-object interactions.
- We propose a deliberative approach that searches for the “best explanation”.

**Technical Details**

- Brute-force search over joint state space is intractable: 4 objects, 100 (x,y) positions, 20 orientations: 2000^4 states.
- **Key idea**: Cost function can be decomposed over objects under a monotonicity constraint: Assign object poses sequentially; ensure penalty accrued for an assigned object never decreases later.
- Constraint results in “non-occluding” order, problem reduces to tree search on this **Monotone Scene Generation Tree**.

- Tree search is still hard. For branching factor of 8000 and tree depth of 4, we have ~4 x10^15 nodes in the tree.
- We use Focal Multi-Heuristic A* (MHA*)[7] for the search, and parallelize child node generation.

**Experiments**

- **Dataset**: Household objects occlusion dataset[2]
- 36 objects models, 82 instances in 23 scenes
- Objects vary only in (x, y, yaw)

- **Search Configuration**
  - Discretization: 4 cm, 22.5 deg
  - ICP at every stage to compensate for discretization artifacts
  - Parallel child node generation (AWS m4.10x, 2x40 virtual cores)
  - Mean search time: 6.5 mins

- **Baselines**
  - OUR-CVFH[3]: Global viewpoint-based feature, robust to occlusions, trained with 642 views of every object
  - Brute force ICP without rendering: slide every model over scene, take best fit over all possible orderings

- **RGB-D Input**

- **Result**

  - **Pose correct if**
    \[ \|(x, y) - (x_{\text{true}}, y_{\text{true}})\| < \Delta \]
    \[ \text{SHORTESTANGULARDIFFERENCE}(\theta_{\text{true}}) < \Delta \]

  - Mean search time: 6.5 mins

- **Scaling Up**

  - 12 chess pieces, 6 unique models

---

This research was sponsored by ARL, under the Robotics CTA program grant W911NF-10-2-0016