Point Pair Features Based Object Detection & Pose Estimation Revisited

Tolga Birdal[1], Slobodan Ilic[2]

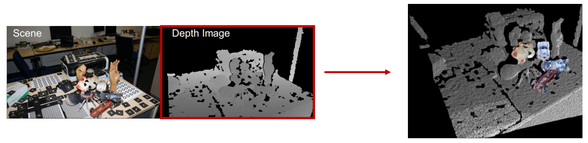
[1] Technical University of Munich, [2] Siemens AG

1. Introduction

Problem

Problem: Robustly locate objects and estimate 3D 6DOF pose

Challenges:
- Noisy Images
- Clutter and Occlusions
- Inaccurate CAD Models
- Lack of descriptive 3D features

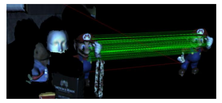


Prior Art

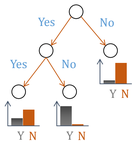
3D Object Detection

Sliding Window
- Template based
- LineMOD
- Requires Depth
- Performs poorly on only depth

Descriptor Based
- **On 3D Data**
- PPF, SHOT, MeshHOG etc.
- Slow to compute
- Non-repeatable



Learning Based
- Random Forests (Brachmann et al)
- Designed for RGBD



Existing Framework [1]

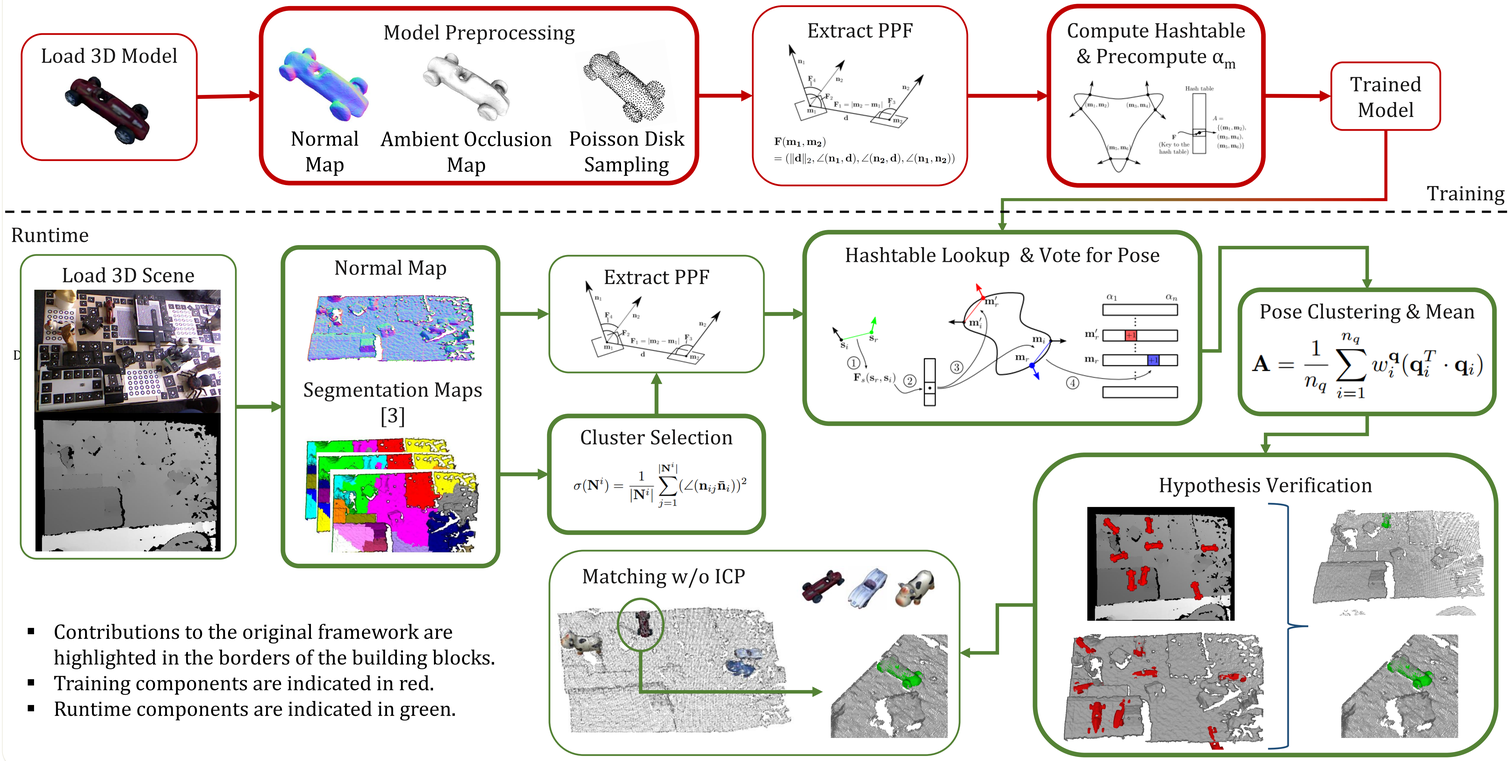
✓ **PROS**
- Operates directly on point clouds (not even meshes)
- Robust to occlusions
- Can find multiple instances of the model
- Reduced search space (votes for angle and model point)
- No plane assumption is made

✕ **CONS**
- Can not handle low detail surfaces
- Can not detect small objects
- Requires ICP verification
- Combinatorial pairing

Contributions

- We augment the entire pipeline
- No modifications to 3D feature
- Operates purely on depth images
- Cleaner Hough Space and more accurate poses
- Importance to model points
- Reduced dependency on ICP
- Integrated hypothesis verification
- Robust to large scenes with small objects

2. Method



- Contributions to the original framework are highlighted in the borders of the building blocks.
- Training components are indicated in red.
- Runtime components are indicated in green.

2.1. Method Details

Hypothesis Verification

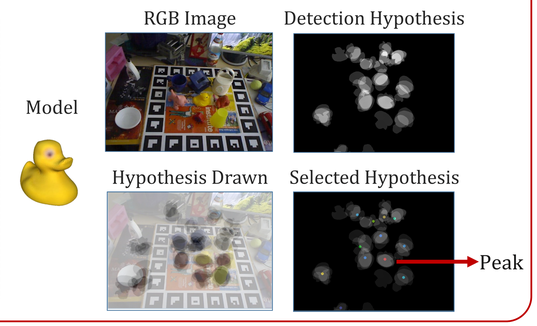
$$E_h(p,m) = D_p - \Phi(p|\mathbf{M}, \Theta_h, \mathbf{K})$$

Model, Pose, Intrinsics; Depth; Projection

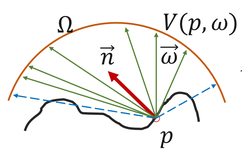
$$p \in \begin{cases} S_m, & \text{if } |E_h(p,m)| \le \tau_m \\ S_o, & \text{if } E_h(p,m) \ge \tau_o \\ S_c, & \text{otherwise} \end{cases}$$

→ On the model; → Occlusion; → Clutter

Score: $S_h = \left(1 - \frac{|p \in S_o|}{N_m}\right) \cdot \frac{|p \in S_m|}{N_m - |S_o|}$



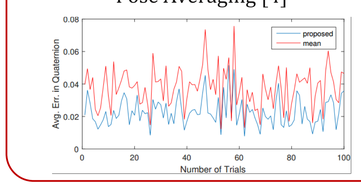
Ambient Occlusion Map



$$A(p) = \frac{1}{\pi}\int_\Omega V(p,\omega)\,\vec{n}\cdot\omega\,d\omega$$

$$W(m_r, m_i) = \lambda\sqrt{A(m_r)A(m_i)}$$

Pose Averaging [4]



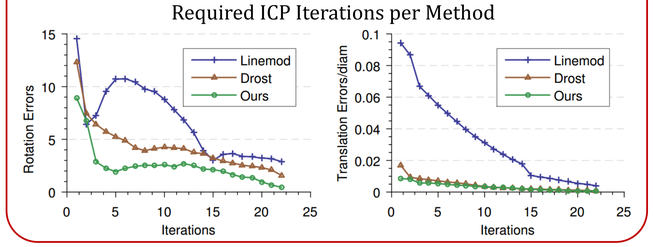
3. Results & Evaluation

Results on ACCV3D Dataset [2]

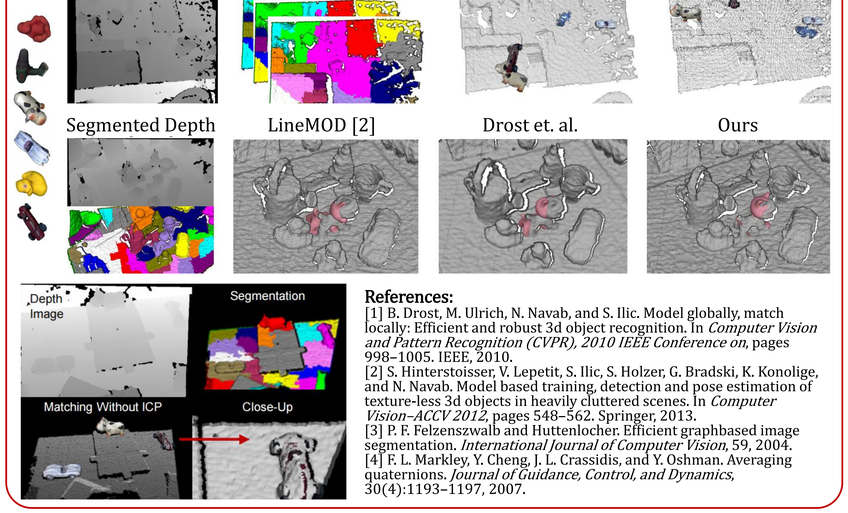
Table 1. Detection results on ACCV3D for different objects.

	LineMod	Drost et al.	Ours
ape	42.88%	65.54%	**81.95%**
cam	68.78%	84.92%	**91.00%**
cat	35.62%	87.30%	**95.76%**
driller	51.52%	81.06%	**81.22%**
iron	35.22%	87.06%	**93.92%**
Average	46.80%	81.18%	**88.77%**
Avg. Runtime	119ms	6.3s	2.9s

Required ICP Iterations per Method



Qualitative Results



References:

[1] B. Drost, M. Ulrich, N. Navab, and S. Ilic. Model globally, match locally: Efficient and robust 3d object recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 998–1005. IEEE, 2010.

[2] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In *Computer Vision–ACCV 2012*, pages 548–562. Springer, 2013.

[3] P. F. Felzenszwalb and Huttenlocher. Efficient graphbased image segmentation. *International Journal of Computer Vision*, 59, 2004.

[4] F. L. Markley, Y. Cheng, J. L. Crassidis, and Y. Oshman. Averaging quaternions. *Journal of Guidance, Control, and Dynamics*, 30(4):1193–1197, 2007.