Robust 6D Object Pose Estimation with Stochastic Congruent Sets

Chaitanya Mitash, Abdeslam Boularias and Kostas E. Bekris

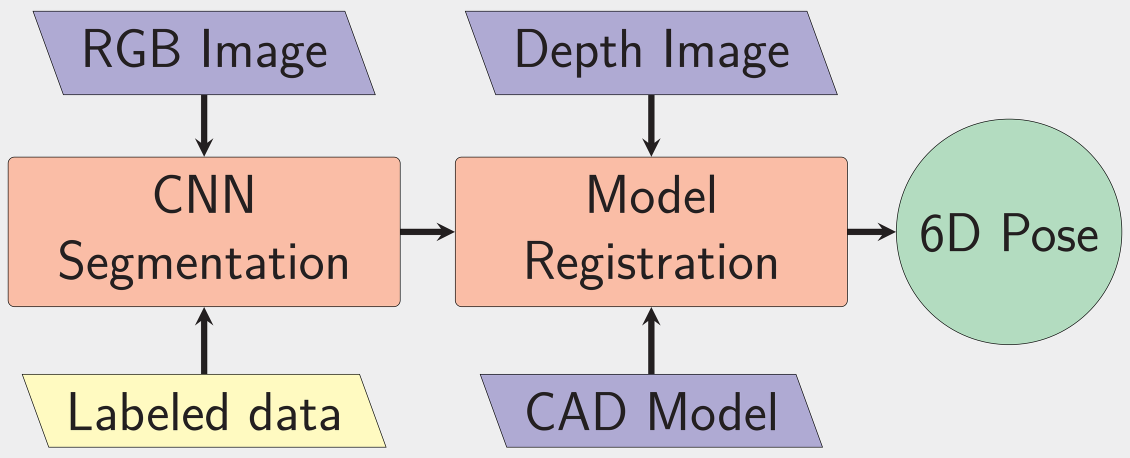
Department of Computer Science, Rutgers, the State University of New Jersey

6D Pose Estimation Problem

Objective:

- Input: RGB-D scene, CAD models for known objects in scene.
- Output: Translation ($t \in R^3$) and Rotation ($R \in SO(3)$) for each object.

Popular Pose Estimation Pipeline:



Motivation

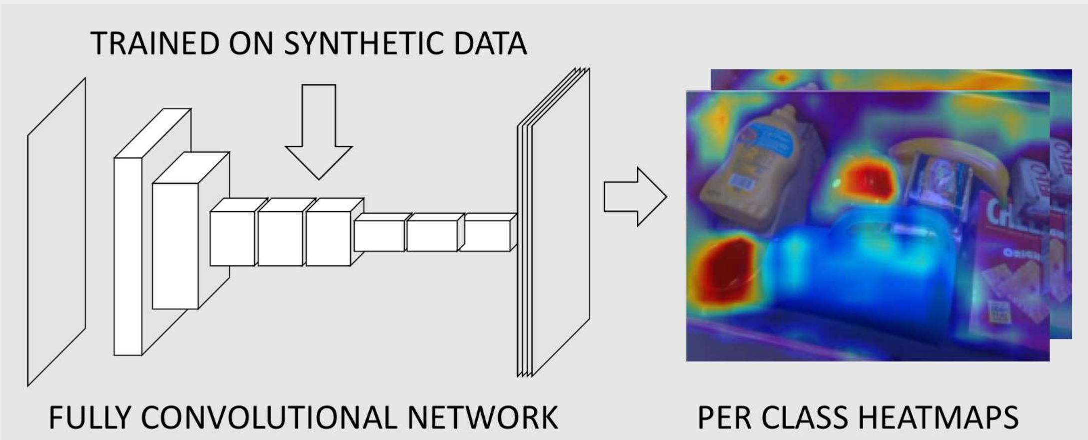
- Collecting labeled training data requires substantial manual effort. Our goal is to train the CNN using synthetic data.
- Often hard segmentation decision results in over/under segmentation due to the domain gap between synthetic and real data.

Prior algorithms like Super4PCS,

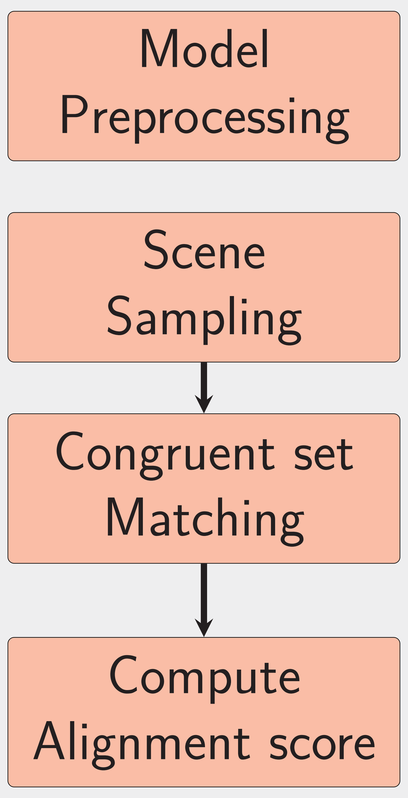
- Use **deterministic segments**, resulting in sub-optimal solutions in the above mentioned error cases.
- **Randomly sample** set of points on the segment and find congruent sets on the object model. This requires several iterations to get a high probability of success.

Proposed Solution

- We propose to use a stochastic representation of the output from FCN for model registration.



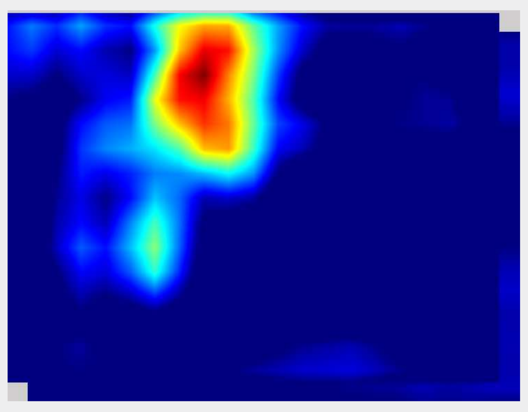
- We combine a global geometric descriptor with the soft segmentation output of the CNN to propose a fast and robust optimization process.



Scene Sampling

- The objective of this step is to sample a set of 4 points B = $\{b_1, b_2, b_3, b_4\}$ on the scene with a high joint probability of these points belonging to the object O_k

Input image (I)



Prior: $\pi(p_i \in O_k)$

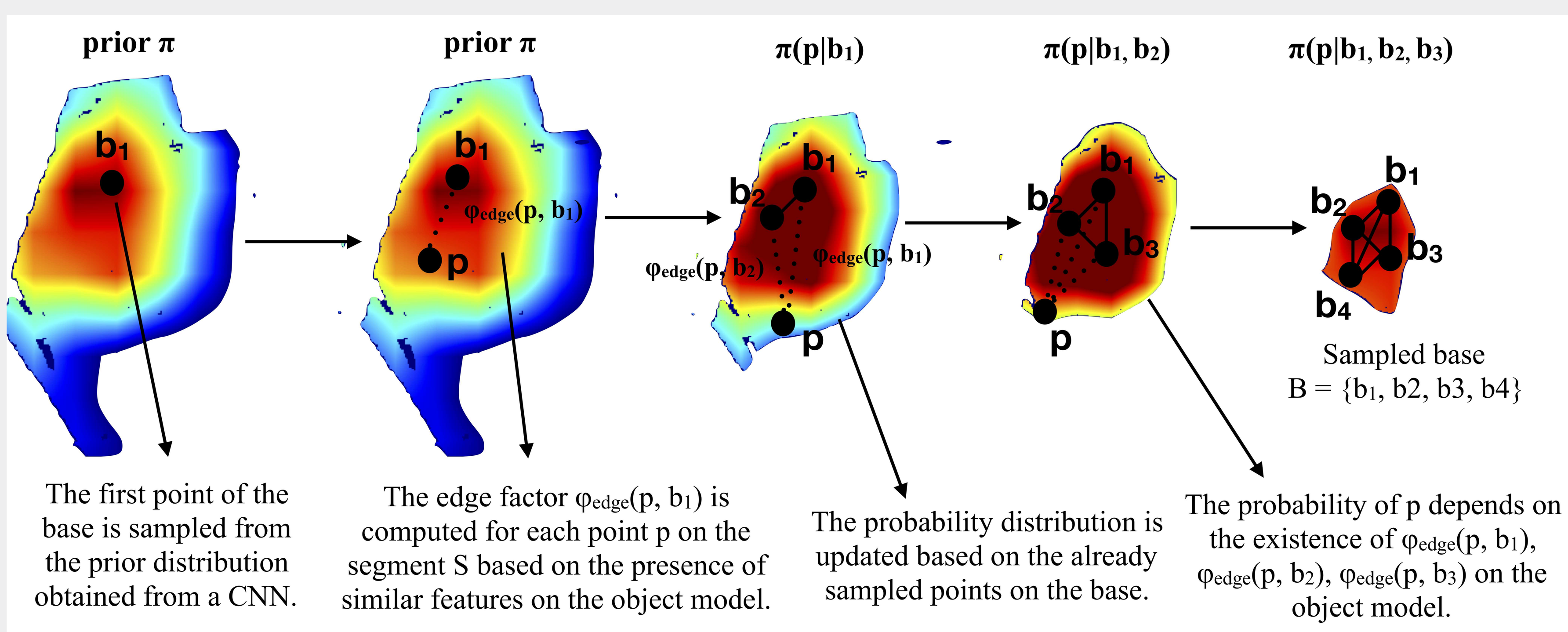
Object models $M_1, \dots M_n$

$$\Pr(B \rightarrow O_k) = \frac{1}{Z}\Pi_{i=1:4}\{\phi_{node}(b_i)\Pi_{j<i}\phi_{edge}(b_i, b_j)\} \quad (1)$$

$$\phi_{node}(b_i) = \pi(b_i \in O_k) \quad (2)$$

$$\phi_{edge}(b_i, b_j) = \begin{cases} 1, & \text{if } \mid MAP(M_k, \mathrm{PPF}(b_i, b_j)) \mid > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

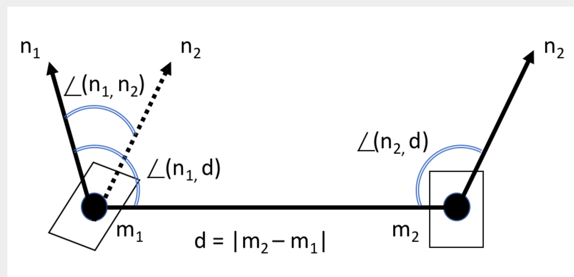
where, ϕ_{node} is the independent pixel probability from FCN, ϕ_{edge} is the pairwise probability computed from lookup table (MAP) generated in pre-processing step.



Model pre-processing

- Point pair features are computed for each pair of points on the object model.

$$PPF(m_1, m_2) = (\|\, d \,\|_2, \angle(n_1, d), \angle(n_2, d), \angle(n_1, n_2))$$

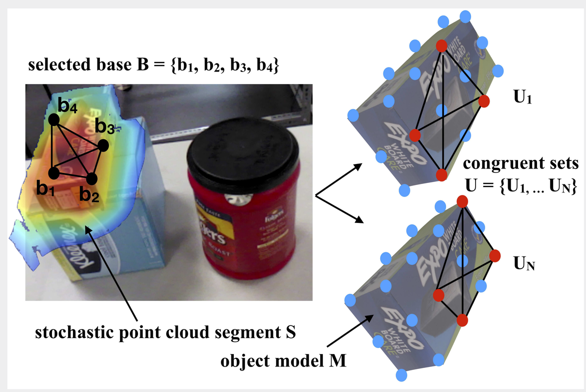


- A lookup table is generated for each object, that maps feature vectors to a set of pair of points that share the same point pair features.

$$MAP(M_k, f) : f \rightarrow \{(m_i, m_j) \in M_k \mid PPF(m_i, m_j) = f)\}$$

Congruent Set Matching

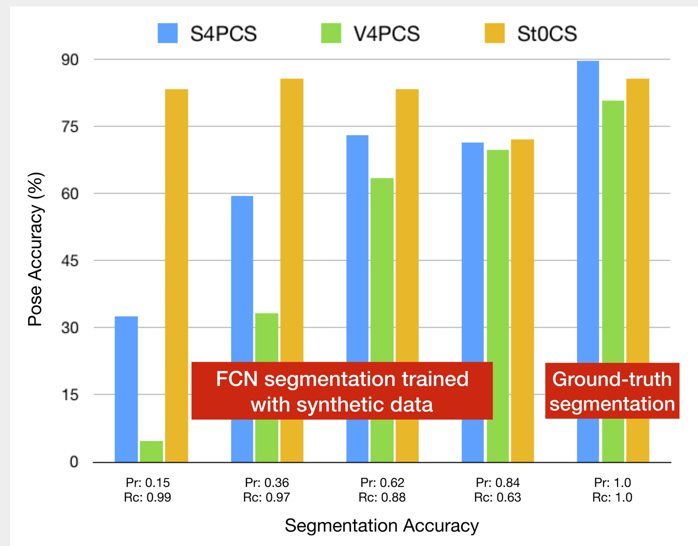
- For the sampled base, a set of congruent 4-points is found on the model.
- Congruency is determined by pair of points sharing the same features and intersecting line segments maintaining affine invariant ratios.
- `Pointset alignment score` is computed for each pose hypothesis generated from congruent sets.



Experiments

- We evaluate the accuracy of the pose with highest pointset alignment score.
- Area under the accuracy-threshold curve (AUC) is used to evaluate pose success on the YCB dataset.

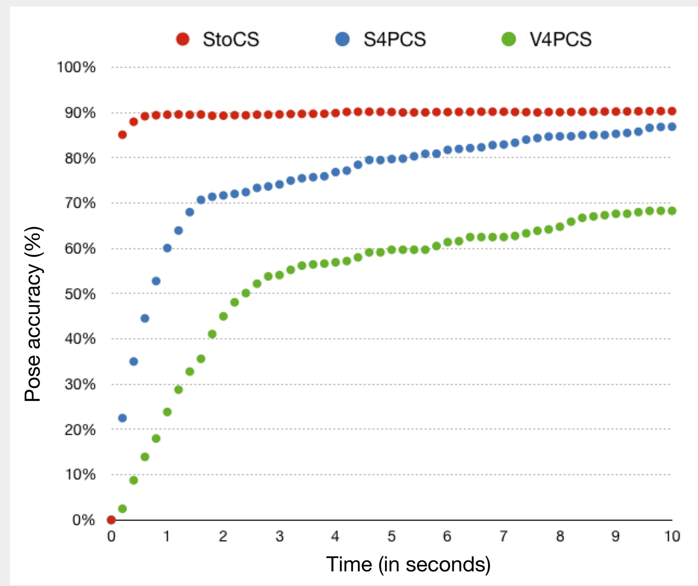
Method	Pose success	Time
PoseCNN	57.37%	0.2s
PoseCNN+ICP	76.53%	10.6s
PPF-Hough	83.97%	7.18s
Super4PCS	87.21%	43s
V4PCS	77.34%	4.32s
StoCS (OURS)	**90.1%**	**0.59s**



Robustness test with varying segmentation accuracy

- Mean rotation and translation error is reported on the Amazon picking challenge (APC) dataset for different registration techniques.

Method	Rot. error	Tr. error	Time
Super4PCS	8.83°	1.36cm	28.01s
V4PCS	10.75°	5.48cm	4.66s
StoCS (OURS)	**6.29°**	**1.11cm**	**0.72s**



Anytime results for different registration techniques

- Computation time for the different components of the registration process.

Method	Base Sampling	Set Extraction	Set Verification	#Set per base
Super4PCS	0.0045s	2.43s	19.98s	1957.18
V4PCS	0.0048s	1.98s	0.36s	46.61
StoCS (OURS)	**0.0368s**	**0.27s**	**0.37s**	**53.52**

{cm1074, ab1544, kb572}@cs.rutgers.edu

We acknowledge the support of JDX (JD.com) Silicon Valley Research Center and NSF grants IIS-1734492 and IIS-1723869.