Learning Robust Features for 3D Object Pose Estimation

Christos Papaioannidis and Ioannis Pitas
Department of Informatics, Aristotle University of Thessaloniki

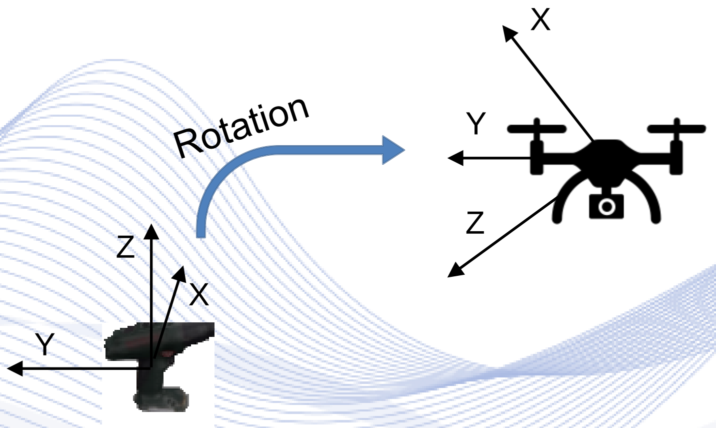
IEEE International Conference on Autonomous Systems (ICAS) 2021

Motivation: Autonomous robots or systems

- Estimate 3D pose of objects of interest in the surrounding environment.
- Use this information to decide on the next action according to a given objective.
  - e.g., grab and pass a tool to a human worker.
- Fast embedded execution.

3D object pose estimation problem

- Estimate the rotation between the object coordinate system and the reference coordinate system (e.g., camera of a UAV).
  - 3D rotation matrix.
  - Unit quaternion.
  - Euler angles.
- Sub-case of 6D object pose estimation.
- Challenges:
  - 3D pose representation.
  - Non-trivial object symmetries.



3D object pose estimation

- 3D object pose regression: directly regress 3D poses.
- 3D object pose classification: classify an object image in a predefined number of orientation classes.
- **3D object pose retrieval**: match extracted 3D pose-related image features with a set of orientation class templates.

3D object pose retrieval

- Advantages over 3D pose regression and classification methods:
  - Only one trained CNN is required.
  - Object classes and 3D poses are predicted simultaneously.
  - A lightweight CNN can be used, enabling real-time execution or execution in embedded systems with limited computational capabilities.

- Disadvantages:
  - Accuracy is limited by the number of the selected orientation class templates.

Proposed method

- A CNN $\boldsymbol{f}$ is trained to extract 3D pose-related features.
- Using the trained CNN, codebook features $\mathbf{f}_{c_i}, i = 1, \dots, K$ are first calculated offline and stored: $\mathbf{f}_{c_i} = \boldsymbol{f}\left(\mathbf{X}_{c_i}\right)$.
- Given a test object image $\mathbf{X}$, the corresponding feature vector is extracted using the same trained CNN: $\mathbf{f} = \boldsymbol{f}(\mathbf{X})$.
- The extracted test image feature vector $\mathbf{f}$ is matched to the most similar $\mathbf{f}_{c_i}, i = 1, \dots, K$ using a Nearest Neighbor algorithm.

Proposed method

- 3D rotations are represented by unit quaternions:

$$\mathbf{q} = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$$

where q_0, q_1, q_2, q_3 are real numbers and $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$.

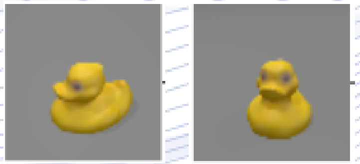
- Advantages:
  - More compact 3D pose representation compared to the rotation matrix.
  - Numerically stable.
  - Avoid the gimbal lock problem.

Dataset construction

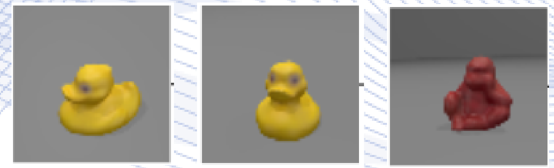
- The CNN is trained using a carefully designed training dataset.
- Training samples $s = \{\mathbf{X}, c, \mathbf{q}\}$, $\mathbf{X}$: RGB-D object image, c: object class label, $\mathbf{q}$: 3D pose quaternion.
- Two separate training sets are constructed: a set P containing sample pairs and a set T containing sample triplets.

Dataset construction

- Each entry of the pair set P consists of training samples s_i, s_j that belong in the same object class and under arbitrary poses.
- Triplet set T entries contain three training samples, s_i, s_j, s_k:
  - s_i, s_j are samples belonging to the same object class,
  - while s_k is a sample coming from any different class.

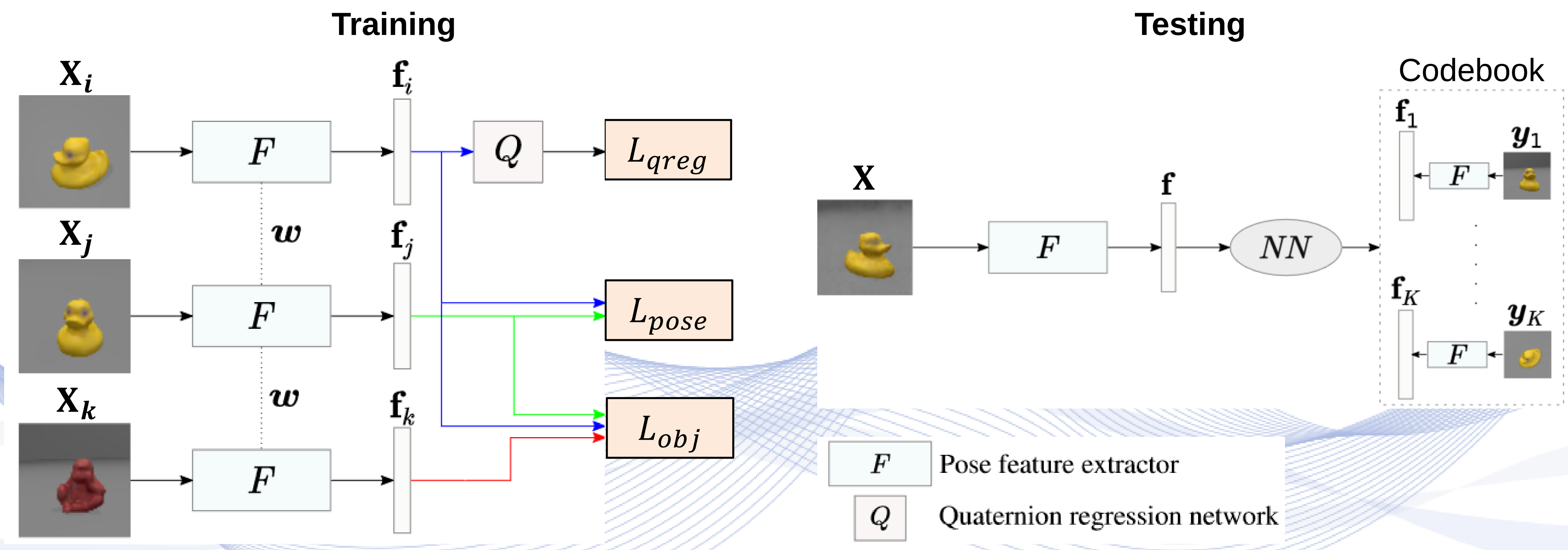


Entry of P



Entry of T

Network architecture



Loss functions

- Objective loss function:

$$L = \lambda_p L_{pose} + \lambda_o L_{obj} + \lambda_r L_{qreg}.$$

- Pairwise loss using entries of P:

$$L_{pose} = \sum_{s_i, s_j} \varphi(\mathbf{q}_i, \mathbf{q}_j) \cdot \{\left\|\mathbf{f}_i - \mathbf{f}_j\right\|_2^2 - 2\arccos(|\mathbf{q}_i^T \mathbf{q}_j|)\}^2.$$

- Symmetry-aware term based on depth images:

$$\varphi(\mathbf{q}_i, \mathbf{q}_j) = \left\|\mathbf{d}_{\mathbf{q}_i} - \mathbf{d}_{\mathbf{q}_j}\right\|_2^2.$$

Training

- Triplet loss using entries of T:

$$L_{obj} = \sum_{s_i, s_j, s_k} \frac{\left\|\mathbf{f}_i - \mathbf{f}_j\right\|_2}{\|\mathbf{f}_i - \mathbf{f}_k\|_2 + \varepsilon}.$$

- Quaternion regression loss:

$$L_{qreg} = 2\arccos(|\mathbf{q}^T\hat{\mathbf{q}}|).$$

Quantitative evaluation

- Evaluation of the proposed method using the angular error in degrees given by:

$$E(\mathbf{q}, \hat{\mathbf{q}}) = 2\arccos(|\mathbf{q}^T \hat{\mathbf{q}}|).$$

- 3D pose estimation accuracy at threshold t: percentage of test samples for which the angular error is below a threshold angle t.

Quantitative evaluation

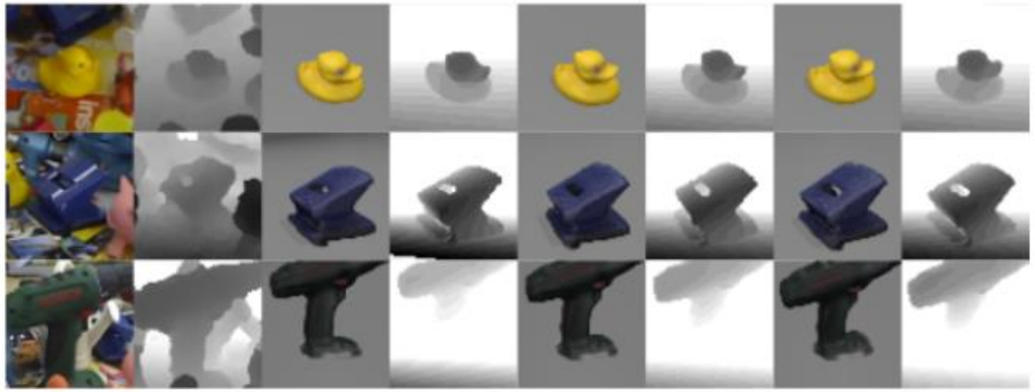
- 3D object pose estimation accuracy on the LineMod [1] dataset.

	Angular threshold t								
	5°	10°	15°	20°	30°	40°	45°	Mean (Median) $\pm$ Std	Object classification
$3DPOD$ [2]	40.15%	72.72%	86.02%	91.76%	95.42%	96.90%	97.34%	$12.75^\circ(7.06^\circ) \pm 24.61^\circ$	98.94%
$PEDM$ [3]	-	60.00%	-	93.20%	-	98.00%	-	-	99.30%
$PGFL$ [4]	41.28%	83.07%	93.98%	97.43%	99.11%	99.52%	99.60%	$6.89^\circ(5.79^\circ) \pm 6.29^\circ$	99.64%
QL [5]	41.37%	82.02%	95.32%	98.49%	99.72%	99.92%	99.94%	$6.64^\circ(5.78^\circ) \pm 5.14^\circ$	99.50%
$ours$	**44.13%**	**84.25%**	**95.76%**	**98.77%**	**99.84%**	**99.93%**	**99.94%**	$\mathbf{6.31^\circ(5.53^\circ) \pm 4.58^\circ}$	**99.68%**

[1] 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," ACCV, 2012.
[2] P. Wohlhart and V. Lepetit, "Learning descriptors for object recognition and 3D pose estimation," CVPR, 2015.
[3] S. Zakharov, W. Kehl, B. Planche, A. Hutter, and S. Ilic,"3D object instance recognition and pose estimation using triplet loss with dynamic margin," IROS, 2017.
[4] V. Balntas, A. Doumanoglou, C. Sahin, J. Sock, R. Kouskouridas, and T.-K. Kim, "Pose guided rgbd feature learning for 3D object pose estimation," ICCV, 2017.
[5] C. Papaioannidis and I. Pitas, "3D object pose estimation using multi-objective quaternion learning," TCSVT, 2019.

Artificial Intelligence & Information Analysis Lab

Qualitative evaluation



Evaluation on LineMod objects.

Evaluation on a previously unseen object.

Artificial Intelligence & Information Analysis Lab

Thank you for your attention!

Contact: cpapaionn@csd.auth.gr

Artificial Intelligence &
Information Analysis Lab