

Pose Guided RGBD Feature Learning for 3D Object Pose Estimation

Vassileios Balntas
Imperial College London

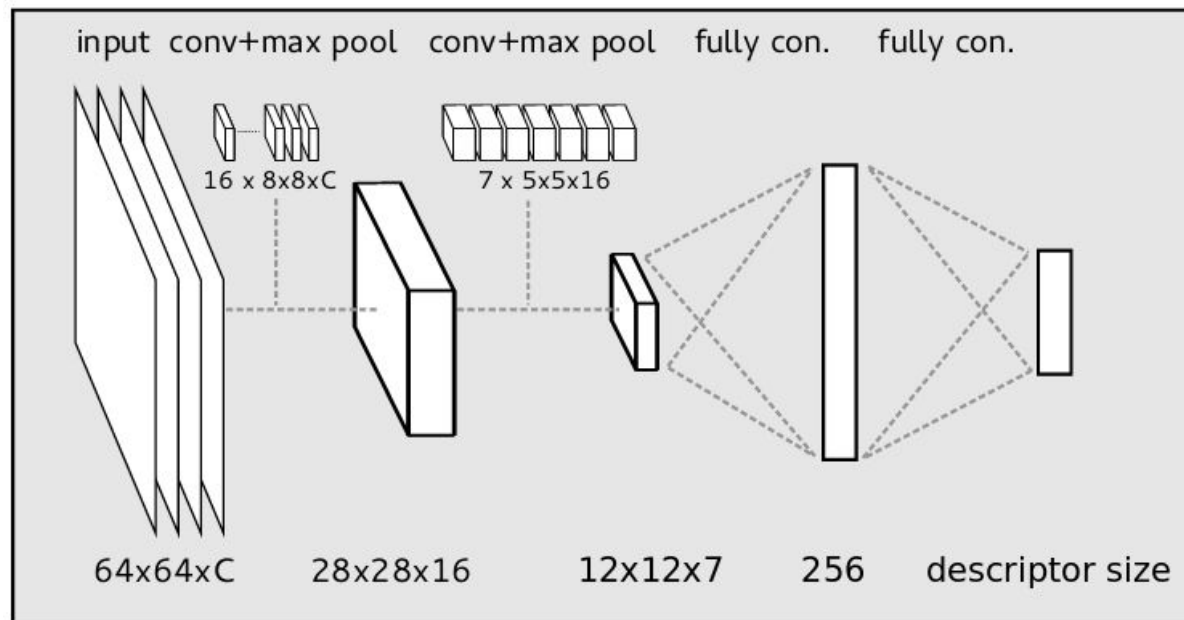
<http://www.robots.ox.ac.uk/~balntas/>

RGBD pose descriptors

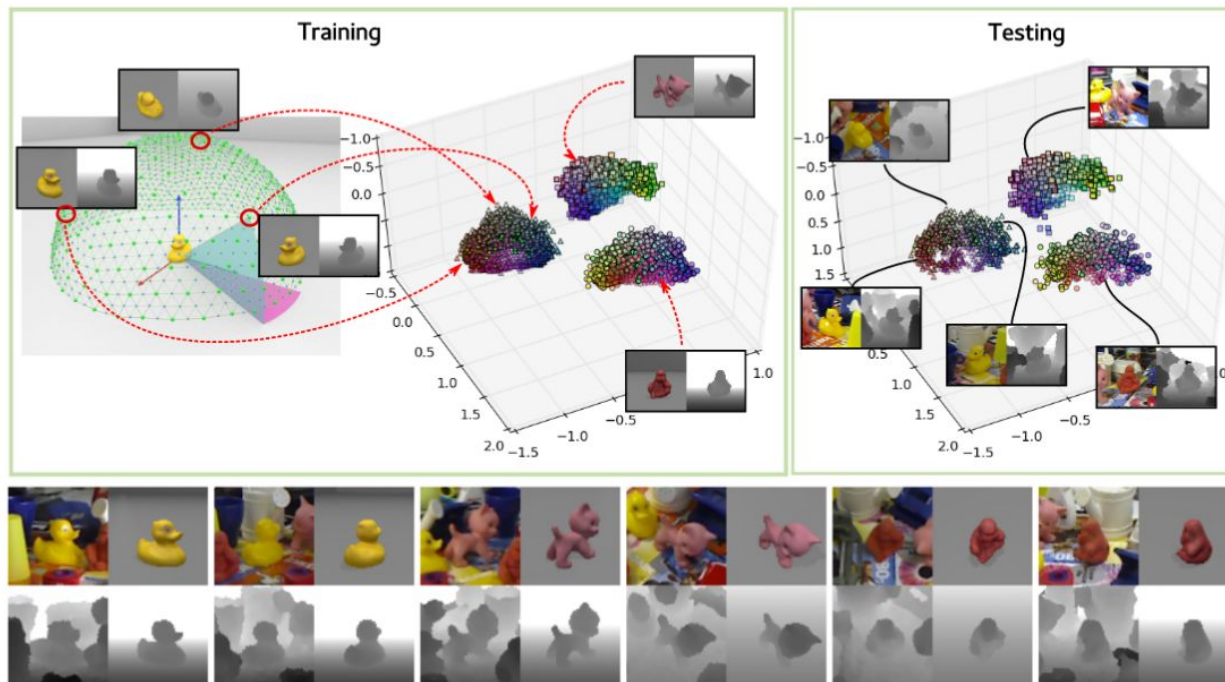


$$I \in \mathbb{R}^{N \times N \times 4} \rightarrow f(I) \in \mathbb{R}^D$$

Learning 3D pose descriptors



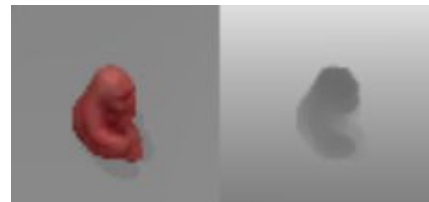
Learning 3D pose descriptors



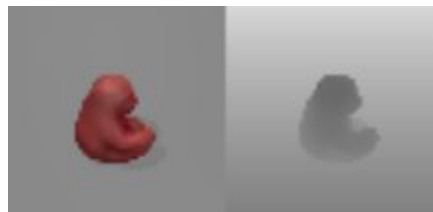
RGBD pose descriptors

set of templates

p_1

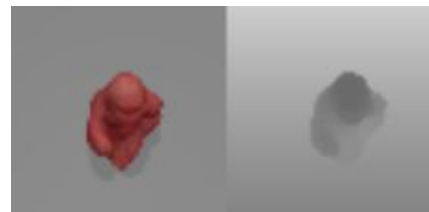


p_2

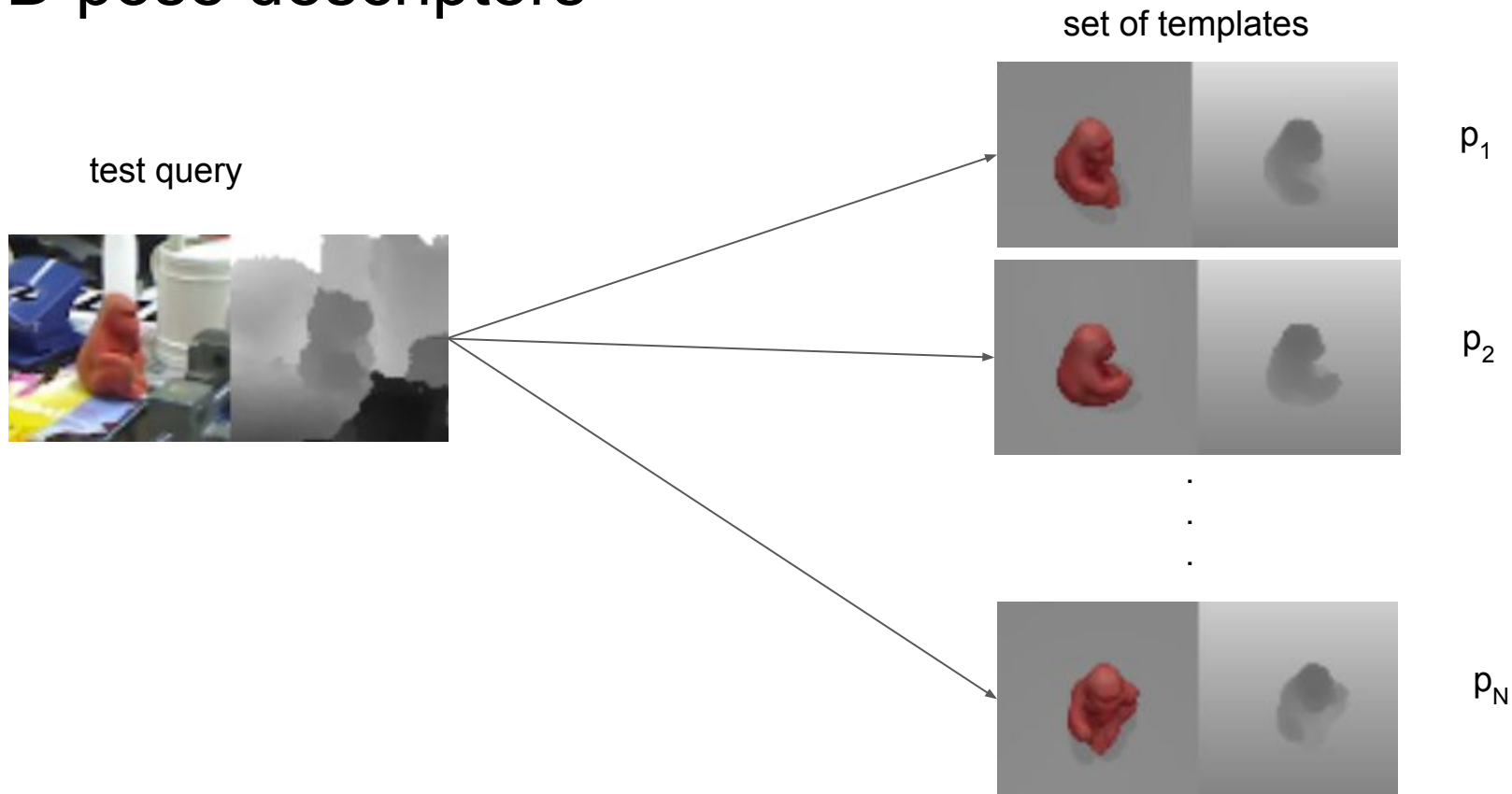


·
·
·

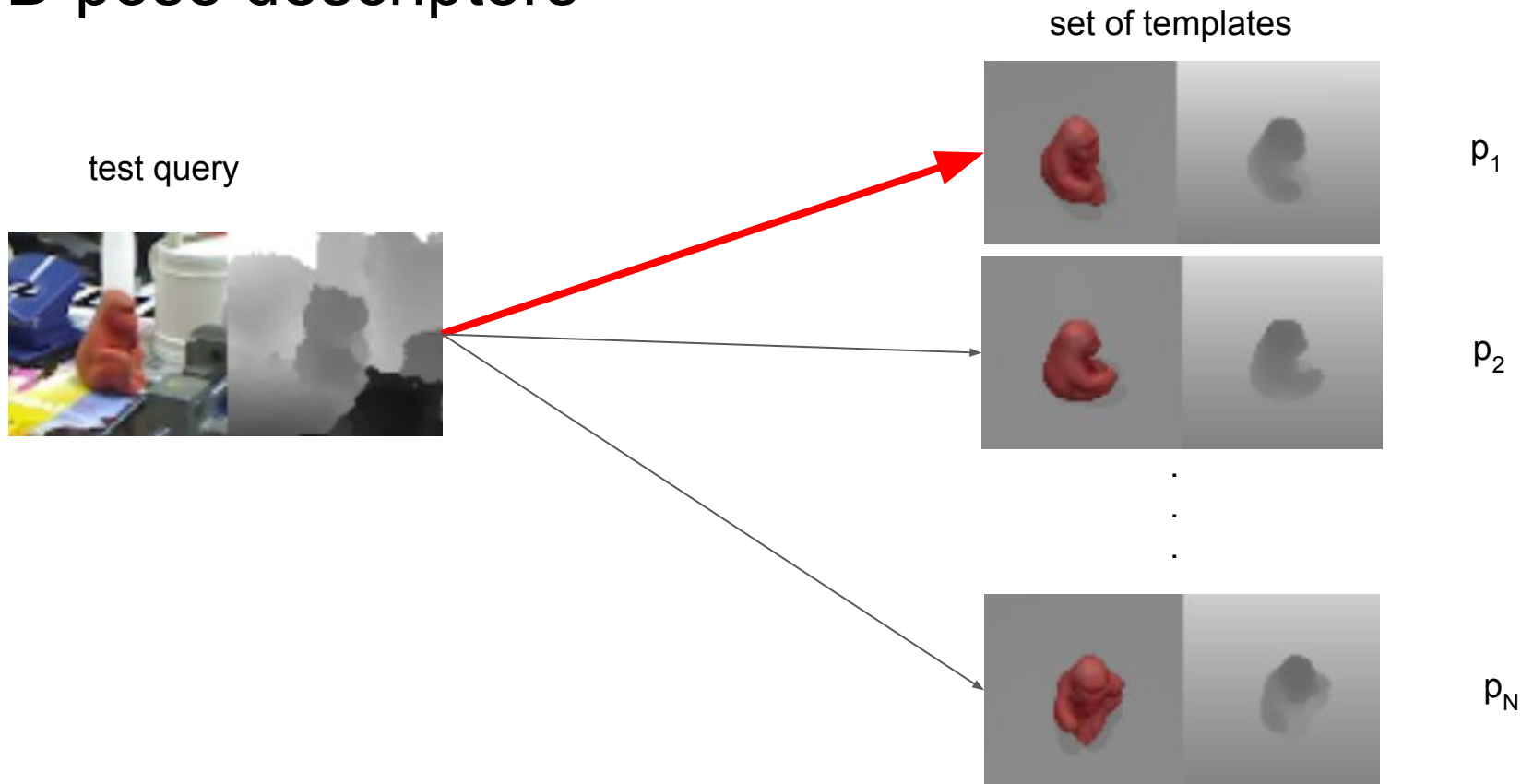
p_N



RGBD pose descriptors



RGBD pose descriptors



Learning 3D pose descriptors

$$\mathcal{L} = \mathcal{L}_{\text{triplets}} + \mathcal{L}_{\text{pairs}} + \lambda ||w'||_2^2$$

$$\mathcal{L}_{\text{triplets}} = \sum_{(s_i, s_j, s_k) \in \mathcal{T}} c(s_i, s_j, s_k)$$

$$\mathcal{L}_{\text{pairs}} = \sum_{(s_i, s_j) \in \mathcal{P}} ||f_w(x_i) - f_w(x_j)||_2^2$$

$$c(s_i, s_j, s_k) = \max \left(0, 1 - \frac{||f_w(x_i) - f_w(x_k)||_2}{||f_w(x_i) - f_w(x_j)||_2 + m} \right)$$

$$\mathcal{L} = \mathcal{L}_{\text{triplets}} + \mathcal{L}_{\text{pairs}} + \lambda ||w'||_2^2$$

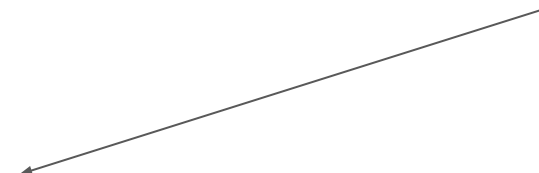


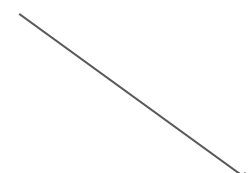
Pose guided feature learning - motivation

- pose is used only as an indicator to form triplets and pairs
- optimiser only learns relationships of limited form by the $<$, $>$ indicators
- no implicit relation between difference in the feature space and pose similarity

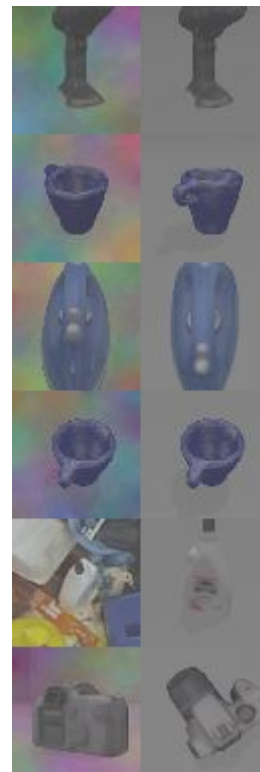
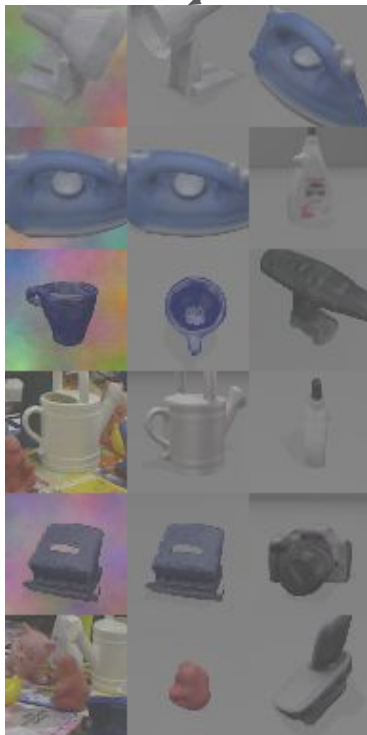
Pose guided feature learning - motivation

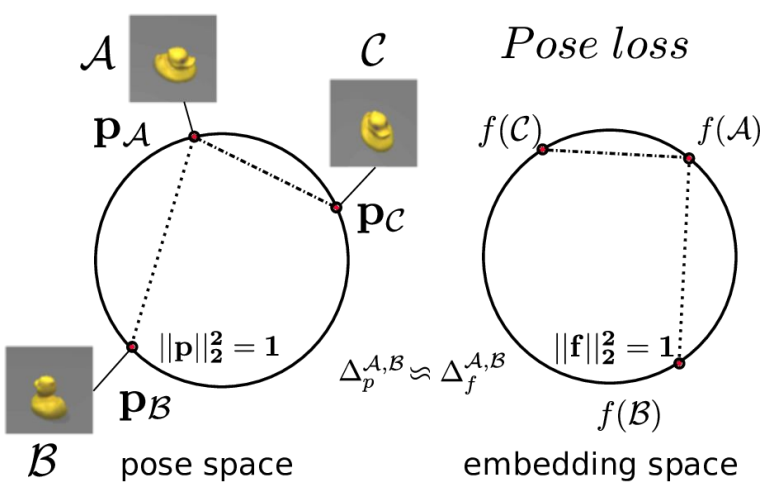
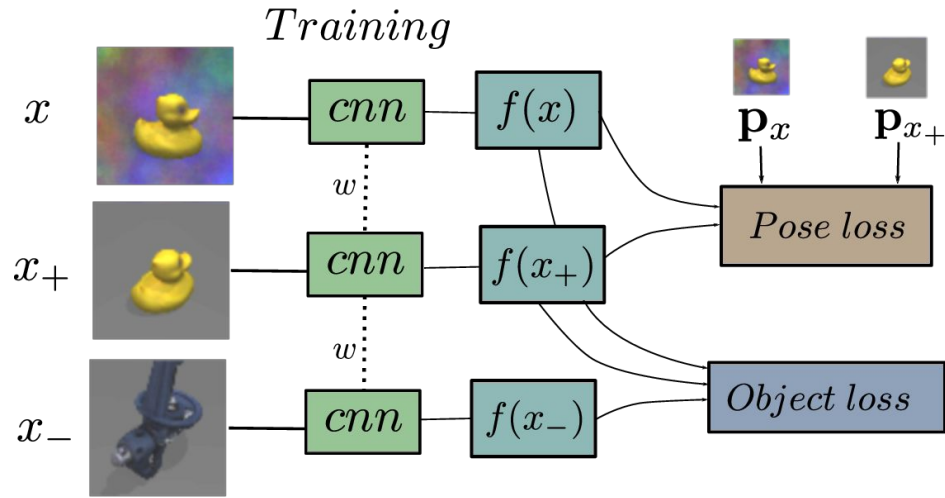
$$\mathcal{L} = \mathcal{L}_{\text{triplets}} + \mathcal{L}_{\text{pairs}} + \lambda ||w'||_2^2$$


$$\mathcal{L}_{\text{triplets}} = \max(0, \mu + ||f_a - f_p||_2^2 - ||f_a - f_n||_2^2)$$

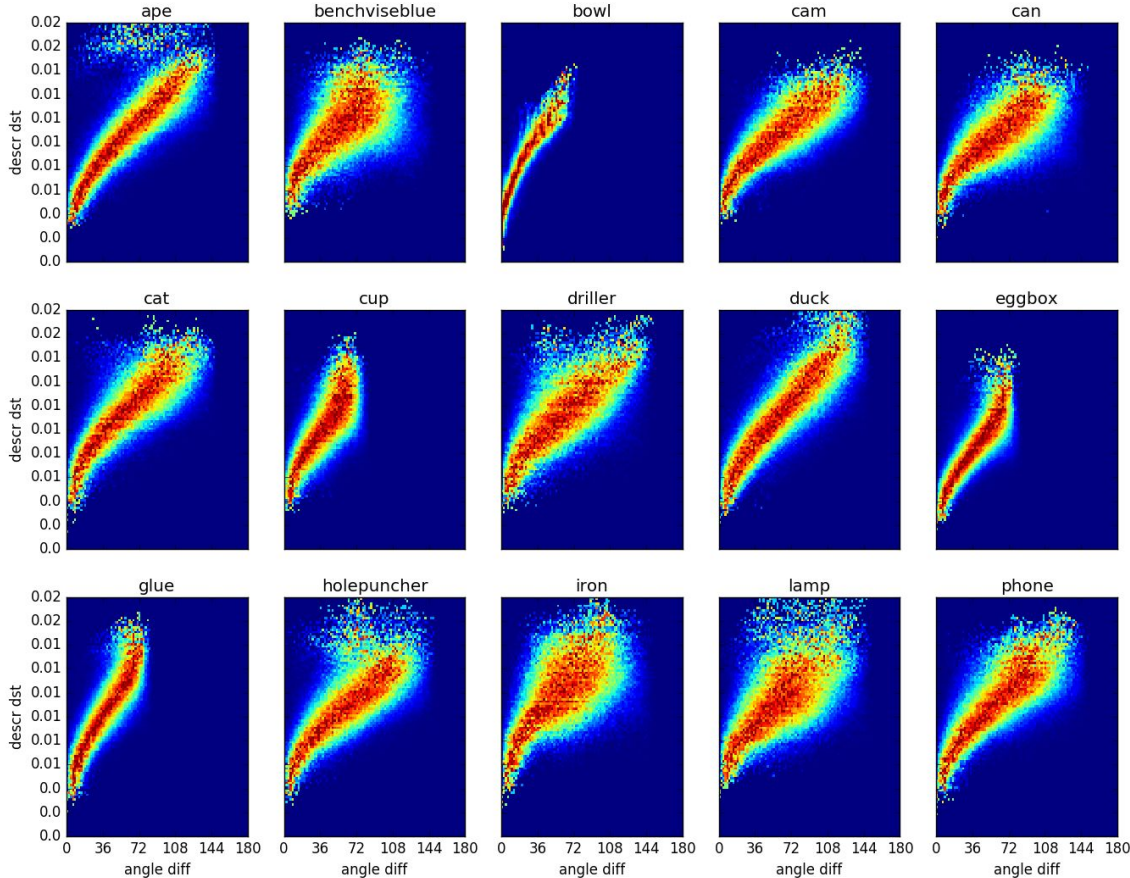

$$\mathcal{L}_{\text{pairs}} = ||f_1 - f_2||_2^2 - \text{sim}(p_1, p_2)^2$$

$$\mathcal{L} = \mathcal{L}_{\text{triplets}} + \mathcal{L}_{\text{pairs}} + \lambda ||w'||_2^2$$

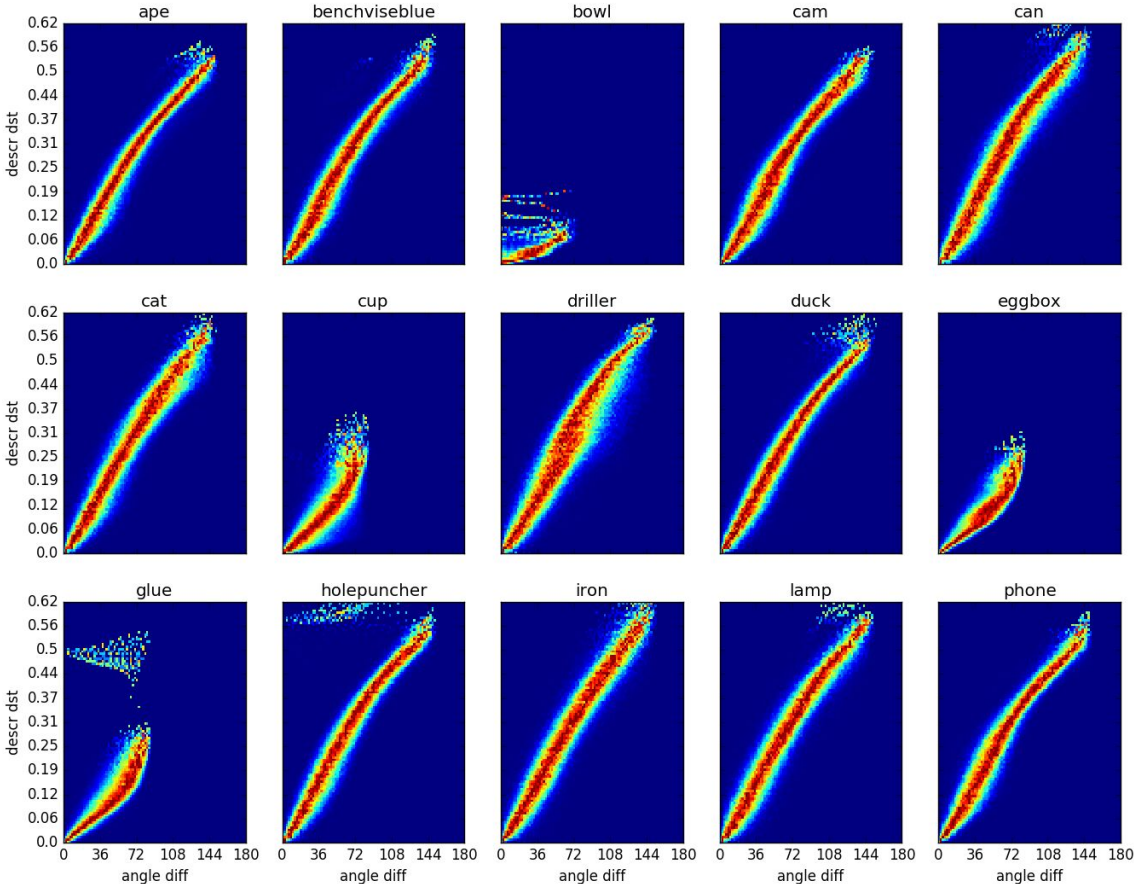




Wohlhart and Lepetit



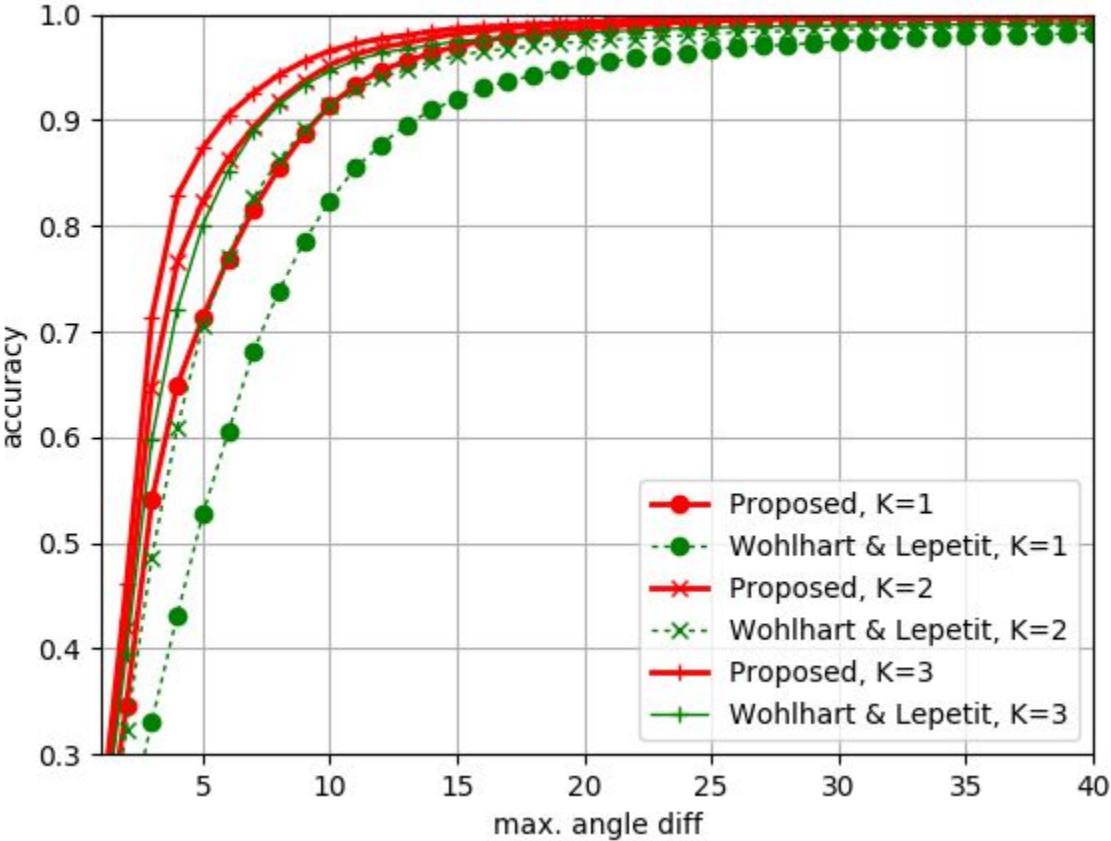
Ours

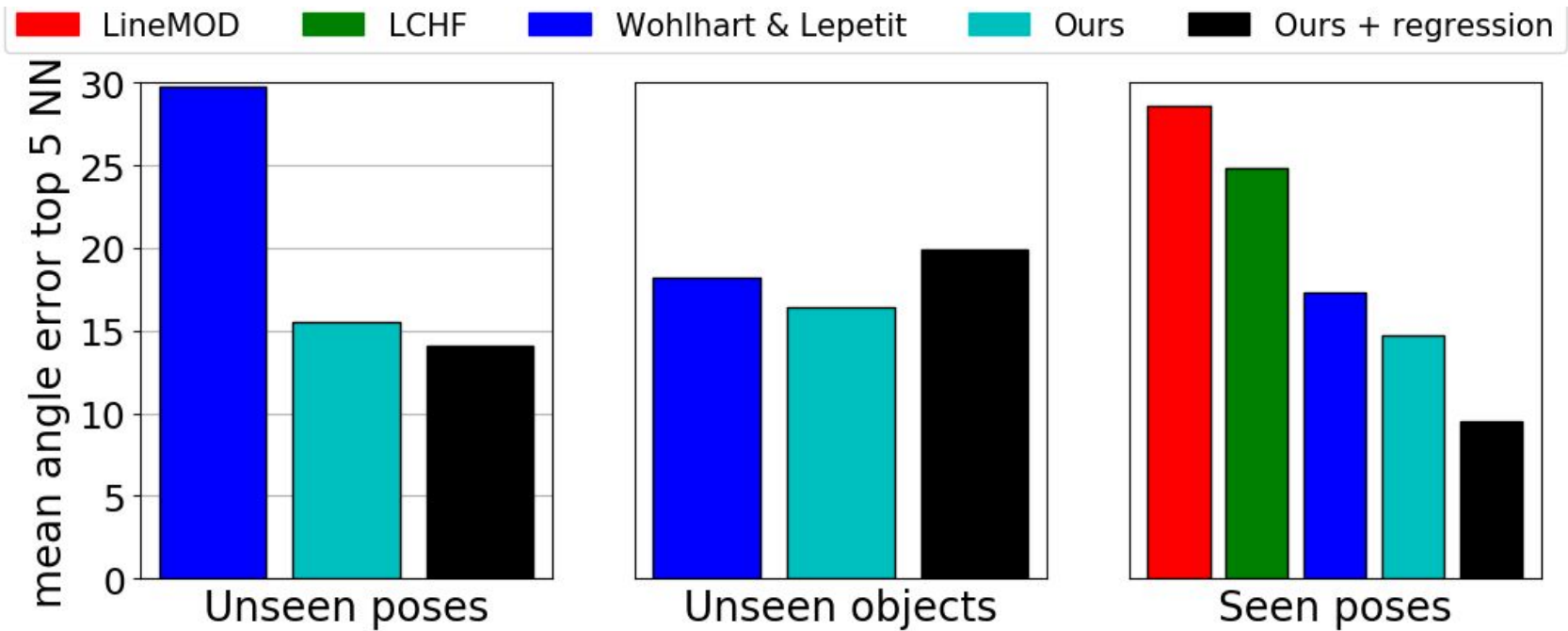


NN retrieval performance

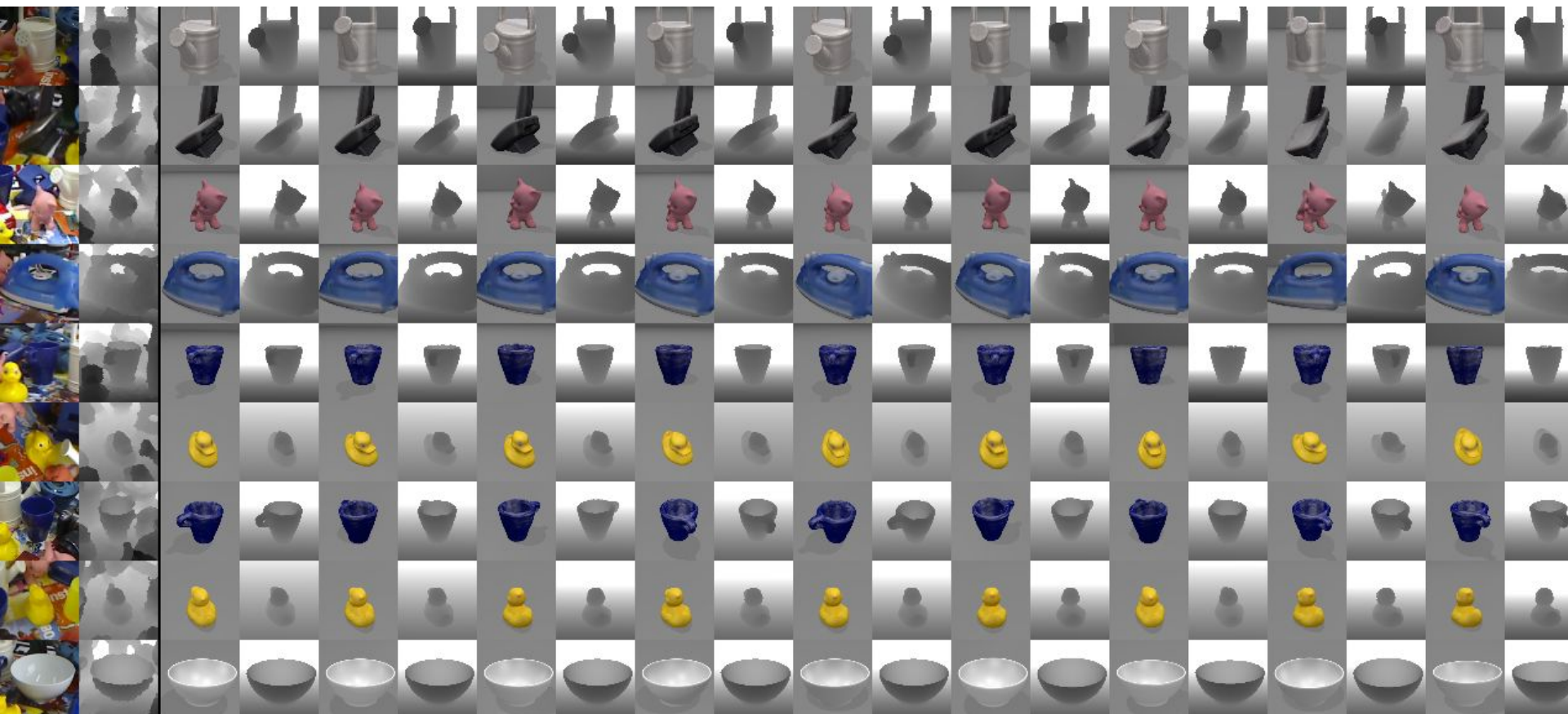
Method	5°	10°	15
Ours	63.43%	82.81%	90.58%
Wohlhart & Lepetit	40.32%	65.13%	80.95%

% of correctly retrieved 1st nearest neighbour for different sensitivity thresholds

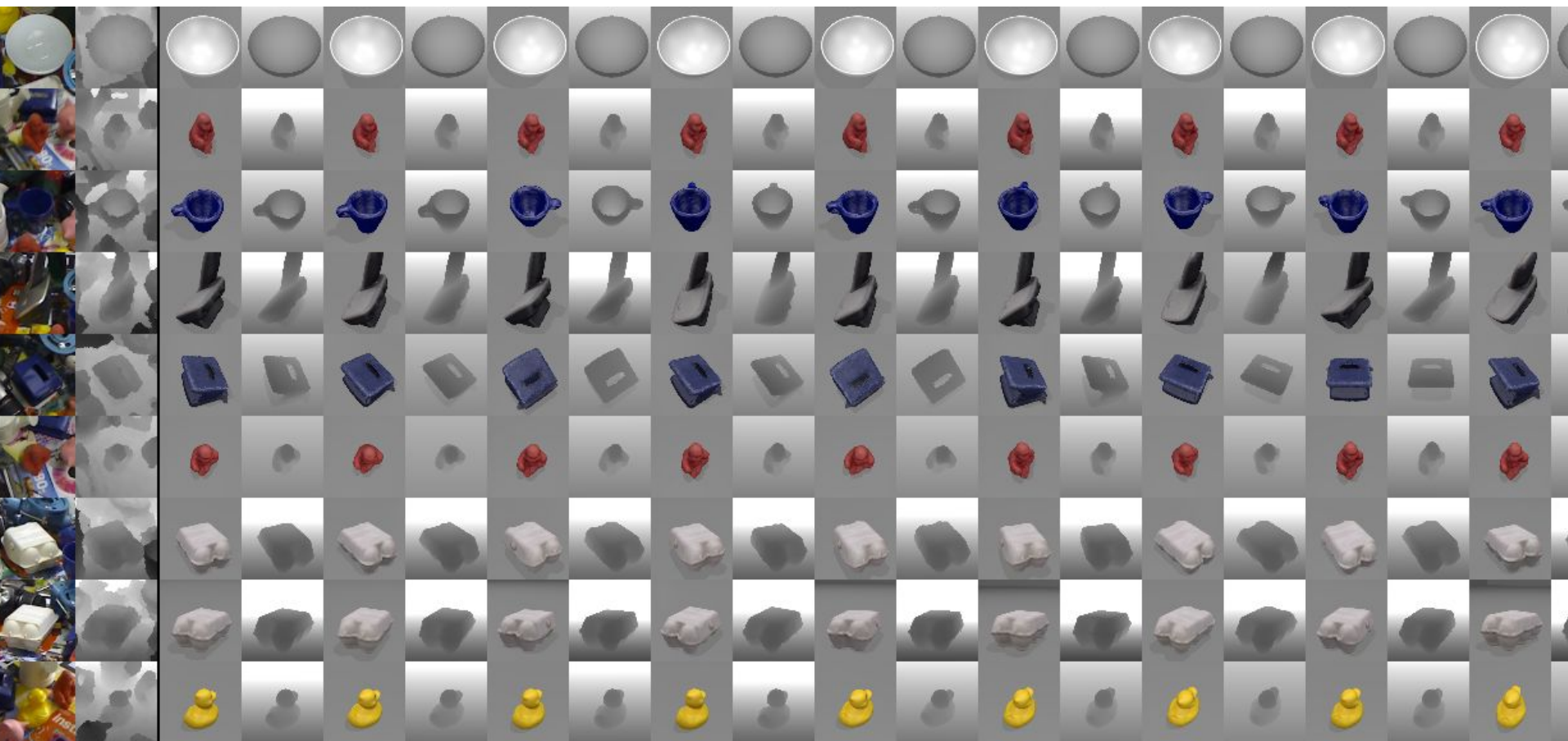




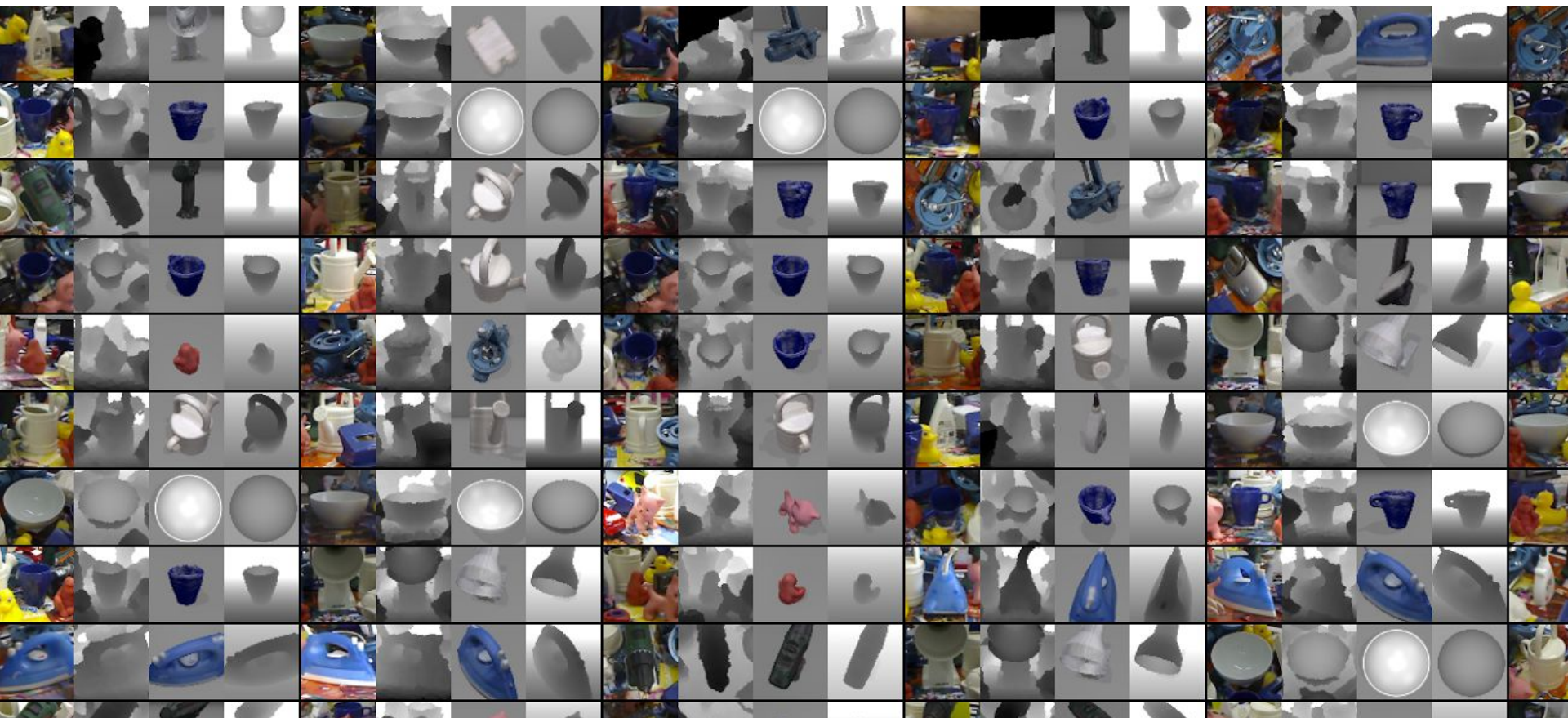
Proposed method / Correct Retrieval Examples



Wohlgart and Lepetit



Proposed method / Incorrect retrieval examples



[illegible]

Conclusions

- A method to optimise RGBD descriptors based on directly relating the difference in the feature space with the pose similarity
- Our method outperforms the previous state of the art in RGBD descriptors by a large margin
- Significant improvement can be seen in high accuracy area ($t=5^\circ$)