Ki-Pode: Keypoint-based Implicit Pose Distribution Estimation of Rigid Objects

Thorbjørn Mosekjær Iversen, Rasmus Laurvig Haugaard, and Anders Glent Buch
SDU Robotics, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, DK

University of Southern Denmark

1. Motivation

Pose estimation is the task of recovering an object's 3D rotation and 3D translation. Prior work has primarily been concerned with the estimation of a single best estimate. However, object poses are often ambiguous e.g. due to symmetry or occlusion. Since pose ambiguity cannot be expressed by a single pose estimate, we estimate an entire probability distribution over object poses. As our primary concern was reliability, our method provides distribution estimates which are expressive, conservative, and interpretable.

Expressive: We express the pose distribution using an implicit formulation since this allows our method to model distributions of arbitrary complexity.

Conservative: Estimates must be conservative since overconfident estimates can lead to false positives, resulting in an unreliable system.

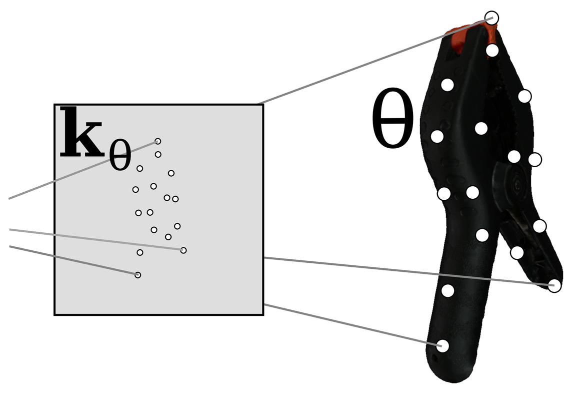
Interpretable: Since our method is trained on synthetic data, it is valuable to have interpretable estimates, to aid in identifying if the sim-to-real domain transfer is successful.

2. Method

Pose from keypoints

The probability distribution over the pose θ can be expressed in terms of a joint distribution over projected keypoints $\mathbf{k}_\theta$. Since the projected keypoints uniquely define a pose, it can be derived that:

$$\mathrm{p}(\theta|I) = \int \mathrm{p}(\theta|\mathbf{k}, I)\mathrm{p}(\mathbf{k}|I)d\mathbf{k} = \mathrm{p}(\mathbf{k}_\theta|I)$$



Joint to marginals

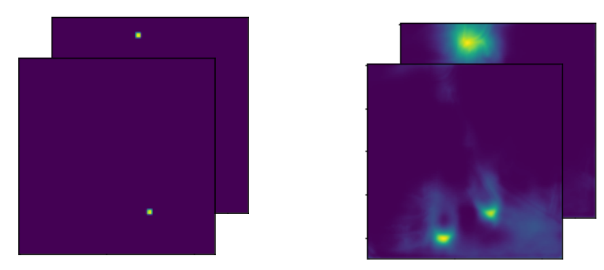
The joint keypoint distribution is approximated from the marginal distributions, using the assumption of **strong correlation** ($\mathrm{p}(k_i, k_j) = \mathrm{p}(k_i)$). This approximation is conservative since error introduced by the approximation will overestimate regions of low probability:

$$\mathrm{p}(\mathbf{k}_\theta|I) \approx C \sqrt[N]{\prod_{i=1}^{N} \mathrm{p}(k_{\theta,i}|I)}$$

Heatmap estimation

The marginal distributions, referred to as a heatmap, are estimated using a **U-Net**[1]. The problem is formulated as **spatial classification** and the network is trained using cross-entropy loss with the target distributions being isotropic Gaussians.

$$L = -\sum_{i=1}^{N} \mathcal{N}(k_i, \sigma) \log \hat{\mathrm{p}}(k_i|I)$$

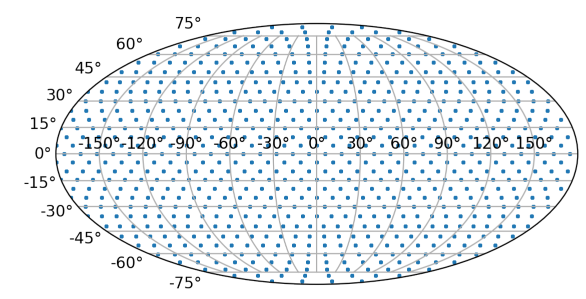


Target Model

Normalization

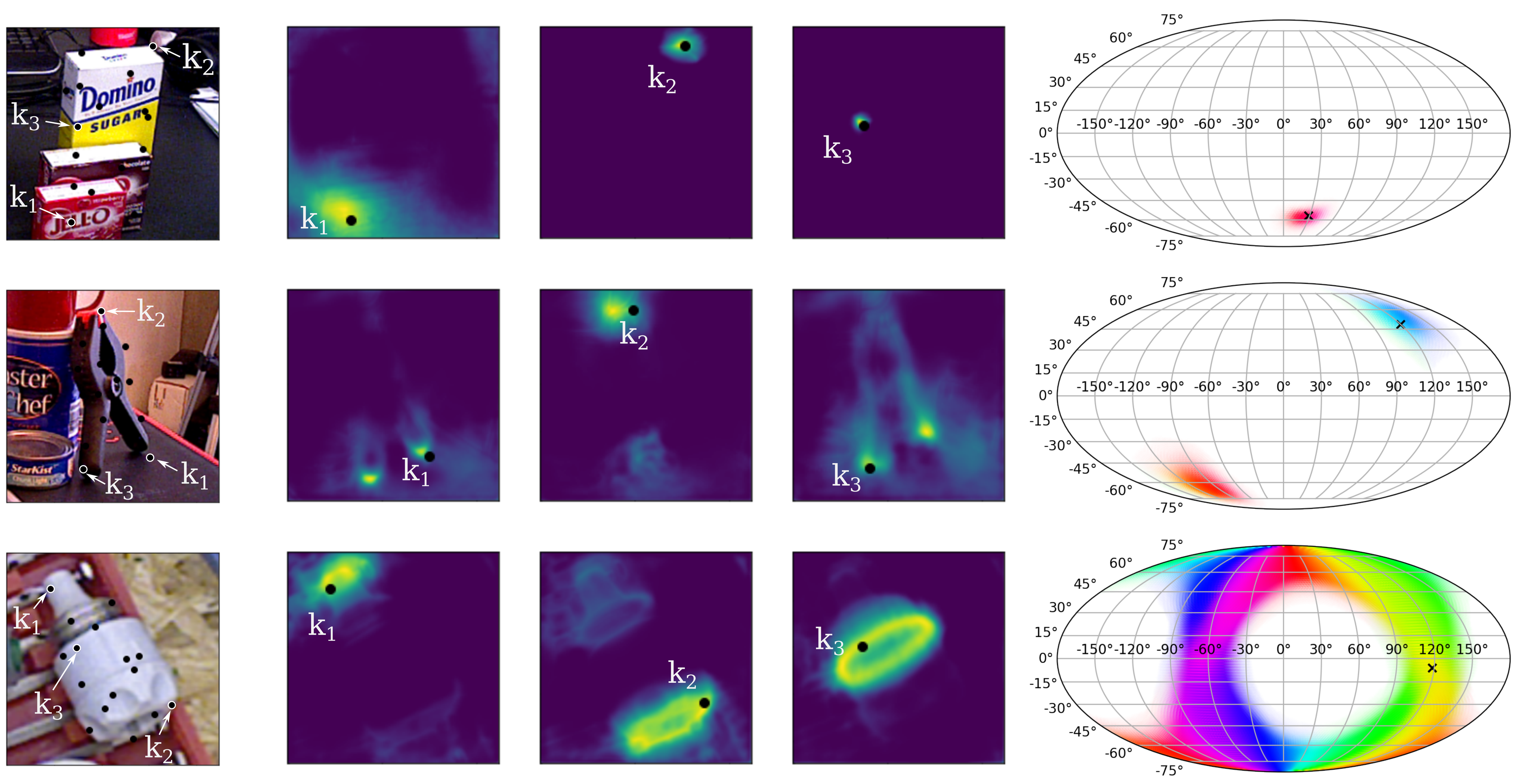
SO(3) is sampled in an **equivolumetric grid** using the Healpix-based method in [2]. The unnormalized likelihood is estimated for the M samples, interpreted as probability, and normalized such that the normalized likelihood can be computed as:

$$\hat{\mathrm{p}}(\theta_i|I) \approx \hat{\mathrm{P}}(\theta_i|I)/(\pi^2/M)$$



3. Results

Our method has been applied to orientation estimation, so it can be compared to existing works. The figure shows our method applied to three objects. With 5 HEALPix recursions (1.8 million samples) the computation time was ~125ms. The mean log-likelihood of the ground truth pose is used as a metric for comparison. Our method achieves state-of-the-art results on the YCB-V dataset and modest results on the T-LESS dataset. Furthermore, our method performs reliably on all objects.



YCB-V

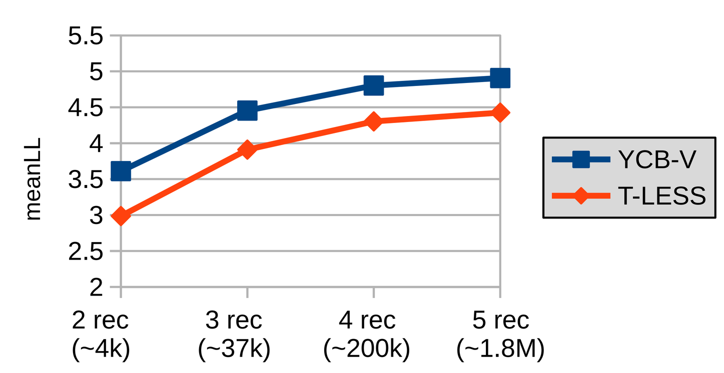
	Summary of Evaluation in [3]				
	I	II	III	IV	Ki-Pode (ours)
master chef can	-0.78	-1.66	*2.32*	0.09	**4.36**
cracker box	*4.03*	3.75	-0.09	3.71	**6.05**
sugar box	3.77	*5.94*	2.62	4.20	**6.47**
tomato soup can	0.90	2.02	2.52	*3.99*	**5.16**
mustard bottle	3.85	4.61	3.02	*4.81*	**5.45**
tuna fish can	-3.12	-0.20	*2.64*	1.23	**5.10**
pudding box	2.18	2.64	3.13	*4.54*	**4.91**
gelatin box	4.65	*6.25*	3.53	5.73	**6.35**
potted meat can	1.18	***3.28***	1.60	3.06	3.11
banana	2.58	0.58	2.27	*3.70*	**4.80**
pitcher base	3.34	4.68	2.35	*4.88*	**5.39**
bleach cleanser	3.91	***4.70***	2.29	3.38	4.02
bowl	***1.37***	-2.77	-1.21	-9.62	-1.50
mug	3.73	2.50	2.75	***4.72***	4.62
power drill	4.31	*5.92*	2.43	4.17	**6.14**
wood block	2.64	-2.09	-0.51	***4.49***	-1.76
scissors	*3.74*	0.51	0.66	1.63	**5.49**
large marker	-7.59	-0.29	*1.02*	-8.13	**3.94**
large clamp	-5.64	-6.67	-1.63	-3.54	**3.07**
extra large clamp	-5.20	-2.92	*0.19*	-5.03	**2.25**
foam brick	-0.26	-2.28	*1.70*	-12.0	**2.44**
All	0.86	1.71	1.74	1.43	**4.09**

Results I - IV published by Okorn et al. [3]

T-LESS

Deng[1]	5.3
Gilitschenski[1]	6.9
Prokudin[1]	8.8
IPDF[1]	9.8
Ki-Pode (ours)	3.3

[1] Results published by Murphy et al. [4]



4. Conclusion

We have presented a novel pose distribution estimation method, which provides estimates that:

- rely on conservative approximations to ensure reliability
- are highly expressive due to being formulated implicitly
- have high interpretability due to the intermediary keypoint representation

5. Acknowledgement

This project was funded in part by Innovation Fund Denmark through the project "MADE FAST - Flexible, Agile, Sustainable manufacturing enabled by Talented employees"

Manufacturing Academy of Denmark

6. References

[1] O. Ronneberger, et al. *U-net: Convolutional networks for biomedical image segmentation*. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2015, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

[2] A. Yershova, et al. *Generating uniform incremental grids on so (3) using the hopf fibration*. The International journal of robotics research, 29(7):801–812, 2010.

[3] B. Okorn, et al. *Learning orientation distributions for object pose estimation*. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 10580–10587. IEEE, 2020.

[4] K. Murphy, et al. *Implicit-pdf: Non-parametric representation of probability distributions on the rotation manifold*. In Proceedings of the 38th International Conference on Machine Learning, pages 7882–7893. PMLR, 2021.

[5] M. Denninger, et al. *Blenderproc*. arXiv preprint arXiv:1911.01911, 2019.