Symmetry Aware Evaluation of 3D Object Detection and Pose Estimation in Scenes of Many Parts in Bulk

Romain Brégier[1,2], Frédéric Devernay[2], Laetitia Leyrit[1], James L. Crowley[2]

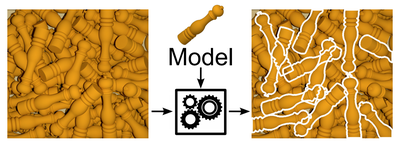
[1]: Siléane, Saint-Etienne, FRANCE [2]: Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, F-38000 Grenoble, FRANCE

International Conference on Computer Vision 2017

Summary

Evaluation methodology suitable for:

- Any rigid object.
- Arbitrary poses.
- Many object instances.
- Exact localization.



Main idea: Accounting for object's symmetries.

Limitations of existing approaches

Existing datasets

Hinterstoisser et al. [2]

- Mainly table-top scenarios.
- **Data redundancy**
  - Hundreds of images from only a few scenes.
  - Cause: pose annotation is tedious.

Results classification

Ground truth pose $\mathcal{G}_t$

Pose hypothesis $\mathcal{P}$

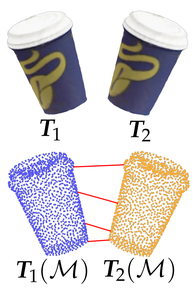
$\rightarrow$ True or False Positive?

Similarity criterion: $\mathrm{d}(\mathcal{G}_t, \mathcal{P}) < \delta$

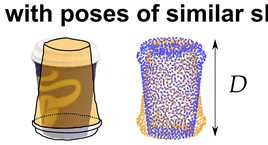
Existing dissimilarity measures d:

- Rigid transformation distance.
- IoU, Visible Surface Discrepancy [Hodaň et al. 2017].
- Pointset distance [2] for symmetric objects:

$$\mathrm{d}_{\text{Hinterstoisser}}(\boldsymbol{T}_1, \boldsymbol{T}_2) \triangleq \underset{x_1 \in \mathcal{M}}{\text{average}} \min_{x_2 \in \mathcal{M}} \|\boldsymbol{T}_2(x) - \boldsymbol{T}_1(x)\|$$



Issue with poses of similar shapes:

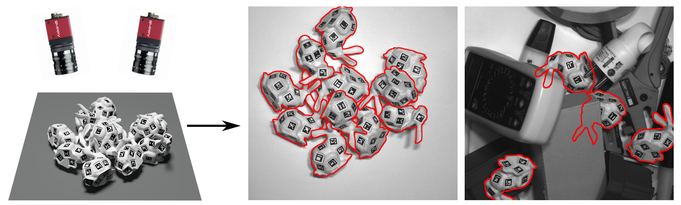


$\mathrm{d}_{\text{Hinterstoisser}}(\boldsymbol{T}_1, \boldsymbol{T}_2) = 4.5\%D$
$< \delta \ (= 10\%D$ [2])

$\rightarrow$ **wrongly classified as True Positive!**

Generating automatically annotated datasets

Marker-based pose annotation



Ground truth retrieved by detecting ArUco markers placed on object instances at known locations from multiple viewpoints.

Pros:

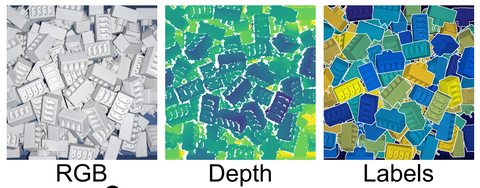
- Automatic annotation.
- Many instances, arbitrary pose.

Cons:

- Intrusive (RGB data).

Synthetic dataset generation

Rigid body and sensor simulation $\rightarrow$



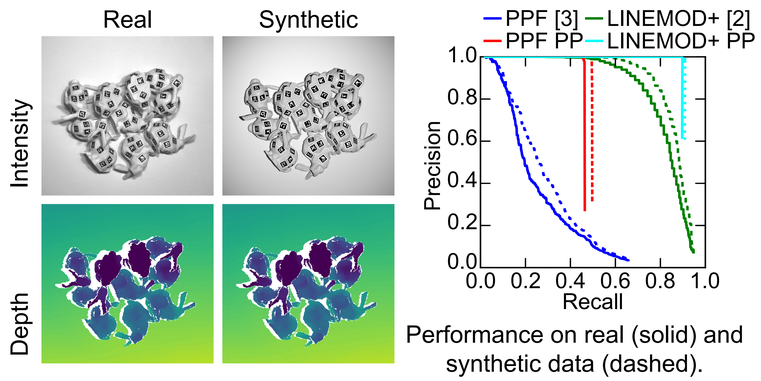
Pros:

- Control of every parameter.
- Annotations for free.

Cons:

- Synthesis of natural scenes may be challenging.

Comparison between the two approaches



Performance on real (solid) and synthetic data (dashed).

Dataset and evaluation tools available at
rbregier.github.io/dataset2017

Accounting for symmetries during evaluation

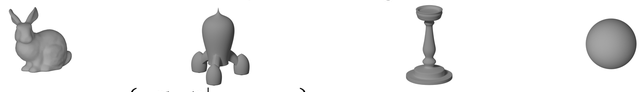
Pose definition

Pose equivalent to a set of rigid transformations:

$$\mathcal{P} \triangleq \{\boldsymbol{T} \circ \boldsymbol{G} | \boldsymbol{G} \in G\}$$

with $G \subset SE(3)$ the *group of proper symmetry* of the object.

Examples of such groups:



$\{\boldsymbol{I}\}$ $\quad \left\{\boldsymbol{R}_z^{2k\pi/3} \middle| k \in [\![0,3[\![\right\}$ $\quad \{\boldsymbol{R}_z^{\alpha} | \alpha \in [0, 2\pi[\}$ $\quad SO(3)$

Which symmetry group for these objects?

$\rightarrow G$ is application-dependent.

Proposed dissimilarity criterion: pose distance [1]

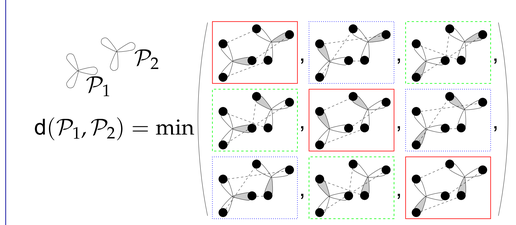
Length of the smallest displacement between two poses:

$$\mathrm{d}(\mathcal{P}_1, \mathcal{P}_2) \triangleq \min_{\boldsymbol{G}_1, \boldsymbol{G}_2 \in G} \mathrm{d}_{\text{no\_sym}}(\boldsymbol{T}_1 \circ \boldsymbol{G}_1, \boldsymbol{T}_2 \circ \boldsymbol{G}_2)$$

where

$$\mathrm{d}_{\text{no\_sym}}(\boldsymbol{T}_1, \boldsymbol{T}_2) \triangleq \sqrt{\frac{1}{S} \int_{\mathcal{S}} \|\boldsymbol{T}_2(\boldsymbol{x}) - \boldsymbol{T}_1(\boldsymbol{x})\|^2 ds}.$$

- Positioning error of the object's surface.
- Computation-friendly (15 multiplications, 25 additions).
- Fast neighborhood queries and pose averaging.
  $\rightarrow$ can be used at test time.
- **Exploiting symmetry at test time improve performance**:
  +50% AP increase in our experiments.



$\mathrm{d}(\mathcal{G}_t, \mathcal{P}) = 70\%D$
$> \delta$

$\rightarrow$ **correctly rejected as False Positive.**

References

[1] R. Brégier, F. Devernay, L. Leyrit, and J. Crowley, "Defining the pose of any 3D rigid object and an associated distance", 2016, under review (available on ArXiv).
[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 ACCV 2012.
[3] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition", in CVPR 2010.