On Evaluation of 6D Object Pose Estimation

Tomáš Hodaň, Jiří Matas, Štěpán Obdržálek

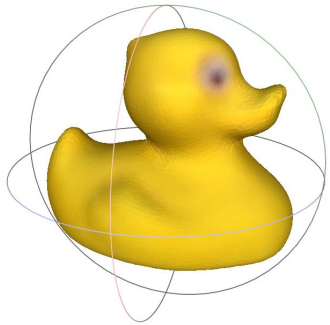
Center for Machine Perception
Czech Technical University in Prague

2nd International Workshop on Recovering 6D Object Pose (ECCV 2016)
9th October 2016, Amsterdam

How to Evaluate a 6D Object Pose Estimate?

Test RGB-D image

Evaluated 6D pose estimator



Training data
(e.g. a 3D model)

Evaluator

How good is the pose?

Images and model from dataset:
Hinterstoisser et al. (ACCV 2012)

Standard Approaches

Estimated pose: $\hat{\mathbf{P}} = (\hat{\mathbf{R}}, \hat{\mathbf{t}})$

Ground truth pose: $\bar{\mathbf{P}} = (\bar{\mathbf{R}}, \bar{\mathbf{t}})$

1. Translational and rotational error

Shotton et al., Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images, CVPR 2013

$$e_{\mathrm{TE}}(\hat{\mathbf{t}}, \bar{\mathbf{t}}) = \left\| \bar{\mathbf{t}} - \hat{\mathbf{t}} \right\|_2$$

$$e_{\mathrm{RE}}(\hat{\mathbf{R}}, \bar{\mathbf{R}}) = \arccos\left(\left(\mathrm{Tr}(\hat{\mathbf{R}}\bar{\mathbf{R}}^{-1}) - 1 \right) / 2 \right)$$

2. Average distance of corresponding model points

Hinterstoisser et al., Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes, ACCV 2012

$$e_{\mathrm{ADD}}(\hat{\mathbf{P}}, \bar{\mathbf{P}}; \mathcal{M}) = \operatorname*{avg}_{\mathbf{x} \in \mathcal{M}} \left\| \bar{\mathbf{P}}\mathbf{x} - \hat{\mathbf{P}}\mathbf{x} \right\|_2$$

Standard Approaches

Estimated pose: $\hat{\mathbf{P}} = (\hat{\mathbf{R}}, \hat{\mathbf{t}})$

Ground truth pose: $\bar{\mathbf{P}} = (\bar{\mathbf{R}}, \bar{\mathbf{t}})$

1. Translational and rotational error

Shotton et al., Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images, CVPR 2013

$$e_{\mathrm{TE}}(\hat{\mathbf{t}}, \bar{\mathbf{t}}) = \left\| \bar{\mathbf{t}} - \hat{\mathbf{t}} \right\|_2$$

$$e_{\mathrm{RE}}(\hat{\mathbf{R}}, \bar{\mathbf{R}}) = \arccos\left(\left(\mathrm{Tr}(\hat{\mathbf{R}}\bar{\mathbf{R}}^{-1}) - 1 \right) / 2 \right)$$

2. Average distance of corresponding model points

Hinterstoisser et al., Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes, ACCV 2012

$$e_{\mathrm{ADD}}(\hat{\mathbf{P}}, \bar{\mathbf{P}}; \mathcal{M}) = \underset{\mathbf{x} \in \mathcal{M}}{\mathrm{avg}} \left\| \bar{\mathbf{P}}\mathbf{x} - \hat{\mathbf{P}}\mathbf{x} \right\|_2$$

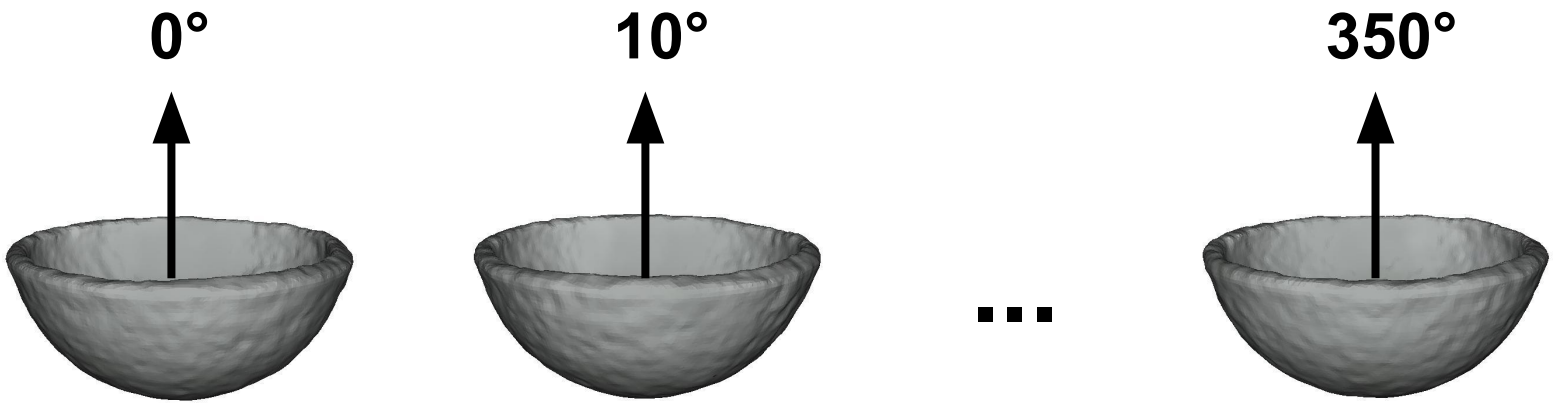
Intuitive, until pose ambiguity enters the game...

Pose Ambiguity

Due to **object symmetries**, multiple poses may be **indistinguishable**

Pose Ambiguity

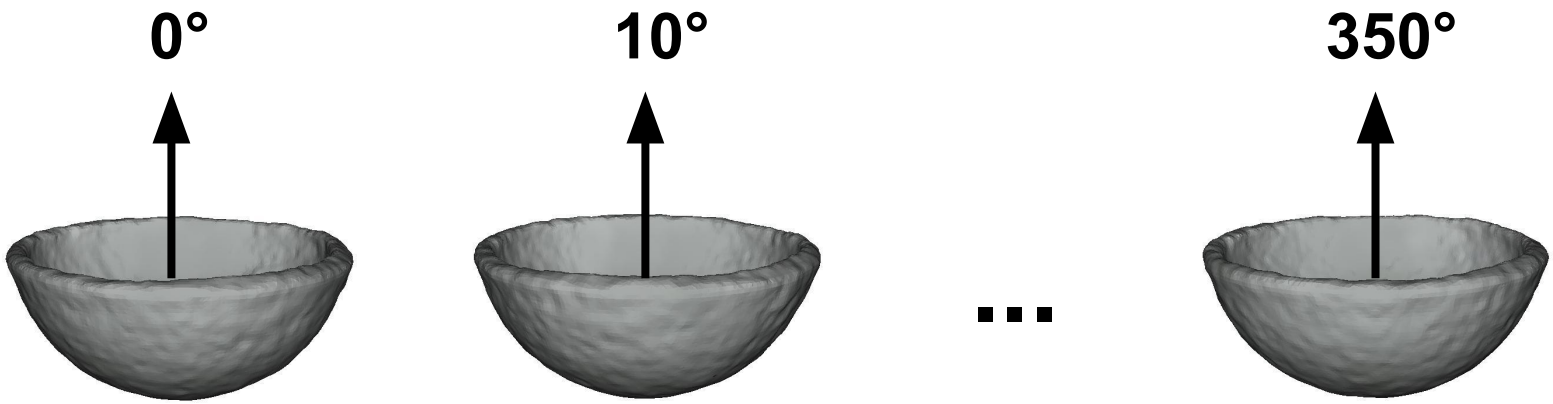
Due to **object symmetries**, multiple poses may be **indistinguishable**



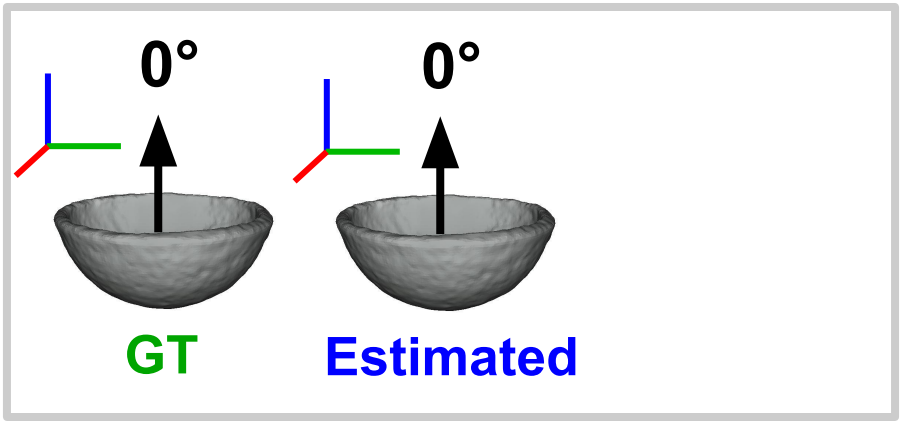
Rotational symmetry

Pose Ambiguity

Due to **object symmetries**, multiple poses may be **indistinguishable**

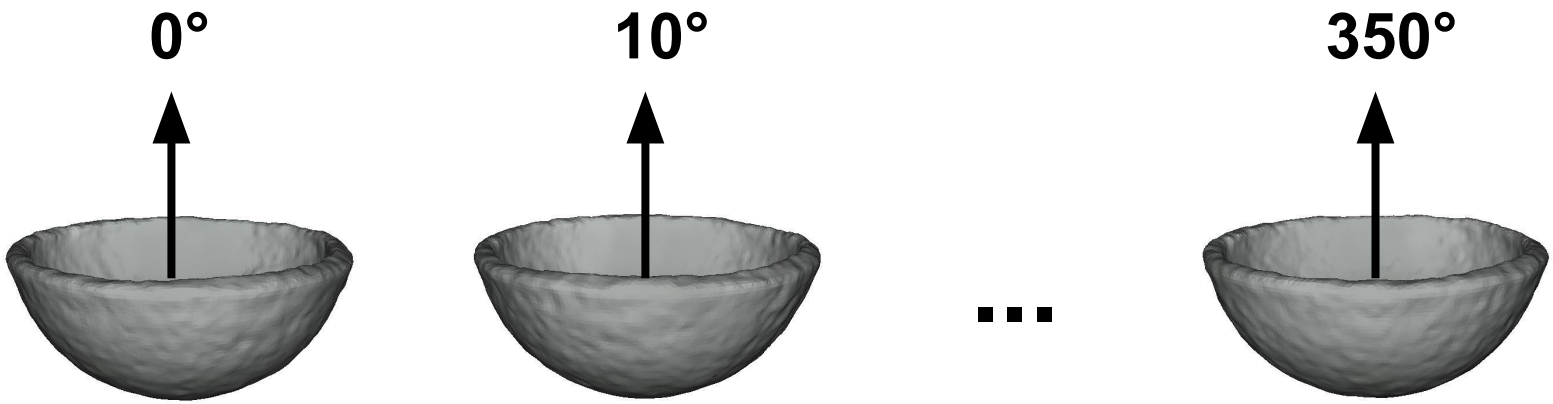


Rotational symmetry

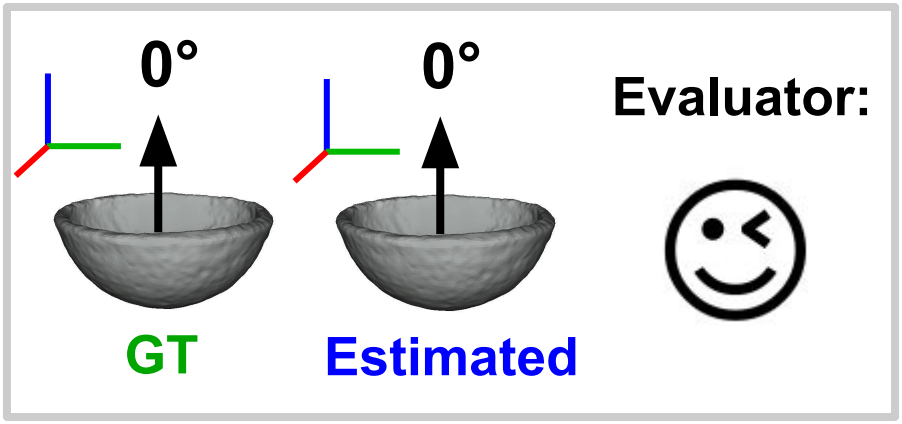


Pose Ambiguity

Due to **object symmetries**, multiple poses may be **indistinguishable**

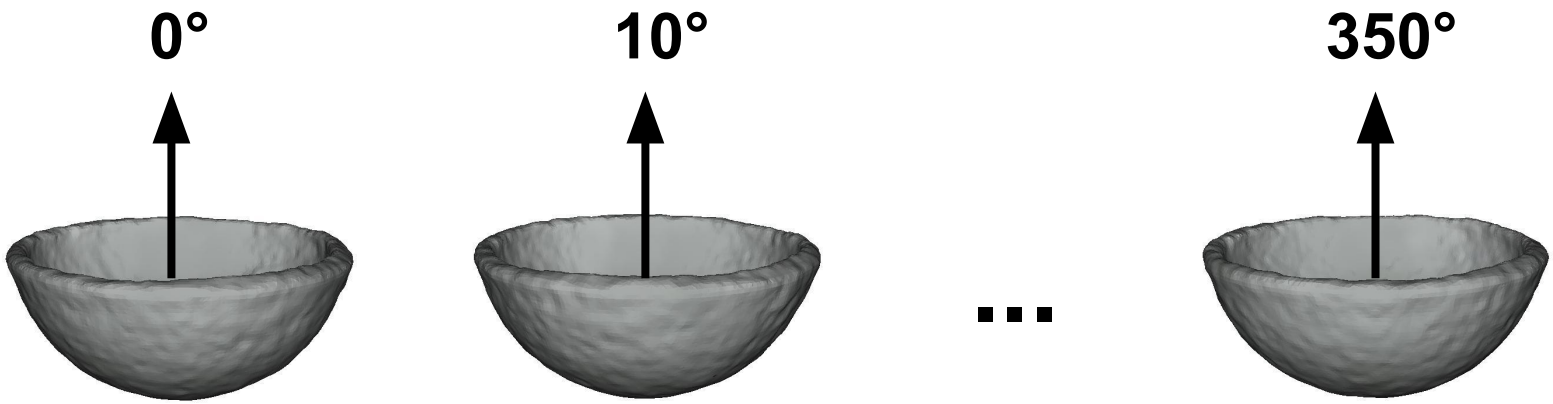


Rotational symmetry

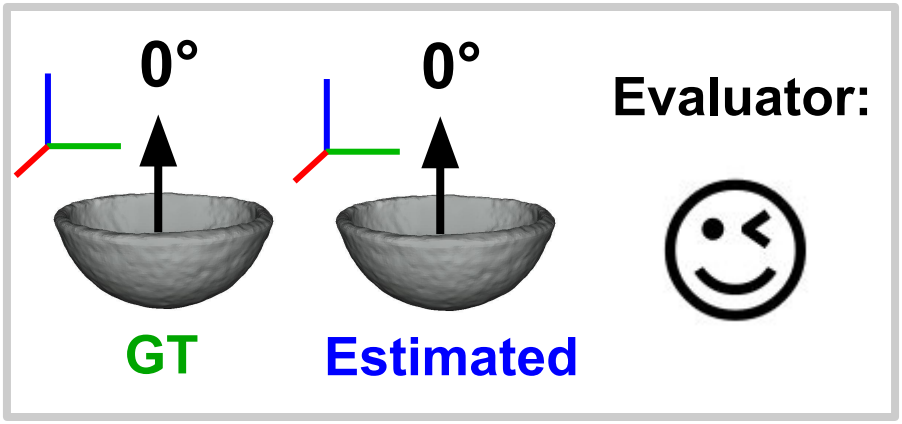


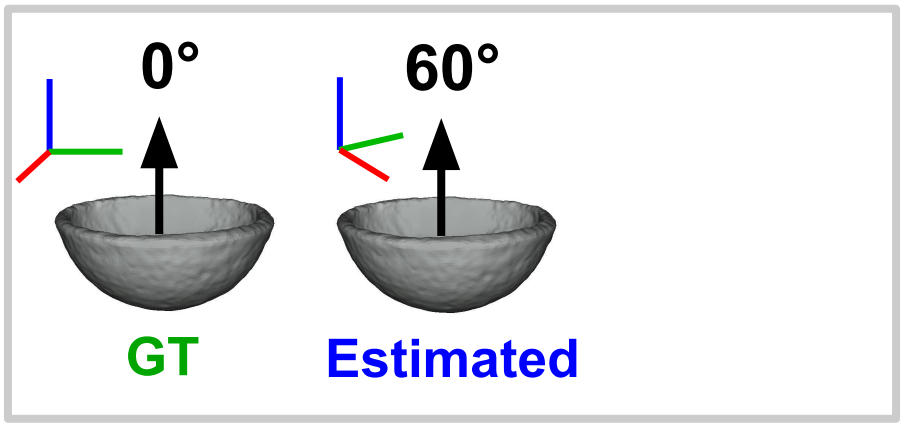
Pose Ambiguity

Due to **object symmetries**, multiple poses may be **indistinguishable**



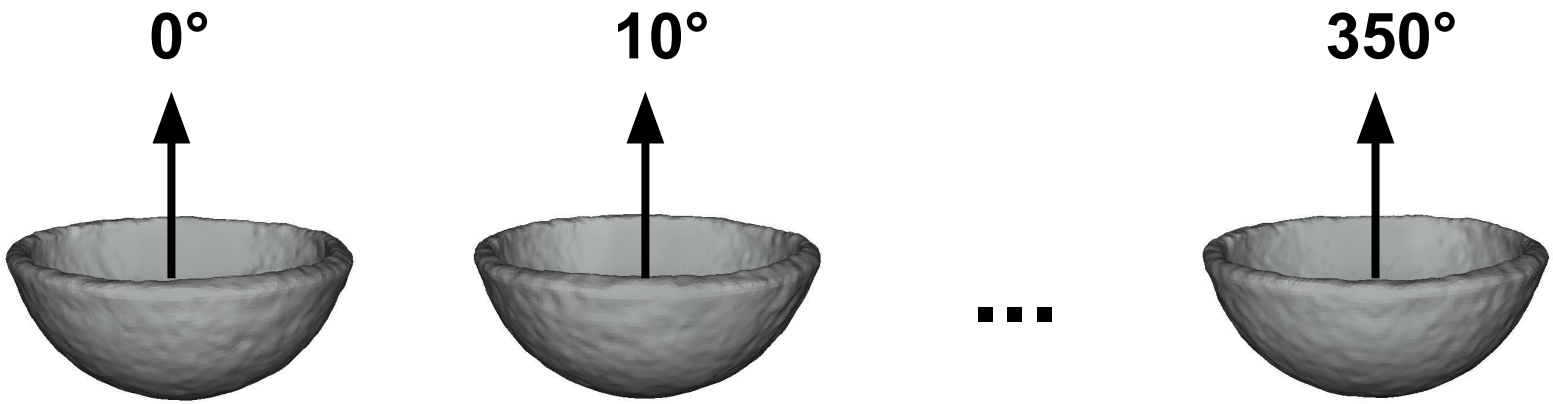
Rotational symmetry



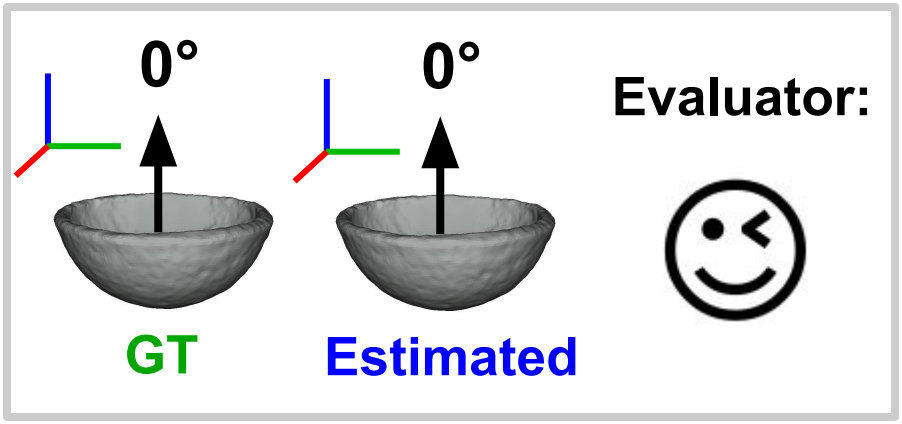


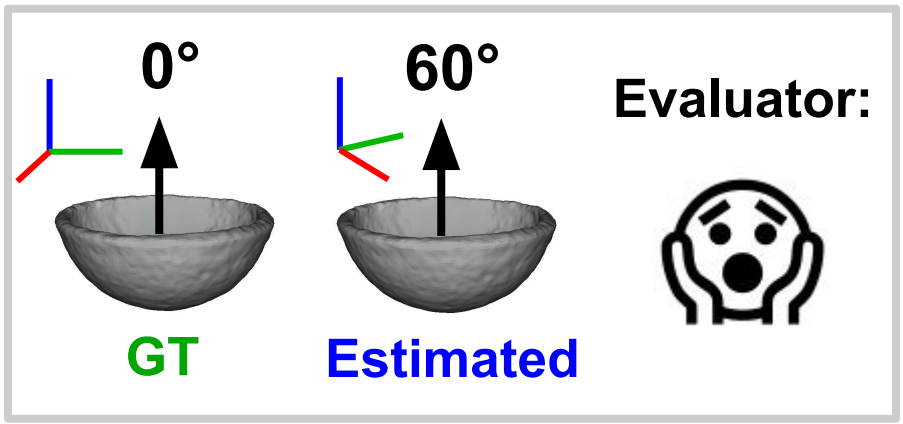
Pose Ambiguity

Due to **object symmetries**, multiple poses may be **indistinguishable**



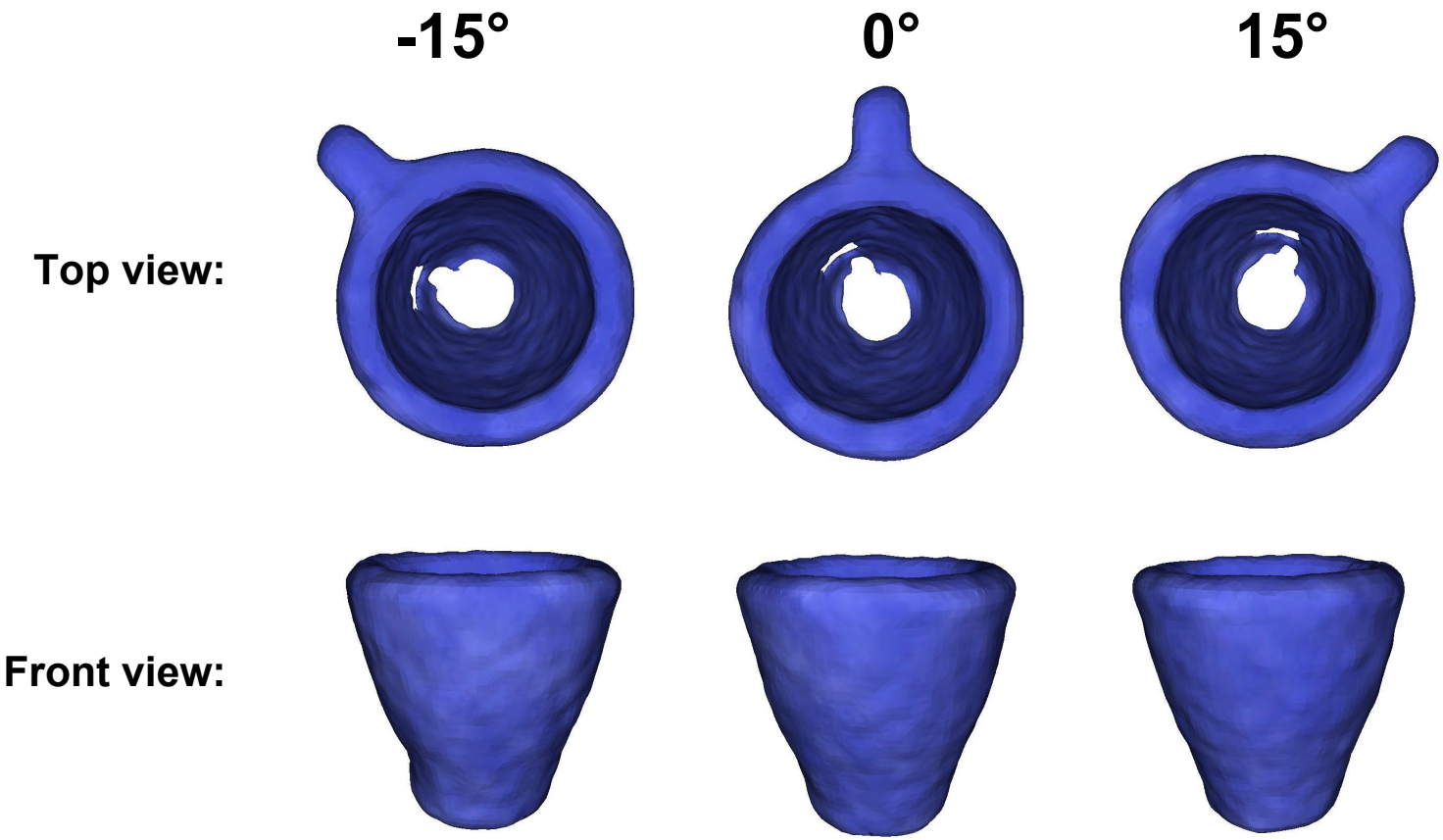
Rotational symmetry





Pose Ambiguity

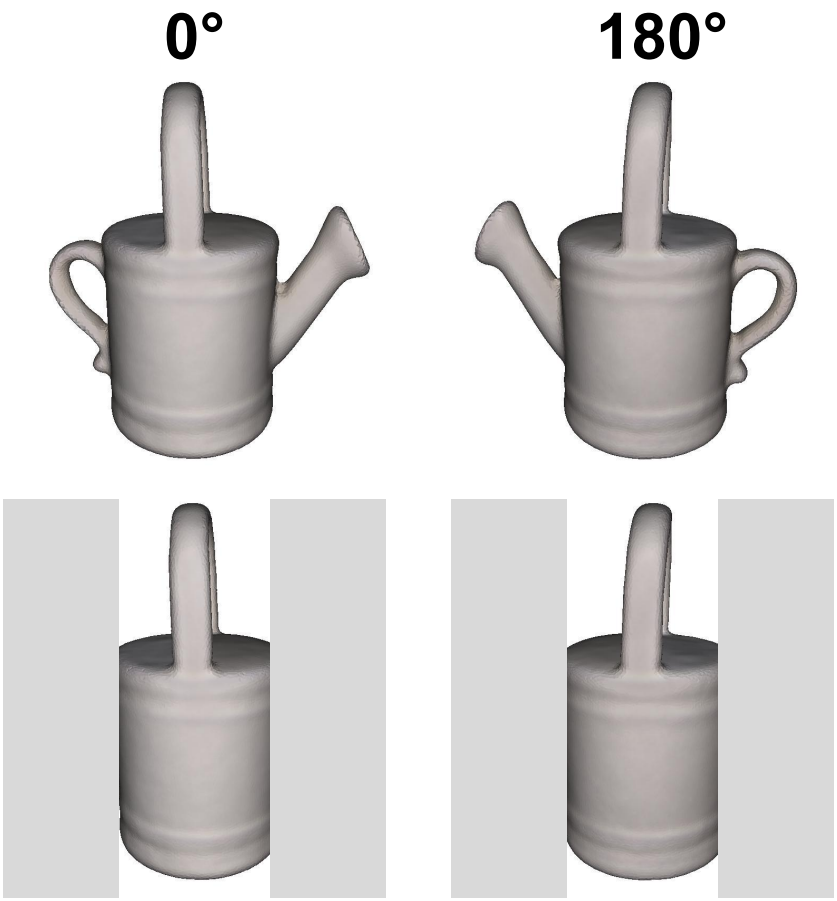
Due to **object symmetries**, multiple poses may be **indistinguishable**



Partial symmetry - self-occlusion

Pose Ambiguity

Due to **object symmetries**, multiple poses may be **indistinguishable**



Partial symmetry - occlusion

Extension of the Standard Errors

1. **Find the indistinguishable poses**
2. Final error given by e.g. **minimum error** over the indistinguishable set

The indistinguishable poses could be found by e.g.:

1. Identification of the visible part of the object surface
2. Finding repetitions of the visible part on the full object surface using:

   Mitra et al., Partial and approximate symmetry detection for 3D geometry, TOG 2006

Extension of the Standard Errors

1. **Find the indistinguishable poses**
2. Final error given by e.g. **minimum error** over the indistinguishable set

The indistinguishable poses could be found by e.g.:

1. Identification of the visible part of the object surface
2. Finding repetitions of the visible part on the full object surface using:

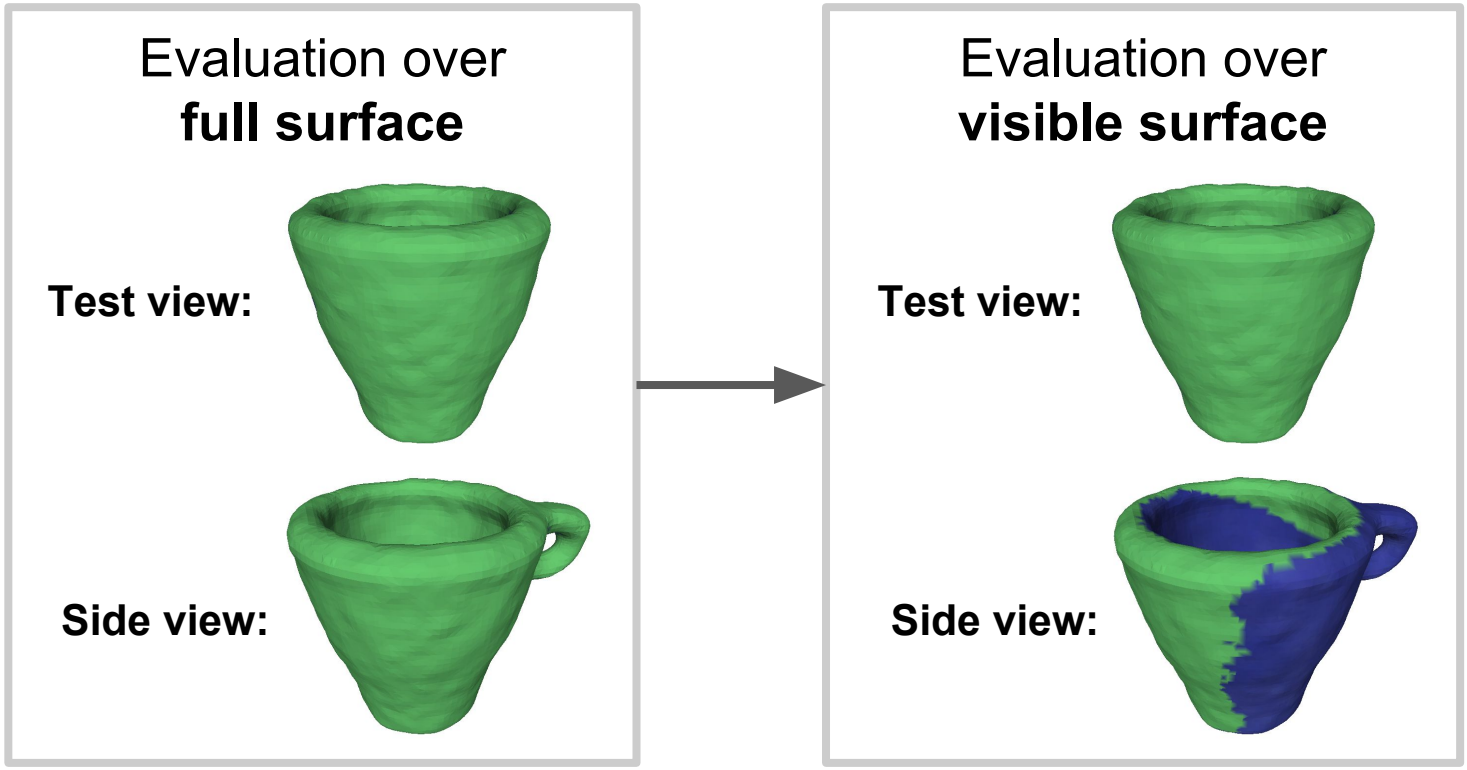
   Mitra et al., Partial and approximate symmetry detection for 3D geometry, TOG 2006

Disadvantages:

- Complicates and slows down the evaluation process
- Problems with representation of sets of poses

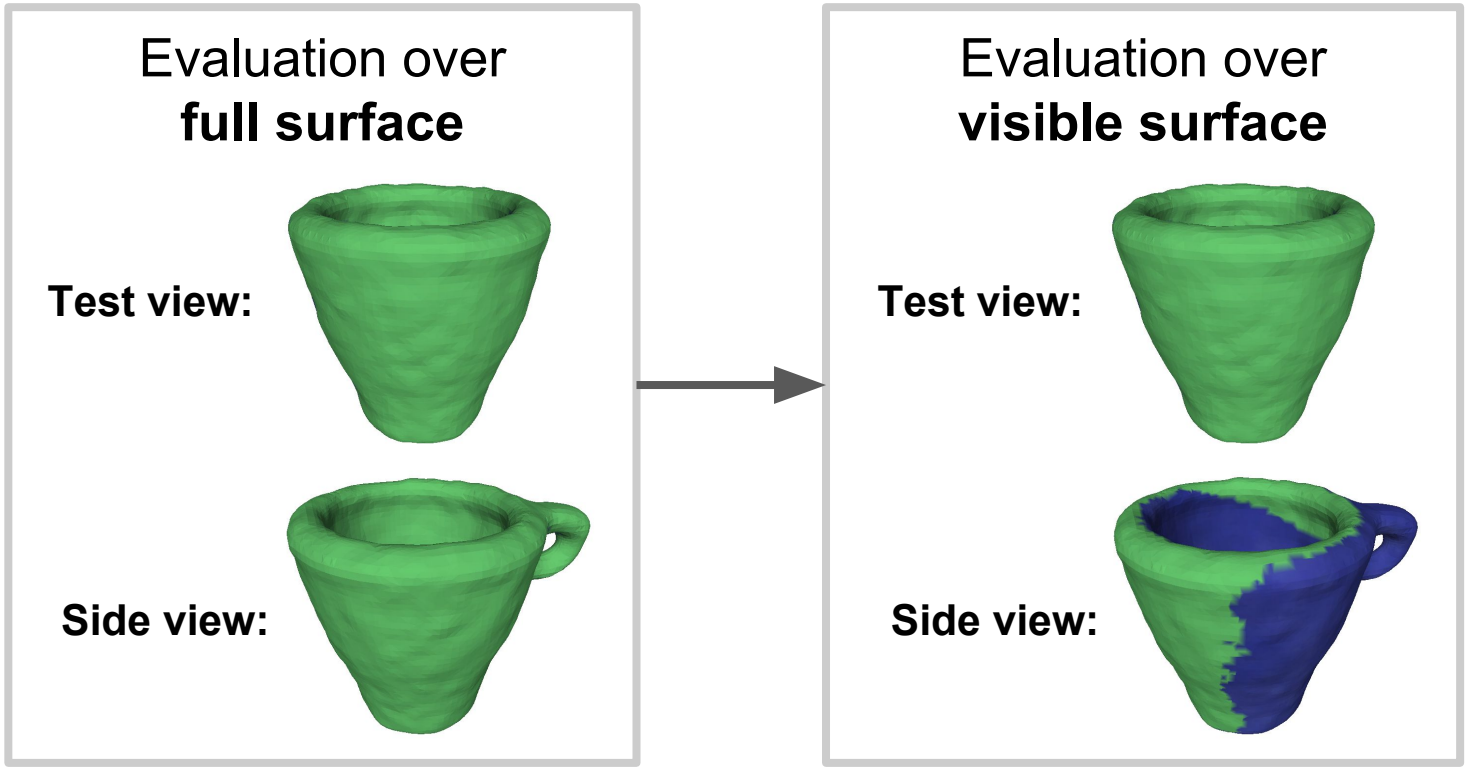
Proposal: Evaluate over Visible Surface

Measure the pose error only over the visible part of the object surface



Proposal: Evaluate over Visible Surface

Measure the pose error only over the visible part of the object surface



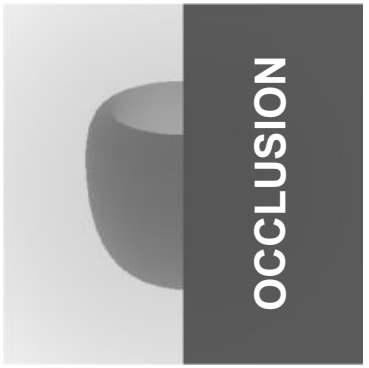
- Inherently invariant under pose ambiguity (the visible surface is the same in all indistinguishable poses)

Visibility

Object surface at a pixel is visible if it is **in front of the scene surface, or at most by a tolerance δ behind**

Visibility

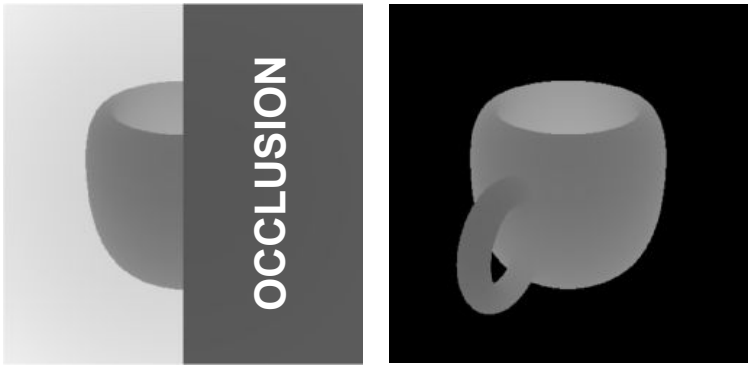
Object surface at a pixel is visible if it is **in front of the scene surface, or at most by a tolerance δ behind**



Test depth im.

Visibility

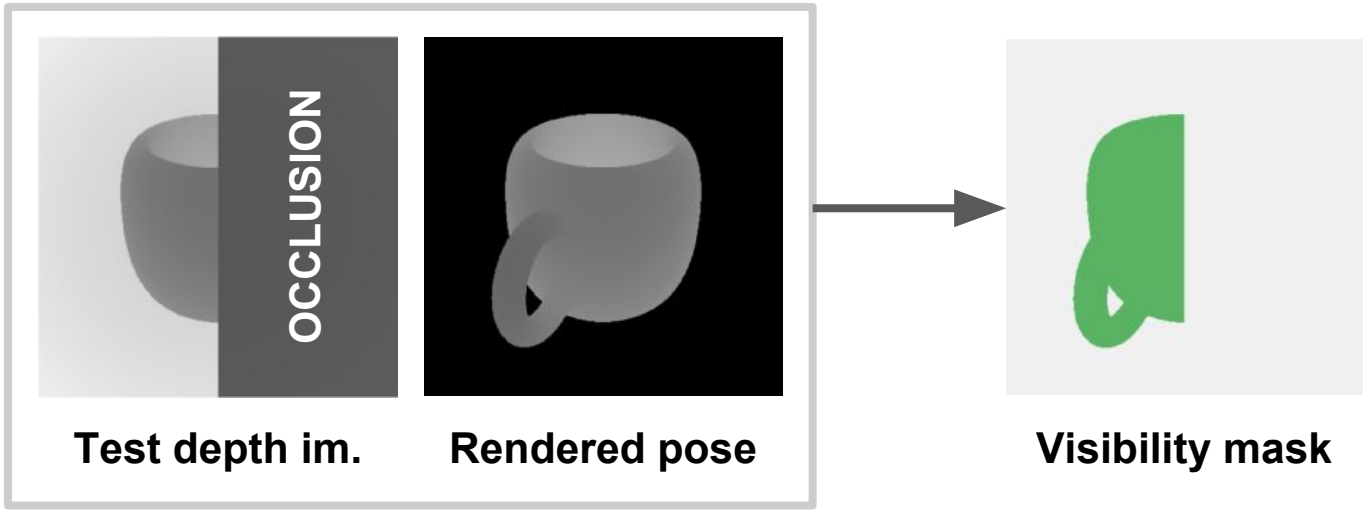
Object surface at a pixel is visible if it is **in front of the scene surface, or at most by a tolerance δ behind**



Test depth im. **Rendered pose**

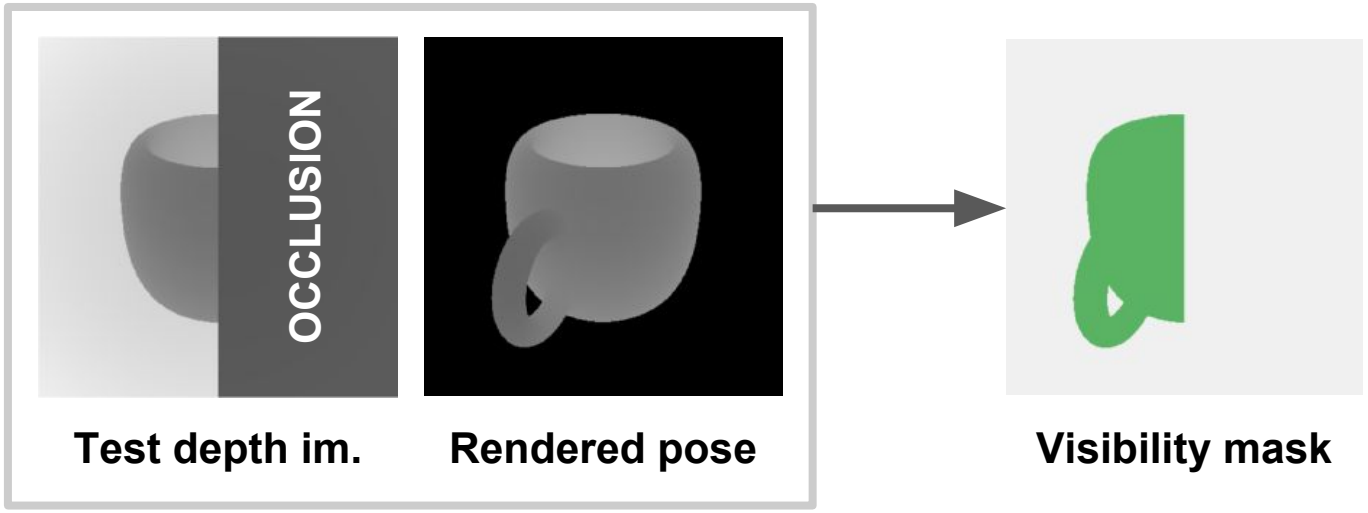
Visibility

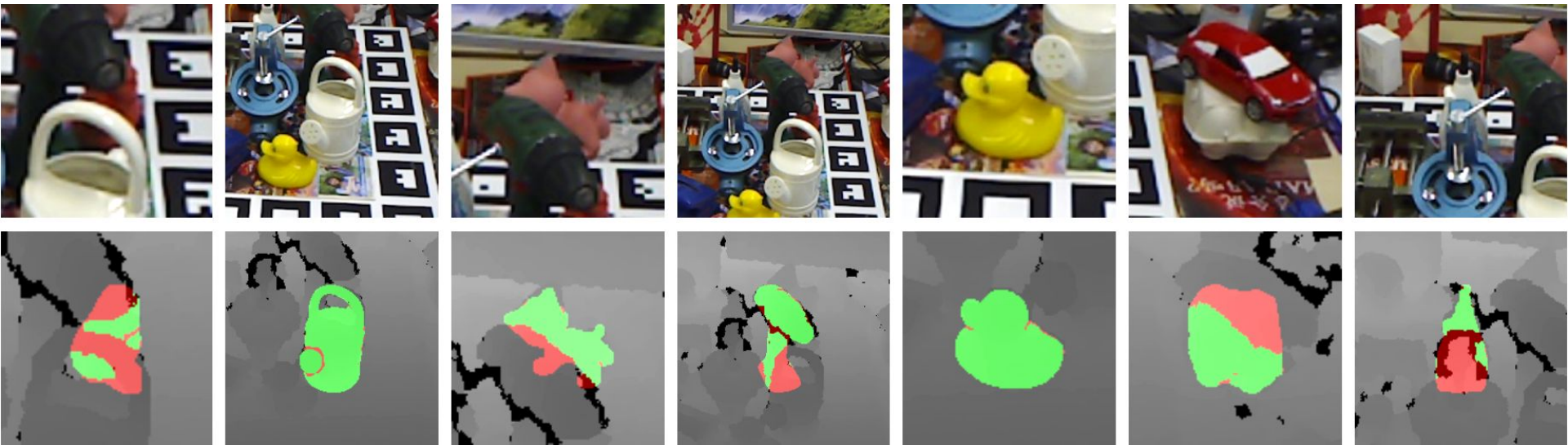
Object surface at a pixel is visible if it is **in front of the scene surface, or at most by a tolerance δ behind**



Visibility

Object surface at a pixel is visible if it is **in front of the scene surface, or at most by a tolerance δ behind**





Visible Surface Discrepancy (VSD)

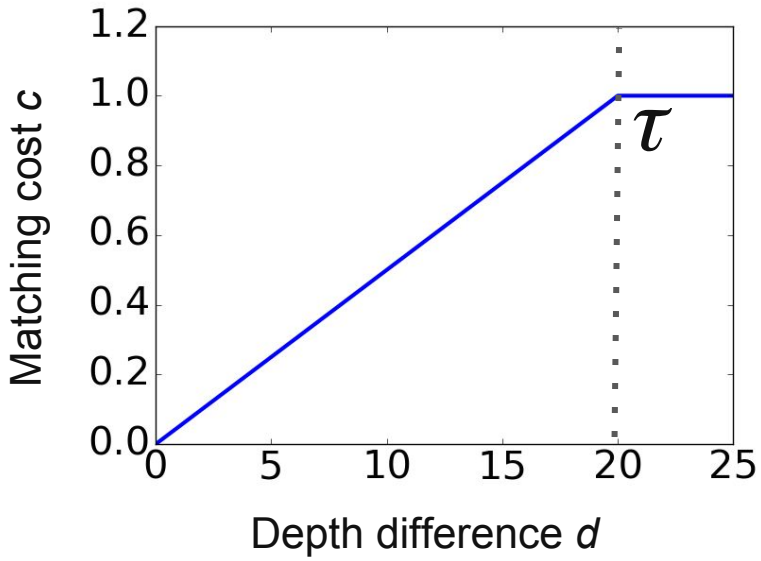
The average pixel-wise matching cost over union of the visibility masks:

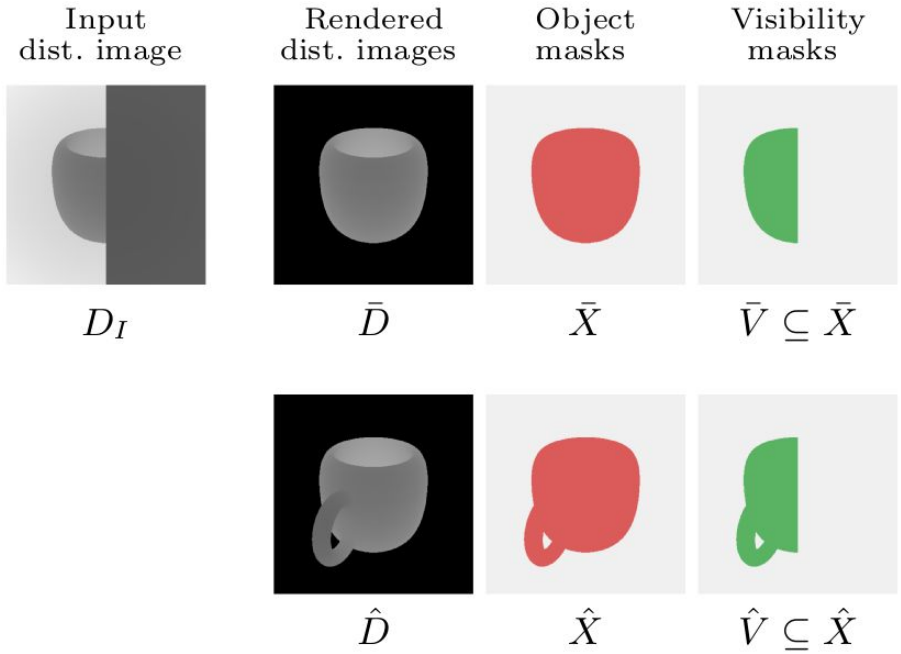
$$e_{\text{VSD}}(\hat{\mathbf{P}}, \bar{\mathbf{P}}; \mathcal{M}, I, \delta, \tau) = \underset{p \in \hat{V} \cup \bar{V}}{\text{avg}} \; c(p, \hat{D}, \bar{D}, \tau)$$

Pixel-wise matching cost:

$$c(p, \hat{D}, \bar{D}, \tau) = \begin{cases} d \,/\, \tau & \text{if } p \in \hat{V} \cap \bar{V} \wedge d < \tau \\ 1 & \text{otherwise,} \end{cases}$$

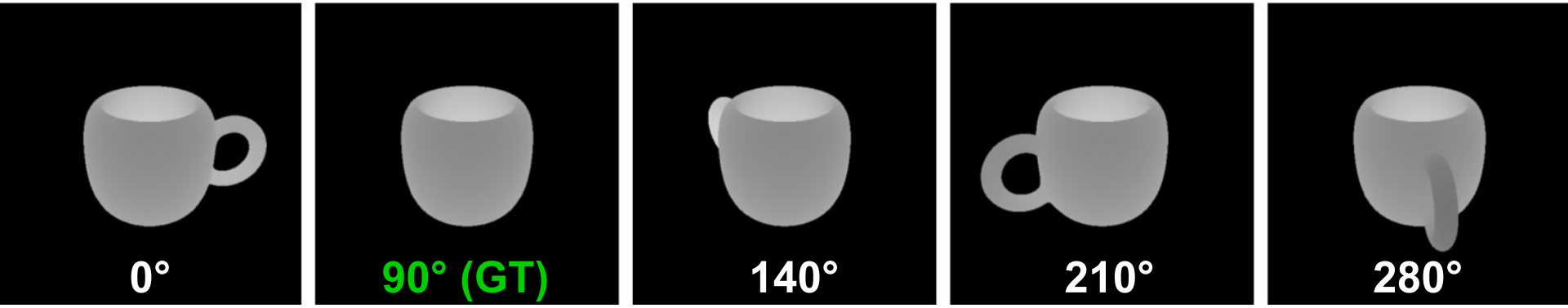
$$d = |\hat{D}(p) - \bar{D}(p)|$$

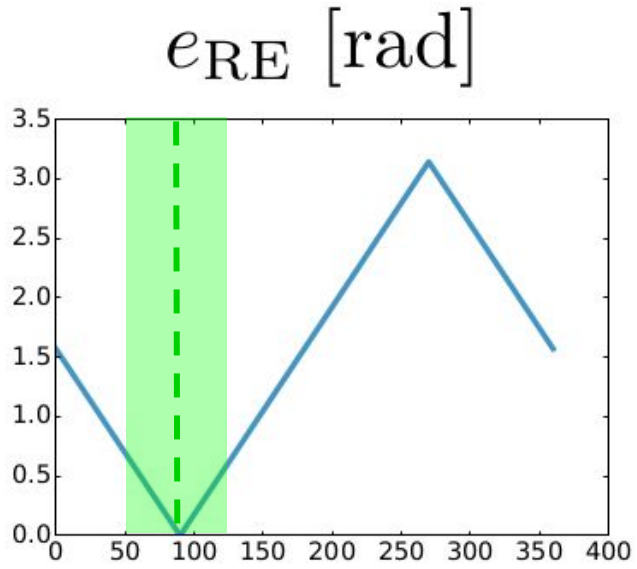


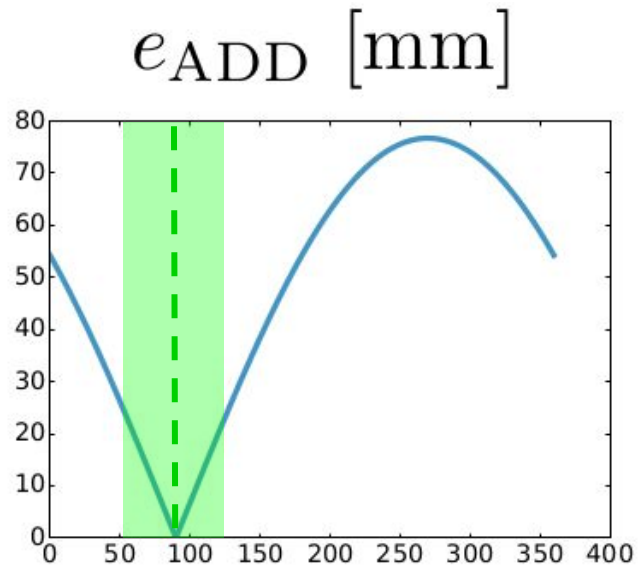


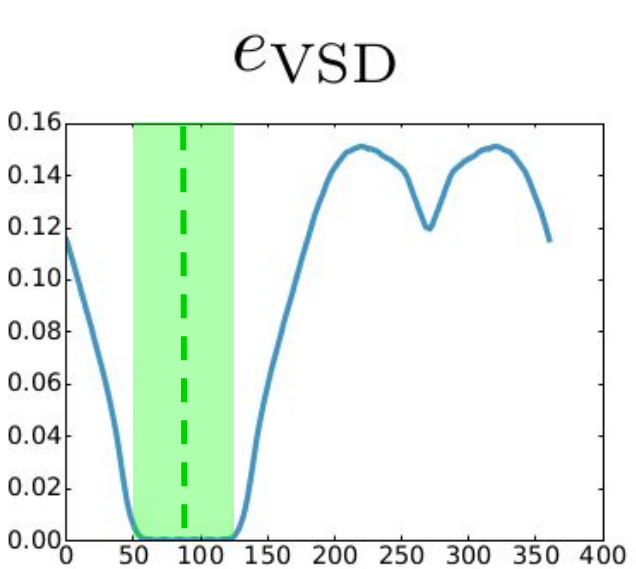
Comparison of Pose Error Functions

- Synthetic sequence (P_0, P_1, …, P_{359}) of 6D poses of a rotating cup
- Pose P_i represents a rotation by i°
- The poses were evaluated against P_{90}, which was set to be the GT







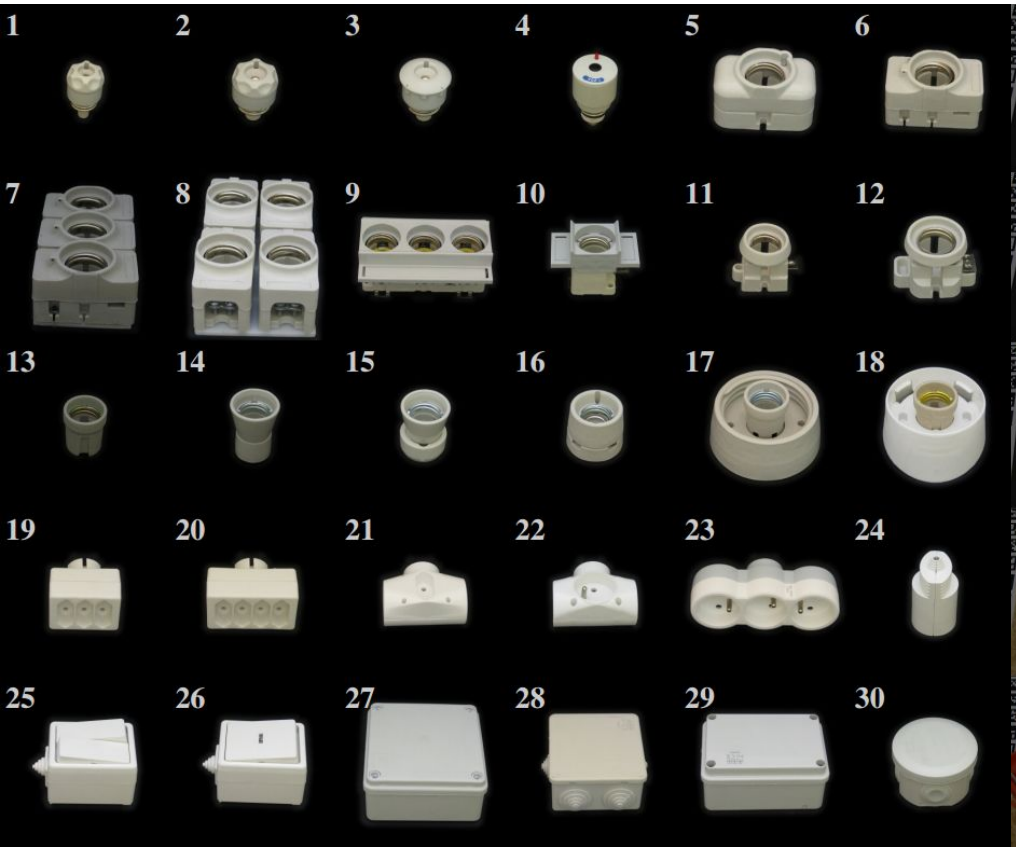


T-LESS

An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects

http://cmp.felk.cvut.cz/**t-less**

- **30 Industry-relevant objects:** No discriminative color, no texture, often similar in shape, some objects are parts of others
- **Training data provided in several forms:** 1) RGB-D templates annotated with 6D object poses, 2) 3D CAD models, and 3) automatically reconst. 3D models
- **Test data includes RGB-D images of 20 scenes with accurate ground truth poses**
- **All images captured with three synchronized sensors:** Primesense CARMINE 1.09, Microsoft Kinect v2, and Canon IXUS 950 IS



Thank you!

Extension of the Standard Errors?

Minimum over: $Q = [\hat{\mathbf{P}}]_{\mathcal{M},I,\varepsilon} \times [\bar{\mathbf{P}}]_{\mathcal{M},I,\varepsilon}$

Indistinguishable poses of the estimated pose

Indistinguishable poses of the ground truth pose

$$e_{\mathrm{ACPD}}(\hat{\mathbf{P}}, \bar{\mathbf{P}}; \mathcal{M}, I, \varepsilon) = \min_{(\hat{\mathbf{P}}', \bar{\mathbf{P}}') \in Q} \operatorname*{avg}_{\mathbf{x} \in \mathcal{M}} \left\| \bar{\mathbf{P}}'\mathbf{x} - \hat{\mathbf{P}}'\mathbf{x} \right\|_2$$

The sets of indistinguishable poses could be found by e.g.:

1. Identification of the visible part of the object surface
2. Finding repetitions of the visible part on the whole object surface using:
   Mitra et al., Partial and approximate symmetry detection for 3D geometry, TOG 2006

Who Finds the Indistinguishable Poses?

Option #1: **The estimator provides the set of indistinguishable poses, instead of a point estimate**

- Relevant to robotic manipulation tasks
- Not provided by the current SOTA methods

Option #2: **The evaluation system finds the indistinguishable poses**

- No extra requirements on the methods
- Slows down and complicates the evaluation process

But do we really need to find the poses?