BOP: Benchmark for 6D Object Pose Estimation

Hodan, Michel, Brachmann, Kehl, Buch, Kraft, Drost, Vidal, Ihrke, Zabulis, Sahin, Manhardt, Tombari, Kim, Matas, Rother

4th International Workshop on Recovering 6D Object Pose
ECCV 2018, September 9th, Munich

State of the art in 6D object pose estimation?

Unclear, because:

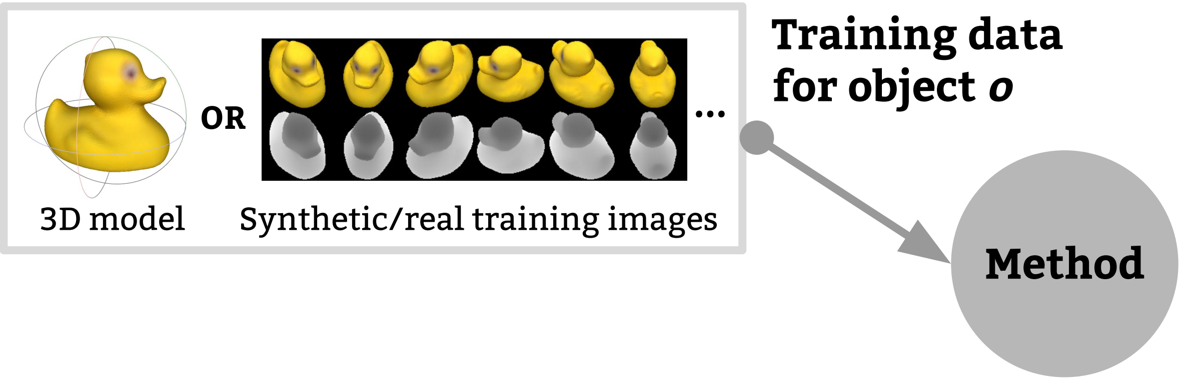
1. No standard evaluation methodology
2. New methods usually compared with only a few competitors on a small number of datasets
3. Scores on the most commonly used Linemod dataset are saturated

The Task

6D localization of a single instance of a single object (SiSo)

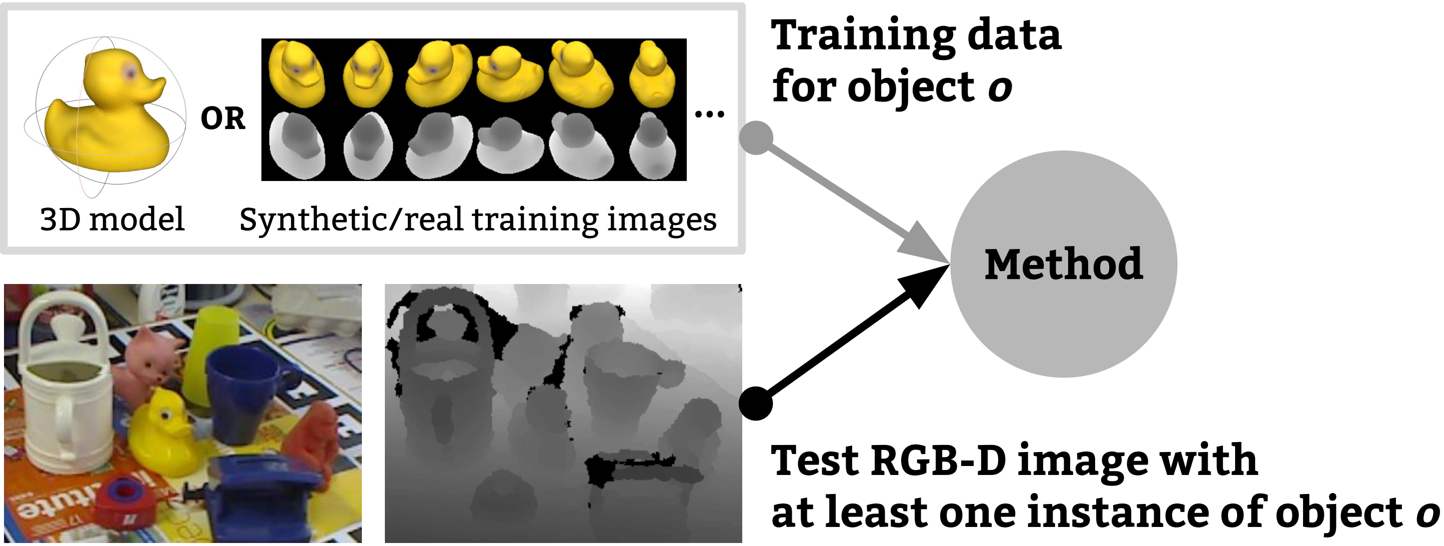
The Task

6D localization of a single instance of a single object (SiSo)



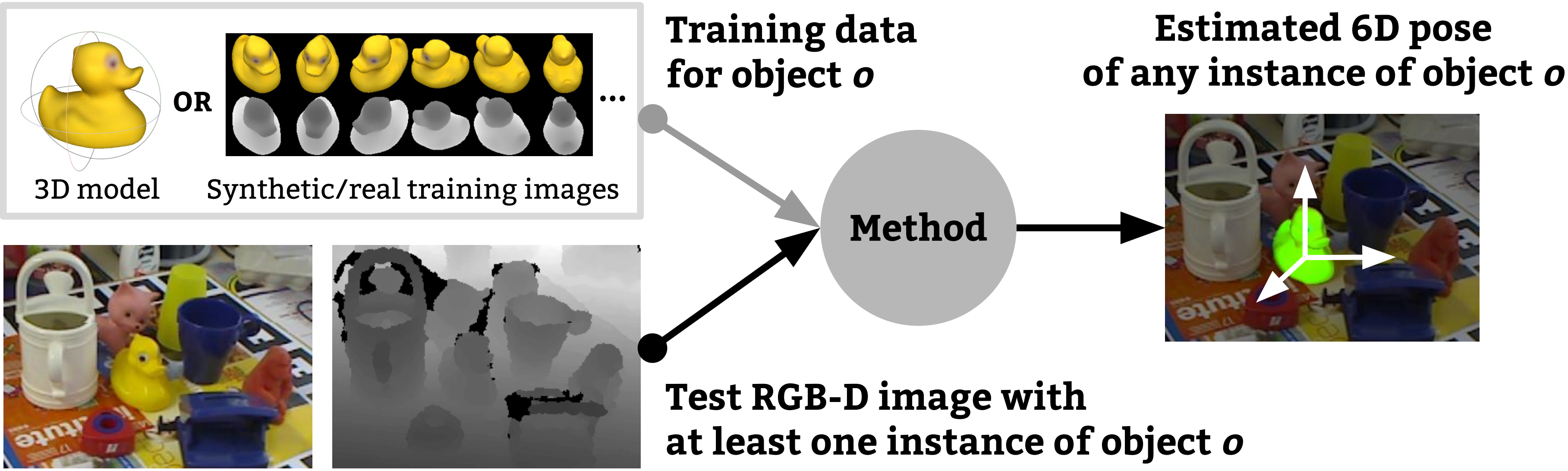
The Task

6D localization of a single instance of a single object (SiSo)



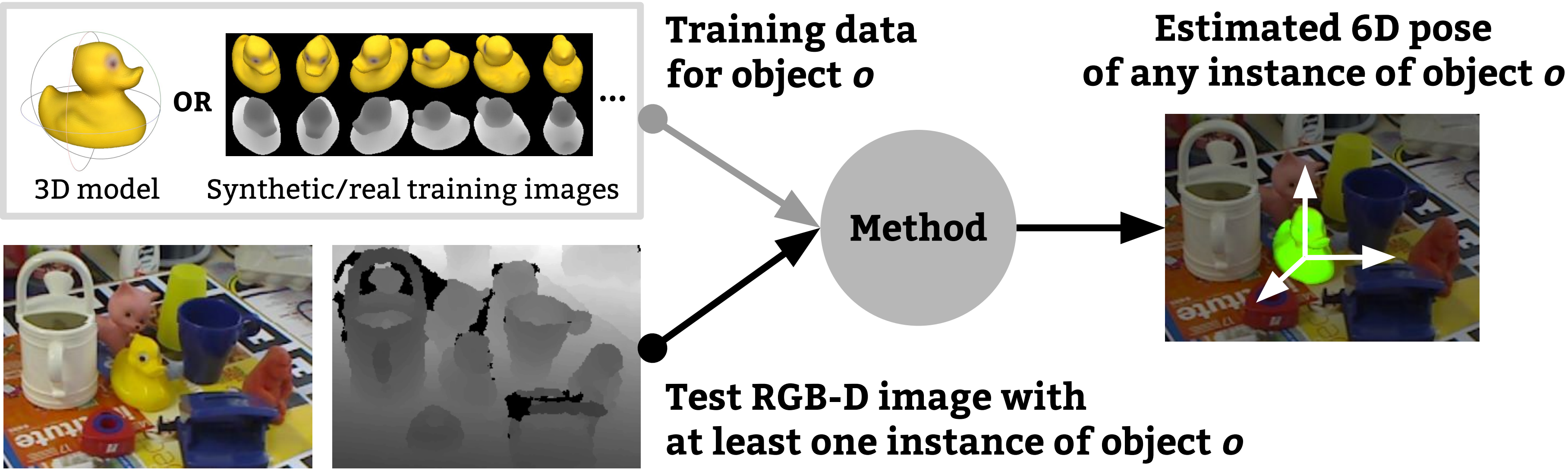
The Task

6D localization of a single instance of a single object (SiSo)



The Task

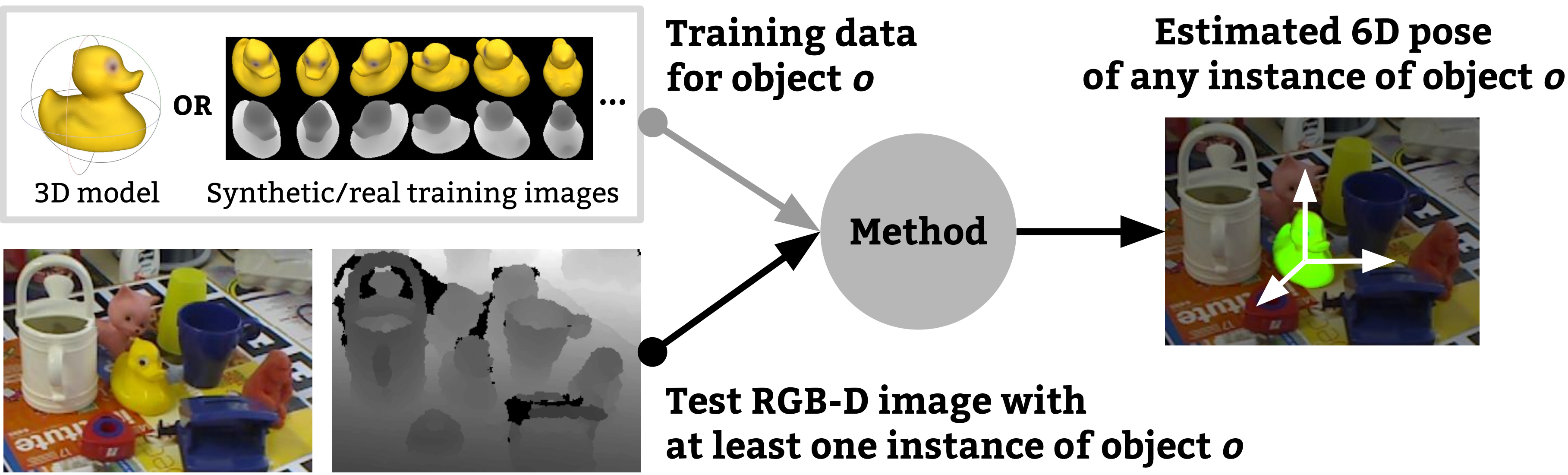
6D localization of a single instance of a single object (SiSo)



- SiSo is **the common denominator** of all 6D localization variants:

The Task

6D localization of a single instance of a single object (SiSo)



- SiSo is **the common denominator** of all 6D localization variants:

- SiSo allows evaluation of all recent methods **out of the box**

Eight datasets in a unified format

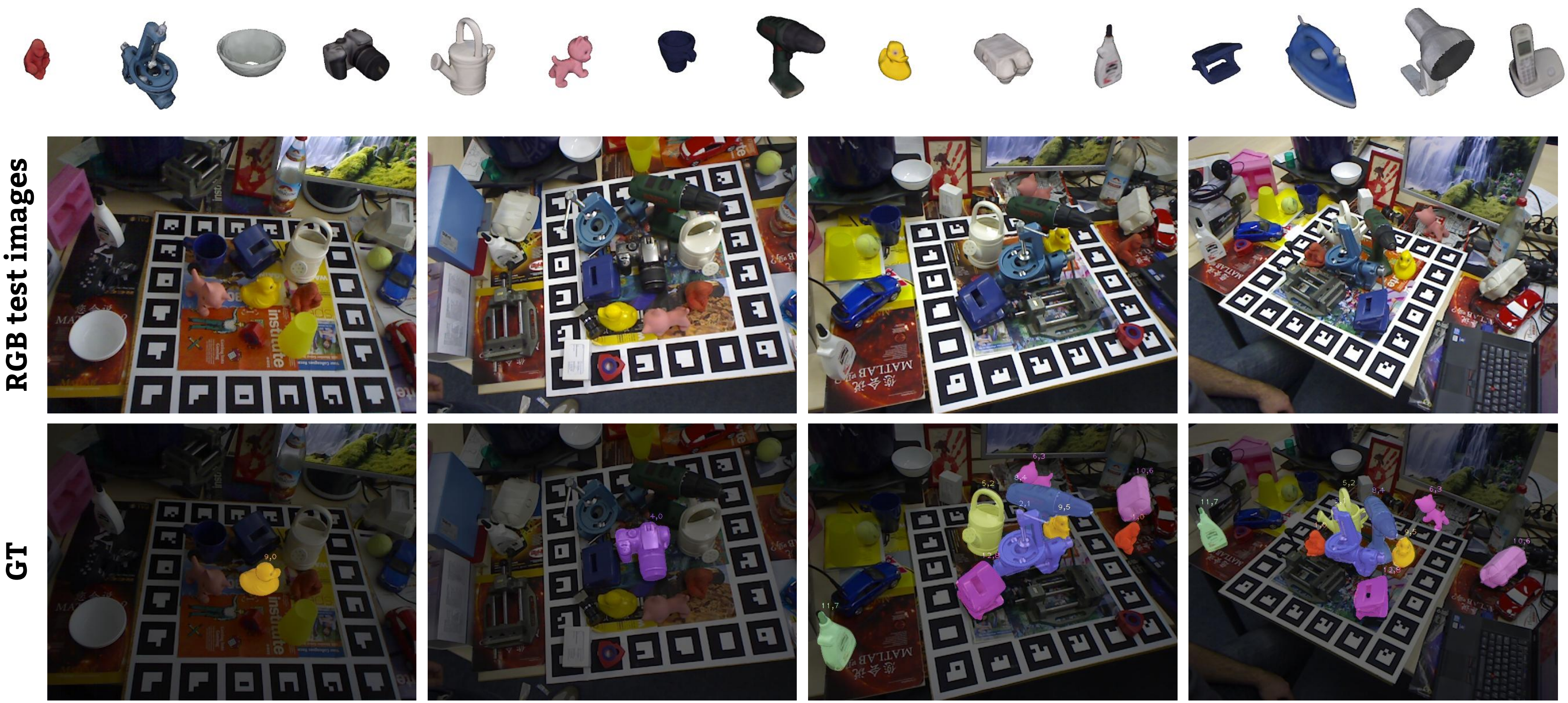
- **Texture-mapped 3D models of 89 objects**
- **277K training RGB-D images** of isolated objects (mostly synthetic images)
- **62K test RGB-D images** of scenes with graded complexity
- **High-quality ground-truth 6D object poses** for all images

Linemod (LM), **Linemod-Occluded** (LM-O)

15 objects, 20K rendered training and 18K test RGB-D images

Texture-less objects with discriminative size, shape or color

Standard benchmark - used for evaluation of most recent methods



Hinterstoisser et al. (ACCV'12), Brachmann et al. (ECCV'14)

T-LESS

30 objects, 38K real and 77K rendered train. images, 10K test images

No significant texture, no discriminative reflectance properties, symmetries and mutual similarities in shape or size

RGB test images

GT

Hodaň et al. (WACV'17)

Rutgers APC (RU-APC) - reduced version

14 objects, 36K rendered training and 6K real test images

Textured objects from the Amazon Picking Challenge

Rennie et al. (RAL'16)

Tejani et al. (IC-MI), Doumanoglou et al. (IC-BIN)

6 objects, 8K rendered training and 2K test RGB-images

Multiple instances of textured and texture-less objects with clutter

RGB test images

GT

Tejani et al. (ECCV'14), Doumanoglou et al. (CVPR'16)

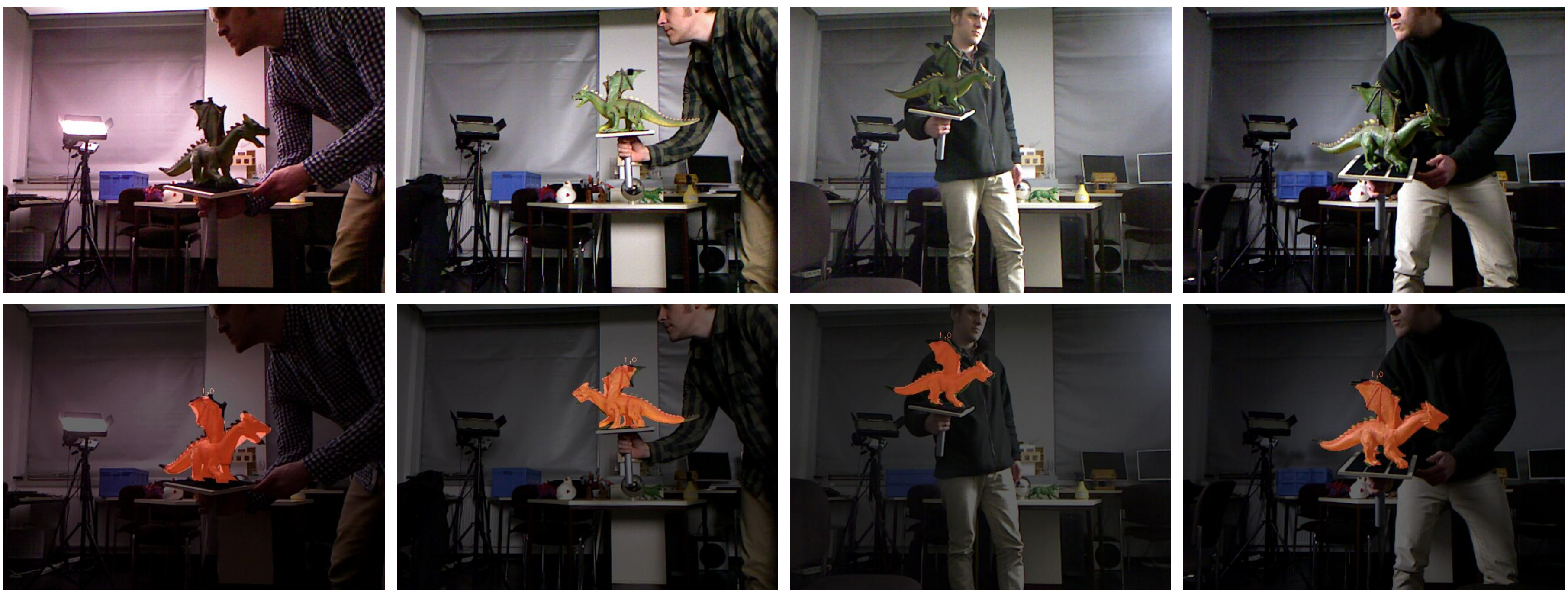
TU Dresden Light (TUD-L) - new

3 objects, 38K real and 5K rendered training images, 24K test images

8 lighting conditions (strong ambient light, strong point light etc.)

RGB test images

GT



Michel et al. Technische Universität Dresden, 2017

Toyota Light (TYO-L) - new

21 objects, 52K rendered training images, 2K test images

5 lighting conditions, 4 backgrounds (textured / texture-less)

RGB test images

GT

Manhardt et al. Technische Universität München, 2017

Visible Surface Discrepancy (VSD)

Test image

RGB

Depth

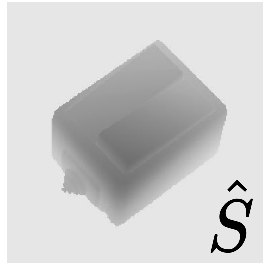
Visible Surface Discrepancy (VSD)

Test image

RGB

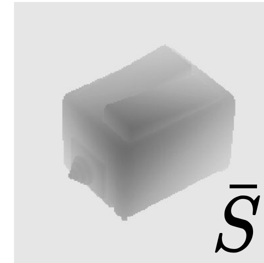
Depth

Estimated pose



Depth

GT pose



Depth

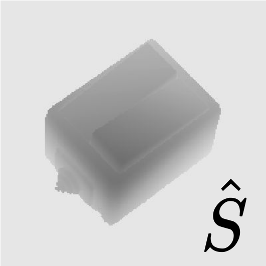
Visible Surface Discrepancy (VSD)

Test image

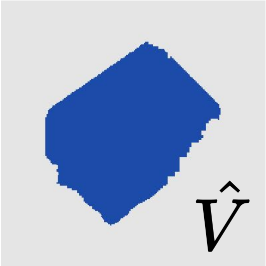
RGB

Depth

Estimated pose

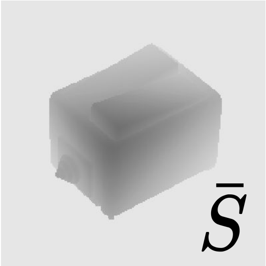


Depth

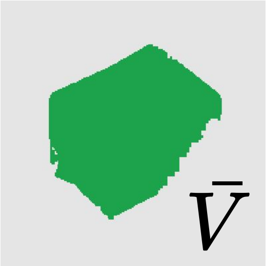


Visibility

GT pose

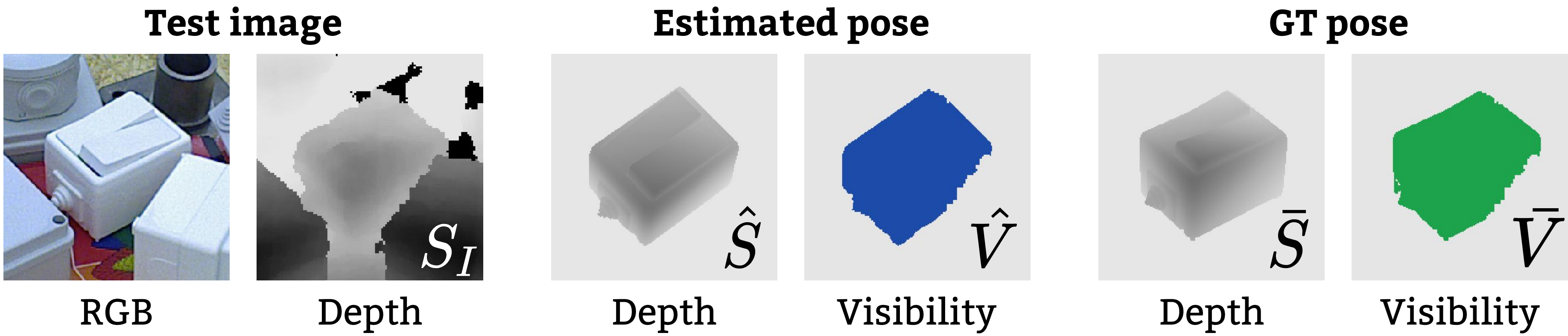


Depth



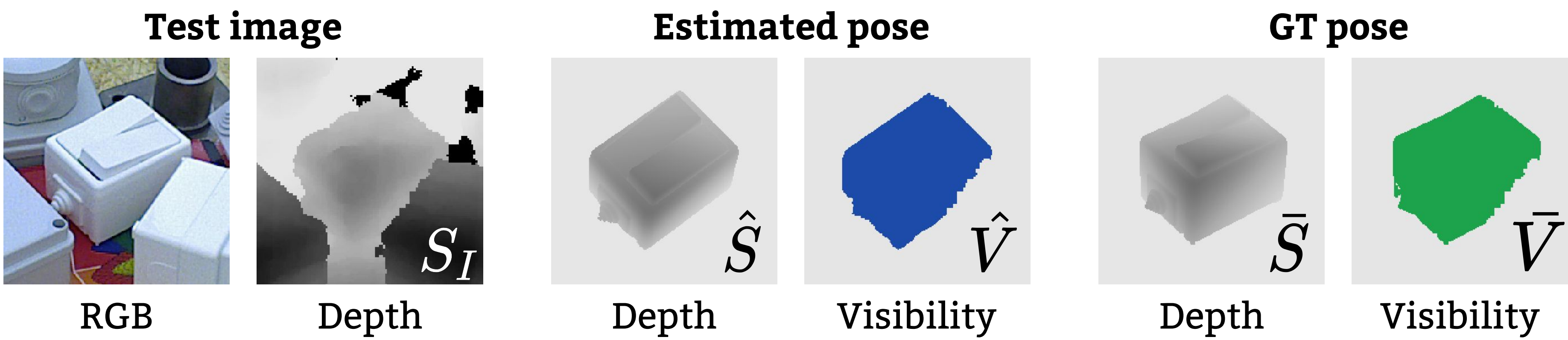
Visibility

Visible Surface Discrepancy (VSD)



- **Visibility masks** are obtained by comparing $\hat{S}$ and $\bar{S}$ with S_I

Visible Surface Discrepancy (VSD)

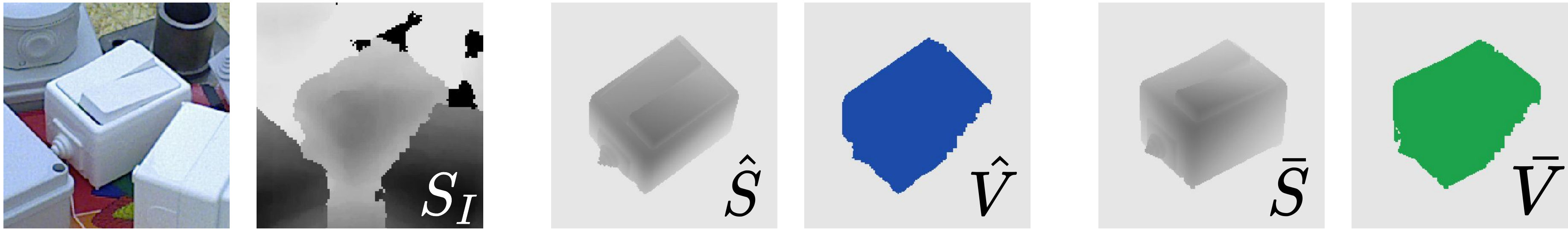


- **Visibility masks** are obtained by comparing $\hat{S}$ and $\bar{S}$ with S_I

$$e_{\mathrm{VSD}}(\hat{S}, \bar{S}, S_I, \hat{V}, \bar{V}, \tau) = \underset{p \in \hat{V} \cup \bar{V}}{\operatorname{avg}} \begin{cases} 0 & \text{if } p \in \hat{V} \cap \bar{V} \ \wedge \ |\hat{S}(p) - \bar{S}(p)| < \tau \\ 1 & \text{otherwise.} \end{cases}$$

Visible Surface Discrepancy (VSD)

Test image | **Estimated pose** | **GT pose**



RGB | Depth | Depth | Visibility | Depth | Visibility

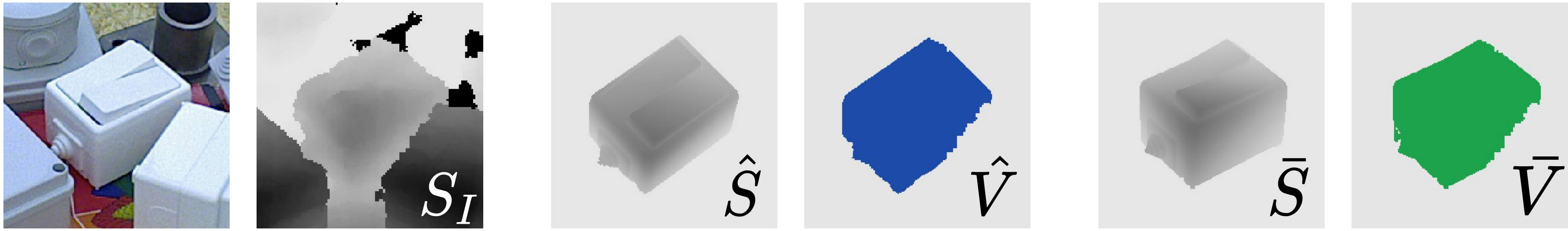
- **Visibility masks** are obtained by comparing $\hat{S}$ and $\bar{S}$ with S_I

$$\boxed{e_{\mathrm{VSD}}(\hat{S}, \bar{S}, S_I, \hat{V}, \bar{V}, \tau) = \underset{p \in \hat{V} \cup \bar{V}}{\mathrm{avg}} \begin{cases} 0 & \text{if } p \in \hat{V} \cap \bar{V} \ \wedge \ |\hat{S}(p) - \bar{S}(p)| < \tau \\ 1 & \text{otherwise.} \end{cases}}$$

- Estimated pose is **considered correct if** $e_{\mathrm{VSD}} < \theta$

Visible Surface Discrepancy (VSD)

Test image | **Estimated pose** | **GT pose**



RGB | Depth | Depth | Visibility | Depth | Visibility

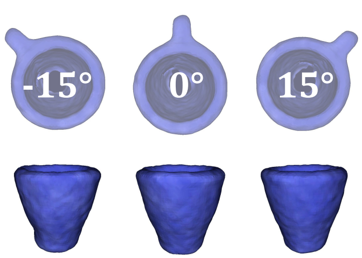
- **Visibility masks** are obtained by comparing $\hat{S}$ and $\bar{S}$ with S_I

$$\boxed{e_{\mathrm{VSD}}(\hat{S}, \bar{S}, S_I, \hat{V}, \bar{V}, \tau) = \underset{p \in \hat{V} \cup \bar{V}}{\mathrm{avg}} \begin{cases} 0 & \text{if } p \in \hat{V} \cap \bar{V} \wedge |\hat{S}(p) - \bar{S}(p)| < \tau \\ 1 & \text{otherwise.} \end{cases}}$$

- Estimated pose is **considered correct if** $e_{\mathrm{VSD}} < \theta$
- Pose error is calculated only over the visible part of the surface
  $\Rightarrow$ **Indistinguishable poses are treated as equivalent**

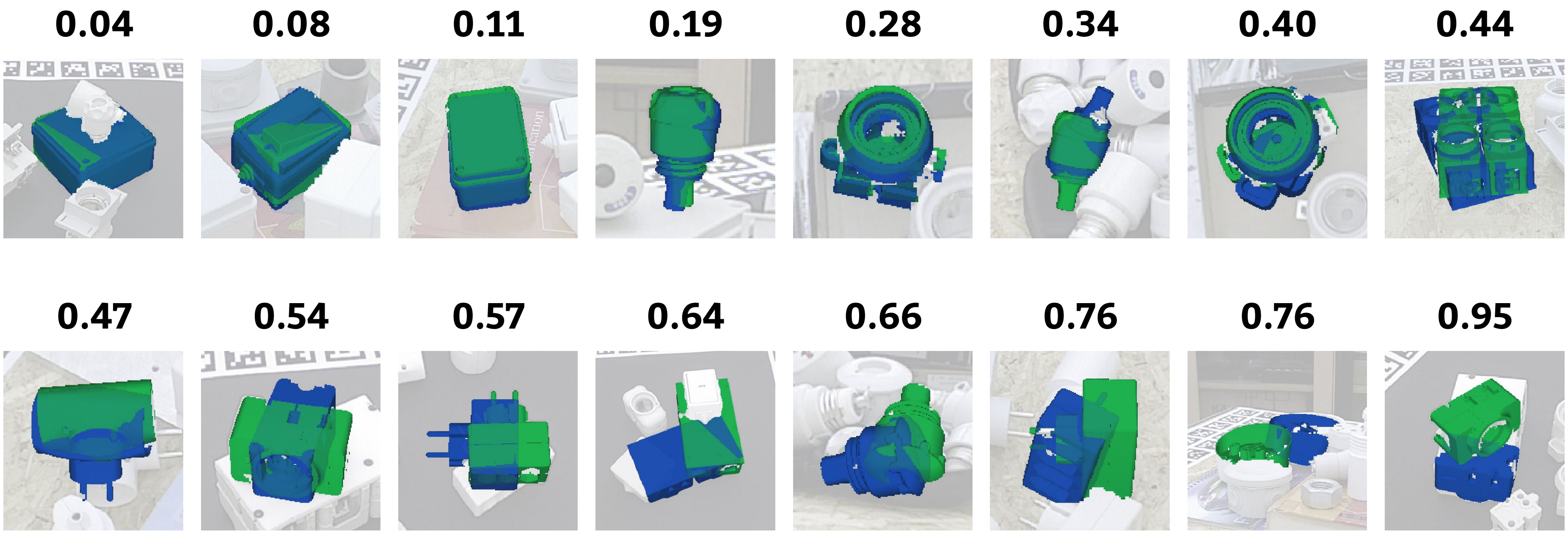
Top view:

Front view:



← Indistinguishable poses

Visible Surface Discrepancy (VSD) – examples



- The estimated pose is in blue, the ground truth in green
- Default parameter settings:
  - misalignment tolerance τ = 20 mm
  - correctness threshold θ = 0.3

Evaluated methods

Methods based on point pair features

- **Drost et al.**, Model globally, match locally: Efficient and robust 3D object recognition, CVPR 2010
- **Vidal et al.**, 6D pose estimation using an improved method based on point pair features, ICCAR 2018

Template matching method

- **Hodan et al.**, Detection and fine 3D pose estimation of texture-less objects in RGB-D images, IROS 2015

Learning-based methods

- **Brachmann et al.**, Learning 6D object pose estimation using 3D object coordinates, ECCV 2014
- **Brachmann et al.**, Uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image, CVPR 2016
- **Tejani et al.**, Latent-class hough forests for 3D object detection and pose estimation, ECCV 2014
- **Kehl et al.**, Deep learning of local RGB-D patches for 3D object detection and 6D pose estimation, ECCV 2016

Methods based on 3D local features

- **Buch et al.**, Local shape feature fusion for improved matching, pose estimation and 3D object recognition, SpringerPlus 2016
- **Buch et al.**, Rotational subgroup voting and pose clustering for robust 3D object recognition, ICCV 2017

Experimental setup

- The methods were **evaluated by their authors**
- **Parameters of each method were fixed** for all objects and datasets
- **Test target** = a pair (I, o), where image I shows at least one instance of object o
- The performance was measured by **recall**, i.e. the fraction of test targets for which a correct object pose was estimated

Evaluation results (1/2)

Methods based on point pair features, Template matching methods, Learning-based methods, Methods based on 3D local features

#	Method	LM	LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L	Average	Time (s)
1.	Vidal-18	87.83	59.31	95.33	96.50	66.51	36.52	80.17	74.60	4.7
2.	Drost-10-edge	79.13	54.95	94.00	92.00	67.50	27.17	87.33	71.73	21.5
3.	Drost-10	82.00	55.36	94.33	87.00	56.81	22.25	78.67	68.06	2.3
4.	Hodan-15	87.10	51.42	95.33	90.50	63.18	37.61	45.50	67.23	13.5
5.	Brachmann-16	75.33	52.04	73.33	56.50	17.84	24.35	88.67	55.44	4.4
6.	Hodan-15-nopso	69.83	34.39	84.67	76.00	62.70	32.39	27.83	55.40	12.3
7.	Buch-17-ppfh	56.60	36.96	95.00	75.00	25.10	20.80	68.67	54.02	14.2
8.	Kehl-16	58.20	33.91	65.00	44.00	24.60	25.58	7.50	36.97	1.8
9.	Buch-17-si	33.33	20.35	67.33	59.00	13.34	23.12	41.17	36.81	15.9
10.	Brachmann-14	67.60	41.52	78.67	24.00	0.25	30.22	0.00	34.61	1.4
11.	Buch-17-ecsad	13.27	9.62	40.67	59.00	7.16	6.59	24.00	22.90	5.9
12.	Buch-17-shot	5.97	1.45	43.00	38.50	3.83	0.07	16.67	15.64	6.7
13.	Tejani-14	12.10	4.50	36.33	10.00	0.13	1.52	0.00	9.23	1.4
14.	Buch-16-ppfh	8.13	2.28	20.00	2.50	7.81	8.99	0.67	7.20	47.1
15.	Buch-16-ecsad	3.70	0.97	3.67	4.00	1.24	2.90	0.17	2.38	39.1

Evaluation results (1/2)

Methods based on point pair features, Template matching methods, Learning-based methods, Methods based on 3D local features

#	Method	LM	LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L	Average	Time (s)
1.	Vidal-18	87.83	59.31	95.33	96.50	66.51	36.52	80.17	74.60	4.7
2.	Drost-10-edge	79.13	54.95	94.00	92.00	67.50	27.17	87.33	71.73	21.5
3.	Drost-10	82.00	55.36	94.33	87.00	56.81	22.25	78.67	68.06	2.3
4.	Hodan-15	87.10	51.42	95.33	90.50	63.18	37.61	45.50	67.23	13.5
5.	Brachmann-16	75.33	52.04	73.33	56.50	17.84	24.35	88.67	55.44	4.4
6.	Hodan-15-nopso	69.83	34.39	84.67	76.00	62.70	32.39	27.83	55.40	12.3
7.	Buch-17-ppfh	56.60	36.96	95.00	75.00	25.10	20.80	68.67	54.02	14.2
8.	Kehl-16	58.20	33.91	65.00	44.00	24.60	25.58	7.50	36.97	1.8
9.	Buch-17-si	33.33	20.35	67.33	59.00	13.34	23.12	41.17	36.81	15.9
10.	Brachmann-14	67.60	41.52	78.67	24.00	0.25	30.22	0.00	34.61	1.4
11.	Buch-17-ecsad	13.27	9.62	40.67	59.00	7.16	6.59	24.00	22.90	5.9
12.	Buch-17-shot	5.97	1.45	43.00	38.50	3.83	0.07	16.67	15.64	6.7
13.	Tejani-14	12.10	4.50	36.33	10.00	0.13	1.52	0.00	9.23	1.4
14.	Buch-16-ppfh	8.13	2.28	20.00	2.50	7.81	8.99	0.67	7.20	47.1
15.	Buch-16-ecsad	3.70	0.97	3.67	4.00	1.24	2.90	0.17	2.38	39.1

- **Occlusion is a challenge** – recall on LM is at least 30% higher than on LM-O

Evaluation results (1/2)

Methods based on point pair features, Template matching methods, Learning-based methods, Methods based on 3D local features

#	Method	LM	LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L	Average	Time (s)
1.	Vidal-18	87.83	59.31	95.33	96.50	66.51	36.52	80.17	74.60	4.7
2.	Drost-10-edge	79.13	54.95	94.00	92.00	67.50	27.17	87.33	71.73	21.5
3.	Drost-10	82.00	55.36	94.33	87.00	56.81	22.25	78.67	68.06	2.3
4.	Hodan-15	87.10	51.42	95.33	90.50	63.18	37.61	45.50	67.23	13.5
5.	Brachmann-16	75.33	52.04	73.33	56.50	17.84	24.35	88.67	55.44	4.4
6.	Hodan-15-nopso	69.83	34.39	84.67	76.00	62.70	32.39	27.83	55.40	12.3
7.	Buch-17-ppfh	56.60	36.96	95.00	75.00	25.10	20.80	68.67	54.02	14.2
8.	Kehl-16	58.20	33.91	65.00	44.00	24.60	25.58	7.50	36.97	1.8
9.	Buch-17-si	33.33	20.35	67.33	59.00	13.34	23.12	41.17	36.81	15.9
10.	Brachmann-14	67.60	41.52	78.67	24.00	0.25	30.22	0.00	34.61	1.4
11.	Buch-17-ecsad	13.27	9.62	40.67	59.00	7.16	6.59	24.00	22.90	5.9
12.	Buch-17-shot	5.97	1.45	43.00	38.50	3.83	0.07	16.67	15.64	6.7
13.	Tejani-14	12.10	4.50	36.33	10.00	0.13	1.52	0.00	9.23	1.4
14.	Buch-16-ppfh	8.13	2.28	20.00	2.50	7.81	8.99	0.67	7.20	47.1
15.	Buch-16-ecsad	3.70	0.97	3.67	4.00	1.24	2.90	0.17	2.38	39.1

- **Occlusion is a challenge** – recall on LM is at least 30% higher than on LM-O
- **Object symmetries and similarities** (T-LESS) cause problems to methods based on 3D local features and learning-based methods

Evaluation results (1/2)

Methods based on point pair features, Template matching methods, Learning-based methods, Methods based on 3D local features

#	Method	LM	LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L	Average	Time (s)
1.	Vidal-18	87.83	59.31	95.33	96.50	66.51	36.52	80.17	74.60	4.7
2.	Drost-10-edge	79.13	54.95	94.00	92.00	67.50	27.17	87.33	71.73	21.5
3.	Drost-10	82.00	55.36	94.33	87.00	56.81	22.25	78.67	68.06	2.3
4.	Hodan-15	87.10	51.42	95.33	90.50	63.18	37.61	45.50	67.23	13.5
5.	Brachmann-16	75.33	52.04	73.33	56.50	17.84	24.35	88.67	55.44	4.4
6.	Hodan-15-nopso	69.83	34.39	84.67	76.00	62.70	32.39	27.83	55.40	12.3
7.	Buch-17-ppfh	56.60	36.96	95.00	75.00	25.10	20.80	68.67	54.02	14.2
8.	Kehl-16	58.20	33.91	65.00	44.00	24.60	25.58	7.50	36.97	1.8
9.	Buch-17-si	33.33	20.35	67.33	59.00	13.34	23.12	41.17	36.81	15.9
10.	Brachmann-14	67.60	41.52	78.67	24.00	0.25	30.22	0.00	34.61	1.4
11.	Buch-17-ecsad	13.27	9.62	40.67	59.00	7.16	6.59	24.00	22.90	5.9
12.	Buch-17-shot	5.97	1.45	43.00	38.50	3.83	0.07	16.67	15.64	6.7
13.	Tejani-14	12.10	4.50	36.33	10.00	0.13	1.52	0.00	9.23	1.4
14.	Buch-16-ppfh	8.13	2.28	20.00	2.50	7.81	8.99	0.67	7.20	47.1
15.	Buch-16-ecsad	3.70	0.97	3.67	4.00	1.24	2.90	0.17	2.38	39.1

- **Occlusion is a challenge** – recall on LM is at least 30% higher than on LM-O
- **Object symmetries and similarities** (T-LESS) cause problems to methods based on 3D local features and learning-based methods
- **Varying lighting conditions** present a challenge for methods that rely on synthetic training RGB images rendered with fixed lighting

Evaluation results (1/2)

Methods based on point pair features, Template matching methods, Learning-based methods, Methods based on 3D local features

#	Method	LM	LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L	Average	Time (s)
1.	Vidal-18	87.83	59.31	95.33	96.50	66.51	36.52	80.17	74.60	4.7
2.	Drost-10-edge	79.13	54.95	94.00	92.00	67.50	27.17	87.33	71.73	21.5
3.	Drost-10	82.00	55.36	94.33	87.00	56.81	22.25	78.67	68.06	2.3
4.	Hodan-15	87.10	51.42	95.33	90.50	63.18	37.61	45.50	67.23	13.5
5.	Brachmann-16	75.33	52.04	73.33	56.50	17.84	24.35	88.67	55.44	4.4
6.	Hodan-15-nopso	69.83	34.39	84.67	76.00	62.70	32.39	27.83	55.40	12.3
7.	Buch-17-ppfh	56.60	36.96	95.00	75.00	25.10	20.80	68.67	54.02	14.2
8.	Kehl-16	58.20	33.91	65.00	44.00	24.60	25.58	7.50	36.97	1.8
9.	Buch-17-si	33.33	20.35	67.33	59.00	13.34	23.12	41.17	36.81	15.9
10.	Brachmann-14	67.60	41.52	78.67	24.00	0.25	30.22	0.00	34.61	1.4
11.	Buch-17-ecsad	13.27	9.62	40.67	59.00	7.16	6.59	24.00	22.90	5.9
12.	Buch-17-shot	5.97	1.45	43.00	38.50	3.83	0.07	16.67	15.64	6.7
13.	Tejani-14	12.10	4.50	36.33	10.00	0.13	1.52	0.00	9.23	1.4
14.	Buch-16-ppfh	8.13	2.28	20.00	2.50	7.81	8.99	0.67	7.20	47.1
15.	Buch-16-ecsad	3.70	0.97	3.67	4.00	1.24	2.90	0.17	2.38	39.1

- **Occlusion is a challenge** – recall on LM is at least 30% higher than on LM-O
- **Object symmetries and similarities** (T-LESS) cause problems to methods based on 3D local features and learning-based methods
- **Varying lighting conditions** present a challenge for methods that rely on synthetic training RGB images rendered with fixed lighting
- **Noisy depth images** in RU-APC present problems to all methods

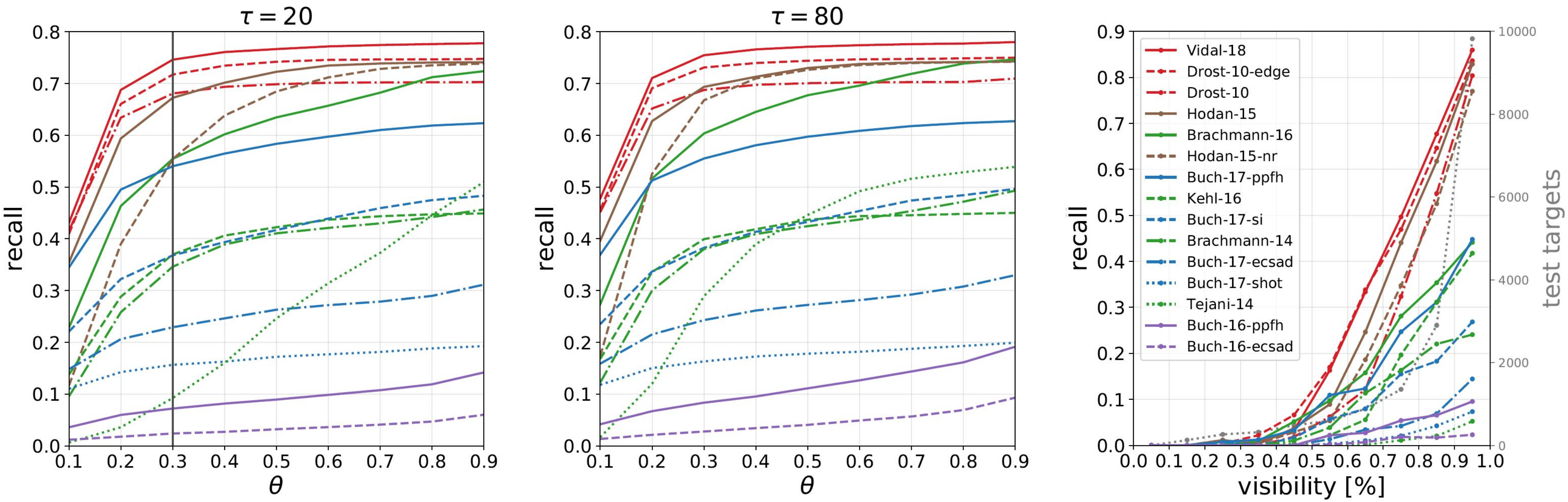
Evaluation results (1/2)

Methods based on point pair features, Template matching methods, Learning-based methods, Methods based on 3D local features

#	Method	LM	LM-O	IC-MI	IC-BIN	T-LESS	RU-APC	TUD-L	Average	Time (s)
1.	Vidal-18	87.83	59.31	95.33	96.50	66.51	36.52	80.17	74.60	4.7
2.	Drost-10-edge	79.13	54.95	94.00	92.00	67.50	27.17	87.33	71.73	21.5
3.	Drost-10	82.00	55.36	94.33	87.00	56.81	22.25	78.67	68.06	2.3
4.	Hodan-15	87.10	51.42	95.33	90.50	63.18	37.61	45.50	67.23	13.5
5.	Brachmann-16	75.33	52.04	73.33	56.50	17.84	24.35	88.67	55.44	4.4
6.	Hodan-15-nopso	69.83	34.39	84.67	76.00	62.70	32.39	27.83	55.40	12.3
7.	Buch-17-ppfh	56.60	36.96	95.00	75.00	25.10	20.80	68.67	54.02	14.2
8.	Kehl-16	58.20	33.91	65.00	44.00	24.60	25.58	7.50	36.97	1.8
9.	Buch-17-si	33.33	20.35	67.33	59.00	13.34	23.12	41.17	36.81	15.9
10.	Brachmann-14	67.60	41.52	78.67	24.00	0.25	30.22	0.00	34.61	1.4
11.	Buch-17-ecsad	13.27	9.62	40.67	59.00	7.16	6.59	24.00	22.90	5.9
12.	Buch-17-shot	5.97	1.45	43.00	38.50	3.83	0.07	16.67	15.64	6.7
13.	Tejani-14	12.10	4.50	36.33	10.00	0.13	1.52	0.00	9.23	1.4
14.	Buch-16-ppfh	8.13	2.28	20.00	2.50	7.81	8.99	0.67	7.20	47.1
15.	Buch-16-ecsad	3.70	0.97	3.67	4.00	1.24	2.90	0.17	2.38	39.1

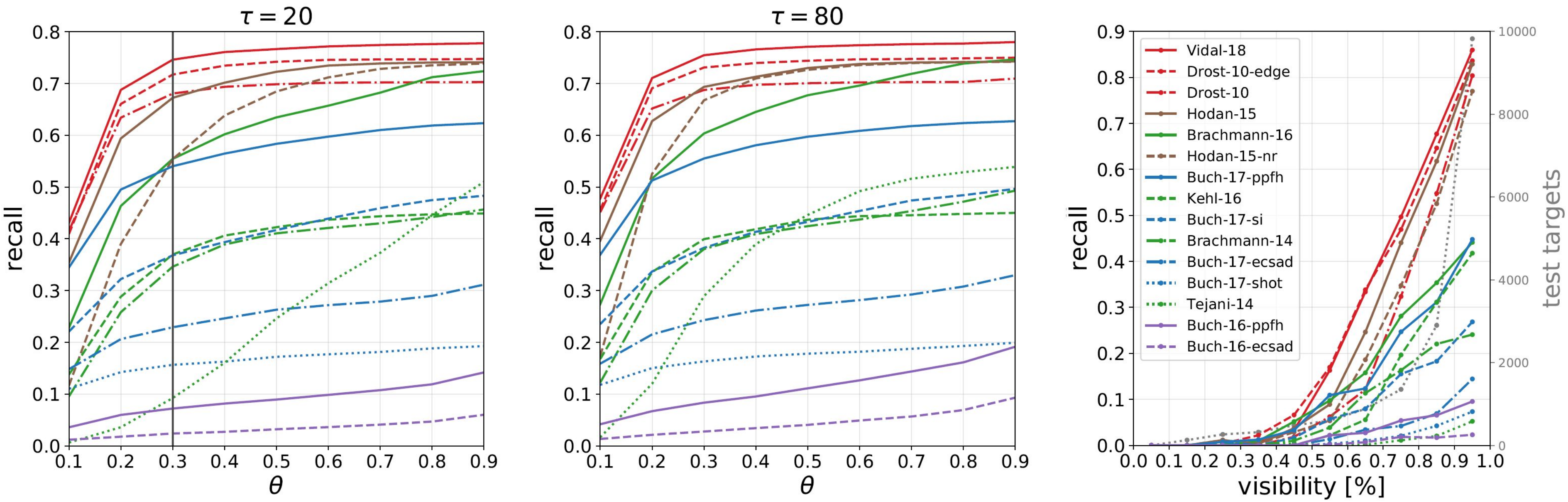
- **Occlusion is a challenge** – recall on LM is at least 30% higher than on LM-O
- **Object symmetries and similarities** (T-LESS) cause problems to methods based on 3D local features and learning-based methods
- **Varying lighting conditions** present a challenge for methods that rely on synthetic training RGB images rendered with fixed lighting
- **Noisy depth images** in RU-APC present problems to all methods
- Methods were **optimized primarily for recall**, not for speed

Evaluation results (2/2)



- Poses estimated by most methods are **either of a high quality or totally off** – recall grows only slightly if τ is increased from 20 to 80 mm, or if $\theta > 0.3$

Evaluation results (2/2)



- Poses estimated by most methods are **either of a high quality or totally off** – recall grows only slightly if τ is increased from 20 to 80 mm, or if $\theta > 0.3$
- Recall scores drop swiftly already **at low levels of occlusion**

Online evaluation system

bop.felk.cvut.cz

Up-to-date leaderboards

Form for continuous submission of new results

Datasets converted to a unified format

Python toolbox