Recognition and 6D Localization of Texture-less Objects

Jiří Matas, Tomáš Hodaň

Center for Machine Perception
Czech Technical University in Prague

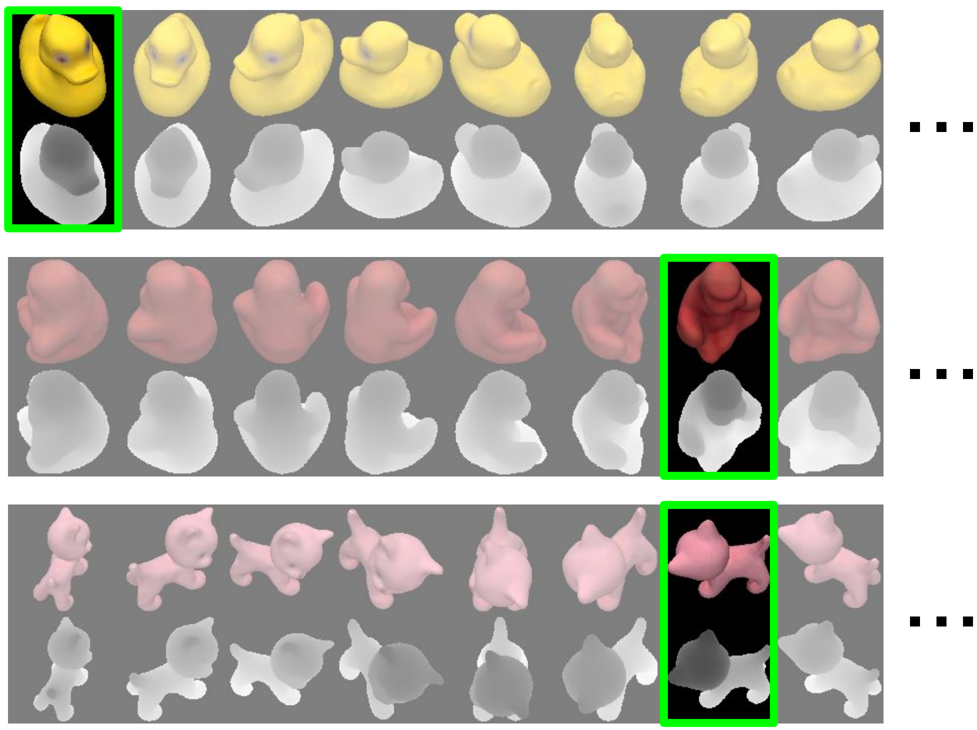
17th December 2015, Chile

Texture-less Object Detection for Robotics

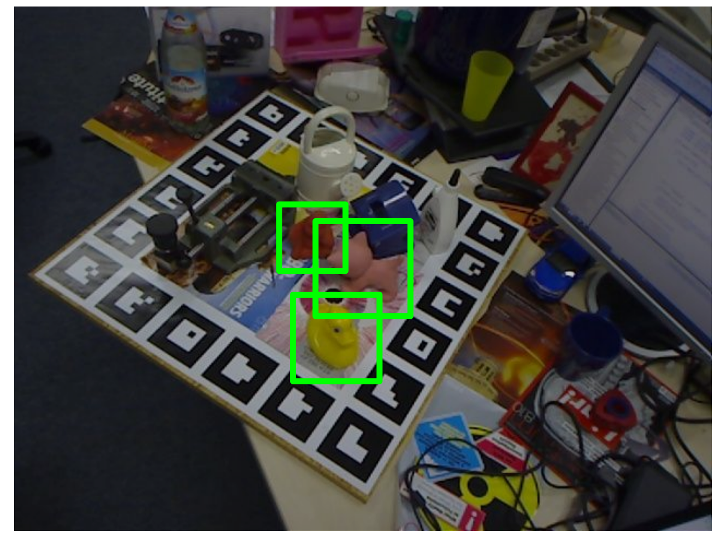
Detection and accurate localization of texture-less or texture-poor objects is commonly required in personal and industrial robotics

Problem Formulation

Given a database of training RGB / RGB-D images annotated with 6D poses or 3D model, **detect all instances of known objects** in a test image and **estimate their 6D poses**



Training RGB / RGB-D images annotated with 6D poses



Test RGB / RGB-D image

T-LESS

A new RGB-D dataset and evaluation protocol for detection and 6D pose estimation of texture-less objects

[http://cmp.felk.cvut.cz/**t-less**](http://cmp.felk.cvut.cz/t-less)

CTU Prague

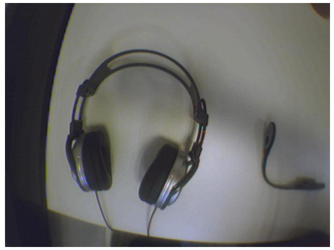
FORTH Heraklion

TU Wien

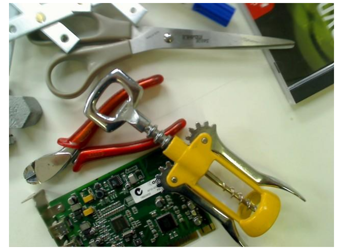
Existing Texture-less Datasets

- **RGB datasets**

CMP Toys



Bristol Tools



D-Textureless

CMU Kitchen Occlusion Dataset

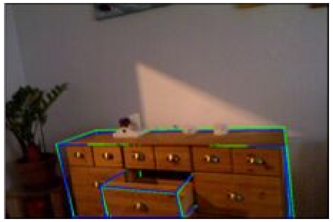
Rios-Cabrera et al.

- **RGB-D datasets**

Hinterstoisser et al.
(extended GT by Brachmann et al.)
3D models for 15 objects,
~1200 test images per object

UoB Highly Occluded Object Dataset (Walas et al.)
25 object categories - 5 objects in each (3D models provided)

Multi-Object Pose Estimation (Tejani et al.)
3D models for 6 objects,
~1000 test images per object

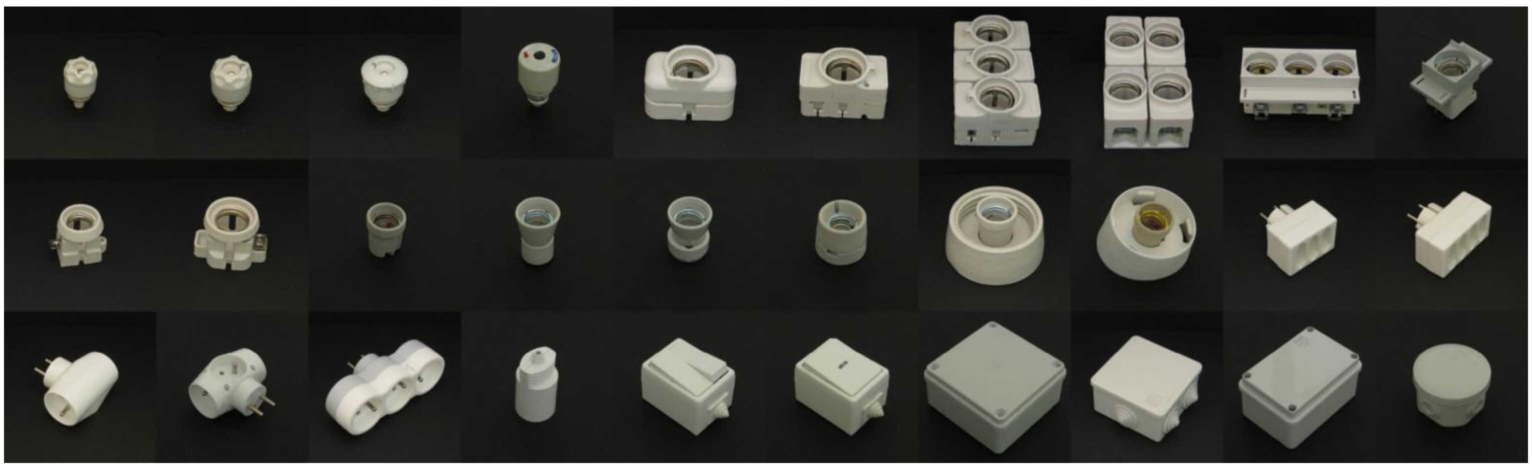


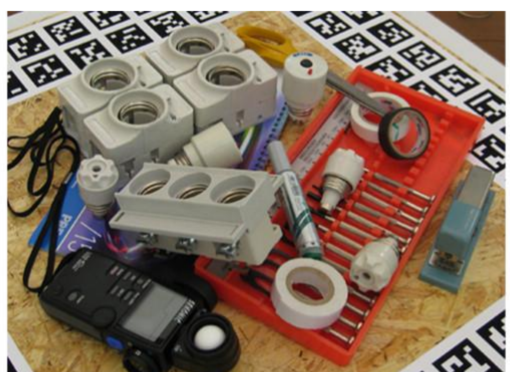
Articulated Objects
(Michel et al.)
4 articulated objects

- Common aspect: **objects often dissimilar in size, shape and color**

T-LESS: Key Features

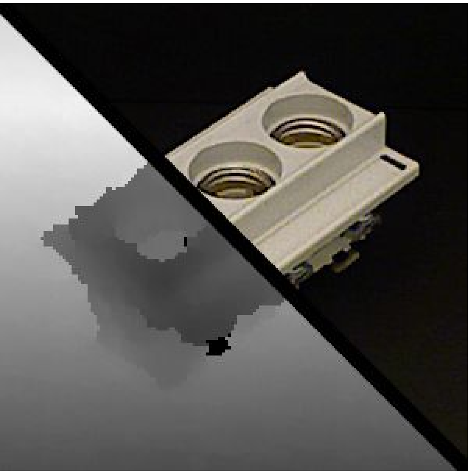
1. Relatively **small objects** often **very similar in shape and color**
2. Test images include significant **clutter and occlusions**
3. Accurate **ground truth 6D pose for all known objects** in each image
4. Data from **three synchronized and mutually calibrated sensors**
   (a structured-light depth sensor, a time-of-flight depth sensor, and a high-resolution camera)



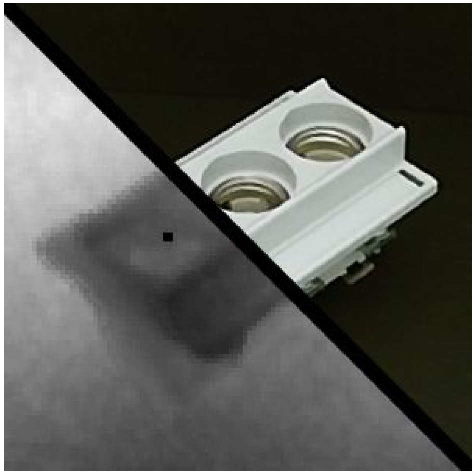


T-LESS: 30 Texture-less Objects

- **RGB-D & RGB training templates** depicting objects from a uniformly sampled full view sphere (10° step in elevation and 5° in azimuth)
  ➡ **1278 templates per object from each sensor**
- Each template is annotated with a **6D pose** of the object
- **Two 3D mesh models** for each object:
  1. Manually created
  2. Automatically reconstructed (TU Wien)

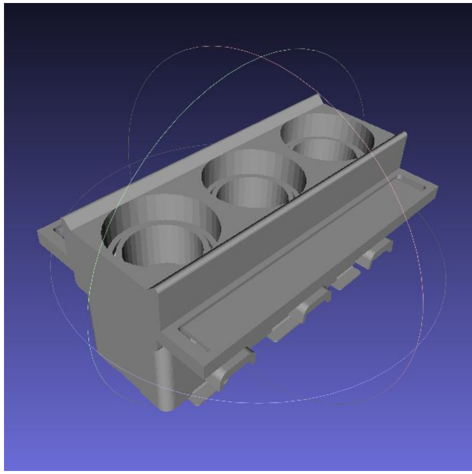


300x300 px RGB-D template from Primesense CARMINE 1.09



300x300 px RGB-D template from Kinect v2

1280x1280 px RGB template from Canon IXUS 950 IS



Manually created 3D model

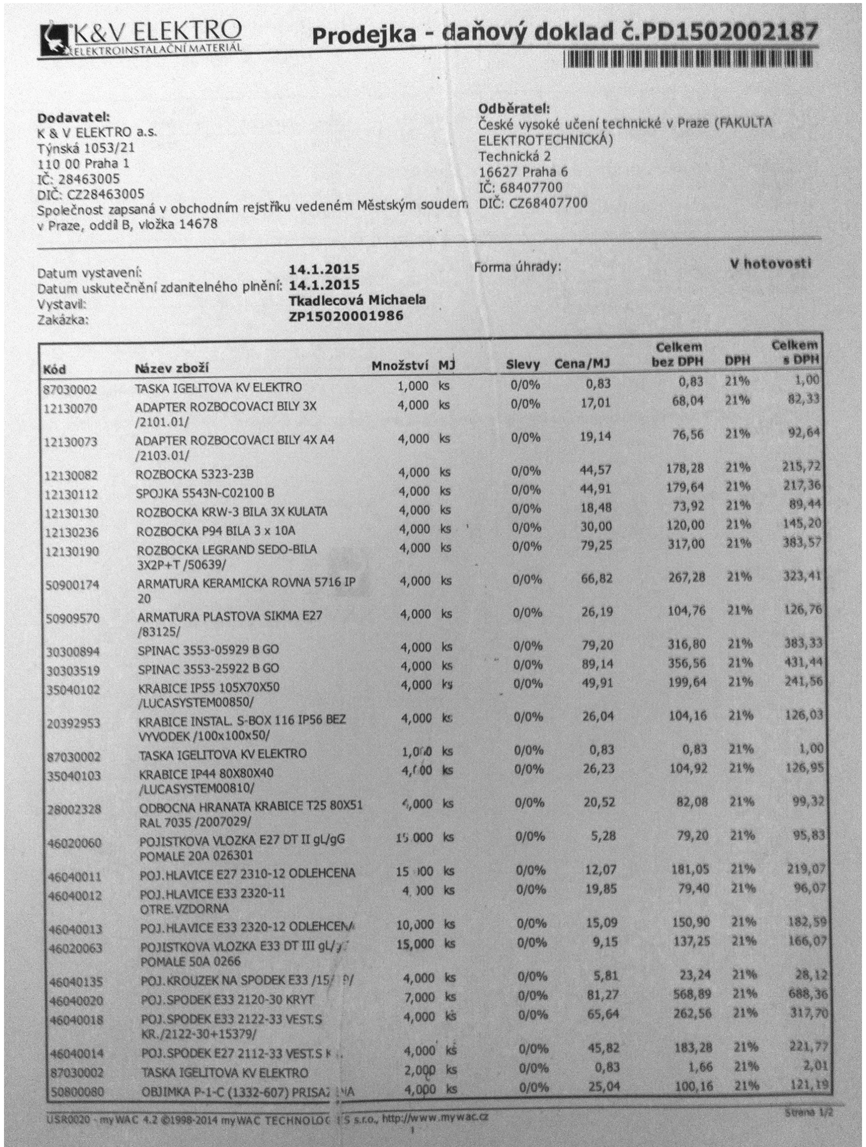
T-LESS: 3D Models vs Real Objects

T-LESS: 21 Test Scenes

- **RGB-D test images** depicting scenes from a uniformly sampled view hemisphere (10° step in elevation and 5° in azimuth)
  ➡ **568 test images from each scene**
- **Ground truth 6D poses** provided for all known objects
- The test scenes vary from simple ones with only few objects and black table top to very challenging ones **containing many similar objects, significant clutter and occlusion**

T-LESS: The Objects Can Be Bought

Buy the objects **for your own experiments** (e.g. grasping)



K&V ELEKTRO
ELEKTROINSTALAČNÍ MATERIÁL

Prodejka - daňový doklad č.PD1502002187

Dodavatel:
K & V ELEKTRO a.s.
Týnská 1053/21
110 00 Praha 1
IČ: 28463005
DIČ: CZ28463005
Společnost zapsaná v obchodním rejstříku vedeném Městským soudem v Praze, oddíl B, vložka 14678

Odběratel:
České vysoké učení technické v Praze (FAKULTA ELEKTROTECHNICKÁ)
Technická 2
16627 Praha 6
IČ: 68407700
DIČ: CZ68407700

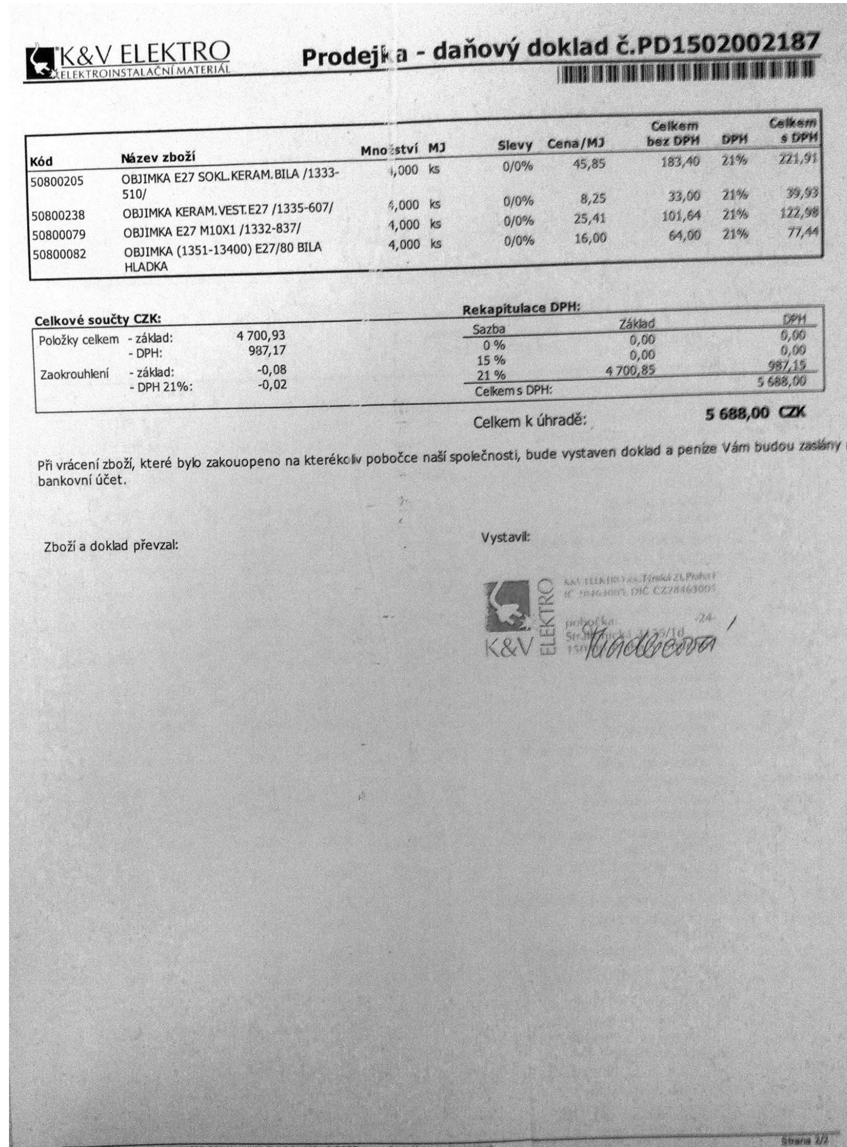
Datum vystavení: **14.1.2015**
Datum uskutečnění zdanitelného plnění: **14.1.2015**
Vystavil: **Tkadlecová Michaela**
Zakázka: **ZP15020001986**

Forma úhrady: **V hotovosti**

Kód	Název zboží	Množství	MJ	Slevy	Cena/MJ	Celkem bez DPH	DPH	Celkem s DPH
87030002	TASKA IGELITOVA KV ELEKTRO	1,000	ks	0/0%	0,83	0,83	21%	1,00
12130070	ADAPTER ROZBOCOVACI BILY 3X /2101.01/	4,000	ks	0/0%	17,01	68,04	21%	82,33
12130073	ADAPTER ROZBOCOVACI BILY 4X A4 /2103.01/	4,000	ks	0/0%	19,14	76,56	21%	92,64
12130082	ROZBOCKA 5323-23B	4,000	ks	0/0%	44,57	178,28	21%	215,72
12130112	SPOJKA 5543N-C02100 B	4,000	ks	0/0%	44,91	179,64	21%	217,36
12130130	ROZBOCKA KRW-3 BILA 3X KULATA	4,000	ks	0/0%	18,48	73,92	21%	89,44
12130236	ROZBOCKA P94 BILA 3 x 10A	4,000	ks	0/0%	30,00	120,00	21%	145,20
12130190	ROZBOCKA LEGRAND SEDO-BILA 3X2P+T /50639/	4,000	ks	0/0%	79,25	317,00	21%	383,57
50900174	ARMATURA KERAMICKA ROVNA 5716 IP 20	4,000	ks	0/0%	66,82	267,28	21%	323,41
50909570	ARMATURA PLASTOVA SIKMA E27 /83125/	4,000	ks	0/0%	26,19	104,76	21%	126,76
30300894	SPINAC 3553-05929 B GO	4,000	ks	0/0%	79,20	316,80	21%	383,33
30303519	SPINAC 3553-25922 B GO	4,000	ks	0/0%	89,14	356,56	21%	431,44
35040102	KRABICE IP55 105X70X50 /LUCASYSTEM00850/	4,000	ks	0/0%	49,91	199,64	21%	241,56
20392953	KRABICE INSTAL. S-BOX 116 IP56 BEZ VYVODEK /100x100x50/	4,000	ks	0/0%	26,04	104,16	21%	126,03
87030002	TASKA IGELITOVA KV ELEKTRO	1,000	ks	0/0%	0,83	0,83	21%	1,00
35040103	KRABICE IP44 80X80X40 /LUCASYSTEM00810/	4,000	ks	0/0%	26,23	104,92	21%	126,95
28002328	ODBOCNA HRANATA KRABICE T25 80X51 RAL 7035 /2007029/	4,000	ks	0/0%	20,52	82,08	21%	99,32
46020060	POJISTKOVA VLOZKA E27 DT II gL/gG POMALE 20A 026301	15,000	ks	0/0%	5,28	79,20	21%	95,83
46040011	POJ.HLAVICE E27 2310-12 ODLEHCENA	15,000	ks	0/0%	12,07	181,05	21%	219,07
46040012	POJ.HLAVICE E33 2320-11 OTRE.VZDORNA	4,000	ks	0/0%	19,85	79,40	21%	96,07
46040013	POJ.HLAVICE E33 2320-12 ODLEHCENA	10,000	ks	0/0%	15,09	150,90	21%	182,59
46020063	POJISTKOVA VLOZKA E33 DT III gL/[illegible] POMALE 50A 0266	15,000	ks	0/0%	9,15	137,25	21%	166,07
46040135	POJ.KROUZEK NA SPODEK E33 /15[illegible]/	4,000	ks	0/0%	5,81	23,24	21%	28,12
46040020	POJ.SPODEK E33 2120-30 KRYT	7,000	ks	0/0%	81,27	568,89	21%	688,36
46040018	POJ.SPODEK E33 2122-33 VEST.S KR./2122-30+15379/	4,000	ks	0/0%	65,64	262,56	21%	317,70
46040014	POJ.SPODEK E27 2112-33 VEST.S K[illegible]	4,000	ks	0/0%	45,82	183,28	21%	221,77
87030002	TASKA IGELITOVA KV ELEKTRO	2,000	ks	0/0%	0,83	1,66	21%	2,01
50800080	OBJIMKA P-1-C (1332-607) PRISA[illegible]NA	4,000	ks	0/0%	25,04	100,16	21%	121,19

USR0020 - myWAC 4.2 ©1998-2014 myWAC TECHNOLOG[illegible]S s.r.o., http://www.mywac.cz

Strana 1/2



K&V ELEKTRO
ELEKTROINSTALAČNÍ MATERIÁL

Prodejka - daňový doklad č.PD1502002187

Kód	Název zboží	Množství	MJ	Slevy	Cena/MJ	Celkem bez DPH	DPH	Celkem s DPH
50800205	OBJIMKA E27 SOKL.KERAM.BILA /1333-510/	4,000	ks	0/0%	45,85	183,40	21%	221,91
50800238	OBJIMKA KERAM.VEST.E27 /1335-607/	4,000	ks	0/0%	8,25	33,00	21%	39,93
50800079	OBJIMKA E27 M10X1 /1332-837/	4,000	ks	0/0%	25,41	101,64	21%	122,98
50800082	OBJIMKA (1351-13400) E27/80 BILA HLADKA	4,000	ks	0/0%	16,00	64,00	21%	77,44

Celkové součty CZK:

Položky celkem	- základ:	4 700,93
	- DPH:	987,17
Zaokrouhlení	- základ:	-0,08
	- DPH 21%:	-0,02

Rekapitulace DPH:

Sazba	Základ	DPH
0 %	0,00	0,00
15 %	0,00	0,00
21 %	4 700,85	987,15
Celkem s DPH:		5 688,00

Celkem k úhradě: **5 688,00 CZK**

Při vrácení zboží, které bylo zakouopeno na kterékoliv pobočce naší společnosti, bude vystaven doklad a peníze Vám budou zaslány n[illegible] bankovní účet.

Zboží a doklad převzal:

Vystavil:

K&V ELEKTRO

Strana 2/2

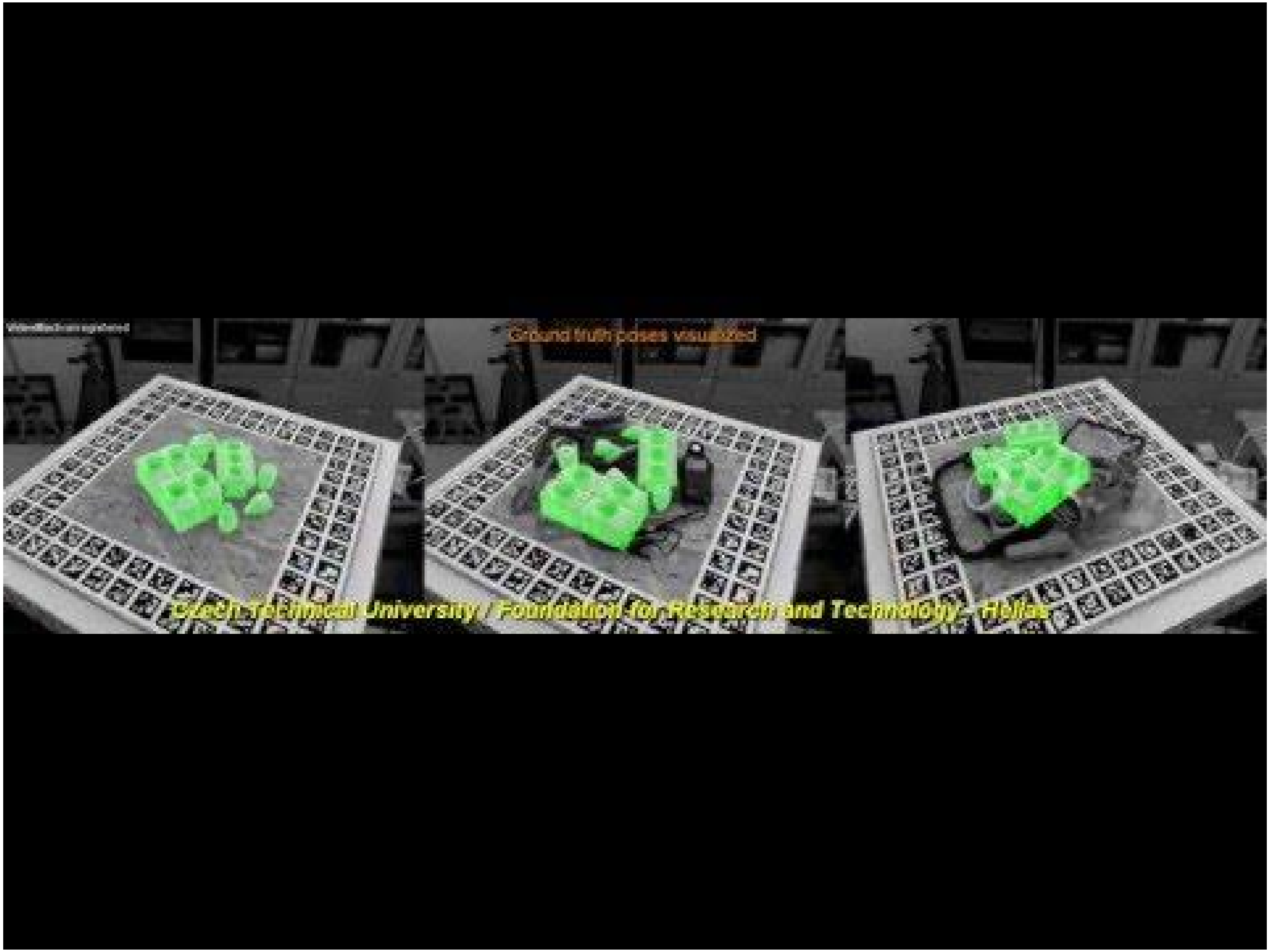
T-LESS: Illumination Conditions

- All training and test data captured in **fixed illumination conditions with dominant ambient light**
- 15 test scenes captured also in **alternative conditions with low ambient light and strong direct light from side**

Dominant ambient light

Low ambient light and strong direct light

T-LESS: Sample Ground Truth 6D Poses

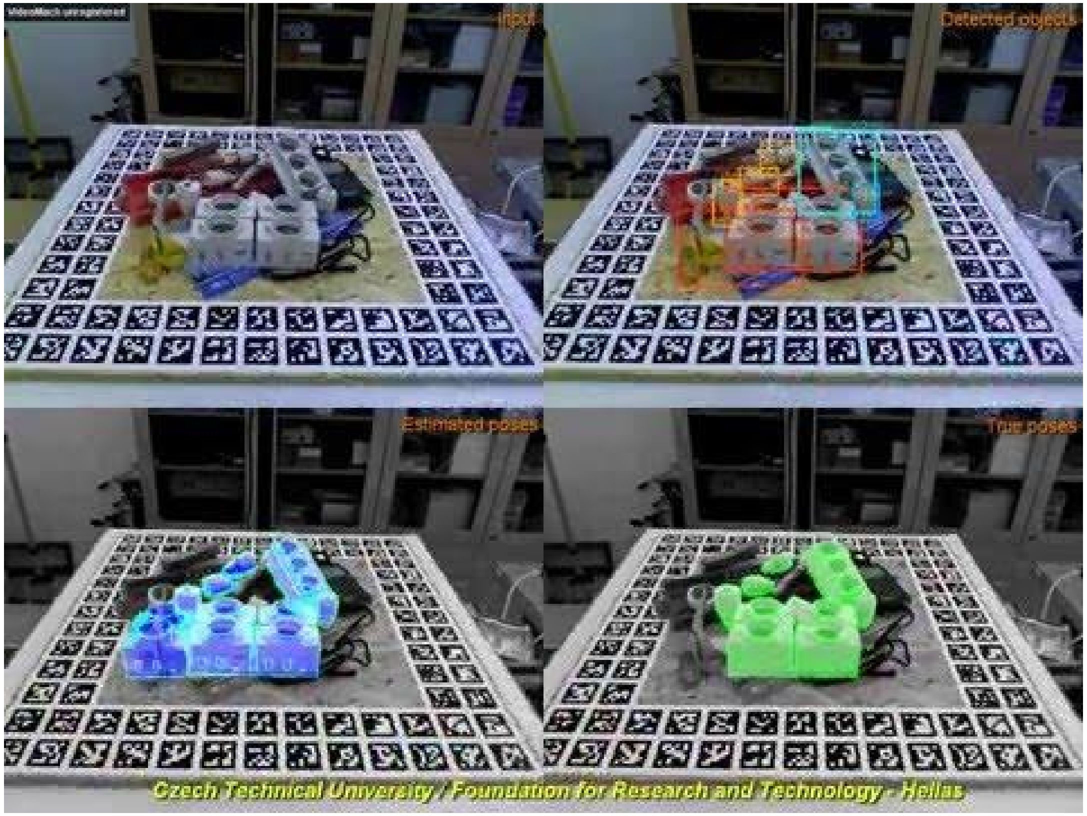


On [YouTube]

T-LESS: Estimation of GT 6D Poses

1. **We manually identify a set of images, in which an object's 6D pose can be accurately estimated** by the recognition and localization method by Hodan et al. (IROS 2015) = RGB-D template matching + 6D pose refinement by particle swarm optimization
2. **Mean of the 6D poses estimated in these images is transformed to all images** using camera poses estimated from the fiducial markers

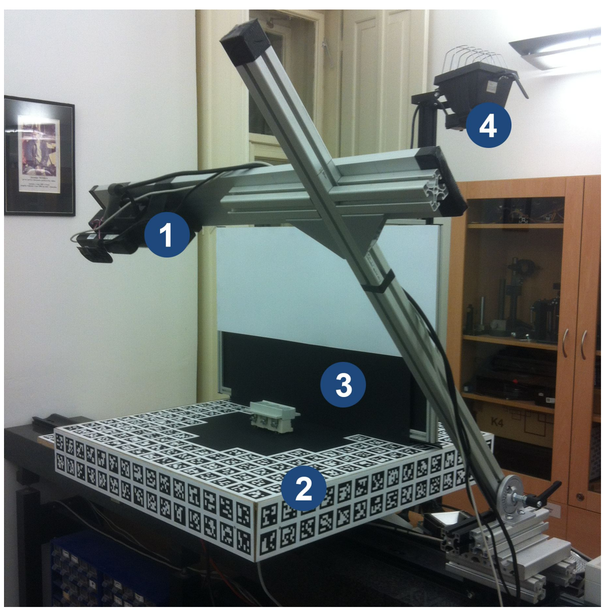
On YouTube



T-LESS: Capturing Setup

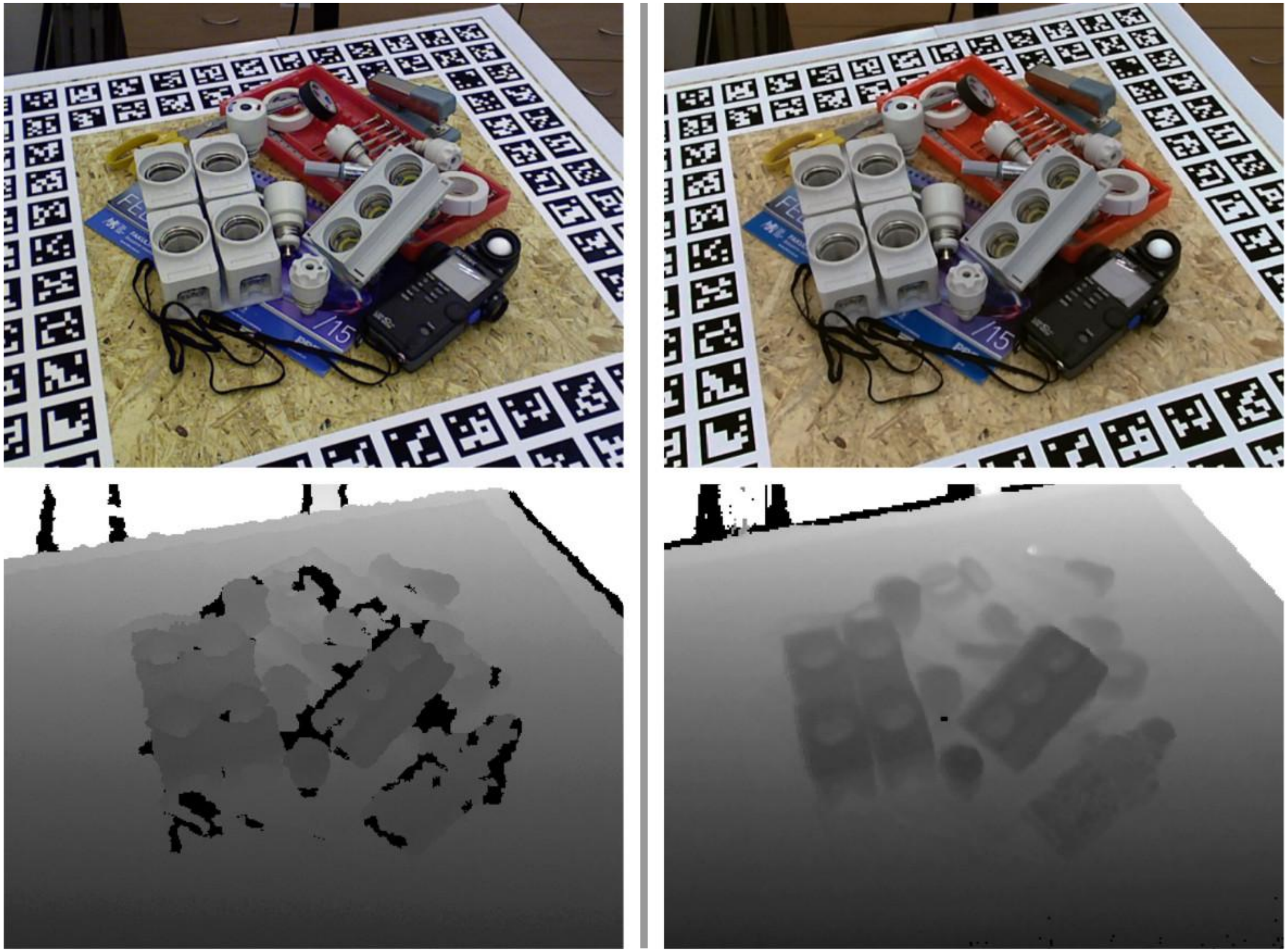
Sensors (synchronized and mutually calibrated):

- **Primesense Carmine 1.09 (Short Range)**
  registered RGB-D images (RGB:1280x1024 px, D:640x480 px)
- **Microsoft Kinect v2**
  registered RGB-D images (1920x1080 px)
- **Canon IXUS 950 IS**
  high resolution RGB images (3264x2448 px)



1. **Sensors** fixed on an arm with adjustable tilt
2. **A turntable with a marker field** for camera pose estimation (the vertical markers enable estimation from low elevations)
3. **A shield to ensure black background** in training templates (it is removed for capturing test data)
4. **A strong reflector** to increase ambient light

T-LESS: Primesense vs Kinect v2

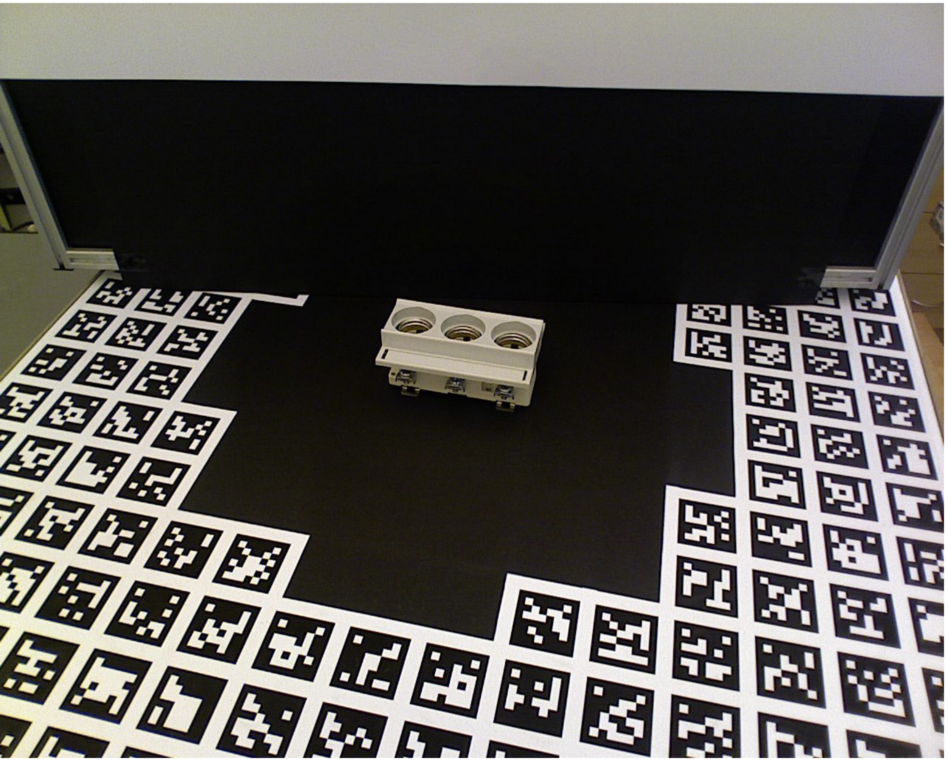


Primesense CARMINE 1.09 | Kinect v2

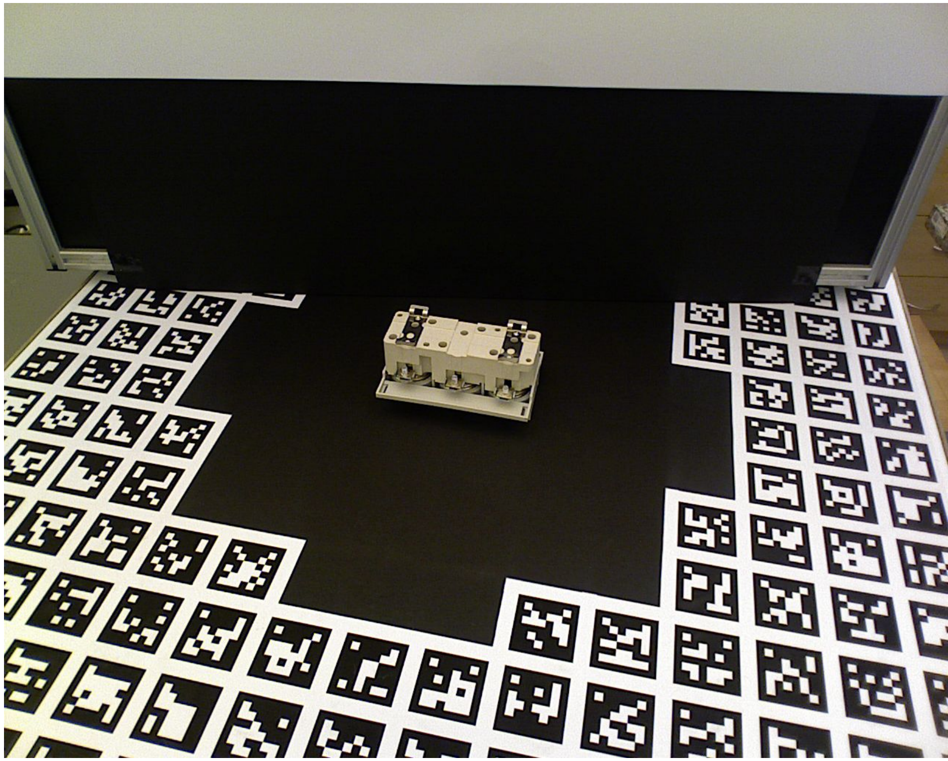
T-LESS: Views From Full Sphere

- To obtain views from the full sphere around the object, each object is captured **1) upright** and **2) upside down**, in both cases from elevations 5° to 85° (10° step in elevation and 5° in azimuth)

Upright



Upside down

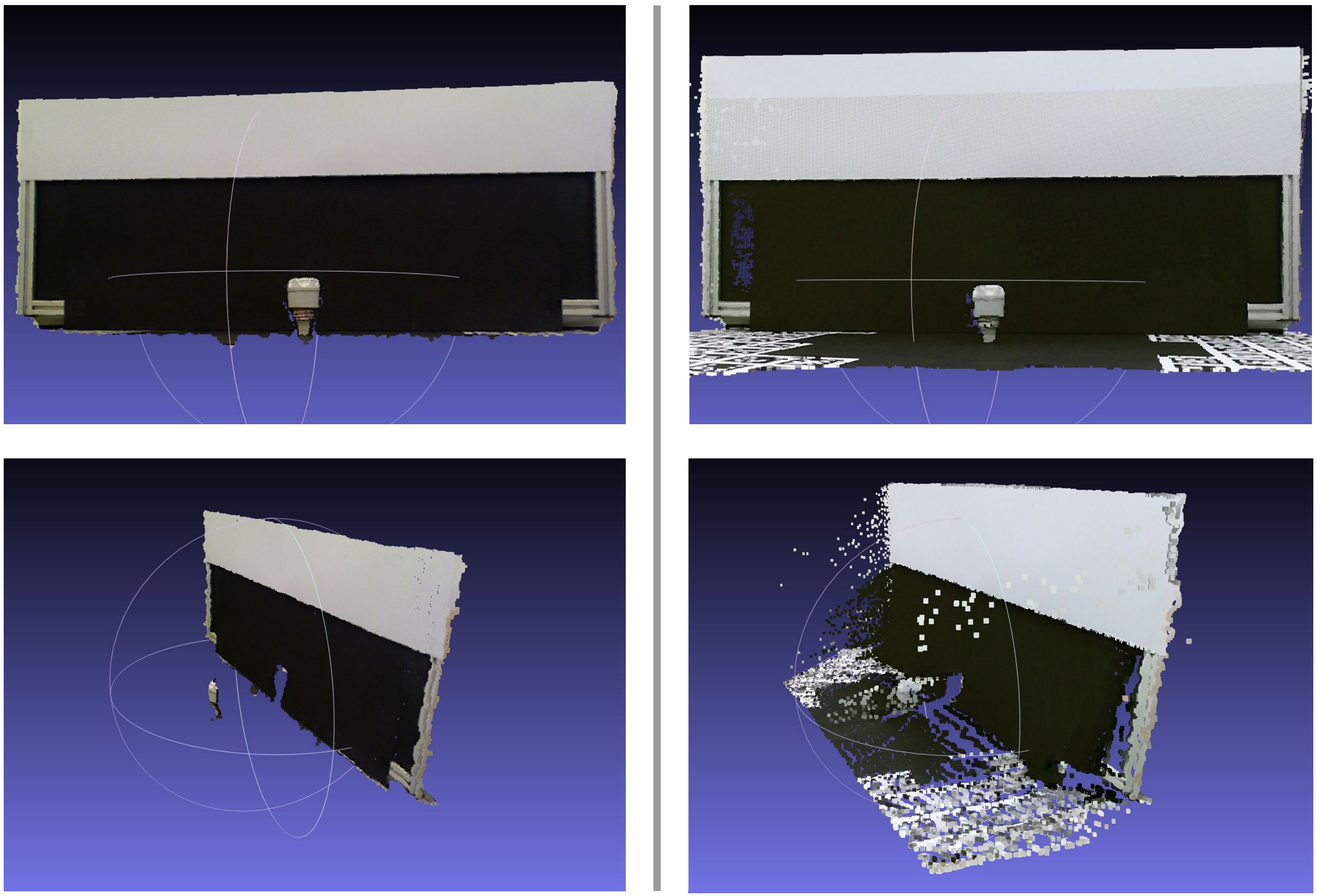


T-LESS: Primesense vs Kinect v2

- Primesense: **less noisy, but more missing values** (at slanted surfaces and around occlusion boundaries)

View 1

View 2



Primesense CARMINE 1.09

Kinect v2

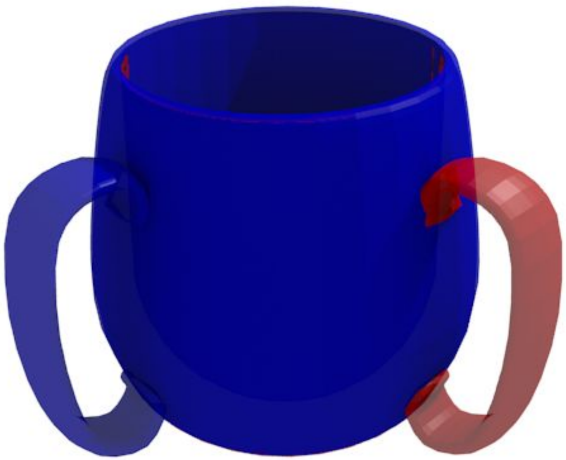
T-LESS: Current State

- Ongoing work:
    1. **Automatic reconstruction of 3D models** (the manually created models available)
    2. **Finalization of ground truth 6D poses**
    3. We are considering adding **new test scenes without markers** (the camera pose could be estimated e.g. from the known texture of the top of the turntable)

- **Expected release of the final version: February 2016**

6D Pose Evaluation

- **Evaluate how well a 3D model in an estimated 6D pose fits the same 3D model in the ground truth 6D pose**



A mug in the **ground truth** and an **estimated** pose

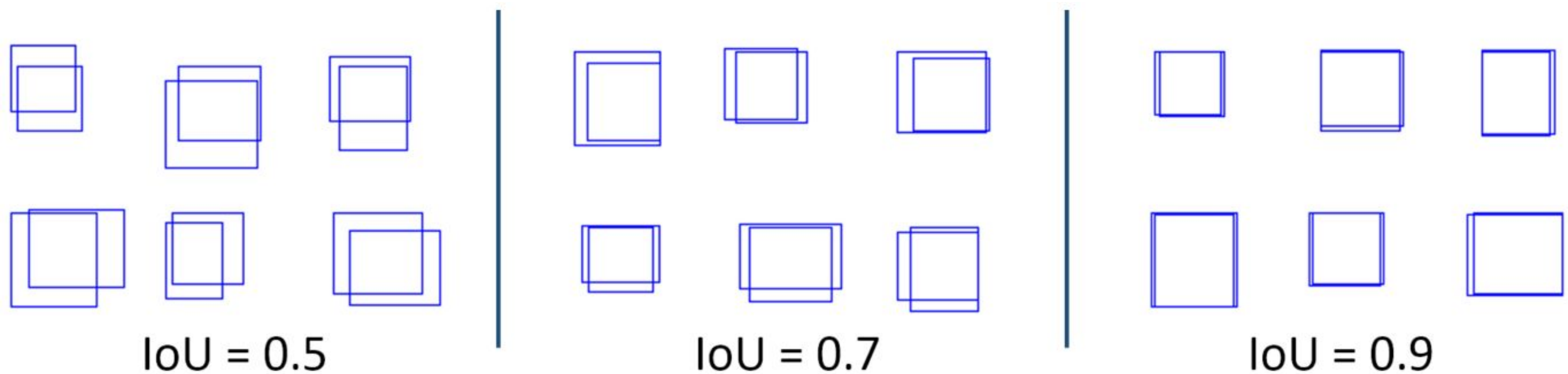
How good is the estimated pose?

- Commonly used evaluation criteria (used in the most of the challenges at this workshop):
  1. **Average distance (AD) criterion** (Hinterstoisser et al.)
  2. **5cm, 5deg** (Shotton et al.)
  3. **2D intersection over union (IoU) criterion** (Everingham et al.)

2D Intersection over Union (IoU) Criterion

M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman: The Pascal Visual Object Classes Challenge. IJCV 2010

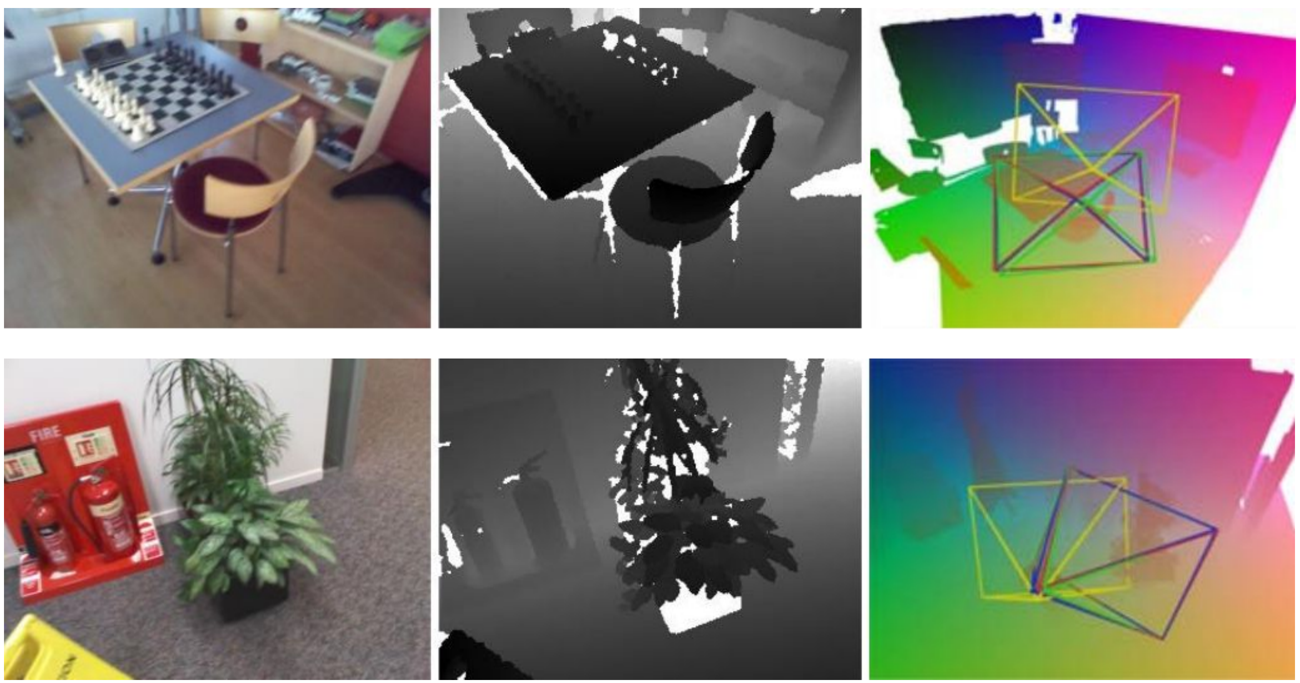
- A pose is considered correct, when **intersection over union of 2D bounding boxes of an object in the estimated and the ground truth pose is above a threshold (e.g. 0.5)**
- Weak, but allows comparison with 2D methods (e.g. Damen et al.)



5cm, 5deg Criterion

J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, A. Fitzgibbon: Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images. CVPR 2013

- A pose is considered correct, when **the translational error is below 5cm and the rotational error is below 5deg**
- Not adaptive to the object size
- Originally used for evaluation of camera pose estimation



Average Distance (AD) Criterion

S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. R. Bradski, K. Konolige, N. Navab: Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes. ACCV 2012

- A pose is considered correct, if **the average distance is below 10% of the object diameter**
- Adaptive to the object size

- Average distance for **non-symmetrical objects** (~ 6D pose distance):

$$d_h\left((\mathbf{R},\mathbf{t}),(\tilde{\mathbf{R}},\tilde{\mathbf{t}});\mathcal{M}\right)=\frac{1}{|\mathcal{M}|}\sum_{\mathbf{x}\in\mathcal{M}}\left\|(\mathbf{R}\mathbf{x}+\mathbf{t})-(\tilde{\mathbf{R}}\mathbf{x}+\tilde{\mathbf{t}})\right\|_2$$

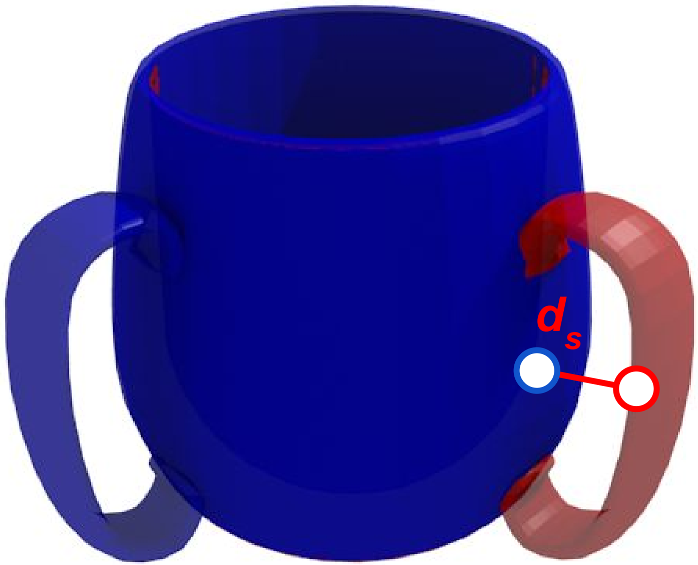
- Average distance for **symmetrical objects** (~ 3D surface distance):

$$d'_h\left((\mathbf{R},\mathbf{t}),(\tilde{\mathbf{R}},\tilde{\mathbf{t}});\mathcal{M}\right)=\frac{1}{|\mathcal{M}|}\sum_{\mathbf{x}_1\in\mathcal{M}}\min_{\mathbf{x}_2\in\mathcal{M}}\left\|(\mathbf{R}\mathbf{x}_1+\mathbf{t})-(\tilde{\mathbf{R}}\mathbf{x}_2+\tilde{\mathbf{t}})\right\|_2$$

Surface Distance (proposed)

$$d_s\left((\mathbf{R},\mathbf{t}),(\tilde{\mathbf{R}},\tilde{\mathbf{t}});\mathcal{M}\right)=\max_{\mathbf{x_1}\in\mathcal{M}}\ \min_{\mathbf{x_2}\in\mathcal{M}}\left\|(\mathbf{R}\mathbf{x_1}+\mathbf{t})-(\tilde{\mathbf{R}}\mathbf{x_2}+\tilde{\mathbf{t}})\right\|_2$$

- **Maximum instead of average**
  - Maximum error is **more relevant** for robotics (grasping, assembly, etc.)
  - There is **no noise**, we are matching the same 3D model in two poses

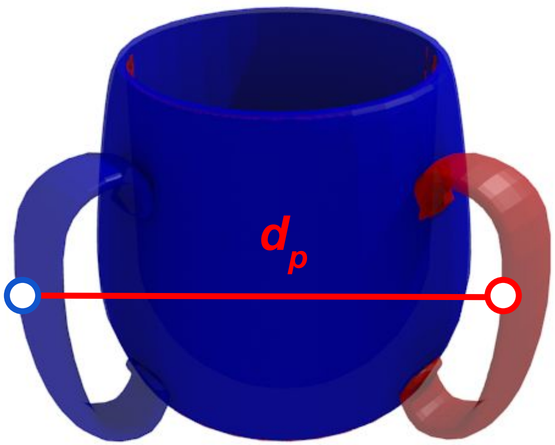


$\boldsymbol{d_s}$ reflects the misalignment better than $\boldsymbol{d'_h}$ (the average distance for symmetrical objects), which is in this case very low, indicating a good fit

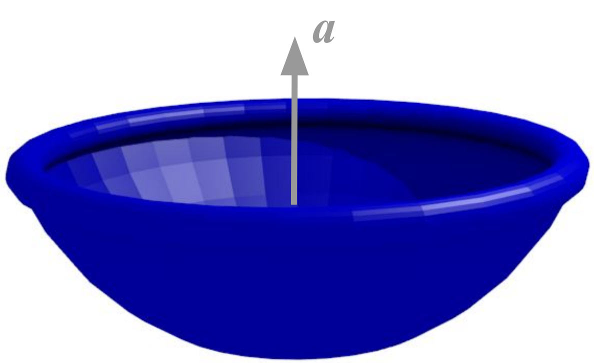
Corresponding Point Distance (proposed)

$$d_p\left((\mathbf{R},\mathbf{t}),(\tilde{\mathbf{R}},\tilde{\mathbf{t}});\mathcal{M}\right)=\min_{(\hat{\mathbf{R}},\hat{\mathbf{t}})\in[(\mathbf{R},\mathbf{t})]}\ \max_{\mathbf{x}\in\mathcal{M}}\left\|(\hat{\mathbf{R}}\mathbf{x}+\hat{\mathbf{t}})-(\tilde{\mathbf{R}}\mathbf{x}+\tilde{\mathbf{t}})\right\|_2$$

- Considers all poses from the equivalence class $[(\mathbf{R},\mathbf{t})]$ of the ground truth pose (given by pre-defined **symmetries of the object**)



a non-symmetrical mug



a bowl with rotational symmetry

➡ all poses varying in the rotation around $\boldsymbol{a}$ are considered equivalent

Definition of Evaluation Tasks

1. **6D localization** (multiple classes, multiple instances)
   - Generalization of the Hinterstoisser's task (one instance per image)
   - **Input:**
     - a test RGB-D image and training data of known objects
     - a list of pairs *(present object class, number of instances)*
   - **Output:**
     - a list ***R*** of tuples *(object class, estimated 6D pose, score)*

2. **Detection and 6D localization** (multiple classes, multiple instances)
   - **Input:**
     - a test RGB-D image and training data of known objects
     - no prior knowledge about the present object instances
   - **Output:**
     - a list ***R*** of tuples *(object class, estimated 6D pose, score)*

Detection and Fine 3D Pose Estimation of Texture-less Objects in RGB-D Images

Tomáš Hodaň[1], **Xenophon Zabulis**[2], **Manolis Lourakis**[2], **Štěpán Obdržálek**[1], **Jiří Matas**[1]

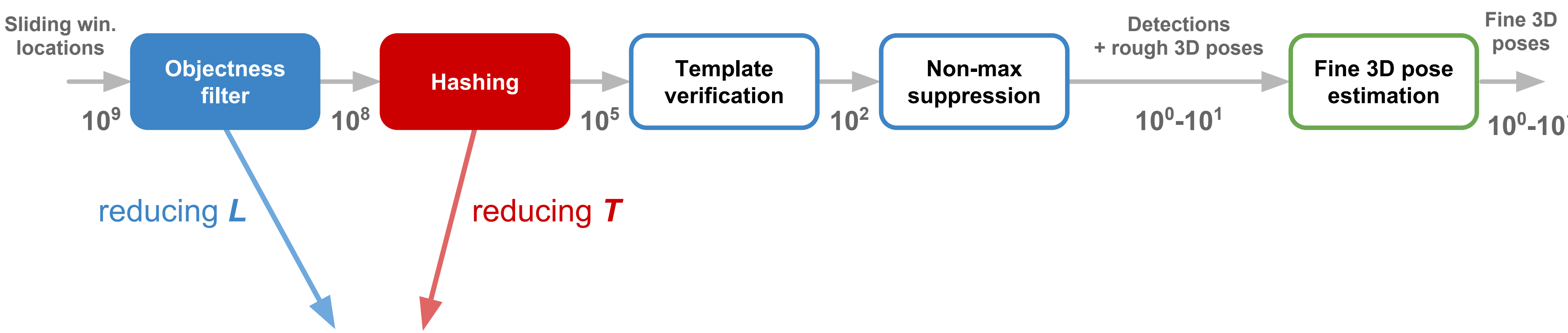
[1] Center for Machine Perception, CTU in Prague, CZ

[2] Institute of Computer Science, FORTH, Heraklion, GR

Published at IROS 2015

The Proposed Method

- Multi-scale **sliding window**
- **Efficient cascade-style evaluation** of each location
- The window has a **fixed size**, the same as the templates
- Stochastic optimization used to **refine the 3D pose**

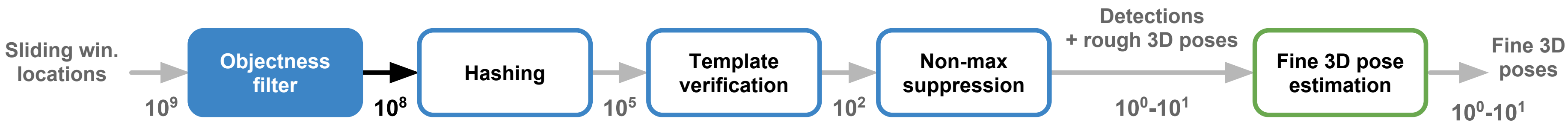


$O(LT)$ = complexity of an exhaustive template matching

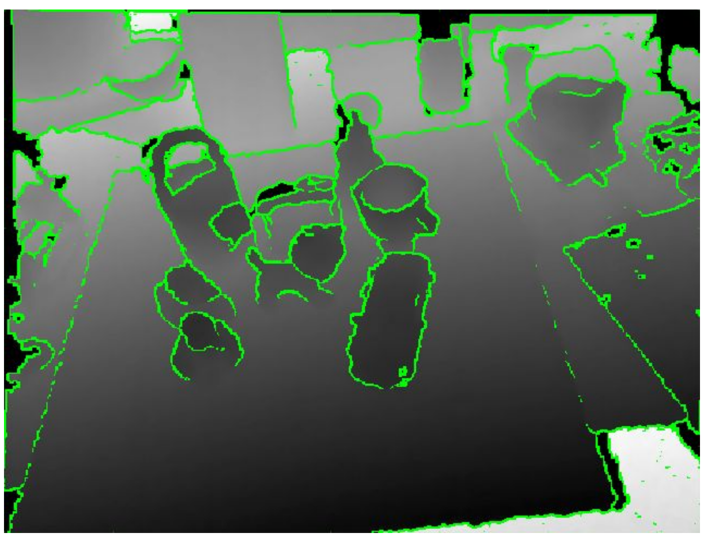
L = the number of **sliding window locations**

T = the number of **training templates**

Objectness Filter

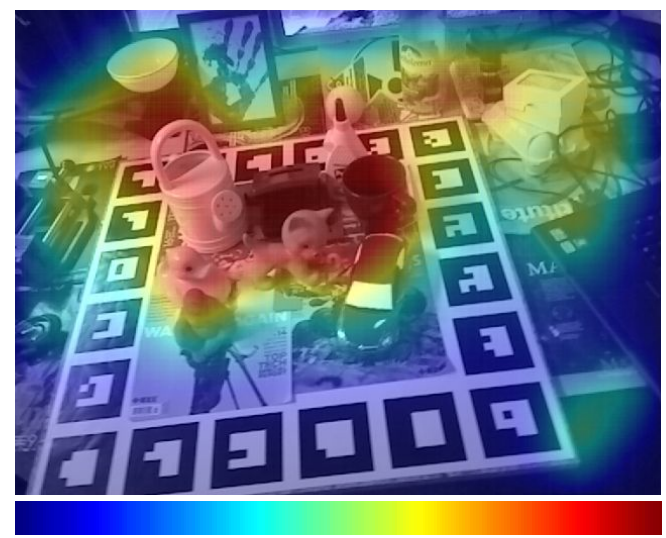


- Based on the **number of depth edges**
- The number of depth edges in a window is required to be **at least 30% of the minimum from the training templates**
- For false negative rate = 0, **60-90% of locations are pruned**
- Other window proposal methods (e.g. Edge-boxes) are being considered



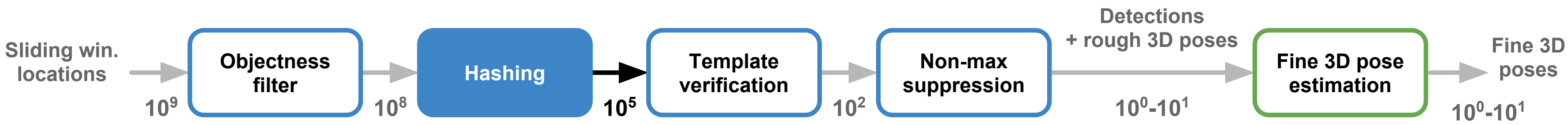
Detected depth edges

Number of detection candidates: 1.7 x 10^8



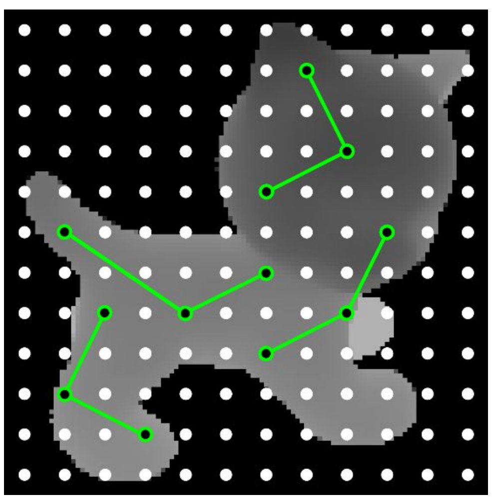
Density of detection candidates
detection candidate = (tpl. id, x, y, scale)

Hashing

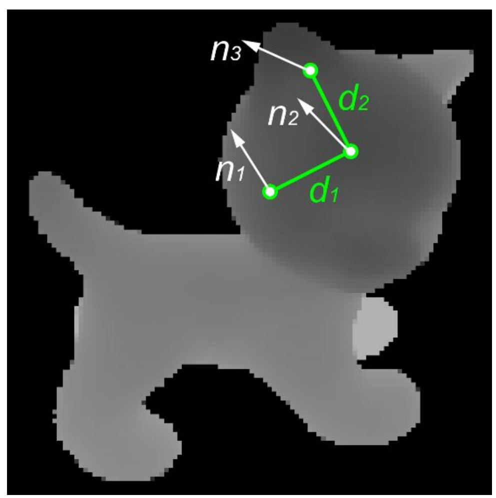


- Voting procedure based on **hashing descriptors of trained triplets of reference points** located on a grid
- Each triplet is described by **surface normals and depth differences**
- **Up to N templates with the most votes** are selected per location

Typically: N = 100, 8 bins for surface normal orientation, 5 bins for depth difference, i.e. $5^2 8^3$ = 12800 hash table bins

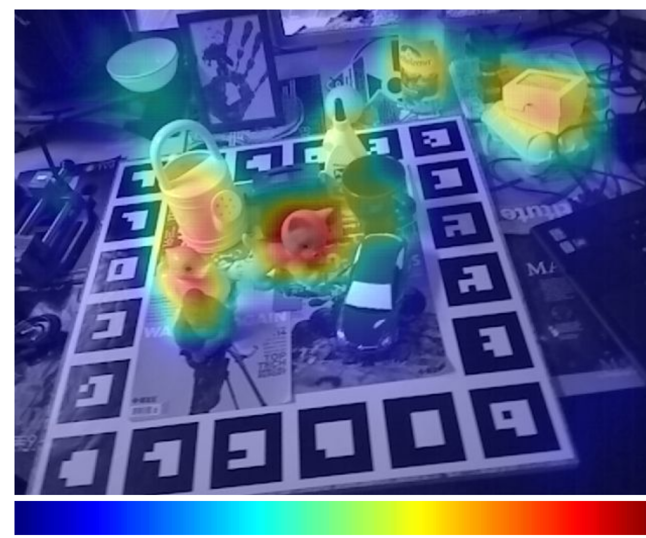


Sample triplets



Triplet description

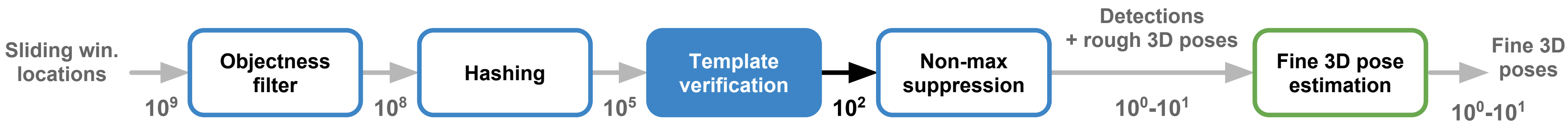
Number of detection candidates: 5.2 x 10^5



Density of detection candidates

detection candidate = (tpl. id, x, y, scale)

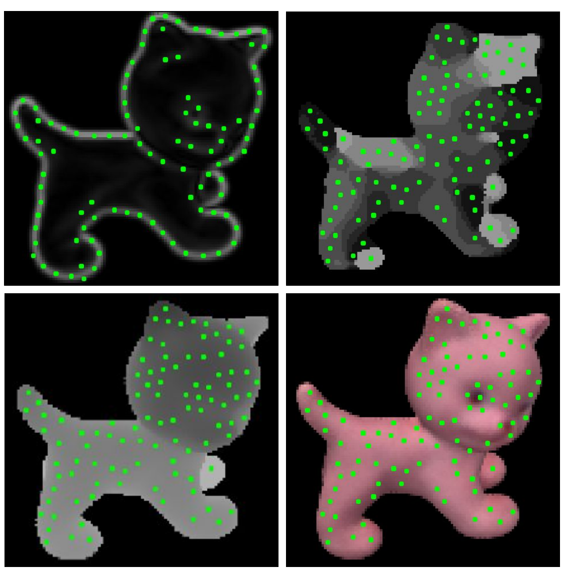
Multimodal Template Verification



- **A sequence of tests** evaluating consistency of:
  a. Object size and the measured depth
  b. Surface normals
  c. Image gradients
  d. Depth
  e. Color (HSV)

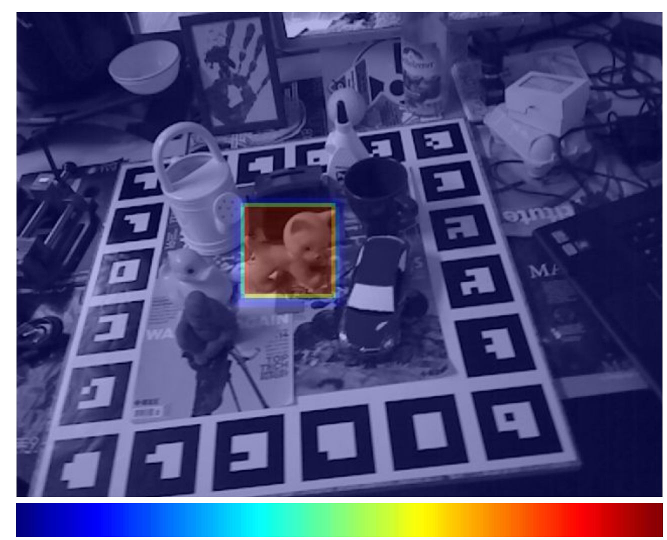
Evaluated on learnt feature points

Based on: Hinterstoisser et al., "Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes", ICCV, 2011



Learnt feature points in different modalities

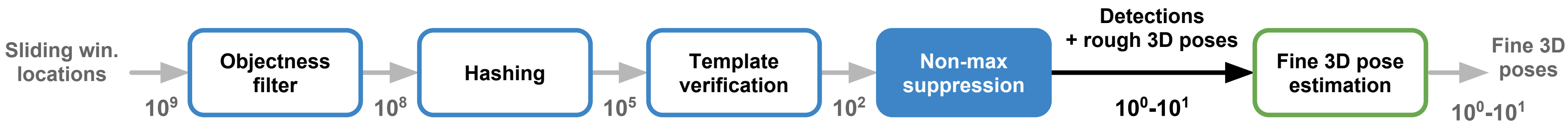
Number of detections: 44



Density of detection candidates

detection candidate = (tpl. id, x, y, scale)

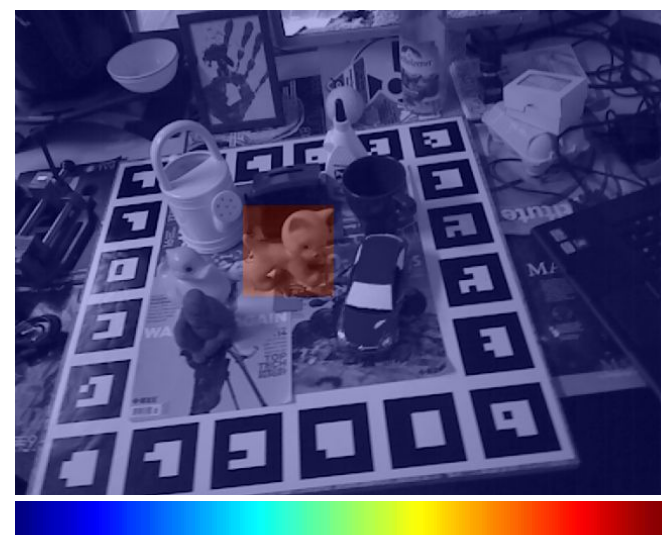
Non-maxima Suppression



- Detection candidates with **locally highest score are retained**
- The 3D poses associated with the detected templates are used as **initial poses** in the pose refinement procedure

Rendering of the 3D pose associated with the detected template

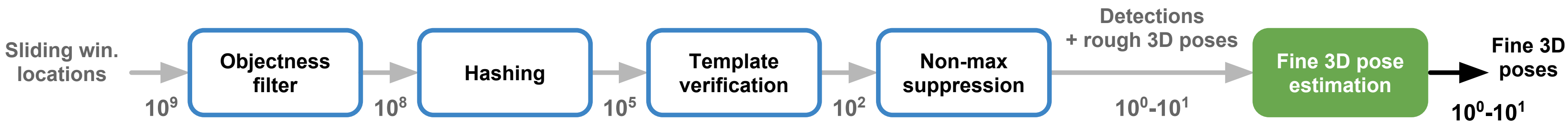
Number of detections: 1



Density of detection candidates

detection candidate = (tpl. id, x, y, scale)

Fine 3D Pose Estimation



- The rough initial 3D pose is refined using a hypothesize and test scheme based on **Particle Swarm Optimization** (PSO)
- PSO stochastically evolves a population of candidate poses over multiple iterations
- Candidate poses are evaluated by comparing their rendered depth images to the input depth image (using a cost function measuring similarity in **depth, surface normals and depth edges**)
- Pose refinement using PSO is **less sensitive to local minima compared to ICP**

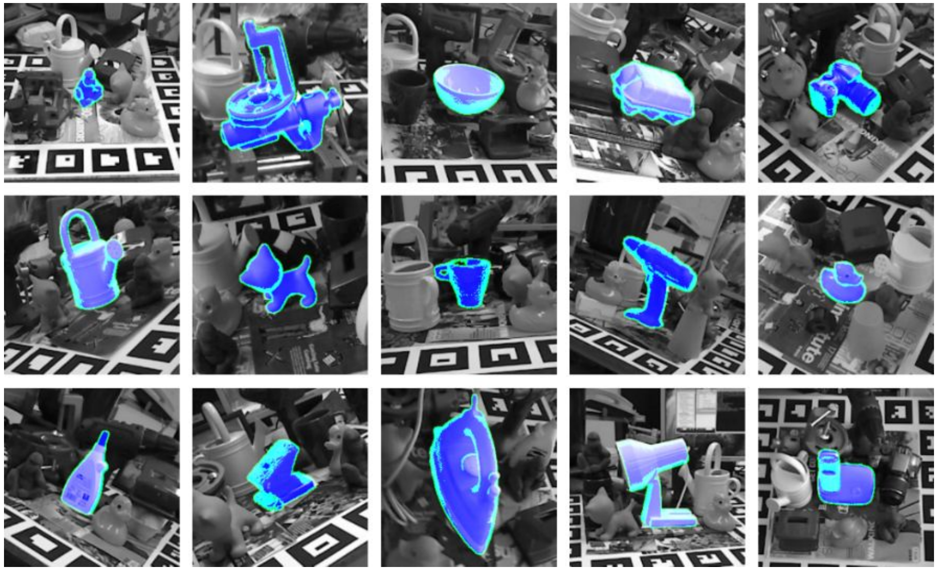
Details in: Zabulis, Lourakis and Koutlemanis, ”3D Object Pose Refinement in Range Images”, Intl Conf. on Computer Vision Systems, ICVS, 2015

Recognition Rate

- Evaluation on the **dataset of Hinterstoisser** [1]:
  - 15 texture-less objects, 1200 RGB-D test images for each
  - **Object localization**: detect the given object and estimate its pose
- The recognition rate (recall) of our method is **comparable to SOTA**

Sequence	Our method	LINEMOD++	LINEMOD	Drost et al.
1. Ape	93.9	**95.8**	69.4	86.5
2. Benchvise	**99.8**	98.7	94.0	70.7
3. Bowl	98.8	**99.9**	99.5	95.7
4. Box	**100.0**	99.8	99.1	97.0
5. Cam	95.5	**97.5**	79.5	78.6
6. Can	**95.9**	95.4	79.5	80.2
7. Cat	98.2	**99.3**	88.2	85.4
8. Cup	**99.5**	97.1	80.7	68.4
9. Driller	**94.1**	93.6	81.3	87.3
10. Duck	94.3	**95.9**	75.9	46.0
11. Glue	**98.0**	91.8	64.3	57.2
12. Hole punch	88.0	**95.9**	78.4	77.4
13. Iron	97.0	**97.5**	88.8	84.9
14. Lamp	88.8	**97.7**	89.8	93.3
15. Phone	89.4	**93.3**	77.8	80.7
Average	95.4	**96.6**	83.0	79.3

Recognition rates [%]
(LINEMOD and LINEMOD++ are methods from [1])

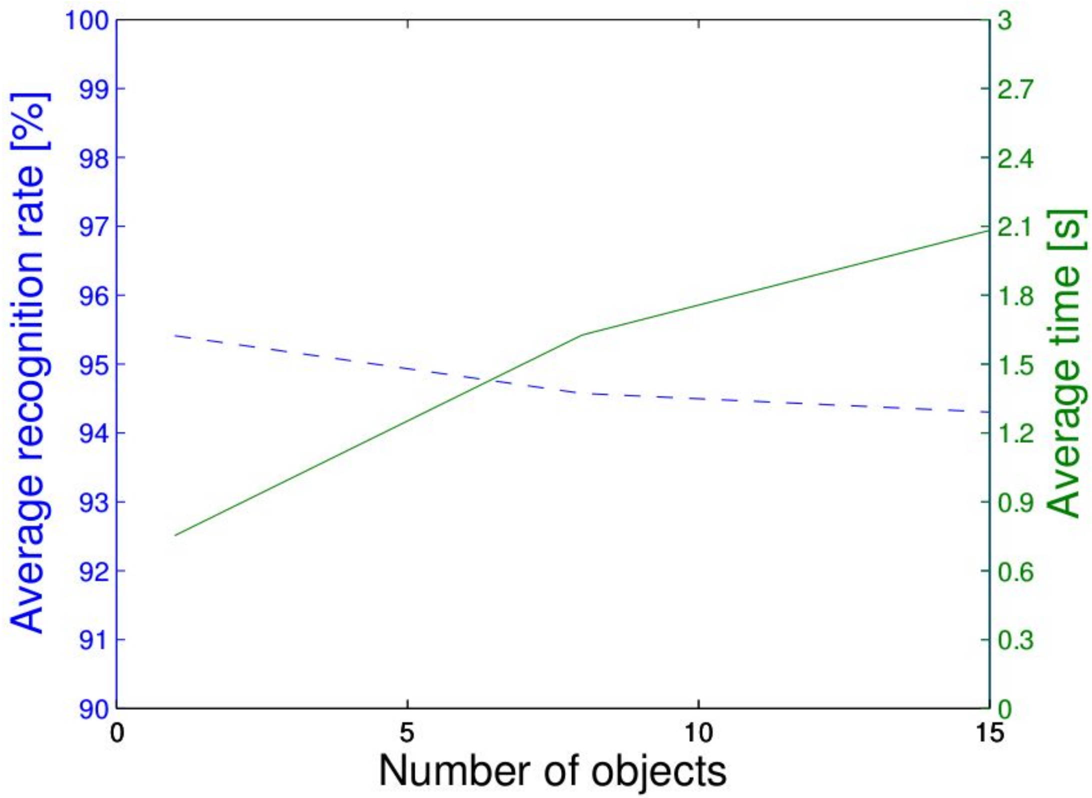


Sample 3D pose estimations

[1] Hinterstoisser et al., “Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes,” ACCV, 2012
[2] Drost et al., “Model globally, match locally: Efficient and robust 3d object recognition,” CVPR, 2010

Scalability and Speed

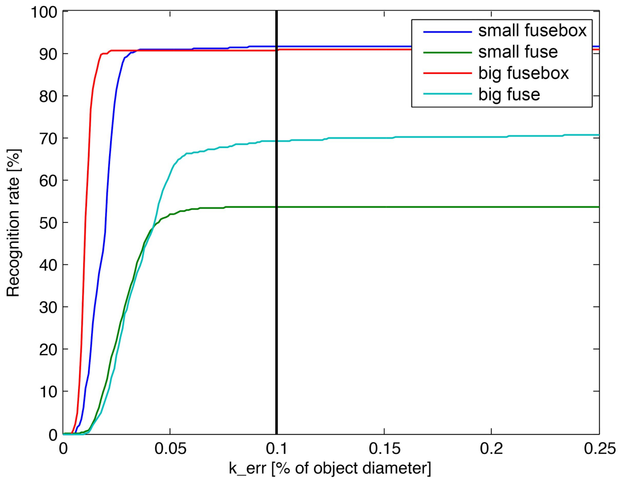
- Time complexity is **sub-linear in the number of templates**
- When the number of loaded templates increased 15 times, the average recognition time increased only less than 3 times:

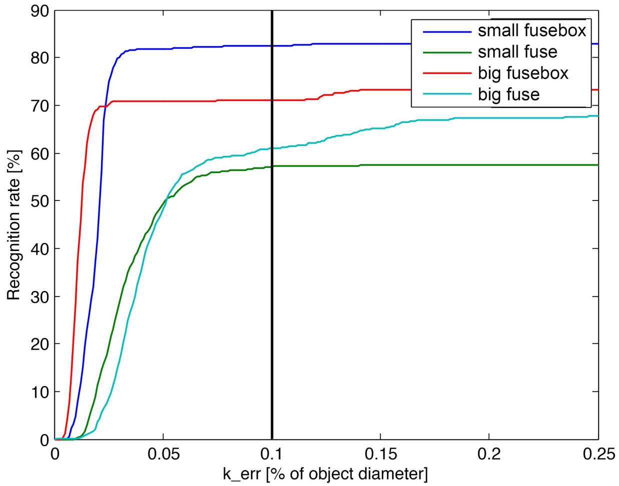


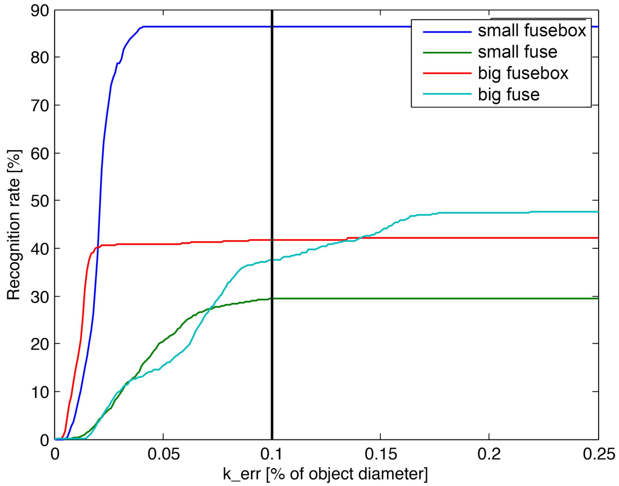
- **0.75 s** per VGA frame (9 image scales) for a single known object

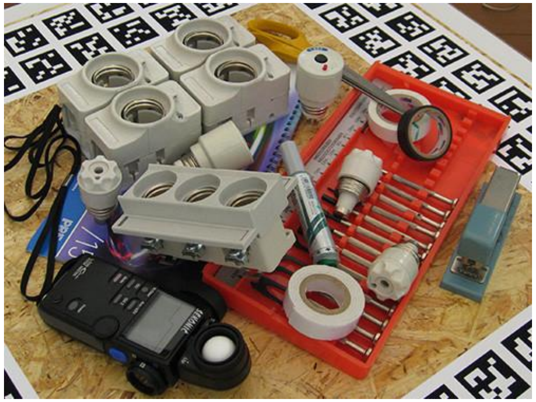
T-LESS: Evaluation on the First 3 Scenes

- Evaluation of **Hodan et al. (IROS 2015)** method
- Hinterstoisser's **average distance (AD) criterion**









Conclusions

1. **T-LESS:** A new industry-relevant RGB-D dataset and evaluation protocol for detection and 6D pose estimation of texture-less objects
   a. Relatively small objects often very similar in shape and color
   b. Significant clutter and occlusions
   c. Accurate GT 6D poses for all known objects
   d. Data from three synchronized and mutually calibrated sensors
2. **Difficulty of the T-LESS dataset** was confirmed in the first evaluation of the method by Hodan et al. (IROS 2015)
3. **Definition of evaluation tasks:**
   a. 6D localization
   b. Detection and 6D localization
4. **New 6D pose distances proposed:**
   a. Surface distance
   b. Corresponding point distance

Thank you!

TP/FP Classification

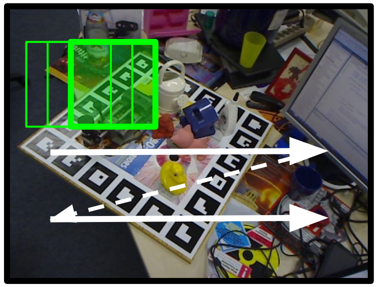
- **Input:**
  - a list ***R*** of tuples *(object class, estimated 6D pose, score)* = an output of the method to be evaluated
  - a list ***G*** of pairs *(object class, ground truth 6D pose)*
- **Output:** TP/FP labeling of the instances from ***R***

- **TP/FP classification algorithm:**
  1. *(only for the 6D localization task)* If there are more than the specified number of instances of some class in the output list ***R***, keep only the ones with the highest score.
  2. From the list ***R*** take the instance with the highest score and compare its pose against ground truth poses of the same class (using the distance ***d*** - the 3D surface distance or the 6D pose distance).
  3. If a match was found (***d* < *th***), classify the estimated pose as a true positive and remove the matched ground truth pose from the list ***G***. Otherwise, classify the estimated pose as a false positive.
  4. Go to step 2.

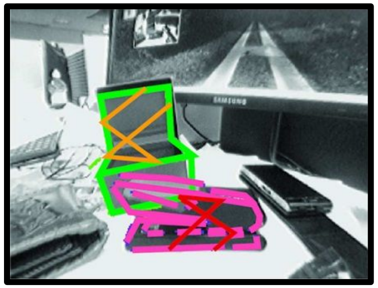
Performance Evaluation

- Calculate “precision vs recall” curve by varying ***th***
- (Calculate the area under the curve)
- ...

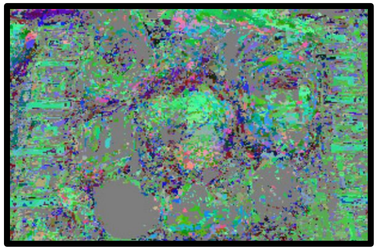
Existing methods



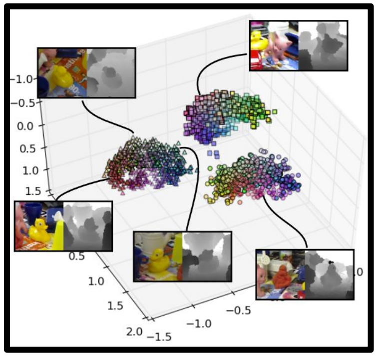
1. **Template matching methods**
   Hinterstoisser (ICCV 2011), Rios-Cabrera (ICCV 2013), Cai (ICVS 2013), Hodan (IROS 2015)



2. **Shape matching methods**
   Damen (BMVC 2012), Tombari (ICCV 2013), Drost (CVPR 2010), Choi (IROS 2012), Hodan (ISMARW 2015)



3. **Methods based on dense features**
   Sun (ECCV 2010), Gall (PAMI 2011), Brachmann (ECCV 2014)



4. **Deep learning methods**
   Wohlhart (CVPR 2015), Held (arXiv 2015), Krull (arXiv 2015)