# Random Forests and Applications

# Object Detection and Pose Estimation

Dr. Rigas Kouskouridas
Computer Vision & Learning Lab
Imperial College London
www.iis.ee.ic.ac.uk/~rkouskou

# Outline

- Object Detection and Pose Estimation
  - What? Why?
  - Challenges
  - How? State-of-the-art

- Latent-Class Hough Forests for 3D Object Detection and Pose Estimation – ECCV 2014

- Concluding remarks - Questions

# Object Detection and Pose Estimation

- What?
  - Detection
    - 2D localisation of objects in form of a bounding box
  - Pose Estimation
    - Full 6 Degrees of Freedom (DoF) estimation w.r.t. a known coordinate system (e.g. camera)

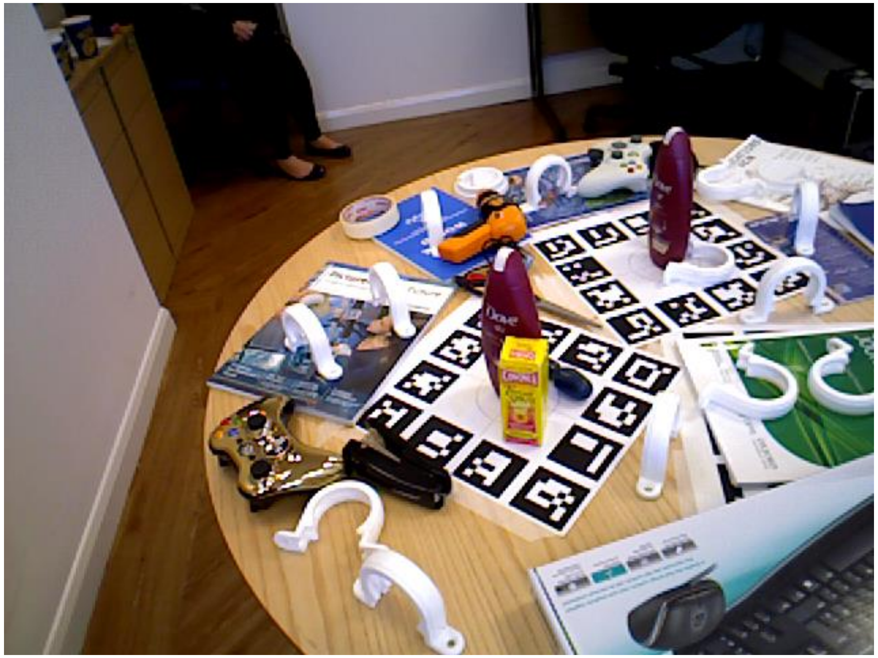

- Why?
  - Robotics…
    - Object manipulation, 3D reconstruction etc.
  - Video surveillance
    - Detect moving objects, abandoned objects

# Challenges

- ❑ Background clutter

# Challenges

- Foreground occlusions



# Challenges

- ❑ Large scale & pose changes

# Challenges

- Multi-instance objects

# How? State-of-the-art

| Template Matching | vs | Part-based Methods |
|---|---|---|
| Holistic representation of the 3D model | | Patch-based multi-voting framework |
| Datasets: CAD, point clouds | | Datasets: Application-dependent |
| Good performance | | Low computational burden |
| Invariant to clutter | | Invariant to occlusions |
| Sensitive to occlusions | | Sensitive to outliers |
| High computational burden | | Fuzzy vote maps |

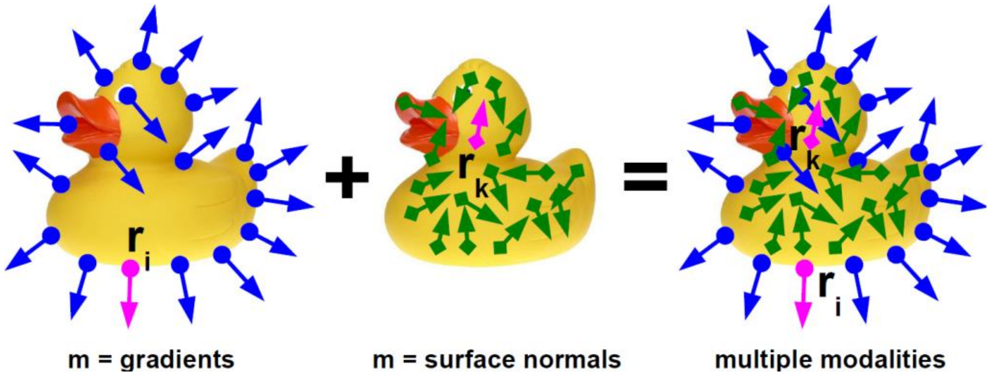
# Latent-Class Hough Forests for 3D Object Detection and Pose Estimation

- Main idea:
  - Integrate Template Matching into Part-based framework
  - Combine state-of-the-art techniques
    - LINEMOD
    - Hough Forests

- Goals:
  - Efficient integration
  - Efficient data split at node levels
  - More challenging scenarios

# Holistic Template Matching – LINEMOD

- Goal:
  - Build holistic 3D descriptors of the model

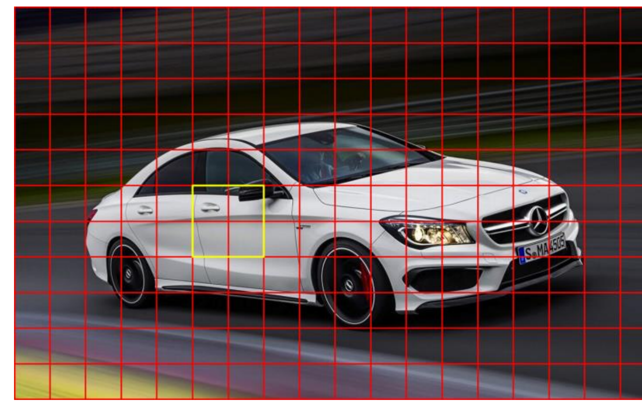



- Modality extraction:
  - RGB cue
  - Depth cue
- Testing:
  - Match extracted model with trained ones (several scales)
  - Similarity measure:
    - Dot product of quantised orientations at different locations

- Indicative references:
  - Hinterstoisser, S. et al. "*Gradient Response Maps for Real-Time Detection of Texture-Less Objects*", PAMI, 2012
  - Hinterstoisser, S. et al. "*Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes*", ACCV 2012

# Part-based Methods – Hough Forests

- Goal:
  - Utilise object parts (spatial information) for classification purposes (foreground/background)
  - Combine spatial information and class information during learning



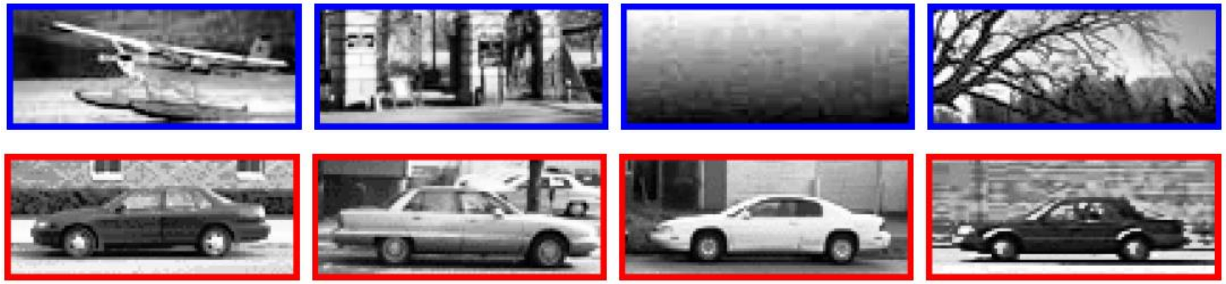
- Patch = (appearance, fg/bg, vote)

- $P_i$=(  , 1, 50 cm from car centroid)

- Indicative references:
  - Gall, J. et al. *“Hough Forests for Object Detection, Tracking, and Action Recognition”*, PAMI, 2011
  - Gall, J and Lempitsky, V. “*Class-Specific Hough Forests for Object Detection*”, CVPR 2009

# Hough Forests in detail…

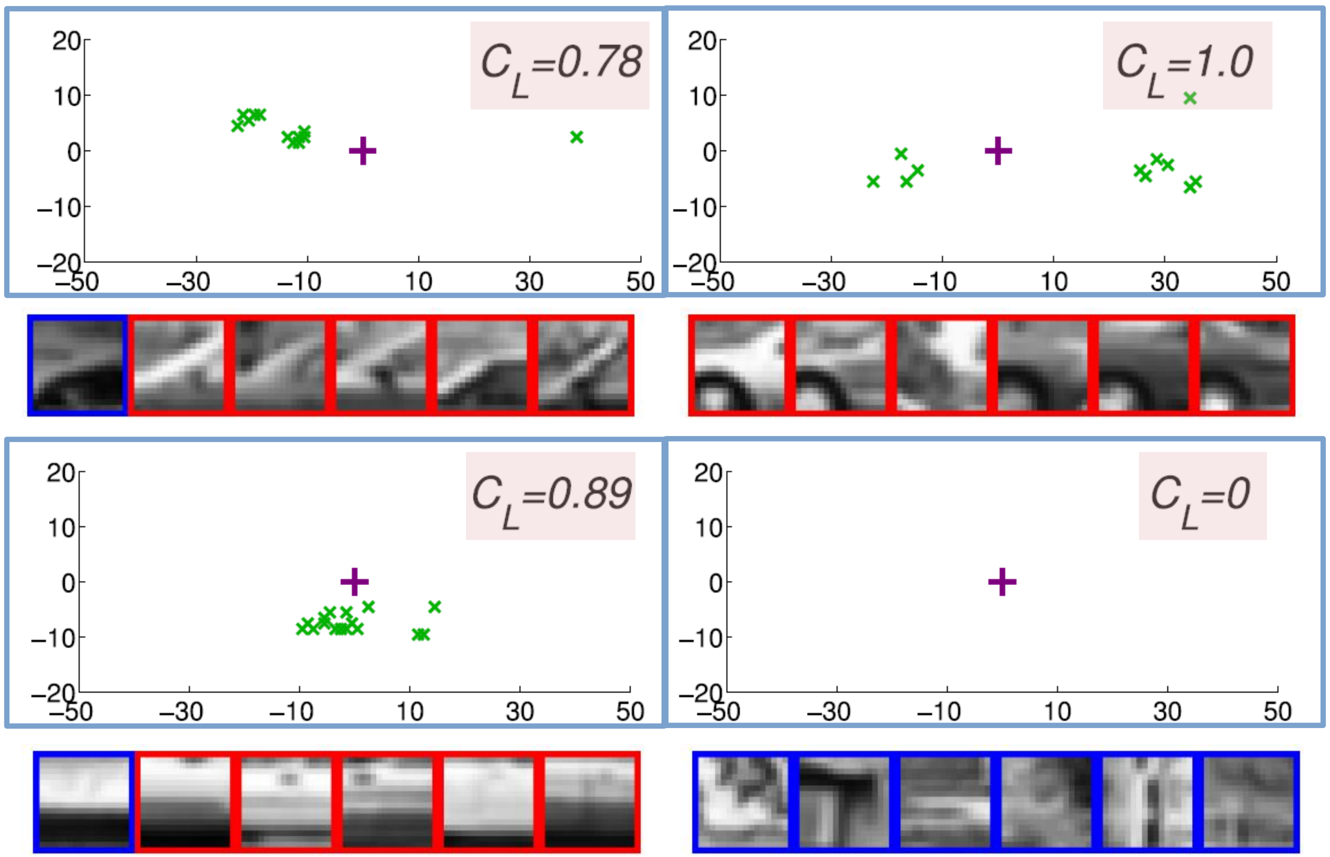
- Training set



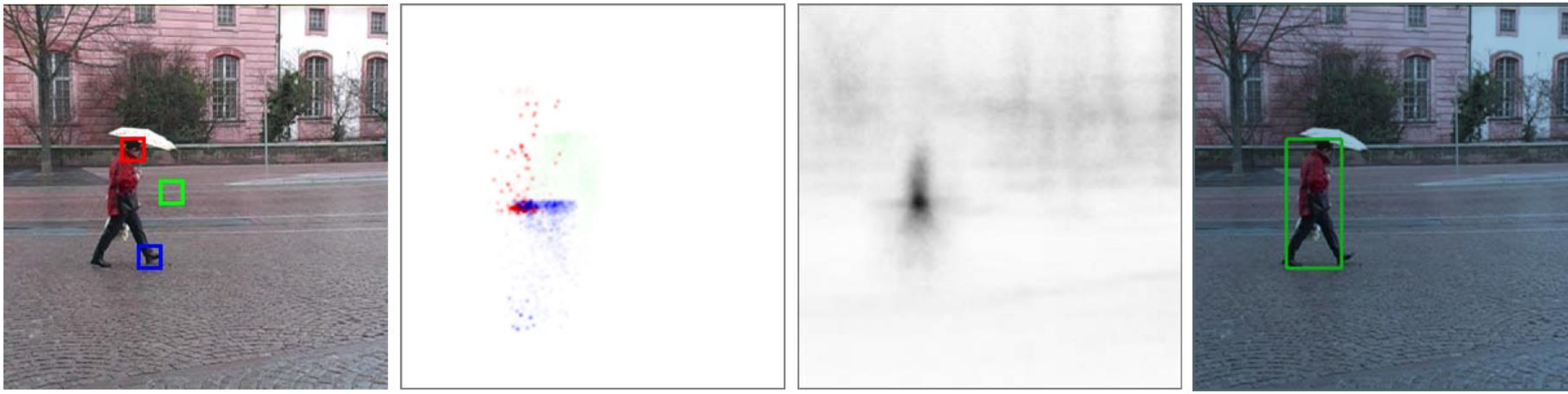
- Measures of uncertainty
  - Class-label uncertainty
    - Impurity of the class labels
  - Offset uncertainty
    - Impurity of the offset vectors

# Hough Forests in detail…

Class probability: $C_L = \dfrac{\left|\{P_i \in L : c_i = 1\}\right| \left|\{P_i \in A : c_i = 0\}\right|}{\left|\{P_i \in L : c_i = 0\}\right| \left|\{P_i \in A : c_i = 1\}\right| + \left|\{P_i \in L : c_i = 1\}\right| \left|\{P_i \in A : c_i = 0\}\right|}$



# Hough Forests results…



*(a)* – Original image with three sample patches emphasized

*(b)* – Votes assigned to these patches by the Hough forest

*(c)* – Hough image aggregating votes from all patches

*(d)* – The detection hypothesis corresponding to the peak in *(c)*

# Latent-Class Hough Forests for 3D Object Detection and Pose Estimation

- Main idea:
  - Hough Forests as patch-based detector
  - LINEMOD as a 3D descriptor for the patches

- Goals:
  - Make LINEMOD scale-invariant (Depth check)
  - Guarantee efficient data split at node levels (Novel split function)
  - Class distributions – latent variables (One-class training)

# Training images – No background

# Training

Set of training patches $\{P_i = (c_i, T_i, D_i, \theta_i)\}$ $\rightarrow$ $(\theta_x, \theta_y, \theta_z, \theta_{ya}, \theta_{pi}, \theta_{ro})$

3D offset from patch centre

$(x_i, y_i)$ central pixel

raw depth map

3 Euler angles

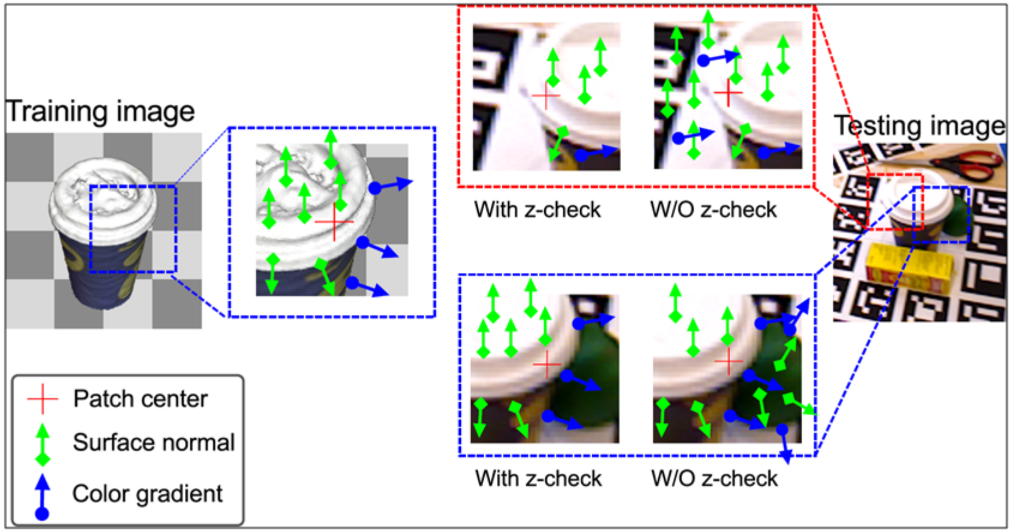
$(\{O_i^m\}_{m \in M}, \Delta_i)$ with $O_i^m$ : aligned reference patches for modality *m*

$\Delta_i$ : discrete set of pairs made up of 2D offsets and modalities

***Split Function***

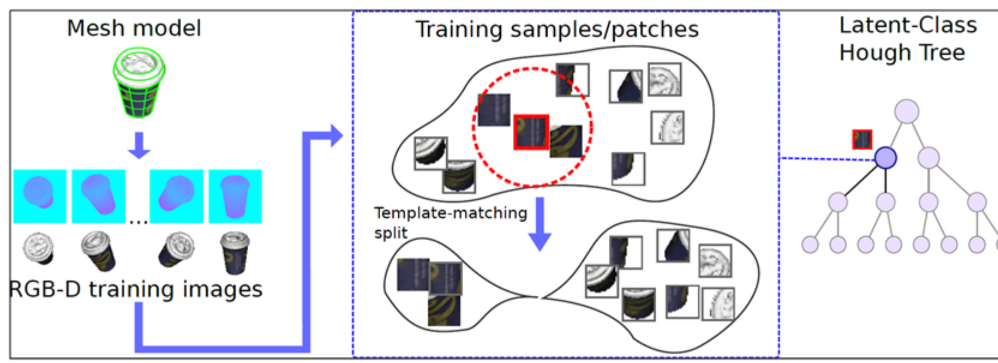
$$\left\{ \begin{array}{c} \varepsilon(P_i, P_j) = \sum_{(r,m)}^{\Delta_i} \left( \max_{t \in R(\varsigma_j(c_j + r))} \omega(P_i, P_j, r) \cdot f_m(\mathrm{O}_i^m(\varsigma_i(c_i + r)), \mathrm{O}_j^m(t)) \right) \\ \omega(P_i, P_j, r) = \delta(\left\| D_i(\varsigma_i(c_i + r)) - D_i(c_i) \right| - \left| D_j(\varsigma_j(c_j + r)) - D_j(c_j) \right\| < z) \\ \varsigma_\chi(c_\chi, r) = c_\chi + \dfrac{r}{D_\chi(c_\chi)} \end{array} \right\}$$

# Split function properties

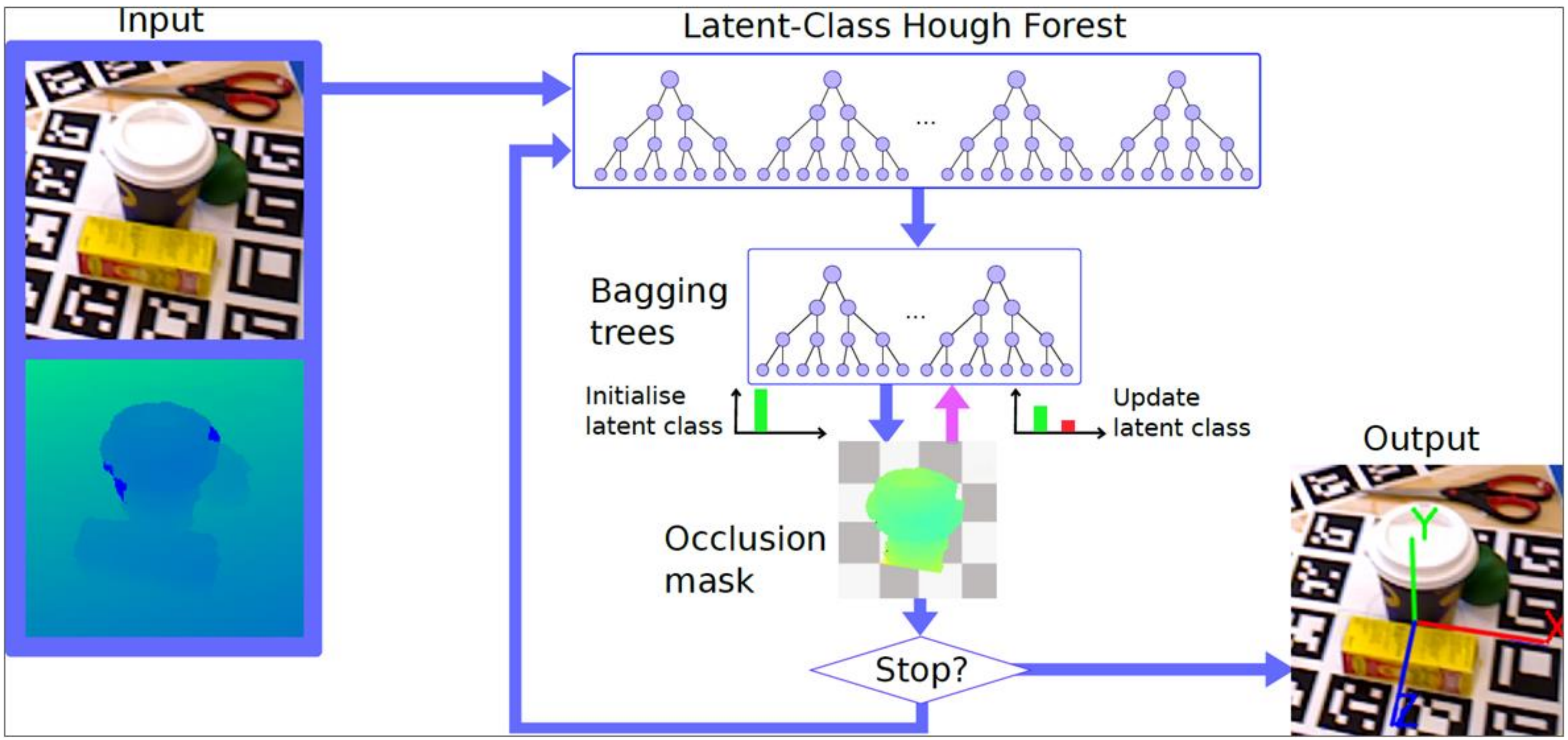



Blue patch: true positive match
Red patch: false positive match

A random patch T (with red frame) is compared with all other patches to produce the correct splits



# Testing…


Input
Latent-Class Hough Forest
...
Bagging trees
...
Initialise latent class
Update latent class
Occlusion mask
Stop?
Output
Y
X
Z

# Testing…

**repeat**

Randomly partition the forest → two partitions (classifiers)

First partition: obtain consensus patch set

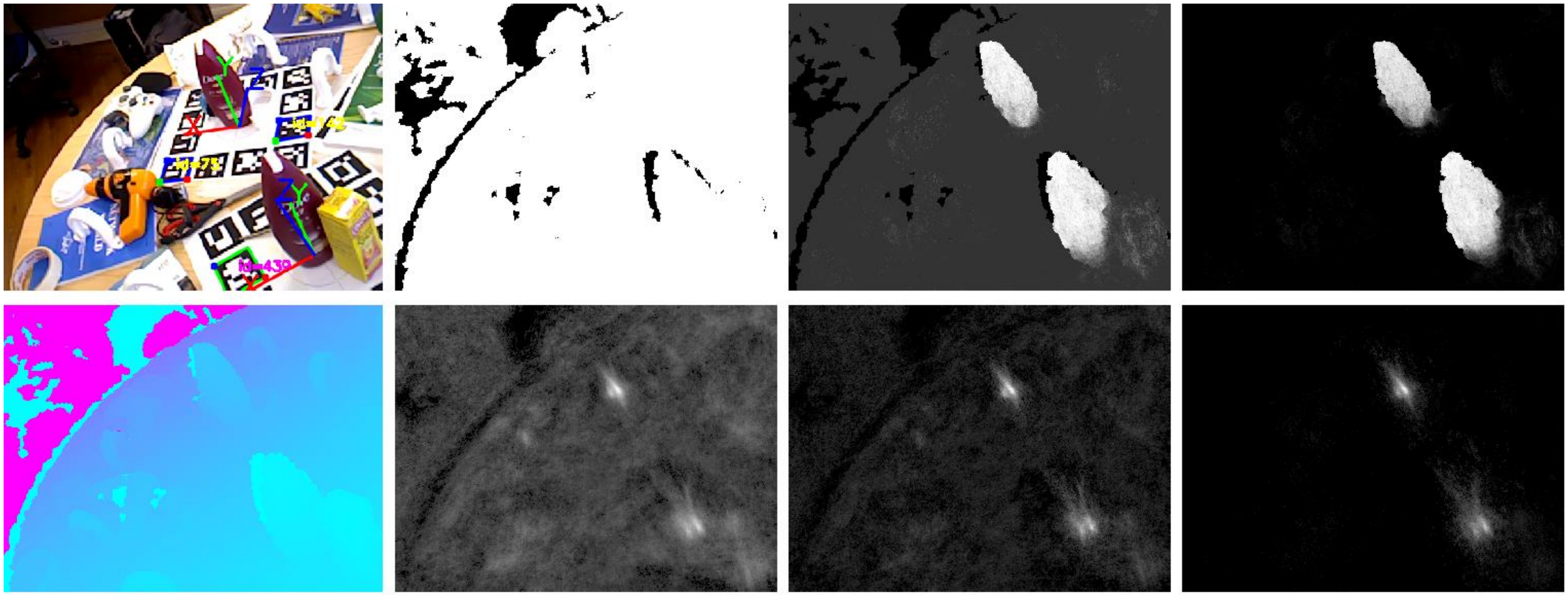
- ❑ Further reduced to consensus pixel set Π (object diameter)
- ❑ All pixels in Π are labelled as foreground and the rest background (two labelled datasets)

Second partition: accumulate patches and update leaf probability distribution

- ❑ Image segmentation mask

**until** *Maximum iteration*

# Testing…



RGB-D images | #iter=1 | #iter=5 | #iter=10

# Evaluation criteria

***Matching score***

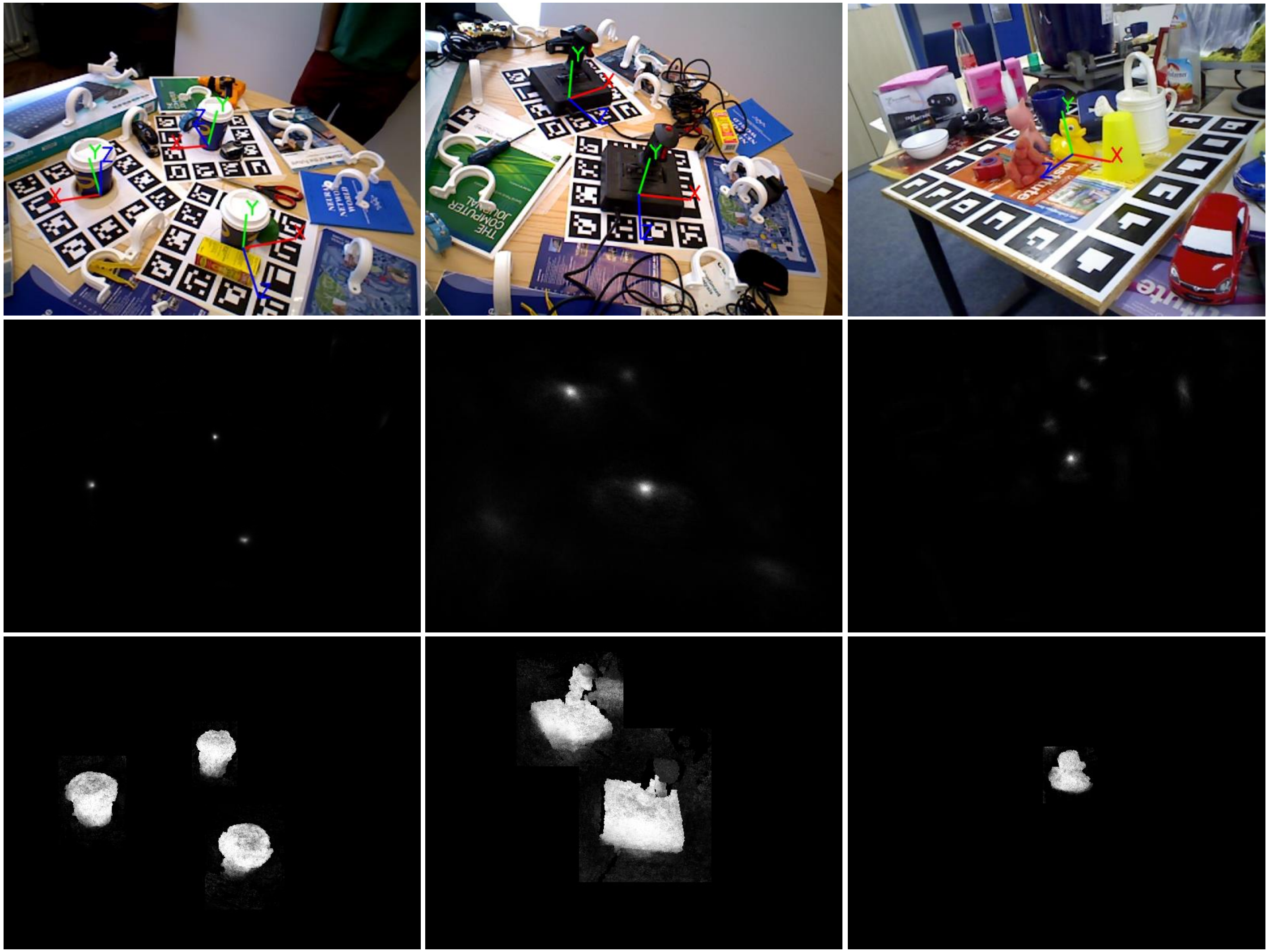
$M$ 3D object model
$R$ ground truth rotation
$T$ ground truth translation
$\hat{R}$ estimated rotation
$\hat{T}$ estimated translation
$k_m$ chosen coefficient
$d$ diameter of object

$$m = \underset{\chi \in \mathrm{M}}{avg} \left\| (R\chi + T) - (\hat{R}\chi + \hat{T}) \right\|$$ Non-symmetric objects

$$m = \underset{\chi_1 \in \mathrm{M}}{avg} \min_{\chi_2 \in \mathrm{M}} \left\| (R\chi_1 + T) - (\hat{R}\chi_2 + \hat{T}) \right\|$$ Symmetric objects

**correct estimation:** $m \le k_m d$

# Results



# Conclusions - Questions

- ❑ Novel object detection and 3D pose estimation technique
- ❑ Random Forest-based framework
  - ❑ Novel split function
- ❑ Robust against occlusions
- ❑ Better than state-of-the-art
- ❑ Extended for robotic bin-picking