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

Alykhan Tejani, Danhang Tang, Rigas Kouskouridas, and Tae-Kyun Kim

Imperial College London

ECCV '14 Zürich, September 6-12th, 2014

This project was supported by the Omron Corporation

Motivation

Holistic Template Matching (Hinterstoisser et al. ACCV 2012)

(+) Invariant to background clutter
(-) Suffer from partial occlusions
(-) Not-tested yet for multi-instance objects

Hough Forests (Gall et al. PAMI 2010)

(+) State-of-the-art patch-based detector
(-) Explicitly exploit classification labels during training

Main Idea & Goals

Integrate Holistic Template Matching into Random Forests framework

Hough Forests as a *patch-based detector* (+) LINEMOD as a *3D descriptor* for the patches

Goals:

(1) Make LINEMOD scale-invariant (Depth check)
(2) Guarantee efficient data split at node levels (Novel split function)
(3) Class distributions are treated as latent variables (One-class training)
(1) + (2) + (3) = Latent-Class Hough Forests

Latent-Class Hough Forests

Training

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

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*

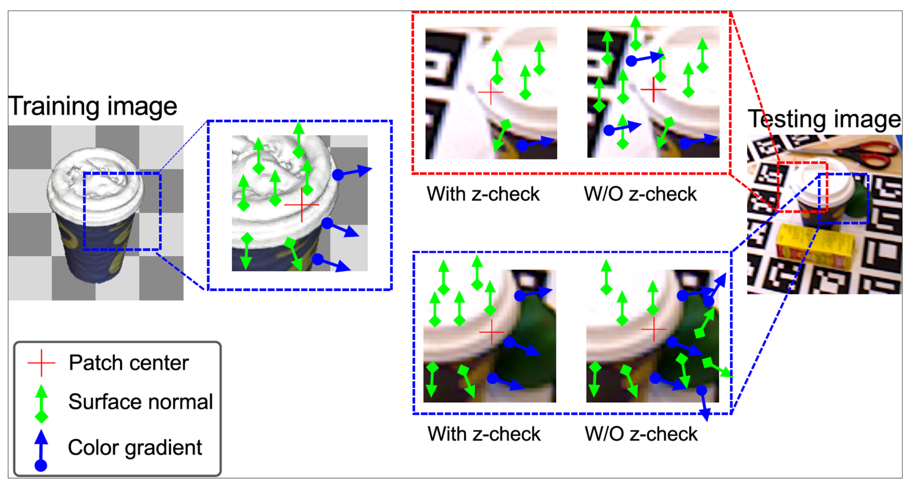
Δ_i : discrete set of pairs made up of 2D offsets and modalities

Split Function

$$\varepsilon(P_i,P_j)=\sum_{(r,m)}^{\Delta_i}\left(\max_{t\in R(\varsigma_j(c_j+r))}\omega(P_i,P_j,r)\cdot f_m(O_i^m(\varsigma_i(c_i+r)),O_j^m(t))\right)$$

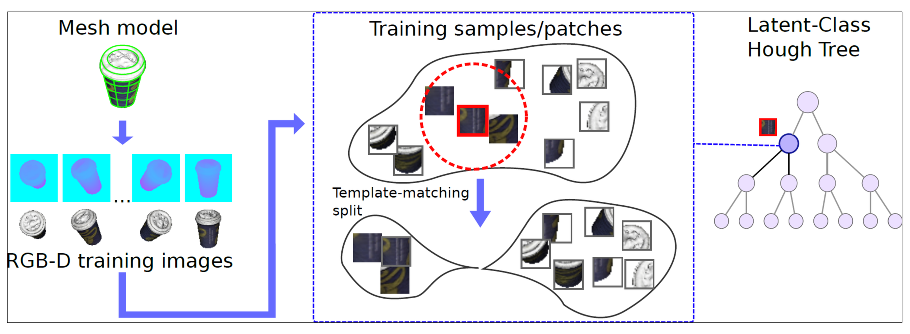
$$\omega(P_i,P_j,r)=\delta(\|D_i(\varsigma_i(c_i+r))-D_i(c_i)|-|D_j(\varsigma_j(c_j+r))-D_j(c_j)\|<z)$$

$$\varsigma_\chi(c_\chi,r)=c_\chi+\frac{r}{D_\chi(c_\chi)}$$

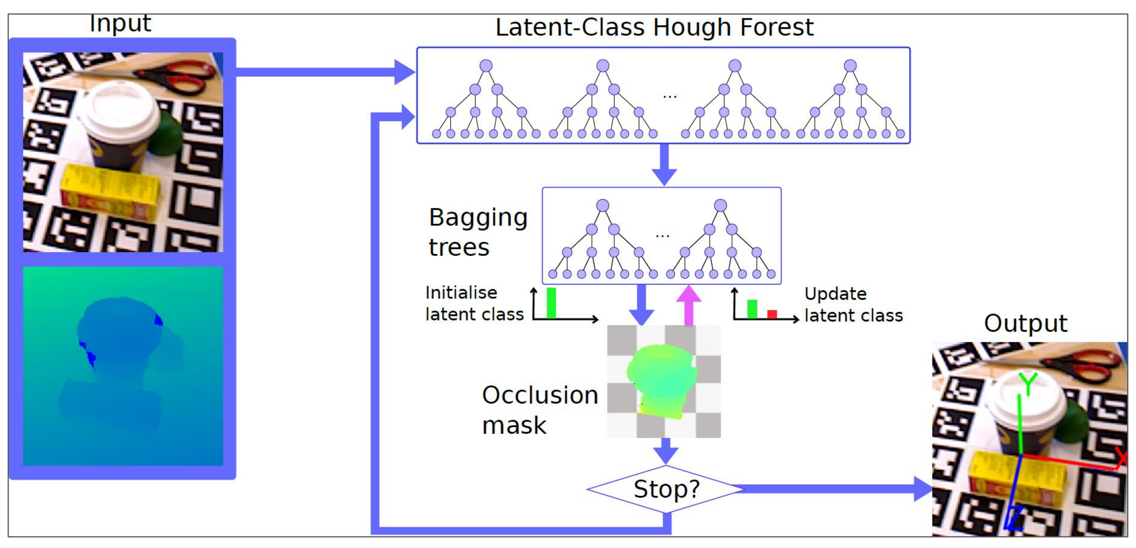


Blue patch: true positive match
Red patch: false positive match

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



Inference



$E(\theta)$ → random event of object's existence under 6 DoF
p_{fg}^l → foreground probability at leaf node l

For a single tree

$$p(E(\theta)\,|\,P;T)=p(E(\theta),p_{fg}^l=1\,|\,P)=\boxed{p(E(\theta),p_{fg}^l=1\,|\,P)}\cdot\boxed{p(p_{fg}^l=1\,|\,P)}$$

$\boxed{p(E(\theta),p_{fg}^l=1\,|\,P)}$ pass patch to forest & accumulate votes at the leaf

$\boxed{p(p_{fg}^l=1\,|\,P)}$ train with positive patches → $p_{fg}^l=1$ all leaf nodes

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)

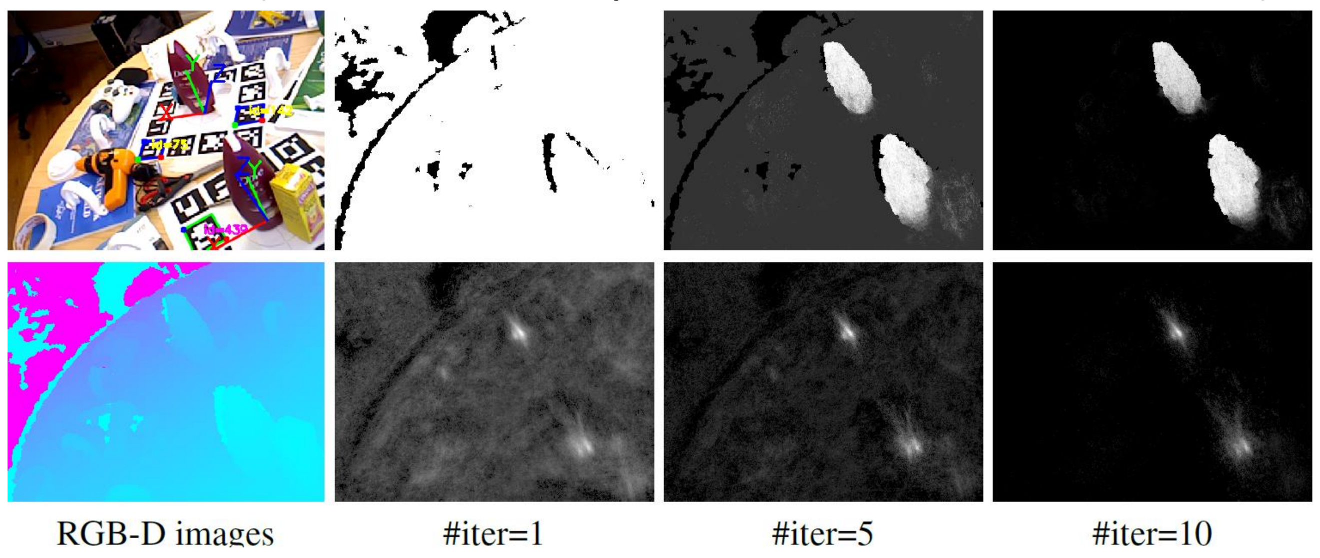
Second partition: accumulate patches and update leaf probability distribution

- Image segmentation mask

until *Maximum iteration*

Key Contributions

- Latent-Class Hough Forests – Novel patch-based approach to 3D object detection and pose estimation
- Joint 3D pose estimation and pixel wise visibility map
- New dataset – *Multi-instance objects, near and far range 2D and 3D clutter, foreground occlusions* (available at http://www.iis.ee.ic.ac.uk/icvl/)



Experimental Results

Matching score

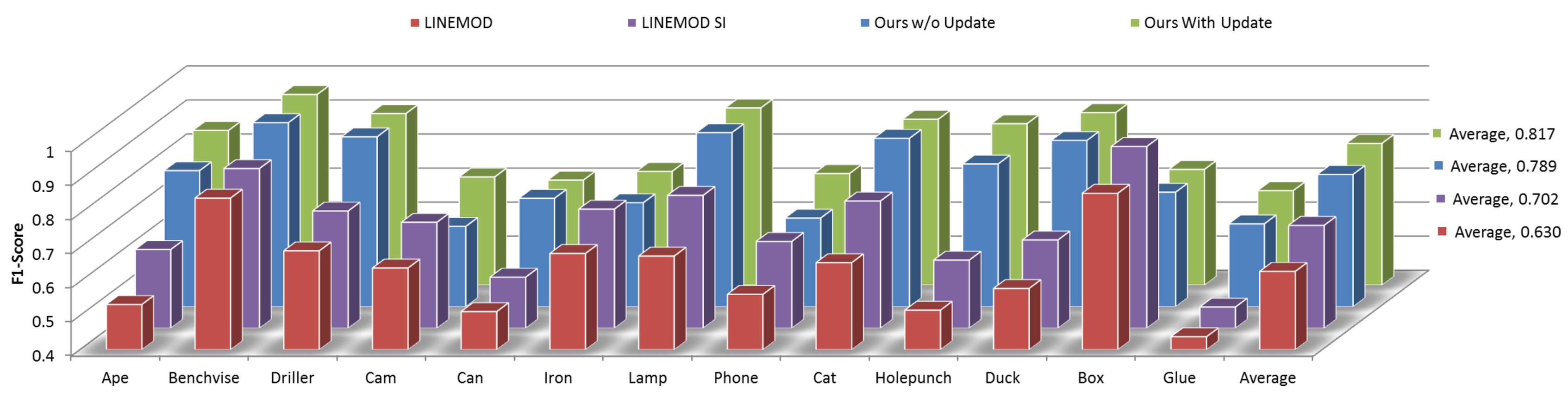
M 3D object model
R ground truth rotation
T ground truth translation
$\hat{R}$ estimated rotation
$\hat{T}$ estimated translation
k_m chosen coefficient
d diameter of object

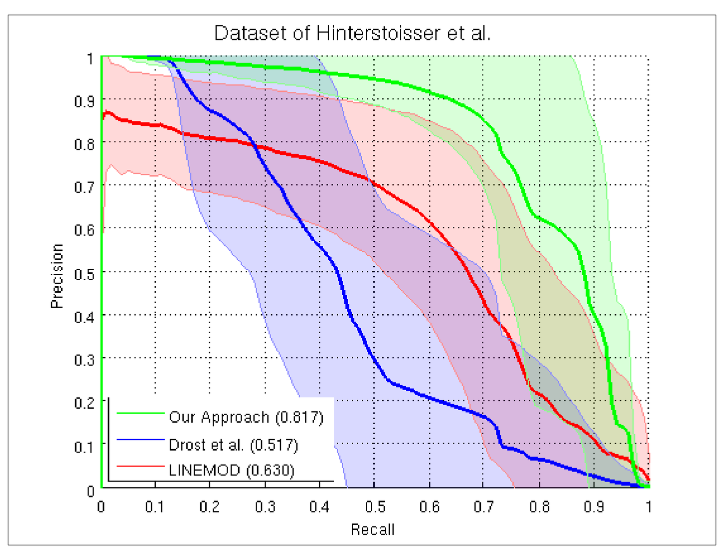
$$m=\underset{\chi\in M}{avg}\left\|(R\chi+T)-(\hat{R}\chi+\hat{T})\right\|$$ Non-symmetric objects

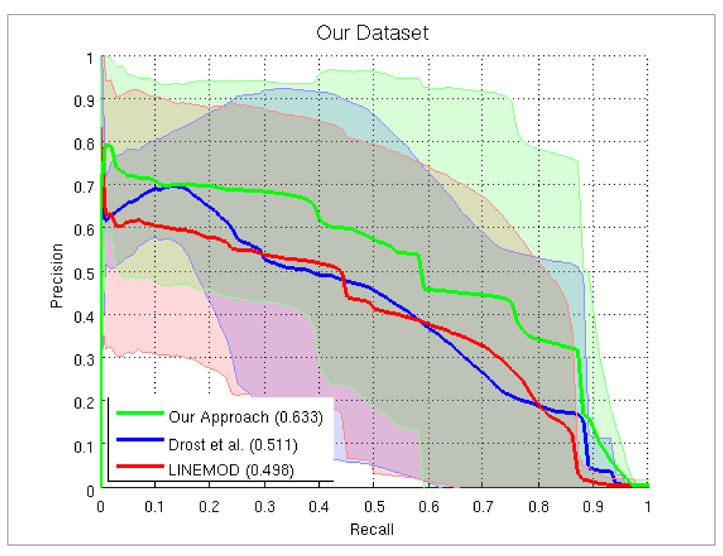
$$m=\underset{\chi_1\in M}{avg}\min_{\chi_2\in M}\left\|(R\chi_1+T)-(\hat{R}\chi_2+\hat{T})\right\|$$ Symmetric objects

correct estimation: $m\le k_m d$

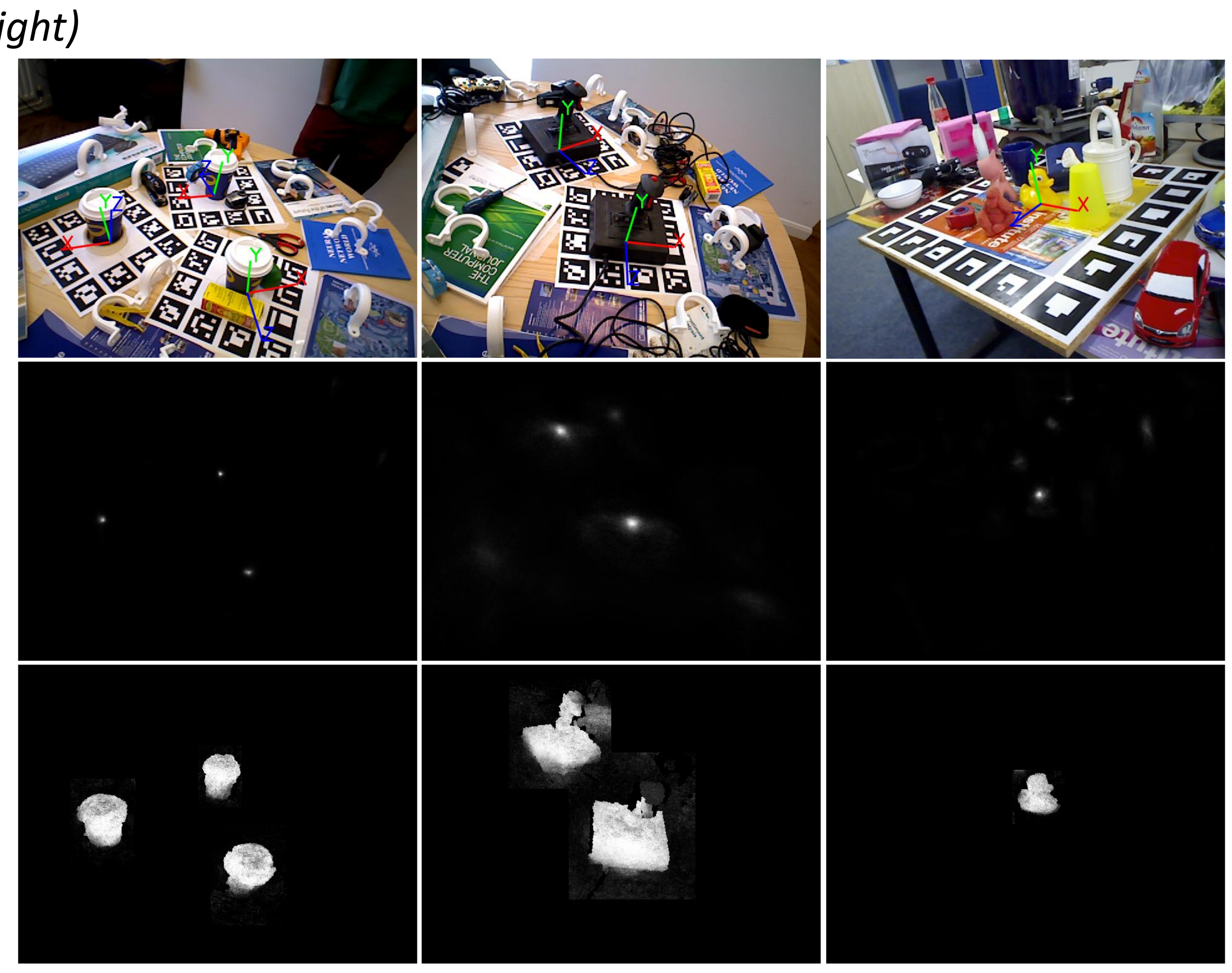
F1-Scores for the 13 objects in the dataset of Hinterstoisser et al.







Average Precision-Recall curves over all objects in the dataset of LINEMOD (left) and our dataset (right)



Augmented 3D axis

Vote map

Segmentation mask