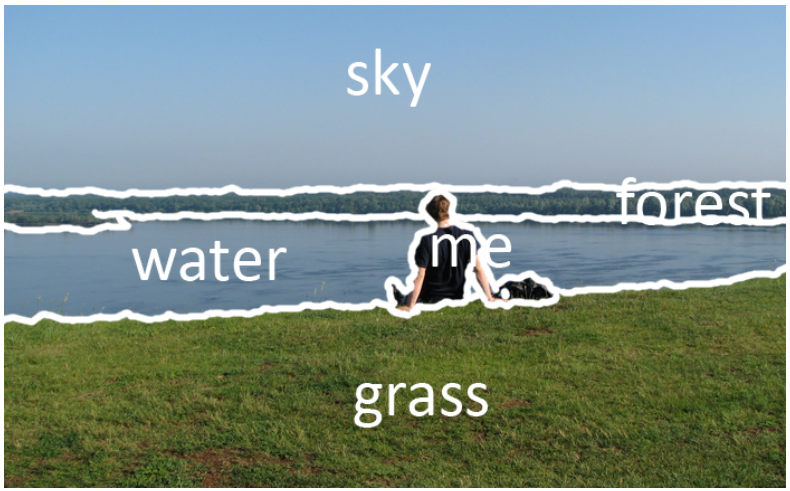
Object Detection with Occlusion

Josef Schulz

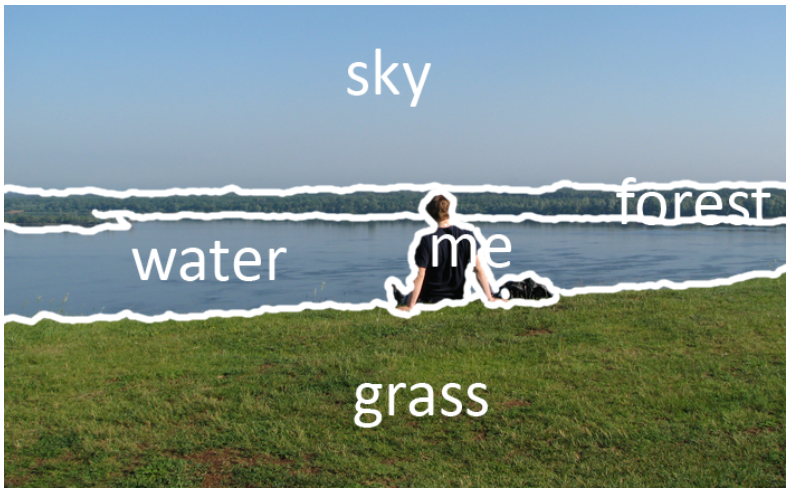
April 7, 2016

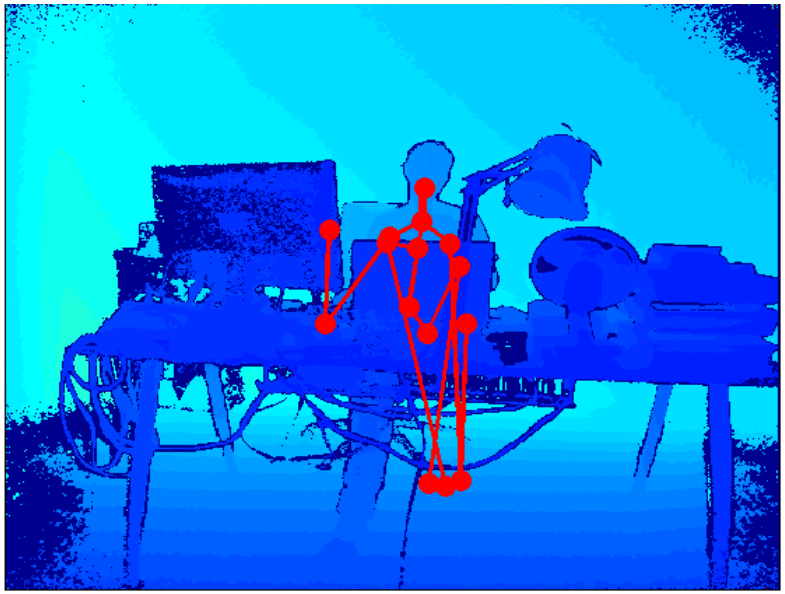
Example Problems

Example Problems

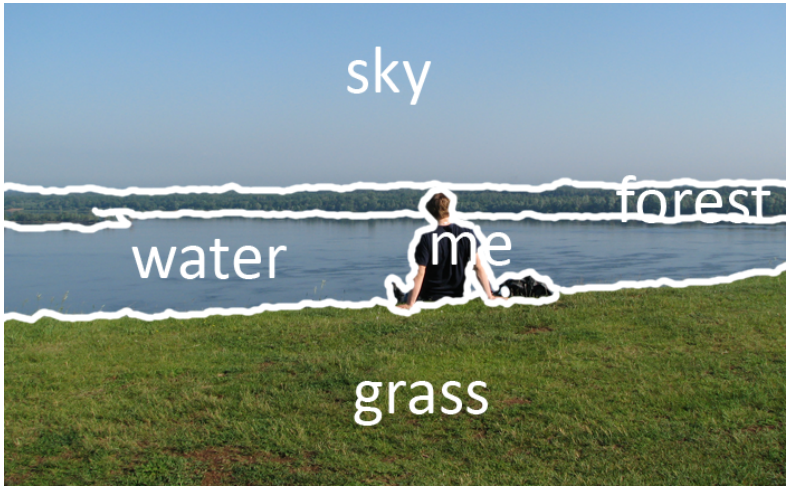


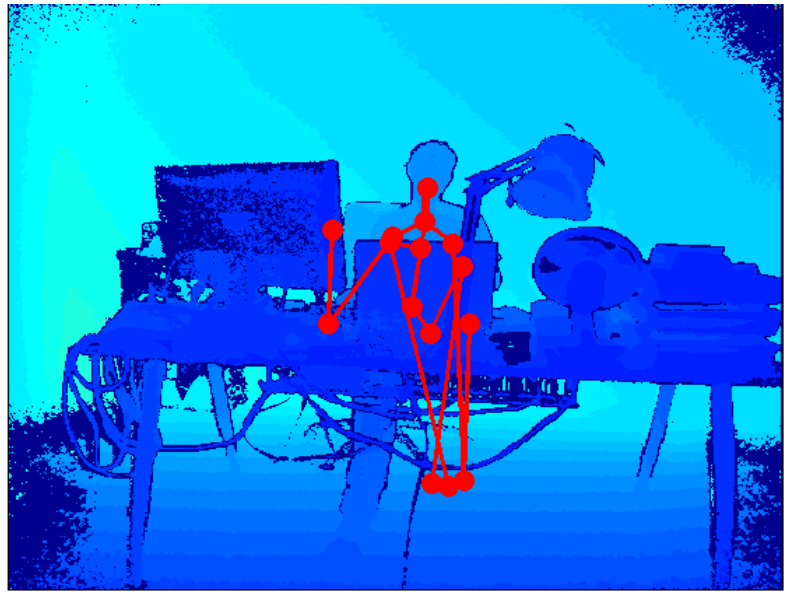
Example Problems





Example Problems





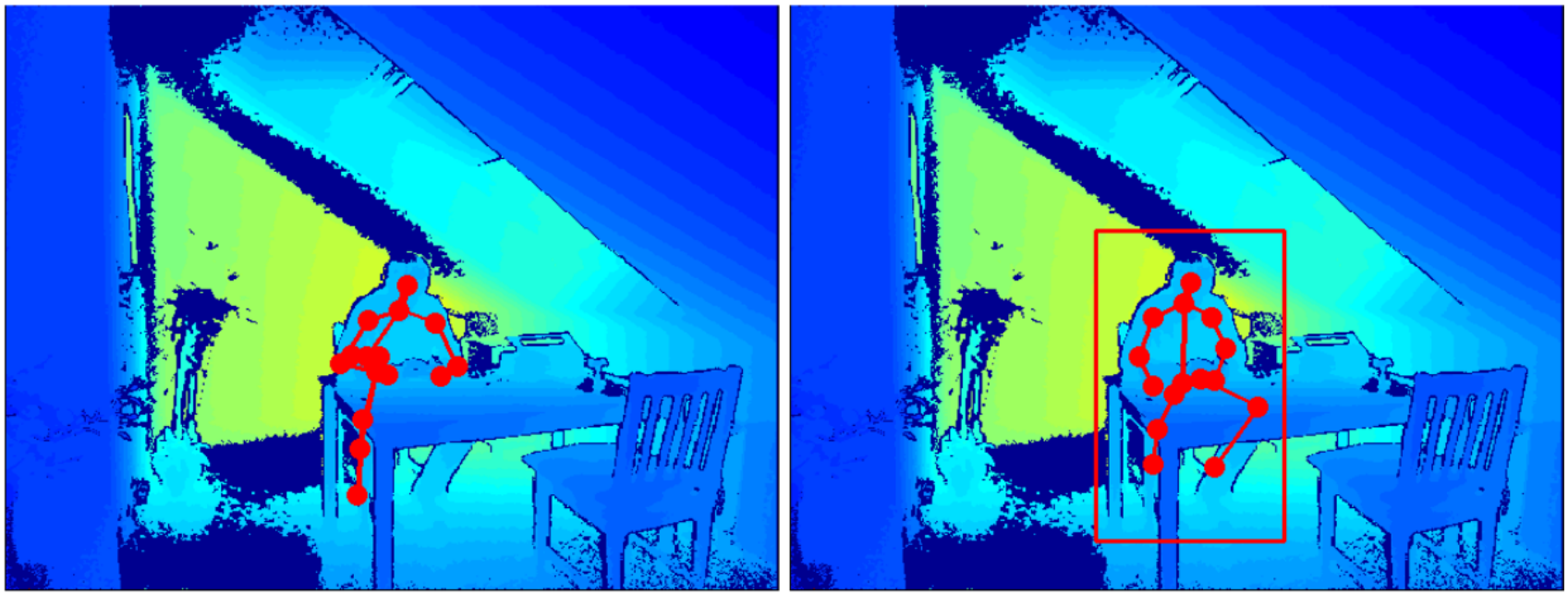
Content

A Semantic Occlusion Model for Human Pose Estimation

Input : Single Depthimage

Output : Estimated Poses of all Parts

3D Human Pose Estimation



Kinect2 SDK

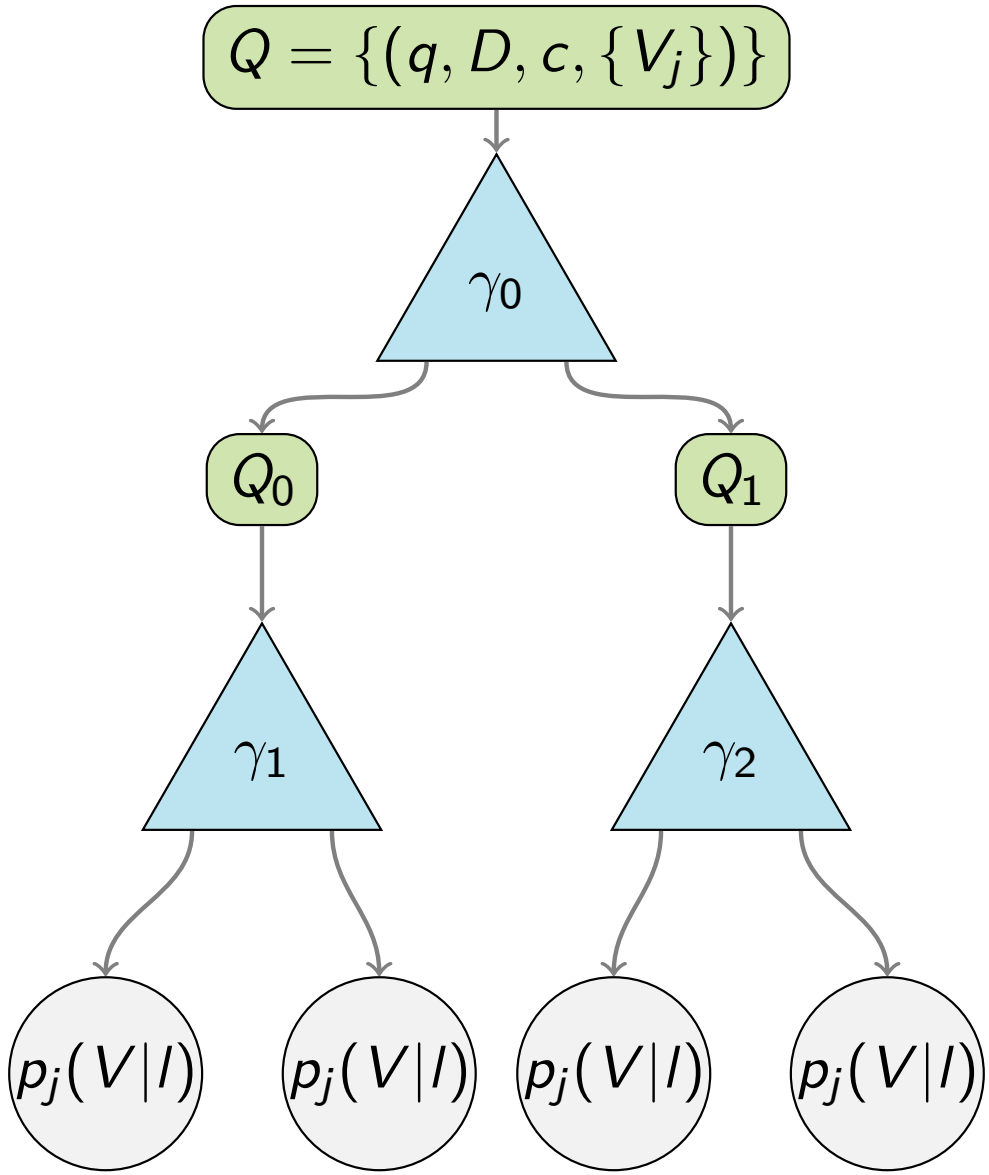
OARF

Trainingsset

$$\{(q, D, c, \{V_j\}), ...\}$$

- q is a pixel position.
- D reference image x.
- c class label, according to a limb
- $\{V_j\}$ is set of vectors $V_j = q_j - q$

Regression Tree



slitnode

$$\Phi_\gamma(q, D) \mapsto \{0, 1\}$$

$$\Phi_\gamma(q, D) = \begin{cases} 1 & \text{if } D(q + \frac{u}{D(q)}) - D(q + \frac{v}{D(q)}) > \tau \\ 0 & \text{else} \end{cases}$$

$$\gamma = (u, v, \tau)$$

choose splitting function

$$\Phi^* = \arg\max_{\Phi} g(\Phi)$$

$$g(\Phi) = H(Q) - \sum_{s \in \{0,1\}} \frac{|Q_s(\Phi)|}{|Q|} H(Q_s(\Phi))$$

$$H(Q) = -\sum_{c} p(c|Q) \log(p(c|Q))$$

leafnode

A leafnode l stores $p_j(V, l)$

V_j is clustered by mean-shift with Gaussian Kernel with bandwidth b

$$p_j(V|l) \propto \sum_{k \in K} w_{ljk} \cdot \exp(-\| \frac{V - V_{ljk}}{b} \|_2^2)$$

w_{ljk} is determined by offset vectors ended in the cluster k.

Pose Estimation

a set of pixels q are sampled in D

$$x_j = q + V_{ljk}$$
$$w_j = w_l jk \cdot D^2(q)$$
$$X_j = \{(x_j, w_j)\}$$

$$p_j(x|D) \propto \sum_{(x_j, w_j) \in X_j} w_j \cdot \exp(-\| \frac{x - x_j}{b_j} \|_2^2)$$

Votes are clustered and only the cluster with the highest summed weights w_j are used for prediction.

Occlusion Aware Regression Forests

$$Q_{ext} = Q \cup Q_{occ} with$$
$$Q_{occ} = \{(q_{occ}, D, c_{occ}, \{v_{jocc}\})\}$$

$$p_j(V_{occ}|I) \propto \sum_{k \in K} w_{ljk} \cdot \exp(-\| \frac{V_{occ} - V_{ljkocc}}{b} \|_2^2)$$

$$X_j = \{(x_j, w_j)\} \cup \{(x_{jocc}, w_{jocc})\}$$

Training Data

Synthetic Data (552 images)

- Human Poses from CMU-Database
- body part labels for each pixel

Real Data (552 images)

- Kinect2 SDK, all fails are discarded

Regression Forest

- 3 regression trees
- max depth $= 20$
- b $= 0.05$m
- 1000 samples per image
- samples 2000 splitting functions per node

Occlusion Patterns for Object Class Detection

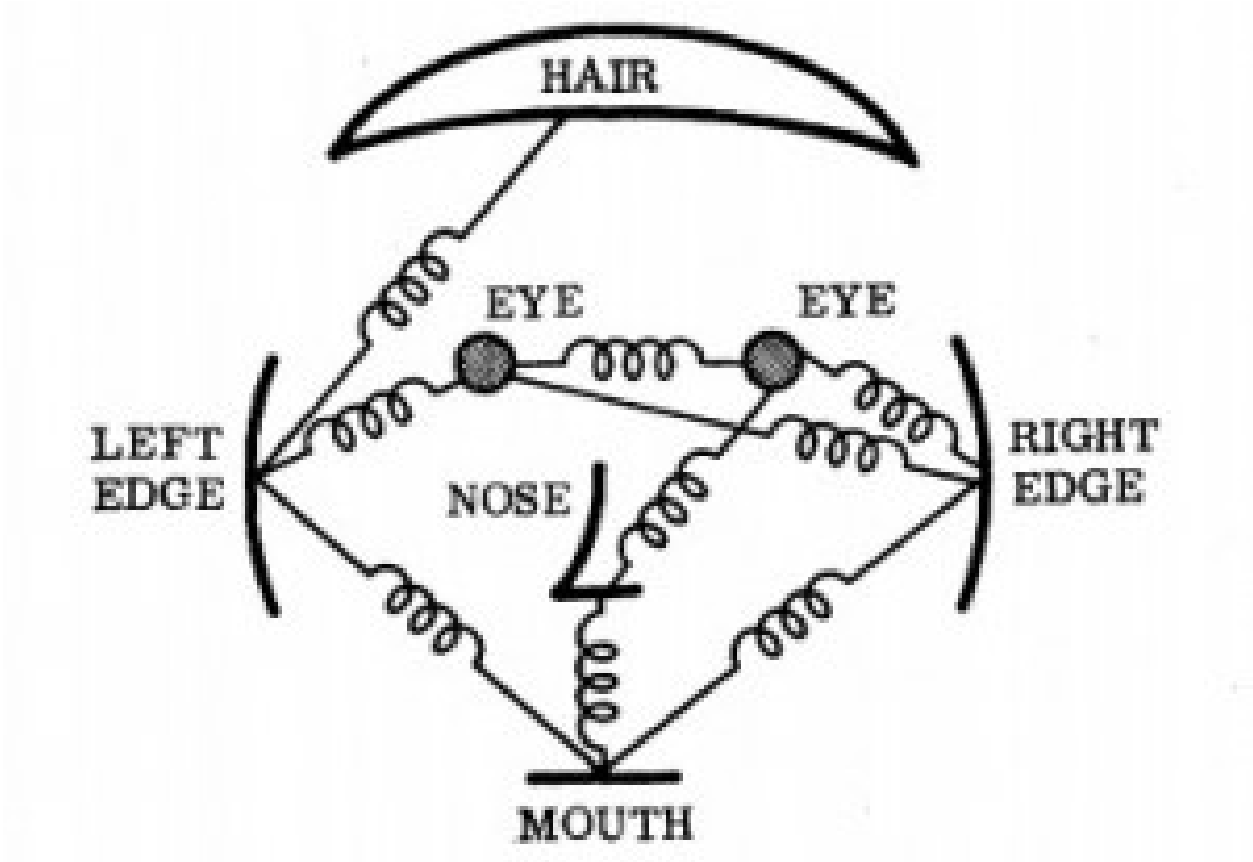
Input : Single RGB-Image

Output : Object-Boundingboxes

Deformable Models approch

- Consider each object as a deformed version of a template
- Compact representation

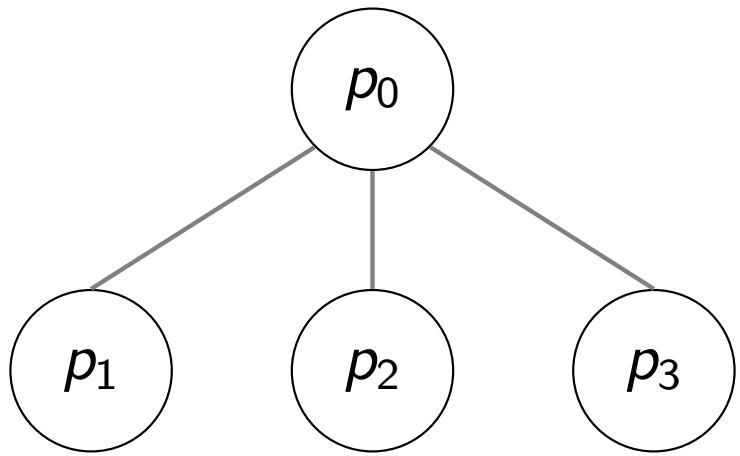
Matching model to image involves joint optimization of part locations "stretch and fit"



Model

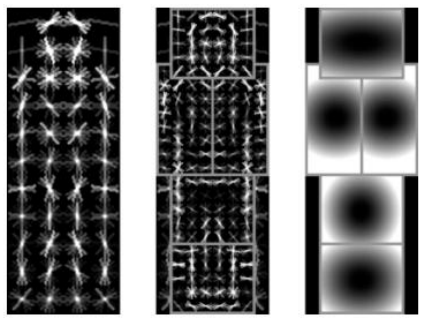
Model is represented by a Graph $G = (V, E)$

- $V = p = \{p_0, ..., p_M\}$ are the parts
- p_i is parameterized through their bounding box (l_i, r_i, t_i, b_i)
- $(p_i, p_j) \in E$



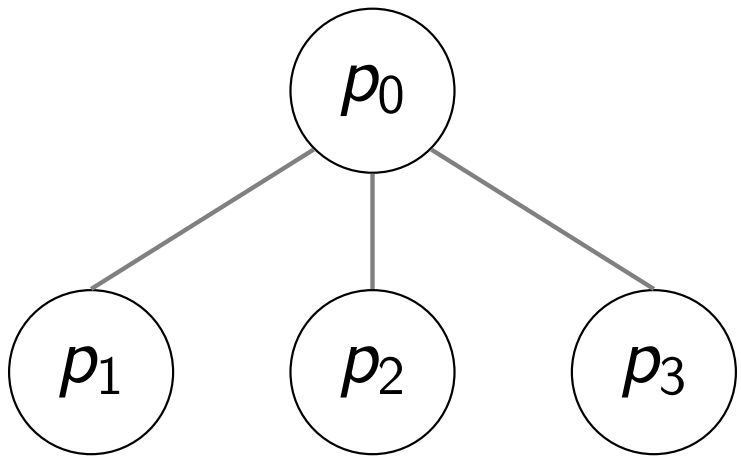
$$E_c(p; I) = \sum_{i=0}^{M} \langle v_i^c, \Phi(p_i; I) \rangle + \sum_{i=1}^{M} \langle w_i^c, \Phi(p_0, p_i) \rangle$$

Training



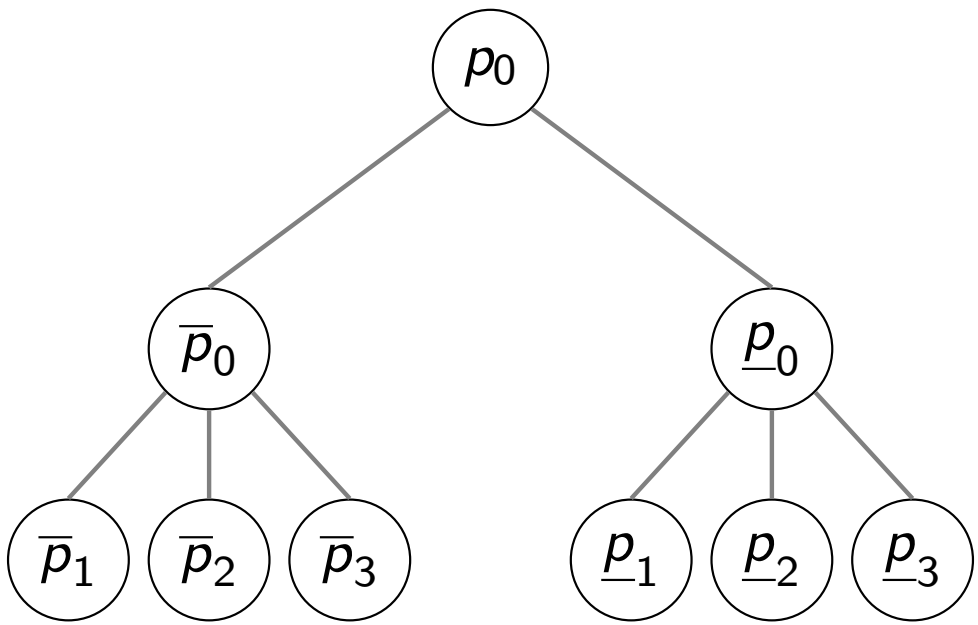
OC-DPM

$$C = \{1, ..., C_{visible}\} \cup C_{invisble}$$

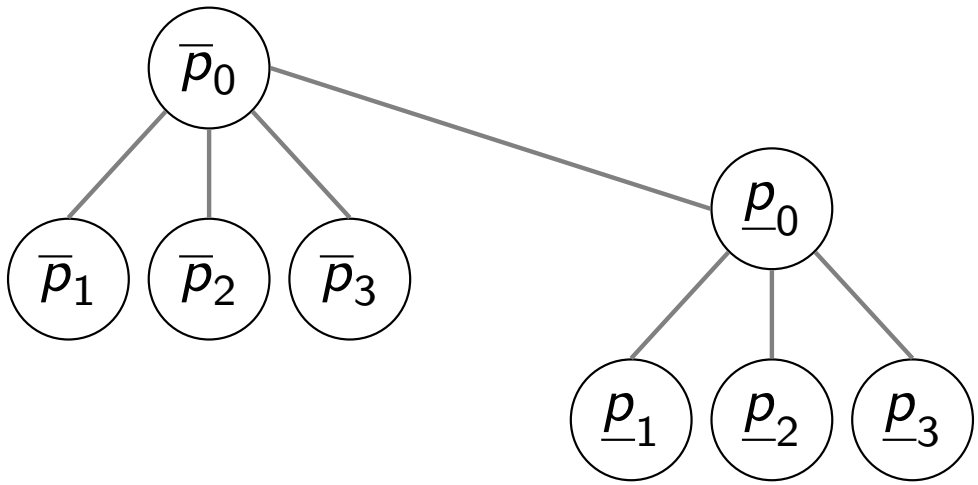


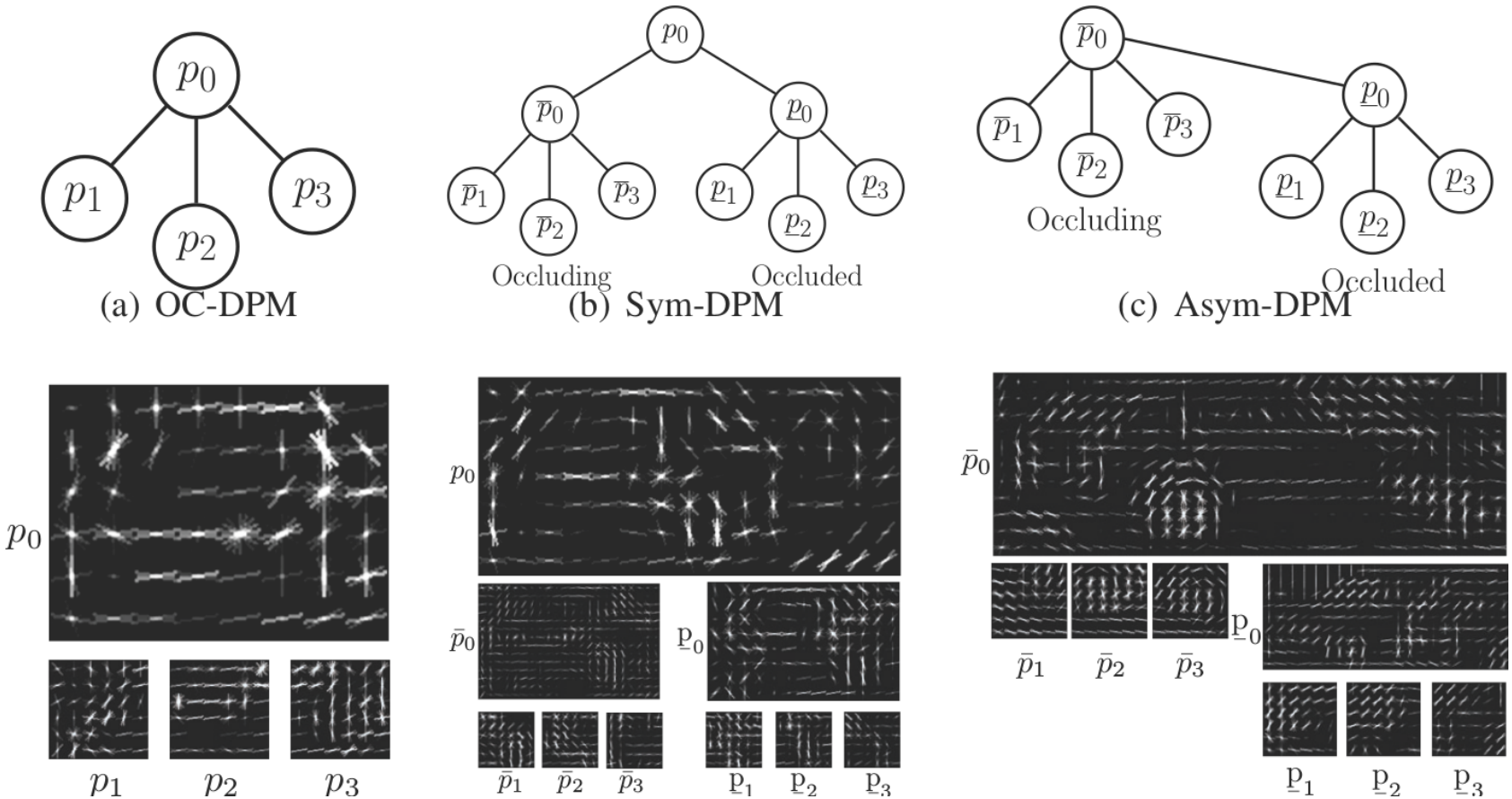
SYM-DPM

$$E'_c(p; I) = \langle v^c_{joint}, \Phi(p_0; I)\rangle + \langle \overline{w}, \Phi(\overline{p}_0, p_{joint})\rangle + \langle \underline{w}, \Phi(\underline{p}_0, p_{joint})\rangle$$
$$+ E_c(\overline{p}_0; I) + E_c(\underline{p}_0; I)$$



ASYM-DPM


$\overline{p}_0$
$\overline{p}_1$
$\overline{p}_2$
$\overline{p}_3$
$\underline{p}_0$
$\underline{p}_1$
$\underline{p}_2$
$\underline{p}_3$


p_0
p_1
p_2
p_3
(a) OC-DPM
$\overline{p}_0$
$\underline{p}_0$
$\overline{p}_1$
$\overline{p}_2$
$\overline{p}_3$
$\underline{p}_1$
$\underline{p}_2$
$\underline{p}_3$
Occluding
Occluded
(b) Sym-DPM
Occluding
Occluded
(c) Asym-DPM

KITTI

KITTI contains 7481 images

	#objects	#occluded objects	%
Car	28521	15231	53.4
Pedest.	4445	1805	40.6
Cycles	1612	772	44.5

components:

visible 6

occluded $16 - 15$

Mining Trainingsdata

Feature Space:

i occluder left/right of occludee

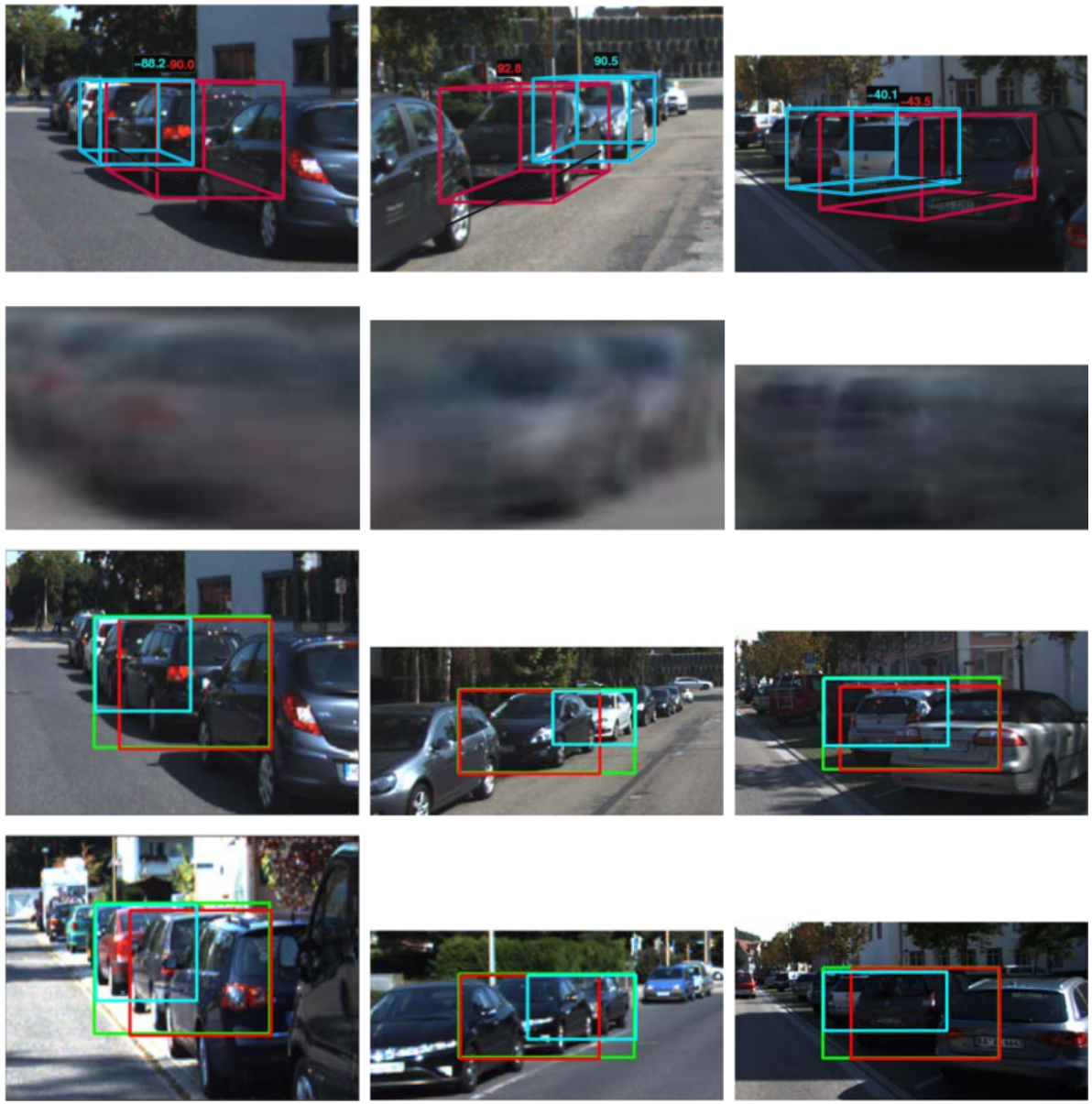
ii orientation of occluder/occludee

iii occluder is/is not occluded

iv degree of occlusion of occludee

Rule-based clustering

Mining Trainingsdata

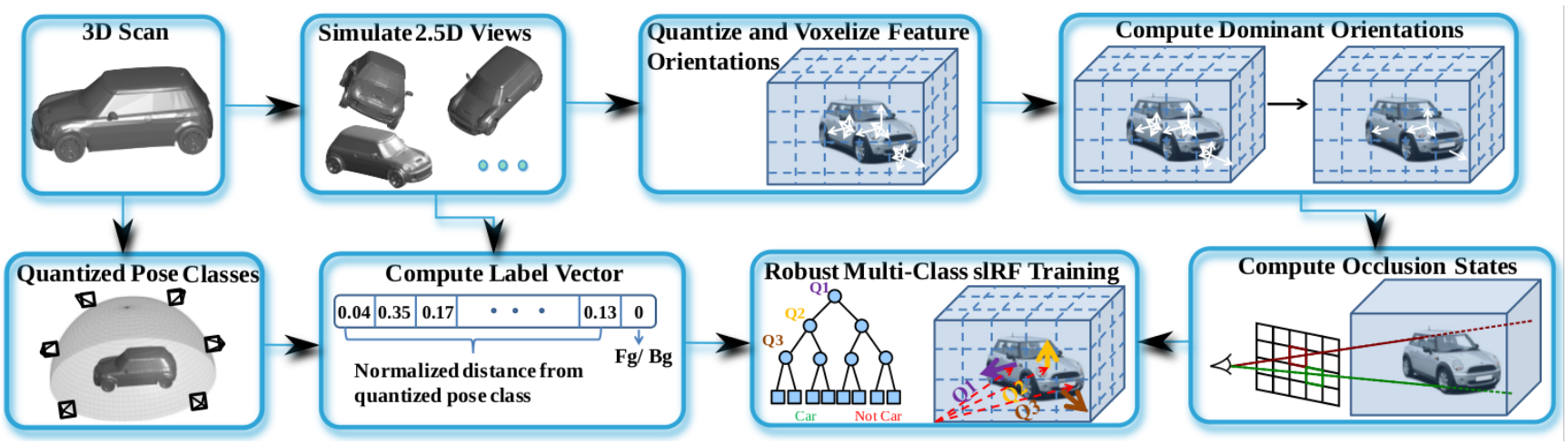


Robust Instance Recognition in Presence of Occlusion and Clutter

Input : 5-10 consecutive frames as one Pointcloud

Output : 6D-Object-Pose

Overview


3D Scan
Simulate 2.5D Views
Quantize and Voxelize Feature Orientations
Compute Dominant Orientations
Quantized Pose Classes
Compute Label Vector
0.04
0.35
0.17
0.13
0
Normalized distance from quantized pose class
Fg/ Bg
Robust Multi-Class slRF Training
Q1
Q2
Q3
Car
Not Car
Compute Occlusion States

Edgelet

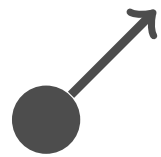
N points per Pointcloud j

FOR ALL $i \in \{1, ..., N\}$

calc λ_1 and λ_2

$r = \frac{\lambda_1}{\lambda_2}$

$r \rightarrow$ curvatureMap



hysteresisThesholding(curvatureMap);

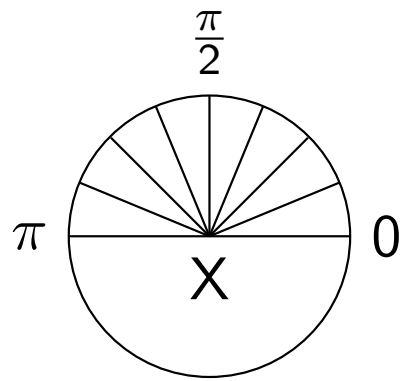
nonmaximalSuppression(curvatureMap);

hysteresisThesholding(depthMap);

projectToPointcloud(curvatureMap, depthMap);

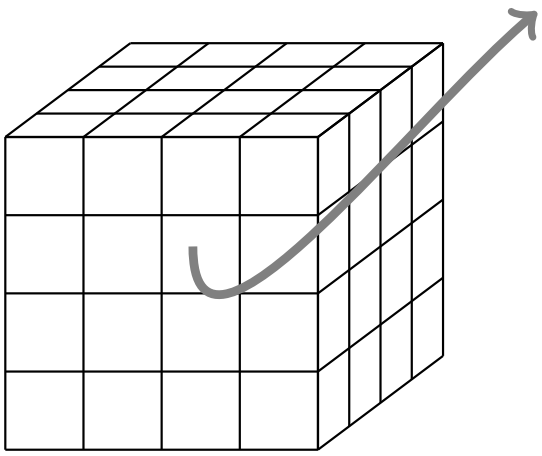
RANSAC line fitting

orientation to 8 bins// (direction%π)



Feature Vector

$w \times h \times d$



Voxel v

1	2	...	8

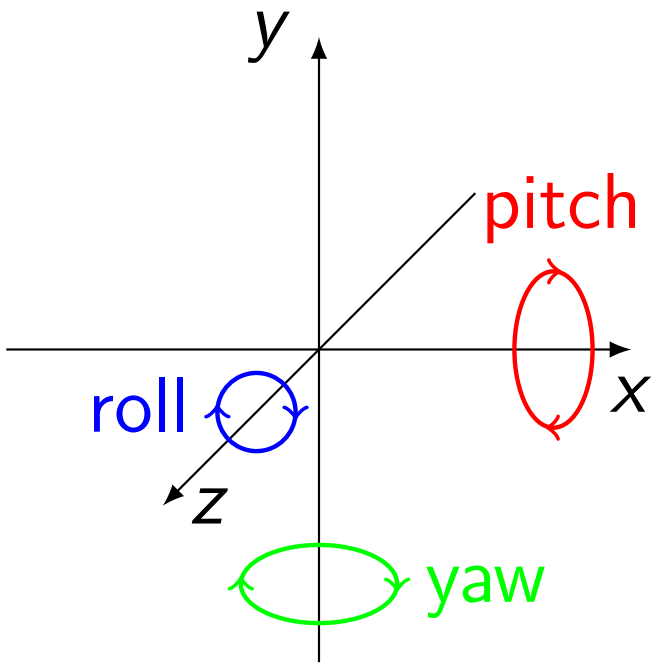
$$x_i^v = \begin{cases} 1 & \text{if } i \text{ is dominant} \\ 0 & \text{else } i\text{is not dominant} \end{cases}$$

$x^v = -1$ if v is not visible

The resulting feature vector is the concatenation of all voxels:

$$w \times h \times d \times 8$$

Soft Label Random Forest

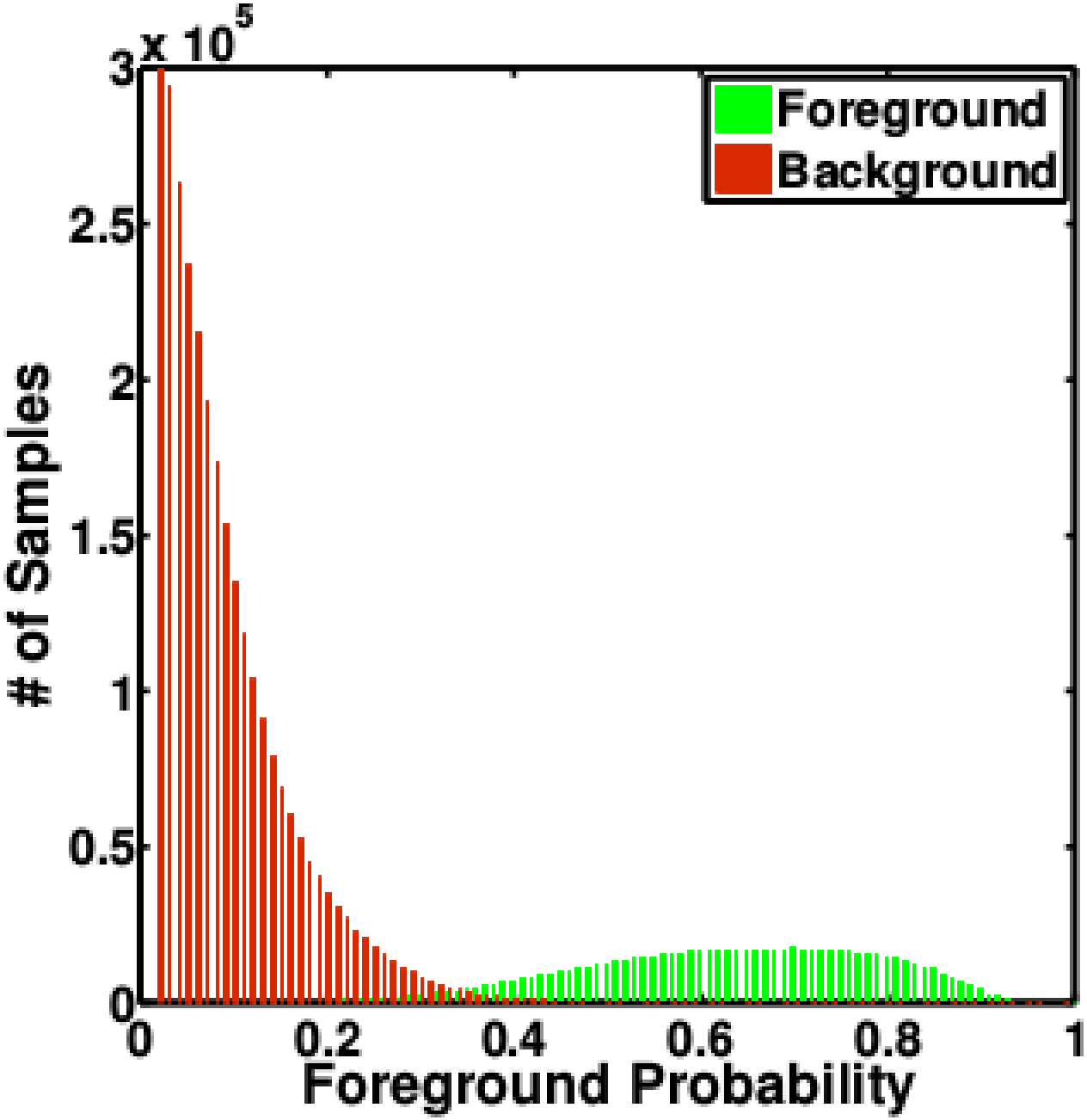


- 16 pose classes
- $+1$ class $= \begin{cases} 1 & \text{if bg} \\ 0 & \text{else} \end{cases}$
- $d_j^i = \|I - R_j^i\|_F$

$$l_j^i = \exp(-{d_j^i}^2), i \in \{1, ..., 16\}$$

$$\text{IF fg THEN } 1 = \sum_{i=1}^{16} l_j^i \text{ ELSE } 0$$

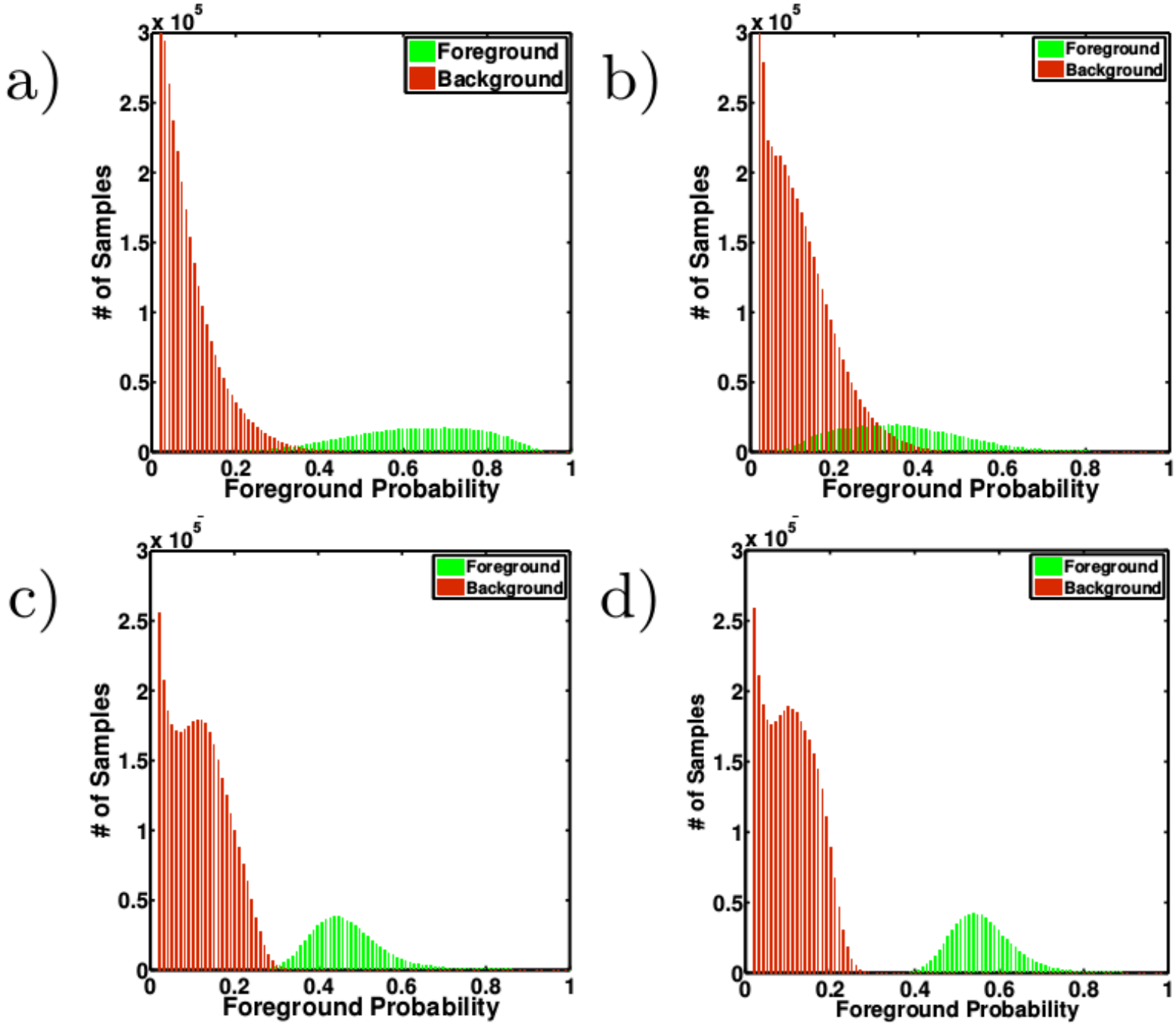
a)


x 10^5
3
2.5
2
1.5
1
0.5
0
# of Samples
0
0.2
0.4
0.6
0.8
1
Foreground Probability
Foreground
Background

training scheme for sLRF

Input: $X = \{x_j, l_j\}$
Output: Learnt sLRF classifier

1. $X_s \subset X$, $|X_s| = |X|/20$
2. Train sLRF with X_s, compute p_{fg} with X
3. add borderline positive (low p_{fg}) and borderline negativ (high p_{fg})
4. add confusing samples
5. compute d_L for all positive samples, add samples with high d_L
6. repeat 2-5 till p_{fg} for all positive data is greater than p_{fg} for all negative data.


a)
b)
c)
d)
x 10^5
Foreground
Background
# of Samples
Foreground Probability
0
0.2
0.4
0.6
0.8
1
0.5
1.5
2
2.5
3

Conclusion

- redundanze
- training with occlusion
- semantic (scene understanding)

Discussion

?


Ujwal Bonde, Vijay Badrinarayanan, and Roberto Cipolla. "Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II". In: ed. by David Fleet et al. Cham: Springer International Publishing, 2014. Chap. Robust Instance Recognition in Presence of Occlusion and Clutter, pp. 520–535. ISBN: 978-3-319-10605-2. DOI: 10.1007/978-3-319-10605-2_34. URL: http://dx.doi.org/10.1007/978-3-319-10605-2_34.


Andreas Geiger. "Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite". In: *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. CVPR '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 3354–3361. ISBN: 978-1-4673-1226-4. URL: http://dl.acm.org/citation.cfm?id=2354409.2354978.


Stefan Hinterstoisser et al. "Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes." In: *ICCV*. Ed. by Dimitris N. Metaxas et al. IEEE Computer Society, 2011, pp. 858–865. ISBN: 978-1-4577-1101-5. URL: http://dblp.uni-trier.de/db/conf/iccv/iccv2011.html#HinterstoisserHCIKNL11.


Alexander Krull et al. "6-DOF Model Based Tracking via Object Coordinate Regression." In: *ACCV (4)*. Ed. by Daniel Cremers et al. Vol. 9006. Lecture Notes in Computer Science. Springer, 2014, pp. 384–399. ISBN: 978-3-319-16816-6. URL: http://dblp.uni-trier.de/db/conf/accv/accv2014-4.html#KrullMBGIR14.


Bojan Pepik et al. "Occlusion Patterns for Object Class Detection." In: *CVPR*. IEEE, 2013, pp. 3286–3293. URL: http://dblp.uni-trier.de/db/conf/cvpr/cvpr2013.html#PepikSGS13.


Umer Rafi, Juergen Gall, and Bastian Leibe. "A Semantic Occlusion Model for Human Pose Estimation From a Single Depth Image". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2015.


Training Deformable Partmodels. *unkown*. [Online; accessed March 18, 2016]. 2016. URL: http://i.stack.imgur.com/Awffz.jpg.


unkown. *unkown*. [Online; accessed April 4, 2016]. unkown. URL: http://3.bp.blogspot.com/_IDEWI0P9RbA/TB6VYXUN_CI/AAAAAAAAACA/DllZ1rwWktA/s1600/semsegm.png.


unkown. *unkown*. [Online; accessed April 4, 2016]. unkown. URL: https://farm6.staticflickr.com/5668/21114849206_759853e42a_z.jpg.