Analyse eines Forschungsthemas

a detection and two pose estimation algoritms deal with occlusion

Josef Schulz

April 21, 2016

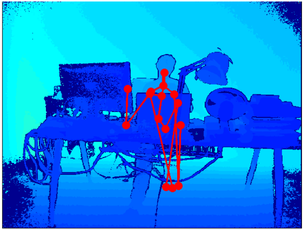
Examples Problems

Examples Problems

DPM: classification and localisation

Examples Problems

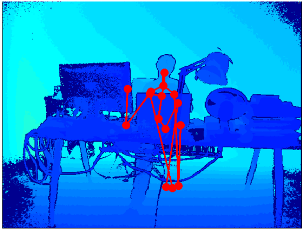
DPM: classification and localisation



joint pose estimation with Kinect2 SDK

Examples Problems

DPM: classification and localisation

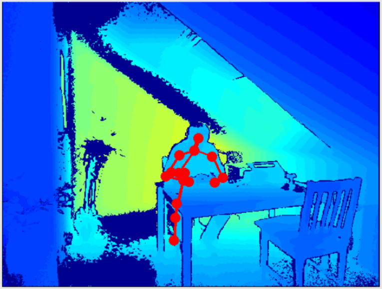


joint pose estimation with Kinect2 SDK

pose estimation with occlusion

Content

A Semantic Occlusion Model For Human Pose Estimation



Kinect2 SDK

Input : single Depth-Image

Output : estimated poses of all parts

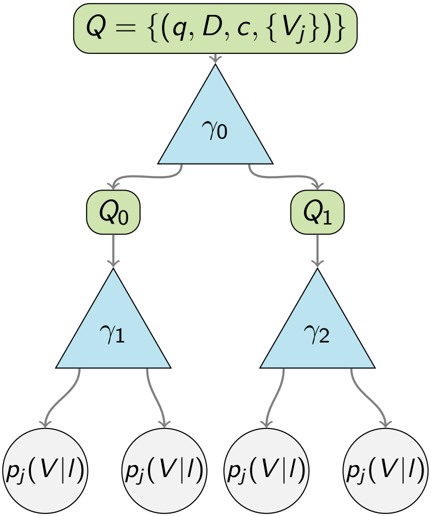
Regression Forest

Training Set

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

- q pixel location
- D reference depth image
- c class label corresponding to limbs
- $\{V_j\}$ is set of 3D offset vectors between q and the joint position q_j:

$$V_j = q_j - q$$



Split Node

$$\gamma = (\boldsymbol{u}, \boldsymbol{v}, \tau)$$

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

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

u, v - offset vectors

τ - threshold

$D(q)$ - depth value

Evaluating The Splitting Functions Information Gain

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

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

$$H(Q) = -\sum_c p(c|Q)\log(p(c|Q))$$

$H(Q)$ - Shannon entropy
$g(\Phi)$ - information gain

Leaf Node

a leaf node stores:

- class c
- two cluster centers V_{ljk} for each joint
- two support values w_{ljk} for each cluster
- probabilities over V_j:

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

K - cluster

w_{ljk} - is determined by offset vectors ended in the cluster $k, k \in K$

V_{ljk} - cluster center

Pose Estimation

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

$X_j \; = \{(x_j, w_j)\}$

x_j - absolute joint position, $x_j = q + V_{ljk}$

w_j - confidence value, $w_j = w_{ljk} \cdot D^2(q)$

The cluster with the highest summed weights w_j are used for prediction.

Occlusion Aware Regression Forests

idea: Add new class labels which does not correspond to the body. The assumption is that if a person is interacting with a specific object, some joint positions could be predicted.

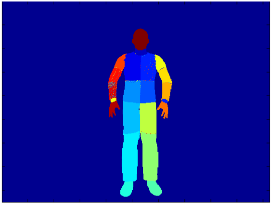
$$Q = Q \cup \{(q_{occ}, D, C_{occ}, \{v_{jocc}\})\}$$

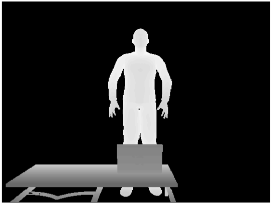
without Semantics

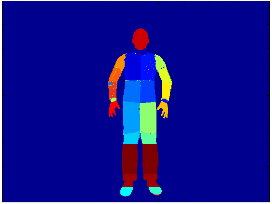
$$C_{occ} = \{c_{occ}\}$$

with Semantics

$$C_{occ} = \{c_{obj1}, c_{obj2}, ...\}$$







Training Data

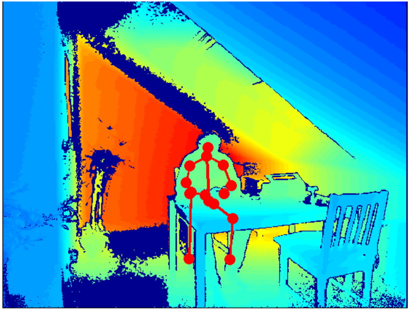
Synthetic Data (552 images)

- Human Poses from CMU-Database [9]
- body part labels for each pixel

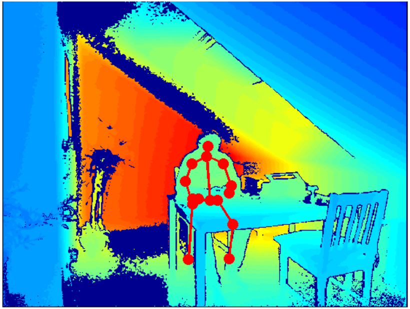
Real Data (552 images)

- Kinect2 SDK, all fails are discarded

With And Without Semantics

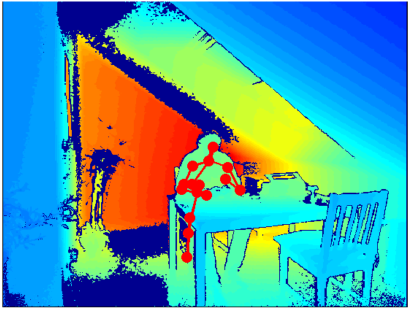


without semantics: the joint at the hip is not good estimated

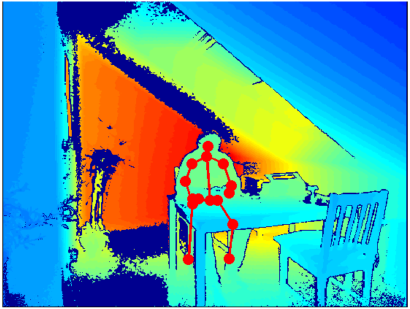


with semantics

OARF vs. Kinect2 SDK



Kinect2 SDK



OARF with semantics

Results

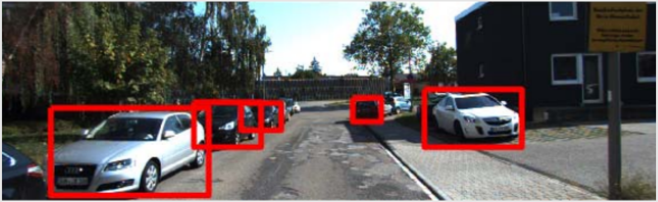
	Occluded Joints	Non Occluded Joints	All Joints
OARF W/O	32.60	55.50	50.66
OARF W	35.77	56.01	51.72
Kinect2 SDK	18.13	66.36	56.94

in %

Testset: real + synthetic data

- real data was taken with the Kinect
- synthetic data is generated by rendered scenes

Occlusion Patterns for Object Class Detection

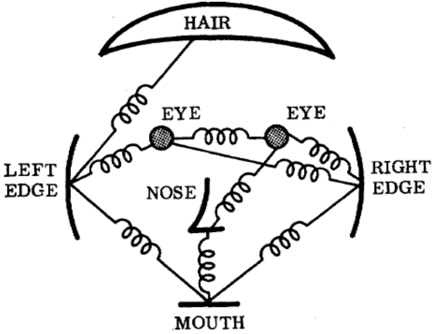


Input : Single RGB-Image

Output : Object-Bounding-Boxes

Deformable Models Approch

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



Pictorial Structure Model [8]

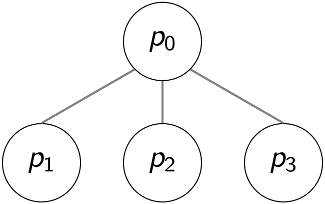
Matching model to image involves joint optimization of part locations "stretch and fit"

Model

Model is represented by a Graph

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

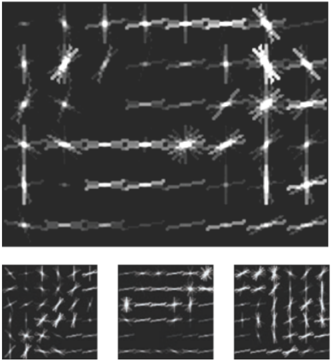
E_c - is the Energy function



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

Filter

- images with bounding boxes
- histograms of oriented gradients (HOG) for placement
- twice the resolution for every level
- Gaussians for deformation



KITTI Data Set

KITTI contains 7481 images

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

Parts:

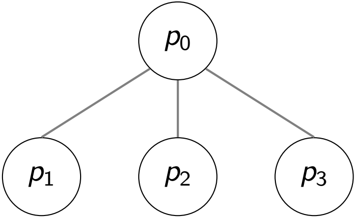
visible 6

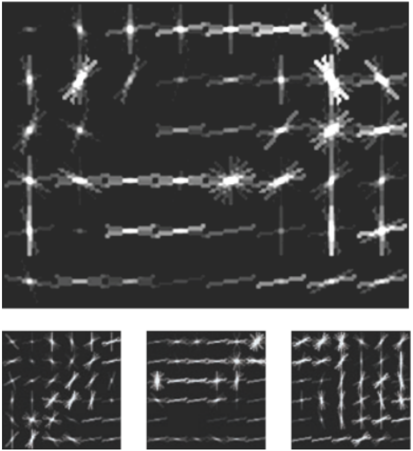
occluded 16 – 15

Single-Object Occlusion Patterns - OC-DPM

Learn class from object instances with distinctive appearances of occlusion.

$$C = \{1, \ldots, C_{visible}\} \cup C_{invisible}$$



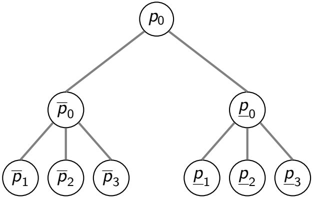


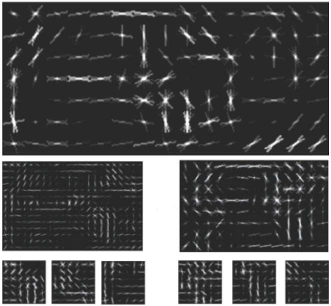
- like standard DPM
- trained with occlusion
- C are components

Double-Objects With Joint Root - SYM-DPM

Represent occluder and occludee at once and combine them with a root joint.

$$\begin{aligned} E'_c(p;I) = \langle v^c, \Phi(p_0;I)\rangle &+ \langle \overline{w}^c, \Phi(p_0,\overline{p}_0)\rangle && + \langle \underline{w}^c, \Phi(p_0,\underline{p}_0)\rangle \\ &+ E_c(\overline{p}_0;I) && + E_c(\underline{p}_0;I) \end{aligned}$$

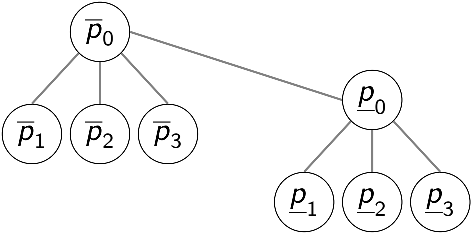


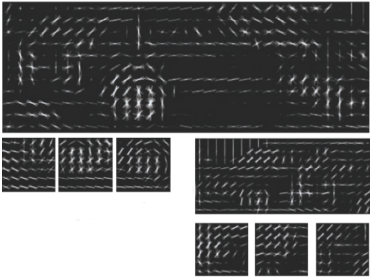


- one root part
- occluder $\overline{p}_0$ root part
- occludee $\underline{p}_0$ root part

Double-Objects Without Joint Root - ASYM-DPM

Both objects are combined to one, based on the intuition that the occluder can be trusted more.





- occluder left, occludee right
- tree structure
- no extra terms

Mining Training Data

Feature Space:

i occluder left/right of occludee

ii orientation of occluder/occludee

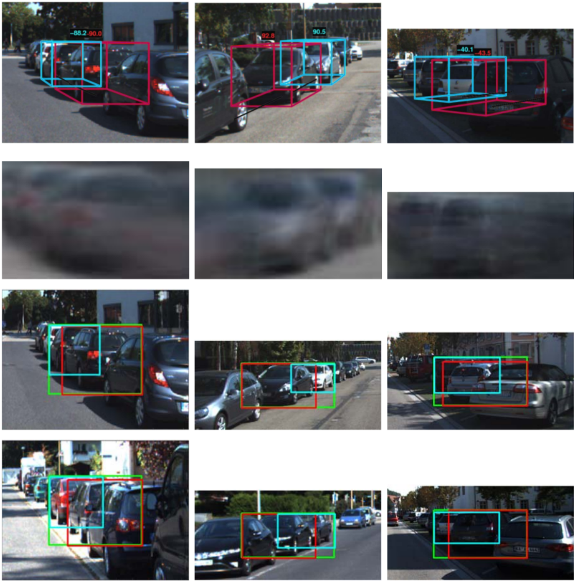
iii occluder is/is not occluded

iv degree of occlusion of occludee

Rule-based clustering

- repeatly splitting the training data
- according to the viewing angle of the occluder

Mining Training Data



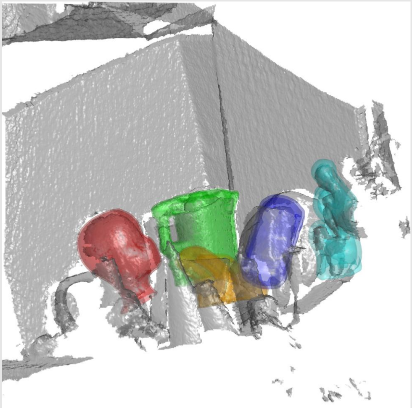
Results

OC-DPM DPM

	DPM	OC-DPM	SYM-DPM	ASYM-DPM
full dataset	62.8	64.4	53.7	52.3
Pedestrian	36.2	37.2	31.4	29.4

in %

Robust Instance Recognition in Presence of Occlusion and Clutter



Input : 5-10 consecutive frames as one Pointcloud

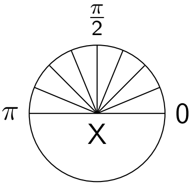
Output : 6D-Object-Pose

Intoduction

- object shape is invariant to changes in illumination or texture
- Kinect sensors generates cheap depth data
- it is easy to synthesize pointcloud data

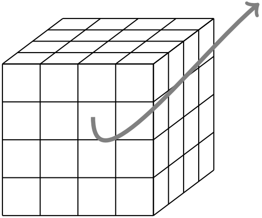
Edgelet

- 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	$\cdots$	8

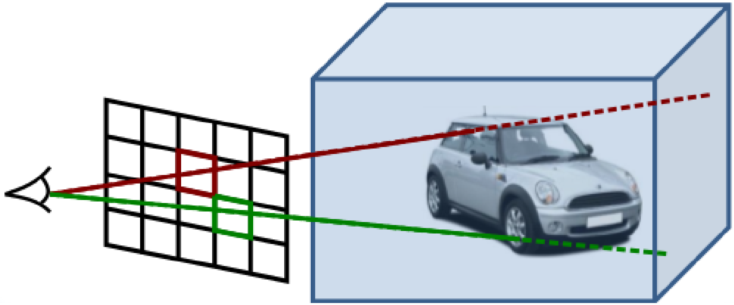
$$x_i^v = \begin{cases} 1 & \text{if } i \text{ is dominant} \\ 0 & \text{else } i \text{is not dominant} \end{cases}$$

$$x^v = -1 \text{ if v is not visible}$$

The resulting feature vector is the concatenation of all voxels:

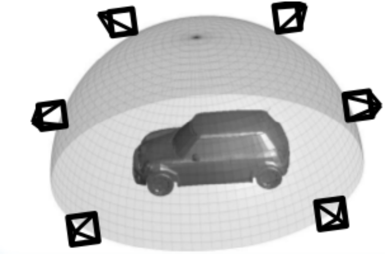
$$w \times h \times d \times 8$$

Box Model For Occluder



- occluder are rectangular
- occluders are restricted to start from the ground plane

Soft Label Random Forest



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

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

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

Occlusion Queries

$$x_j \in \{-1, 0, 1\}$$

(> -1) - visible versus occluded voxel

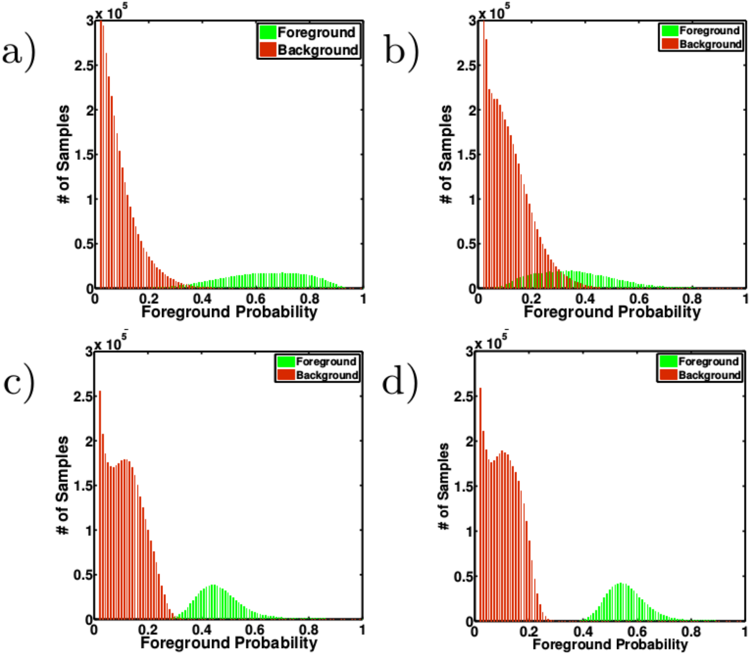
(> 0) - if greater 0 the voxel is dominant

- split questions in the topmost nodes ($\approx 5 - 10$) are restricted to the second type
- training set consists of 27000 simulated views

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

Results

	D-DPM	LineMod	S-Iterative (Edges)	S-Iterative (Occlusion)
L	40.50	30.15	70.70	81.89
L+P	23.72	13.17	52.62	62.11

in %

L - location

P - pose estimation

Robustness Against Occlusion - Conclusion

- semantic (scene understanding)
- mixture models
- multiple representation for one object class
- good feature representation
- training with occlusion

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.

Andreas Geiger et al. "Vision meets Robotics: The KITTI Dataset". In: *International Journal of Robotics Research (IJRR)* (2013).

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.

unknown. *Pictorial Structure Model*. [Online; accessed April 17, 2016]. unknown. URL: http://kenschutte.com/thesis/fischler.png.

unkown. *CMU Mocap*. [Online; accessed April 17, 2016]. unkown. URL: http://mocap.cs.cmu.edu/.