ACIN
AUTOMATION & CONTROL INSTITUTE
INSTITUT FÜR AUTOMATISIERUNGS- & REGELUNGSTECHNIK

TU VIENNA
TECHNISCHE UNIVERSITÄT WIEN
VIENNA UNIVERSITY OF TECHNOLOGY

Object Classification in Domestic Environments

Markus Vincze

Aitor Aldoma, Markus Bader, Peter Einramhof, David Fischinger, Andreas Huber, Lara Lammer, Thomas Mörwald, Sven Olufs, Ekaterina Potapova, Johann Prankl, Andreas Richtsfeld, Robert Schwarz, Karthik Varadarajan, Walter Wohlkinger, Michael Zillich, Kai Zhou

Automation and Control Institute
Technische Universität Wien
vincze@acin.tuwien.ac.at

A Robot in Every Home

- Take robot out of the box and show it your favorite places
- Easy-to-use interface
- Learn places and map
- Humans link places to objects, e.g., room, door, furniture
- Knows about typical objects, their properties and function(s)

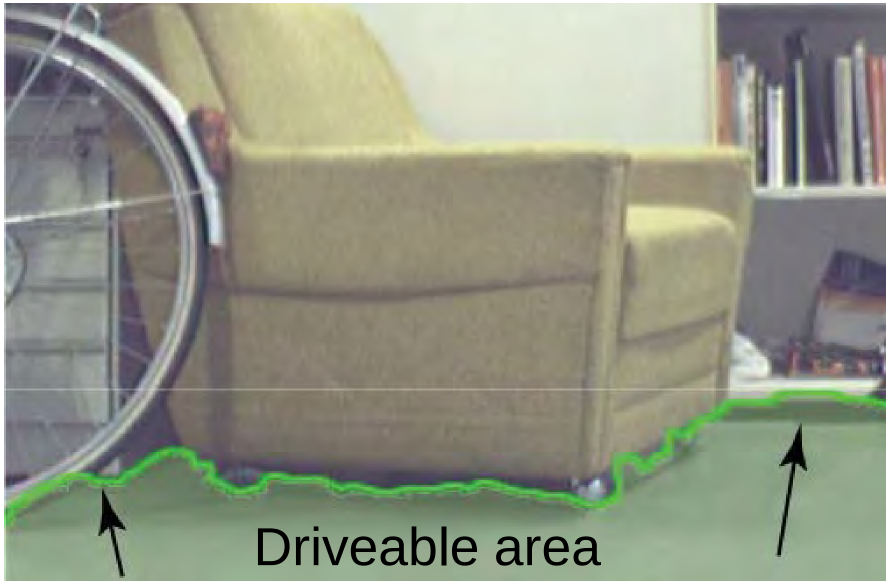
robots@home

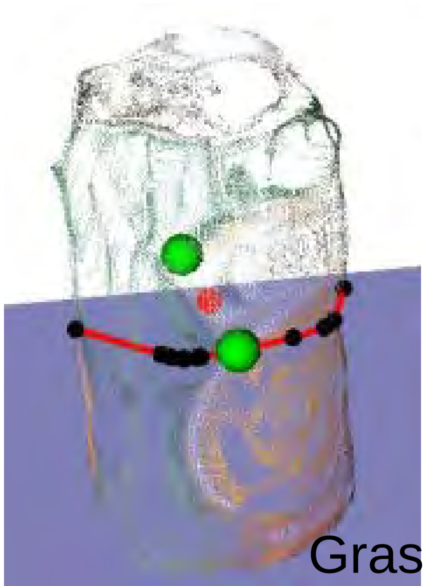
Robot System View

- Robots act in an environment (situated)
- Task is known, task sets context and function
  - E.g., domestic robot, service robot for aging well
- Object(classes) are known ⇒ model

Robot Vision – Embodiment and Envisionment

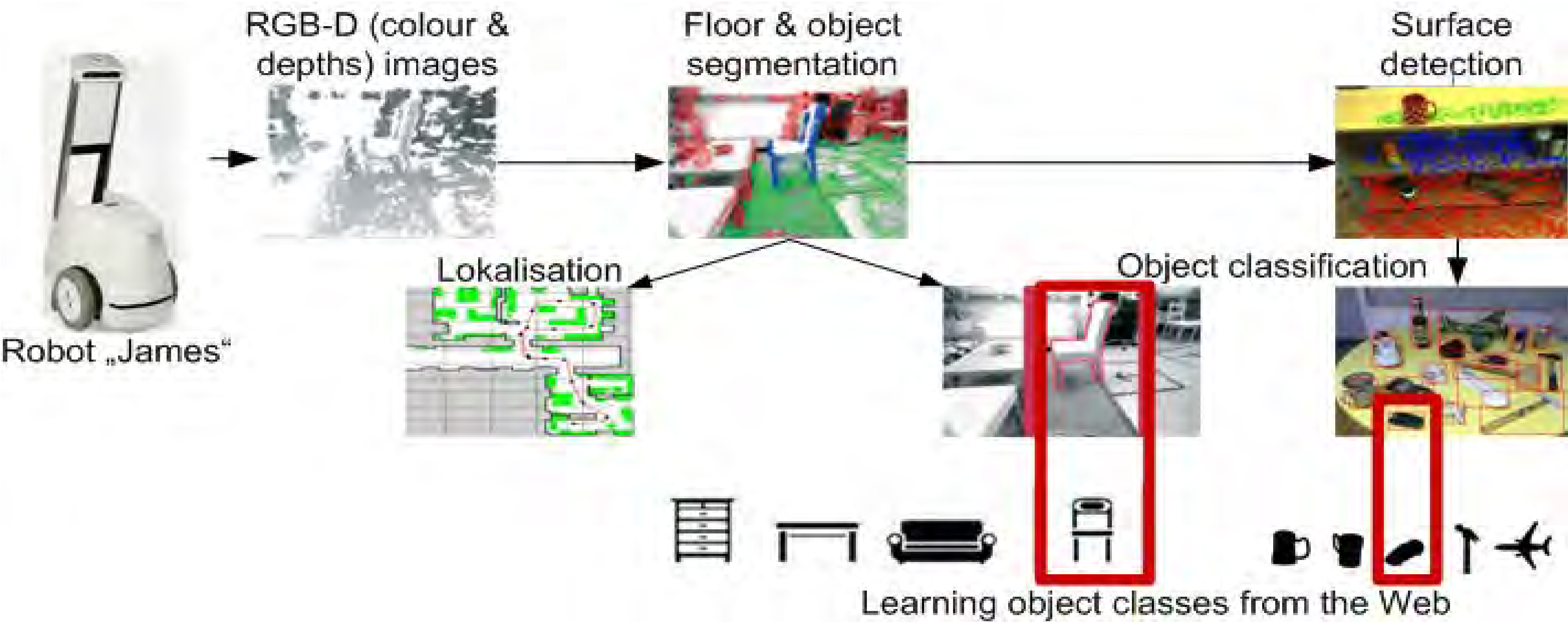
- Embodiment [Pfeifer 2001, 2008]
  - Goal orientation, intended motion of agent structures data
- „Envisionment“ – Situated Vision
  - Intention sets the task of vision





Embodied Object Classification

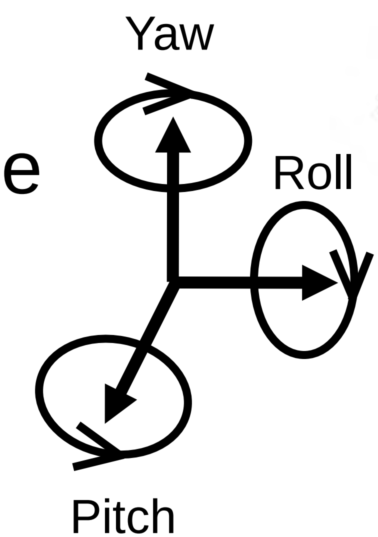
- Interaction of robot structures the environment
- Accumulation of object and scene knowledge

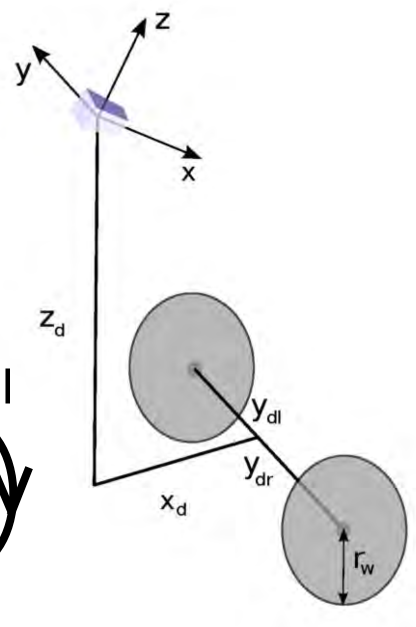


Ground Segmentation

Exploiting known geometry and kinematics

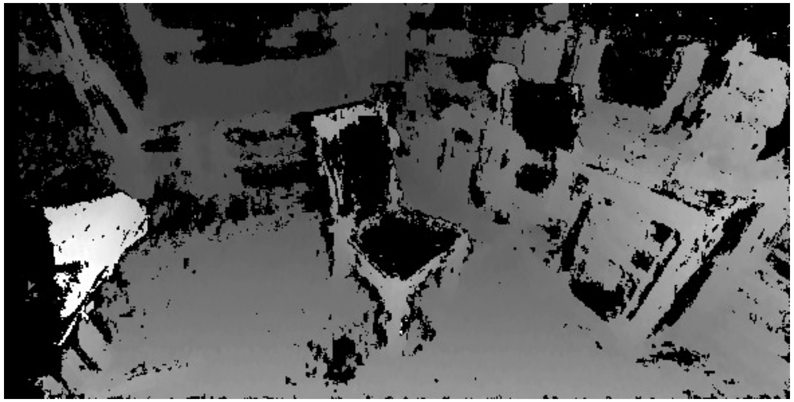
- Calculation of ground point disparities
  - Also obtaining dynamic pitch and roll angles
- Ground plane in 3D from the plane in disparity domain

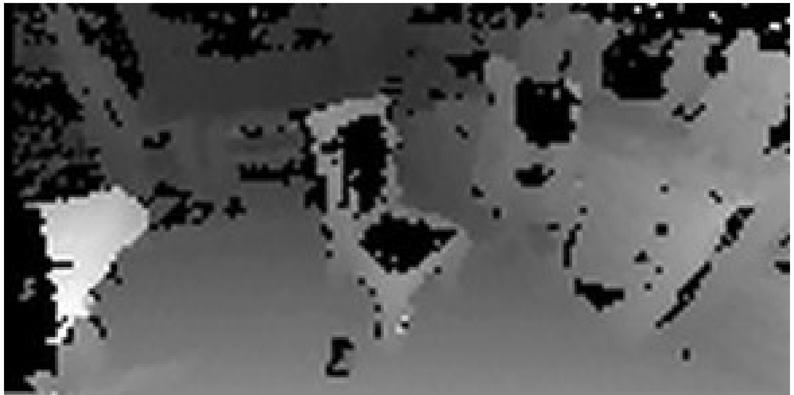




Approach

Noise and data reduction by building histograms over nxn-pixel patches of the disparity image





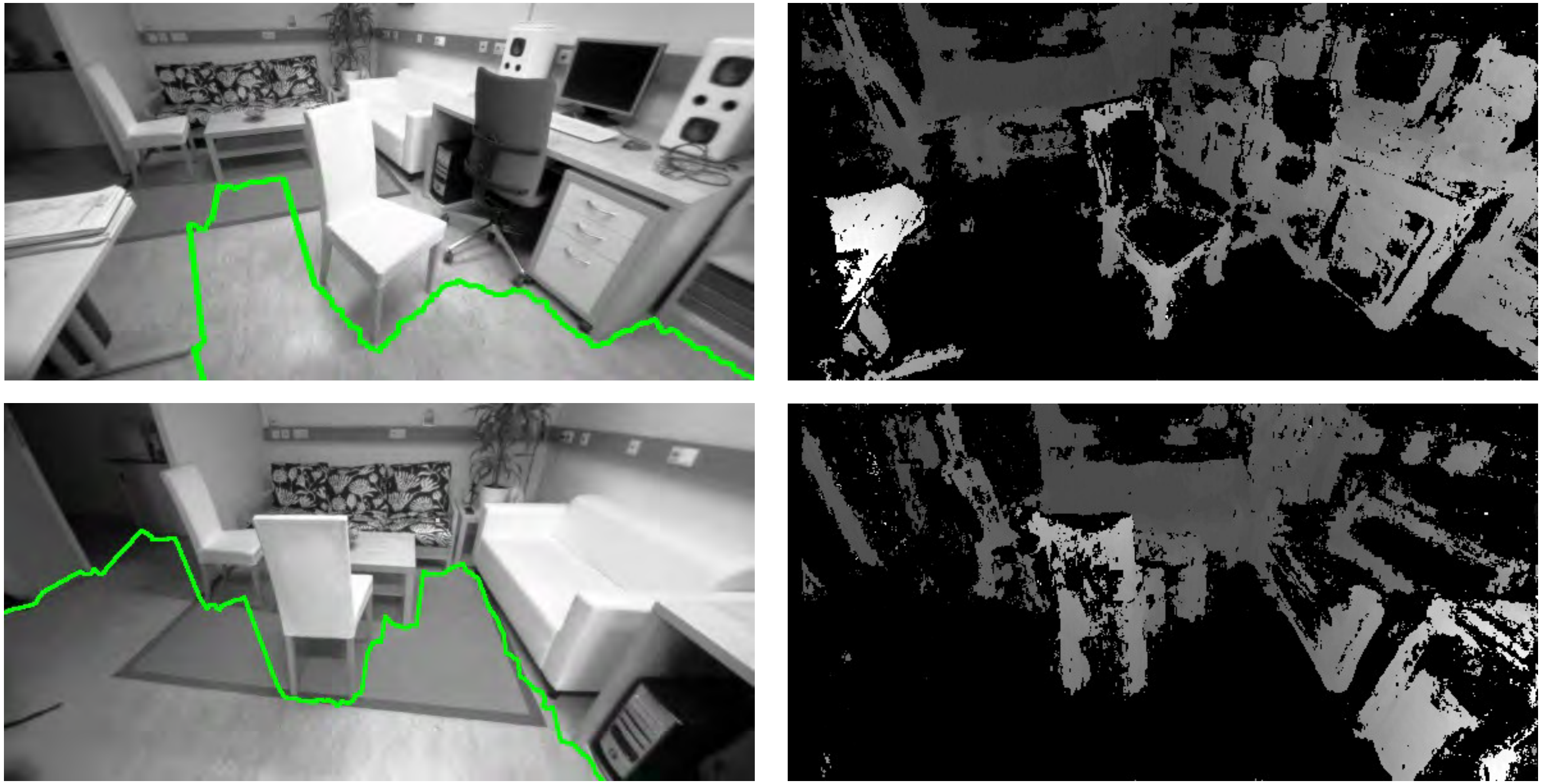
Ground Plane

Labelling of ground points

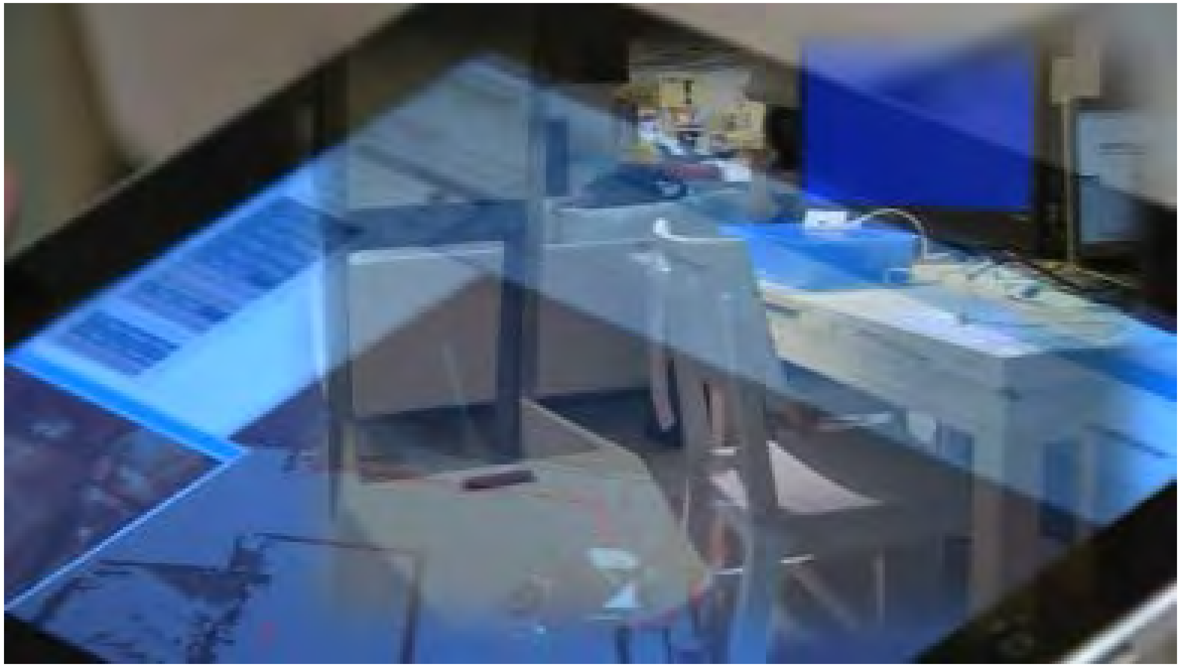
- Select seed points in bottom pixel row
  - Scanning upwards
  - Allow holes to cope with textureless regions
- LS Plane Fit to labelled points
- Labelling of rest of ground

Obstacles – Everything above Ground

- Average computation time for a 600x300 pixel image is ~7ms on a 1.73GHz Intel notebook

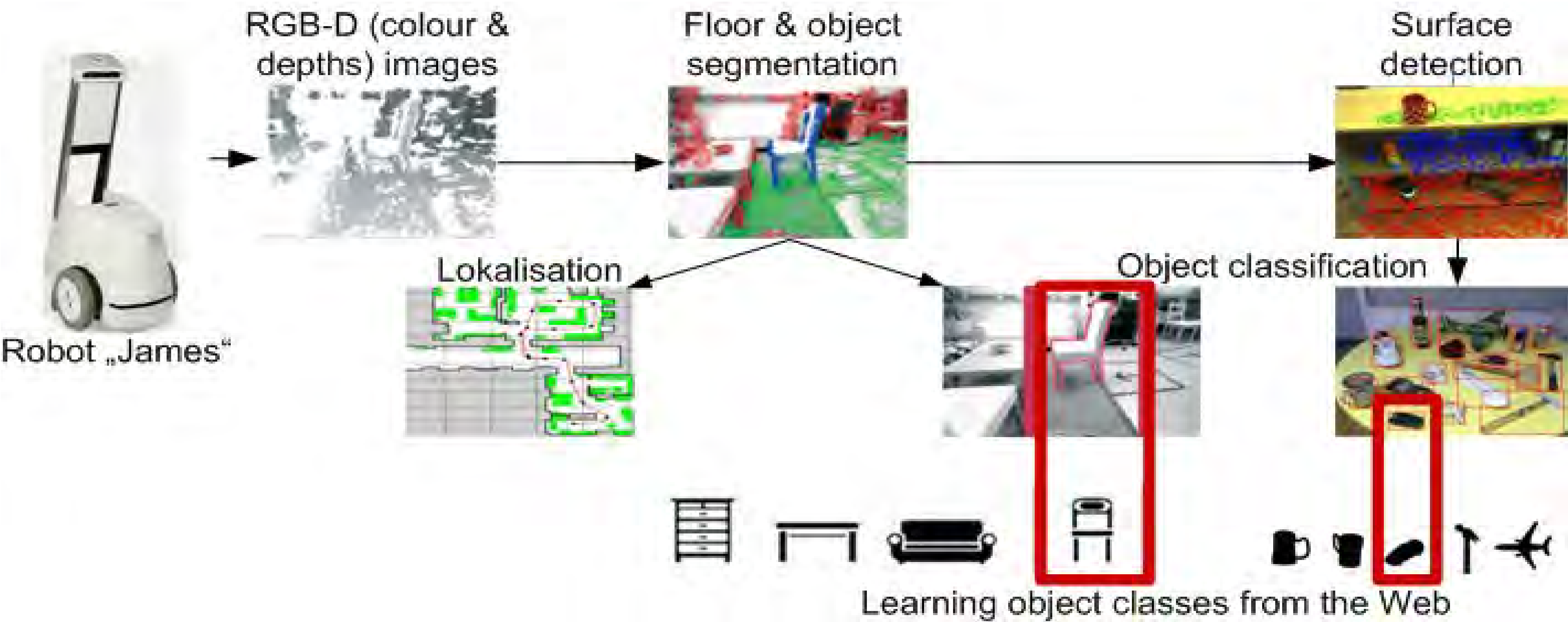


James at Furniture Warehouse



Embodied Object Classification

- Interaction of robot structures the environment
- Accumulation of object and scene knowledge

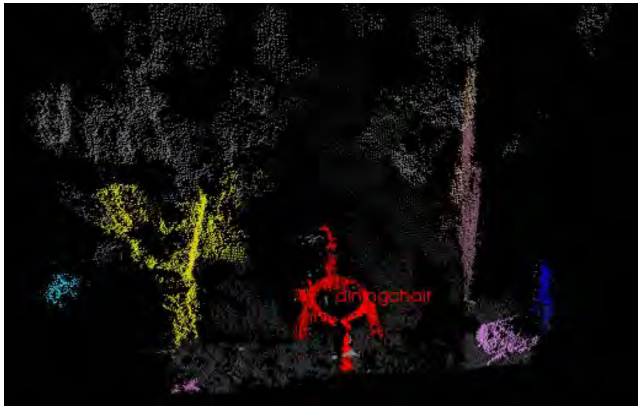


Furniture/Object Classification

of course these are chairs, but...

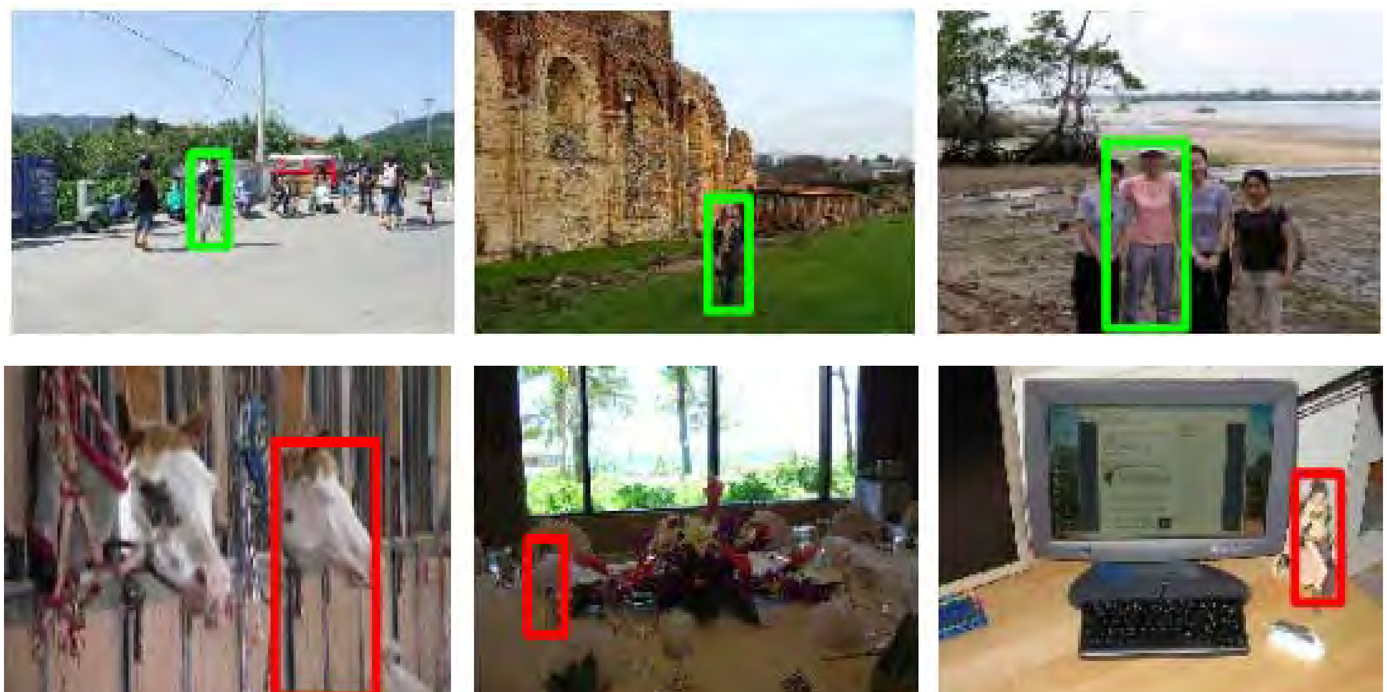
Motivation & Idea

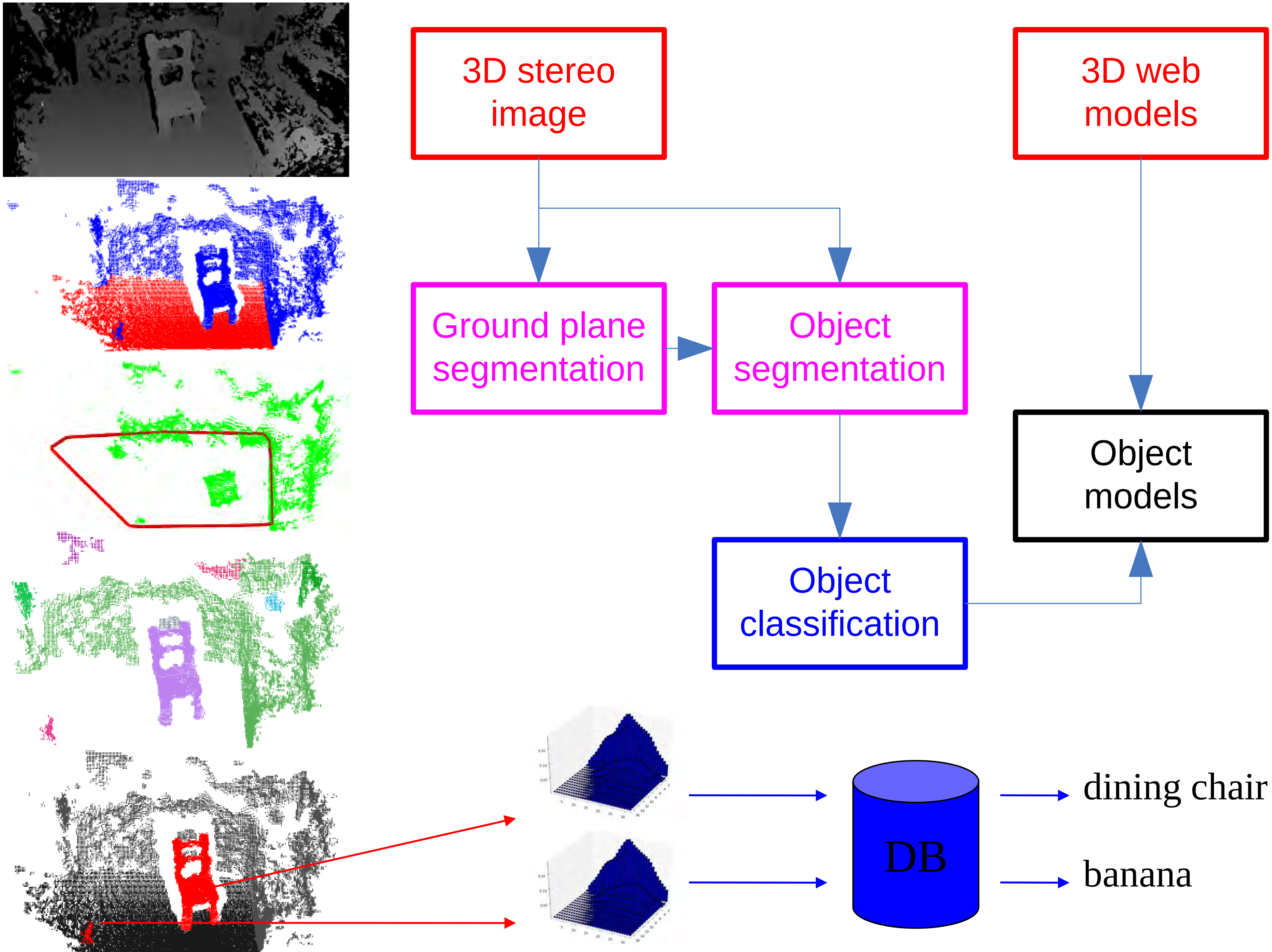
- Easily teach new objects & object classes
- 3D data from depth sensors (stereo camera, Kinect, TOF)
- Include priors (floor, objects on support planes, object-size)



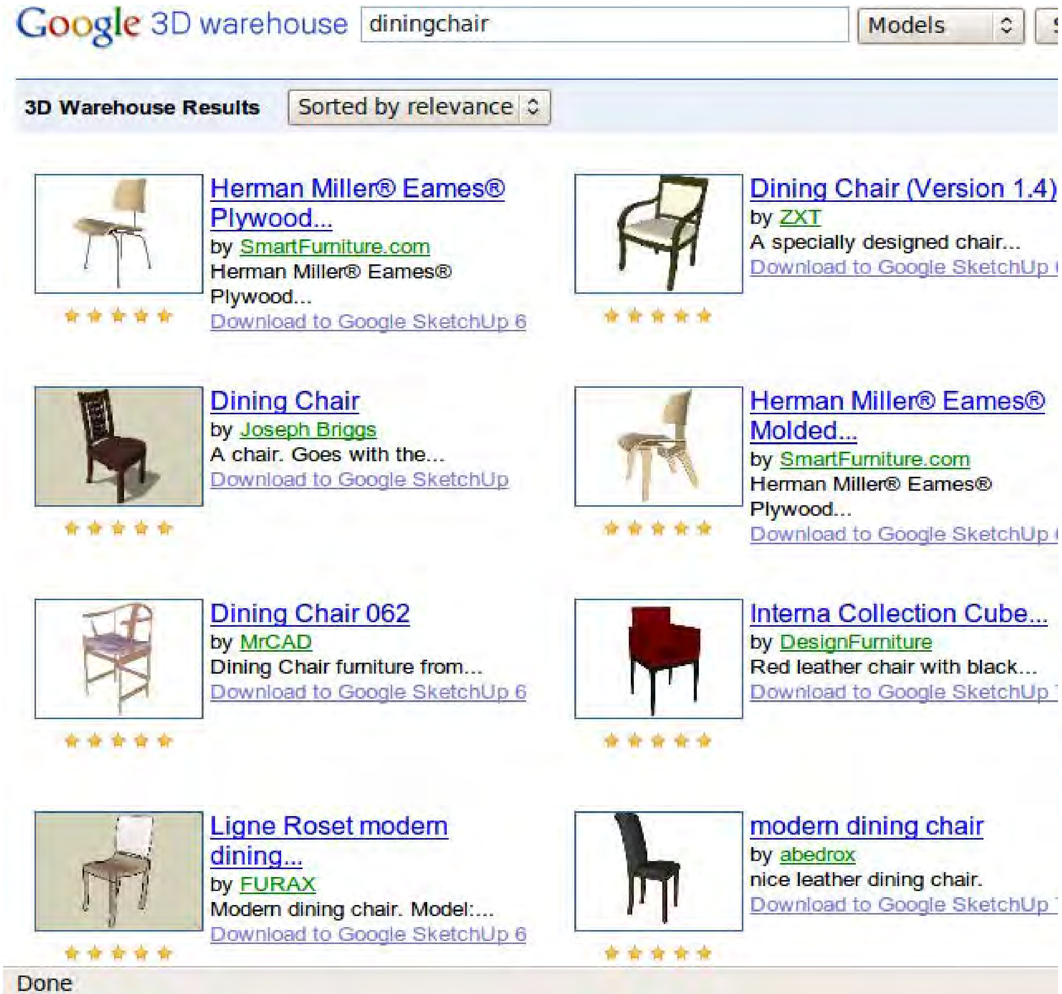
Visual Object Class Detection

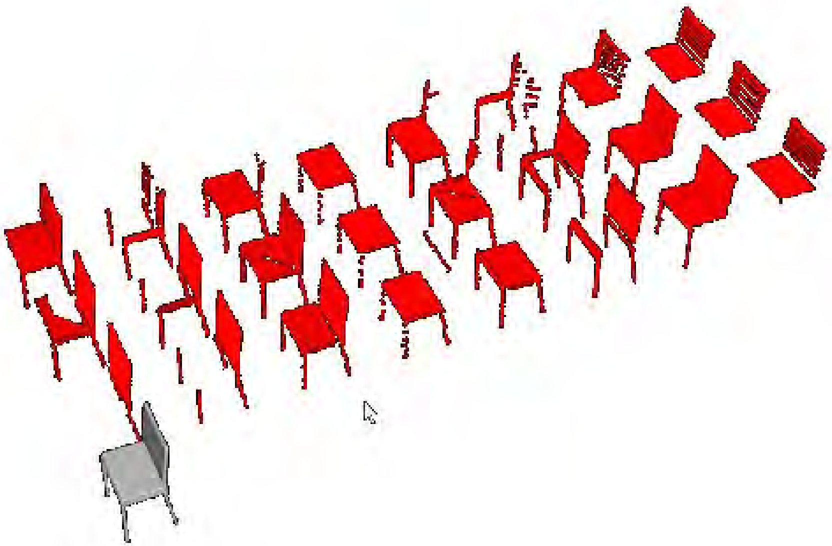
- Pascal VOC, since 2005
  - Learn statistics over feature vector (HoG, HoC, …)
  - 68% of planes, 15% of chairs




3D stereo image
3D web models
Ground plane segmentation
Object segmentation
Object models
Object classification
DB
dining chair
banana

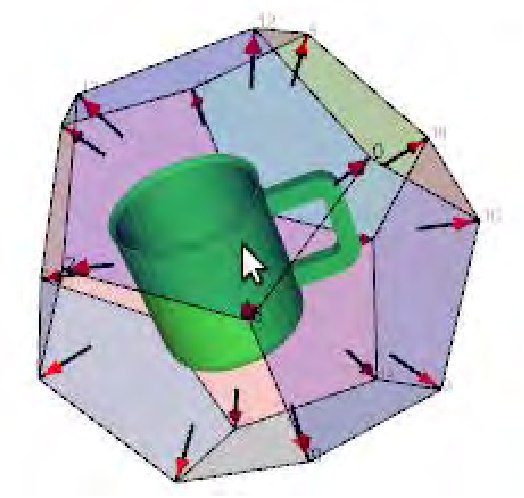
From Web2World: „diningchair"

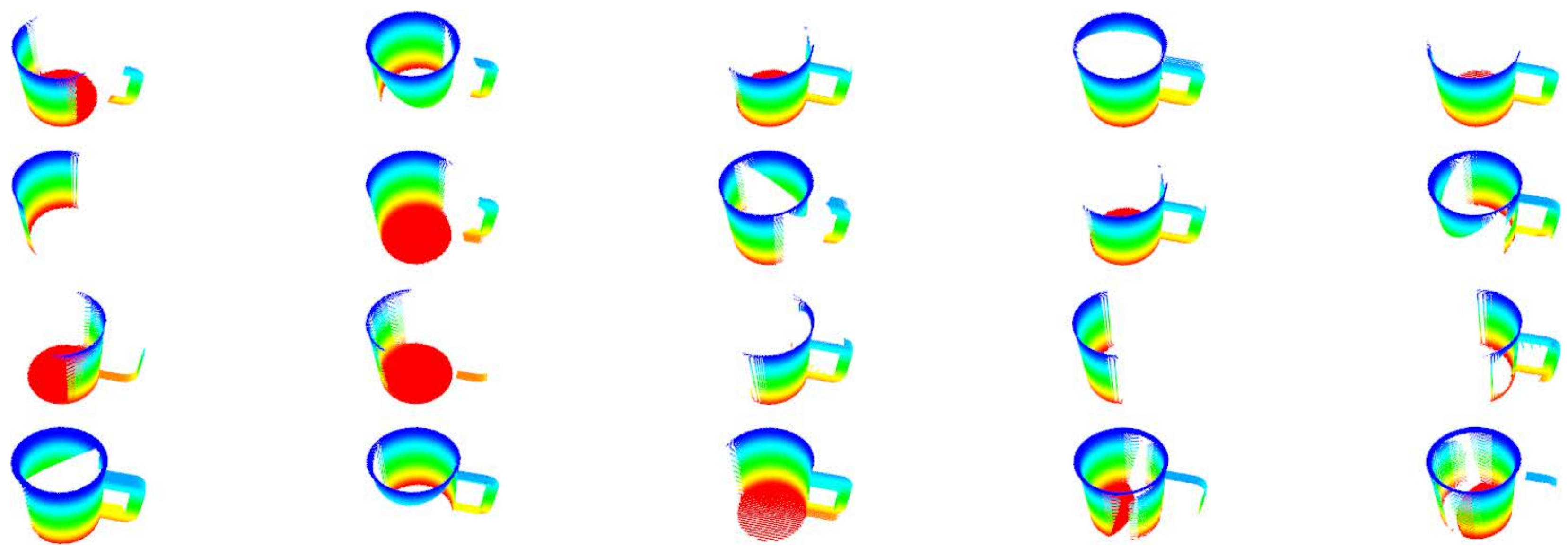




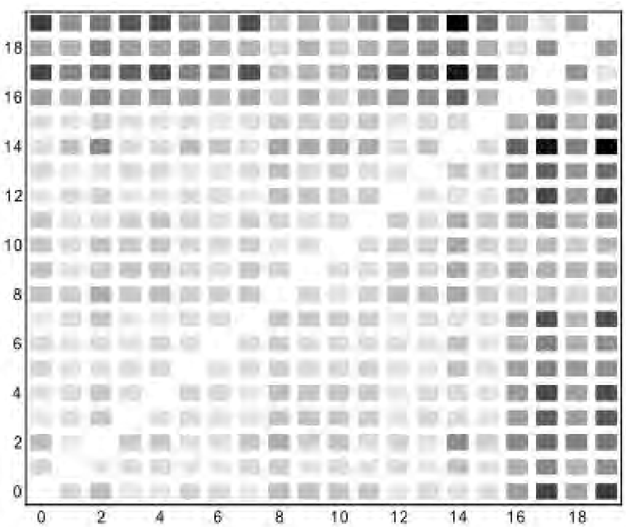
3D Object Classification - Training

- Training DB from the Web (e.g., 3DWarehouse)
- Automatic View Generation to simulate 2D appearance

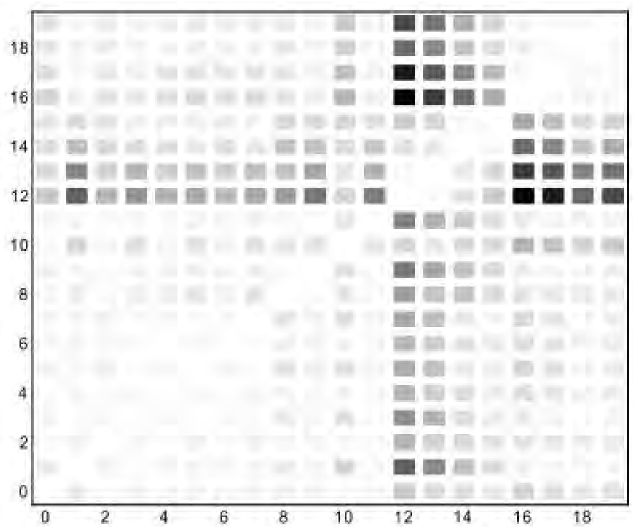




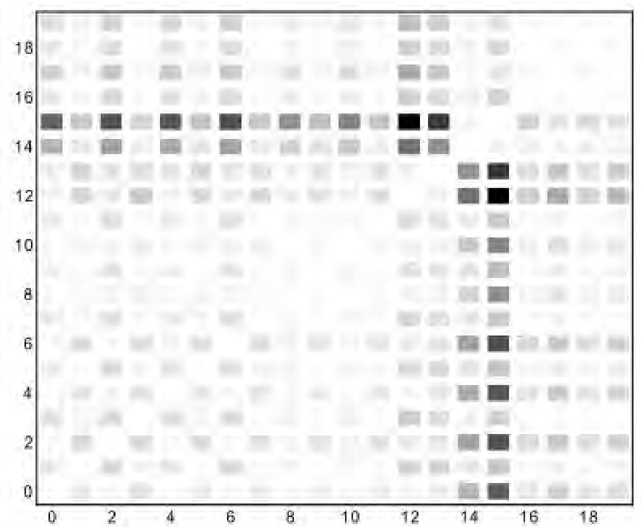
View Similarity



(a) Air plane view similarity matrix with white showing similar views, black dissimilarity.

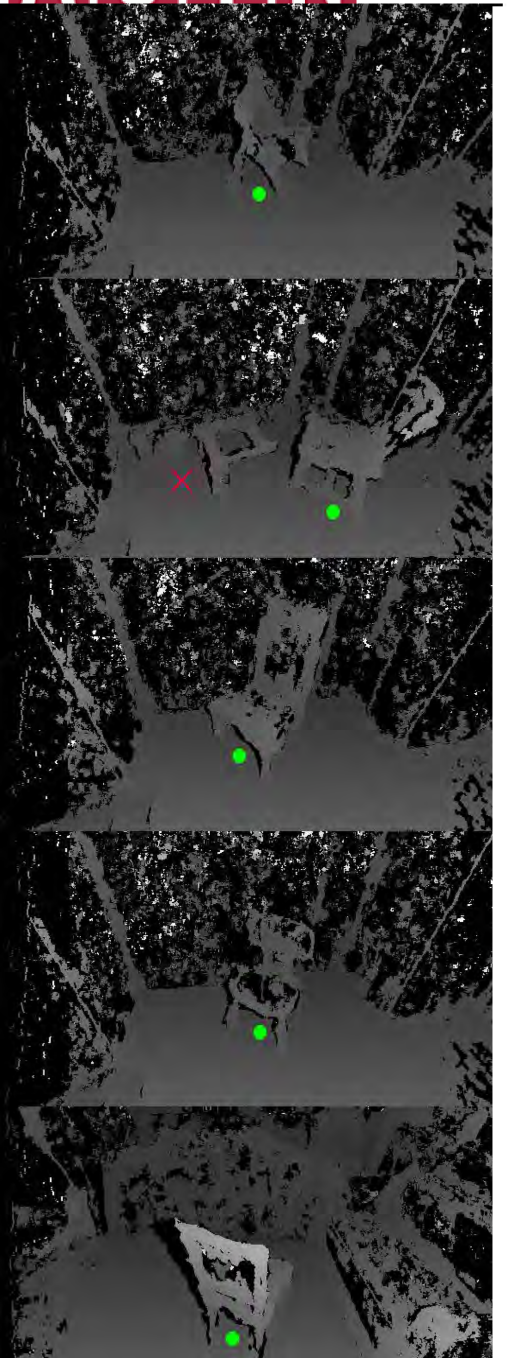


(b) Chair view similarity matrix with self similarity on the diagonal.

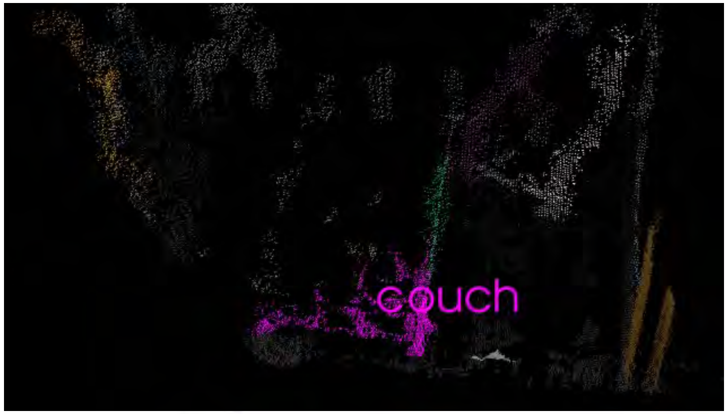


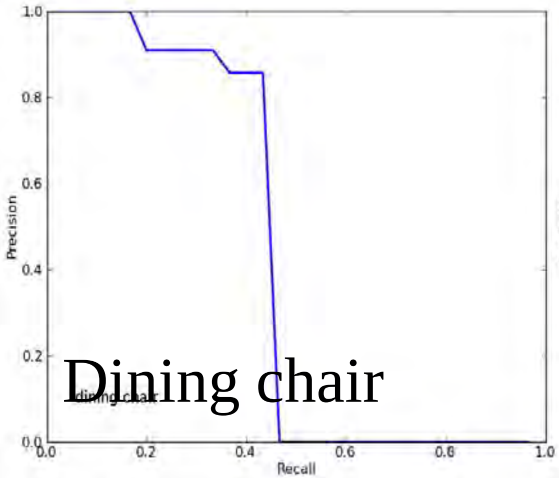
(c) Mug view similarity matrix. Views 0, 2 and 4 are very similar.

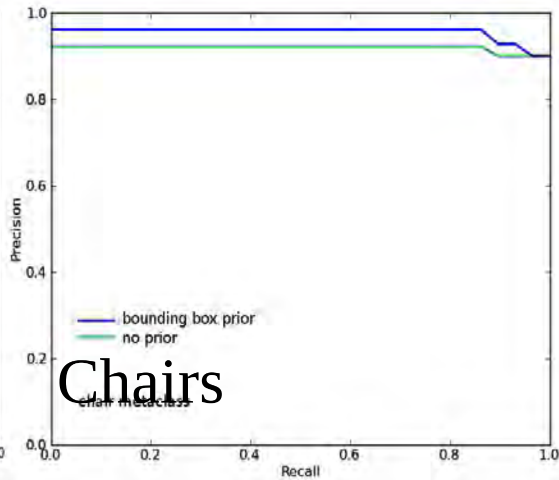
Evaluation on Chairs



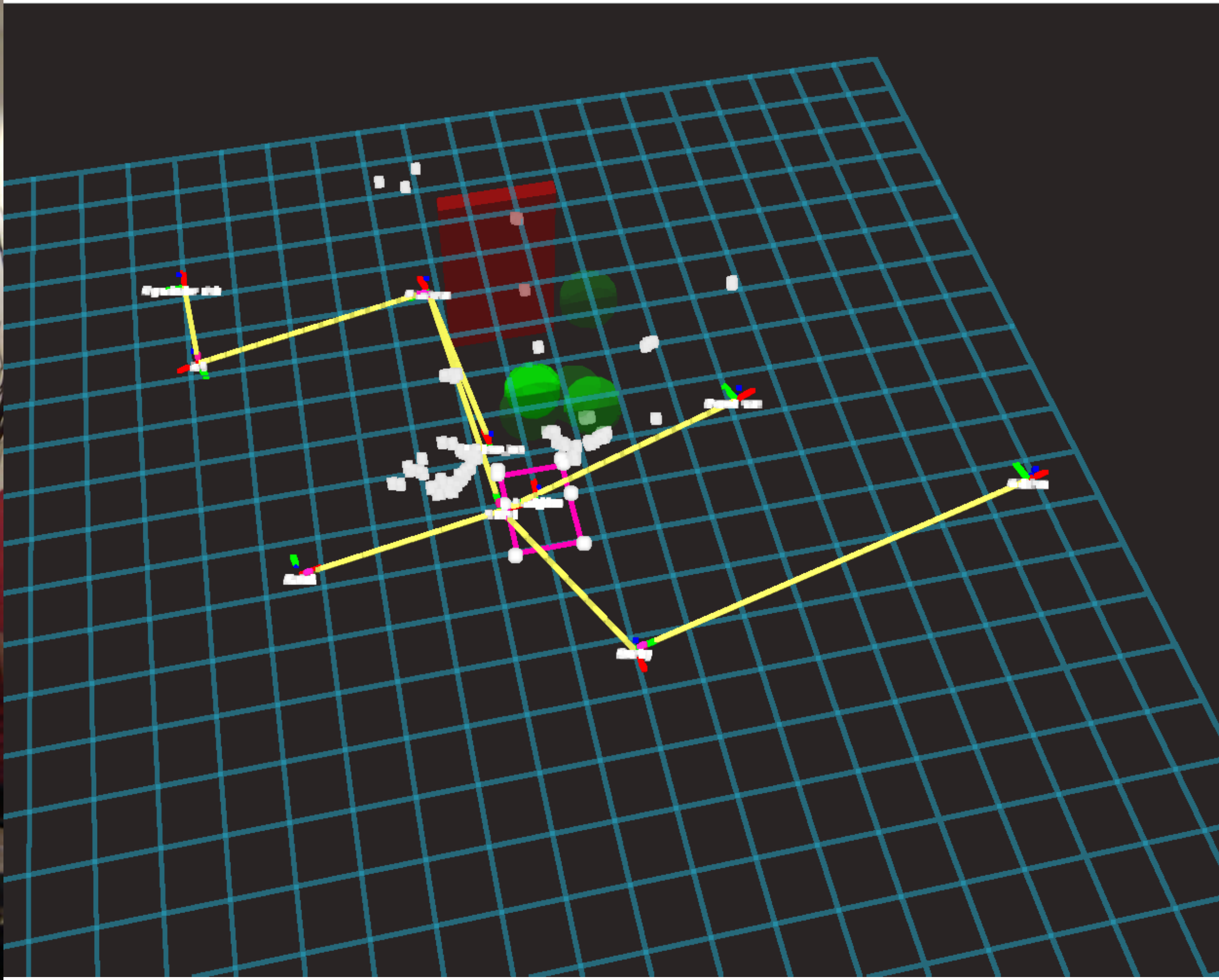

bed


couch


Dining chair
dining chair
Precision
Recall

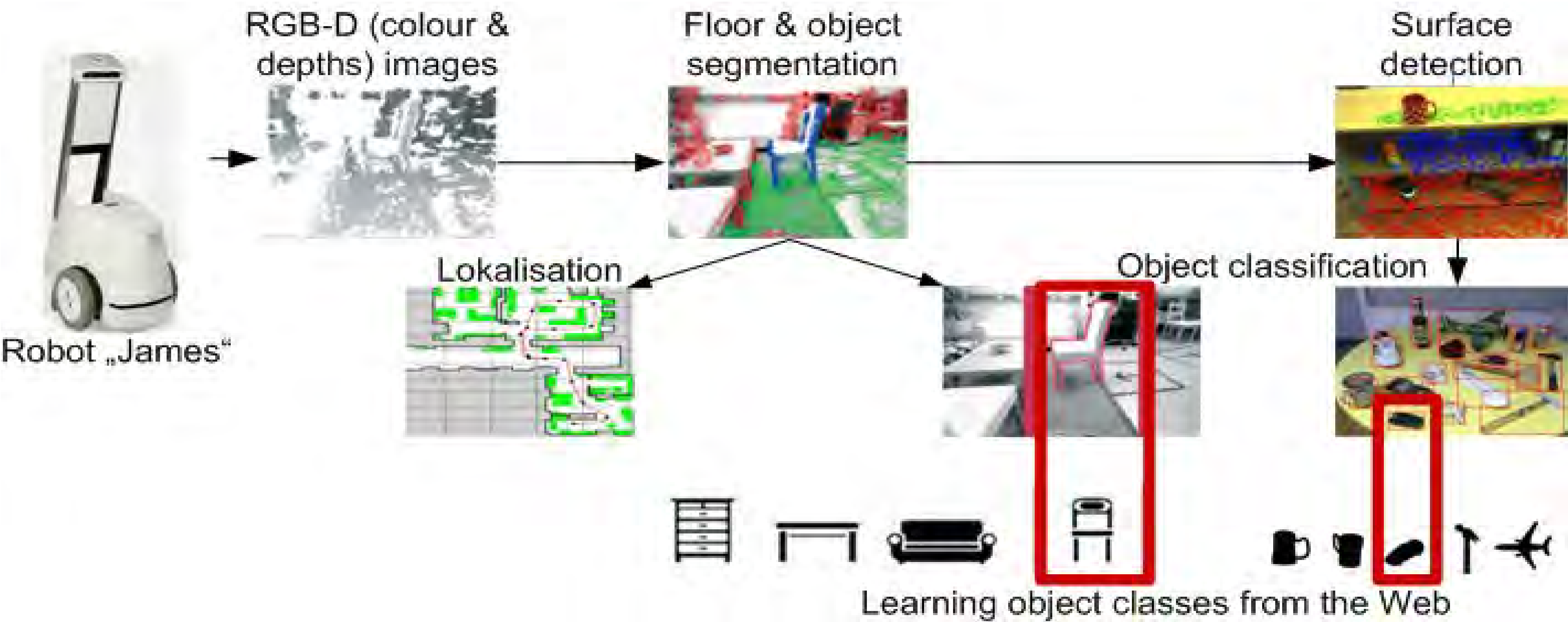

Chairs
bounding box prior
no prior
chair metaclass
Precision
Recall

Topological map



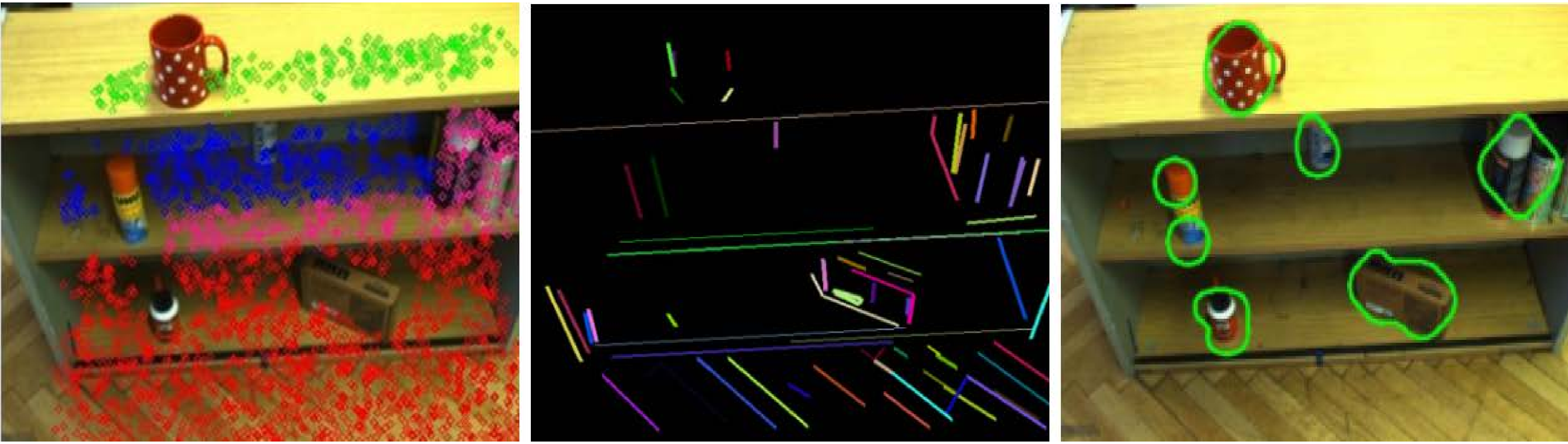
Embodied Object Classification

- Interaction of robot structures the environment
- Accumulation of object and scene knowledge



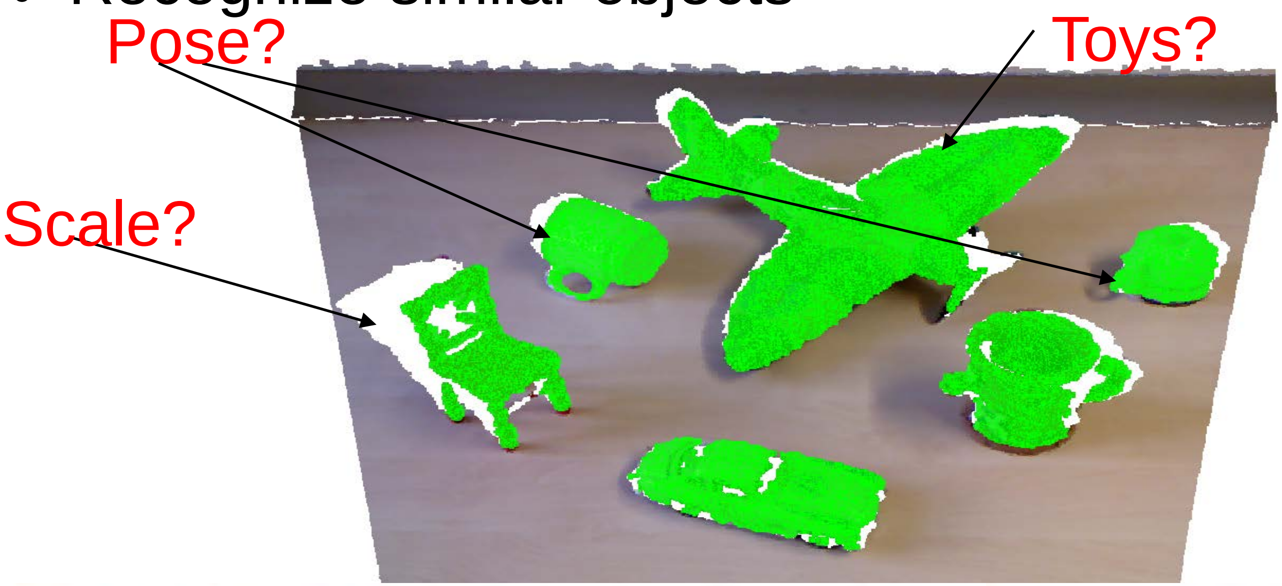
Plane Pop-out

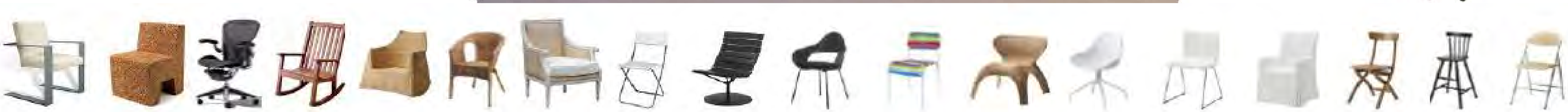
- Horizontal surfaces present many object classes
  - Attention to model (= plane) instances
- Simultaneous exploration of hypothesis space via particles moving in parameter space



Challenge: Learning many Classes

- How to efficiently train large number of categories?
  - Hundreds of classes in homes, pose for manipulation
  - Scale: dining chair to daughters puppet chair
- How to cope with large intra-class variance?
  - Recognize similar objects



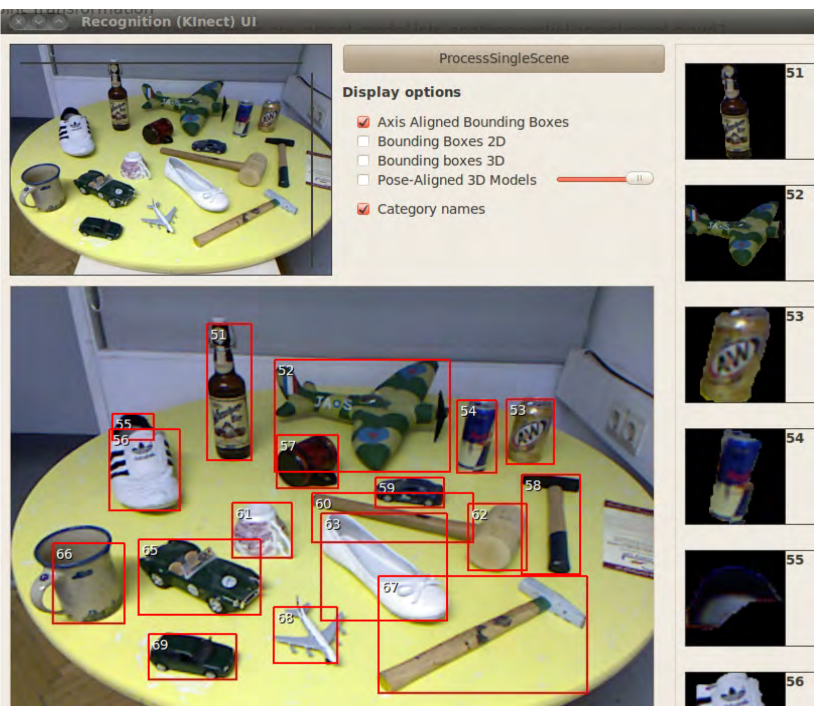


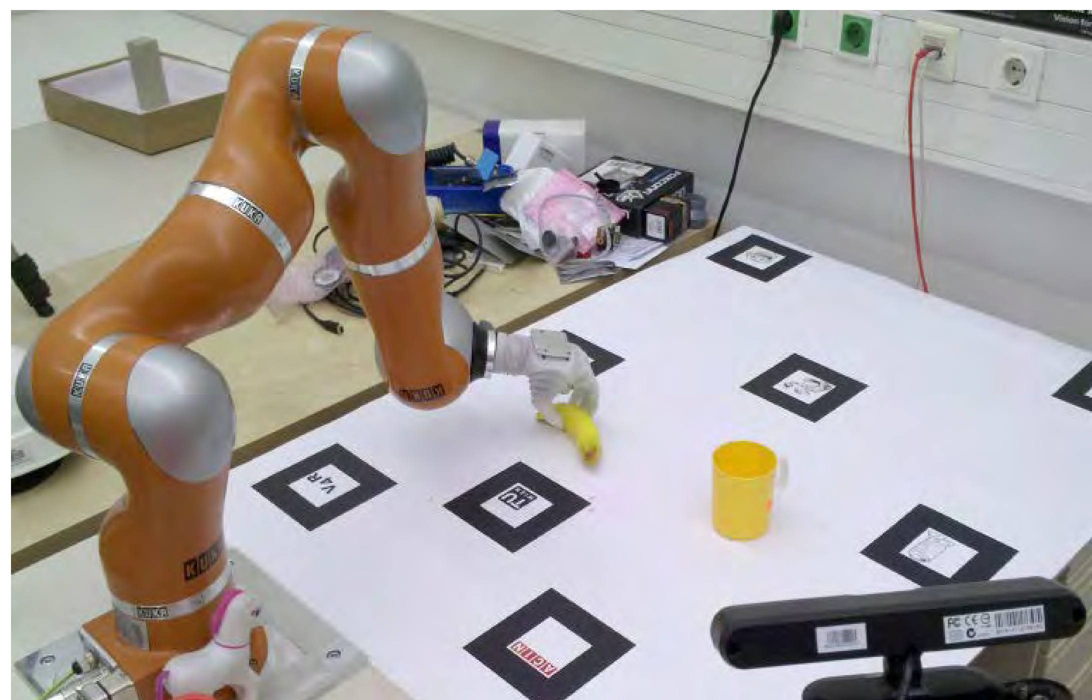
Object Class Detection

200 object classes: http://www.3d-net.org

Class-based object grasping

- Inherit grasp hypothesis from best class model
- Grasp force according to class
- Sort according to superclass





Pose Alignment – Camera Roll Histogram

Class probability is given

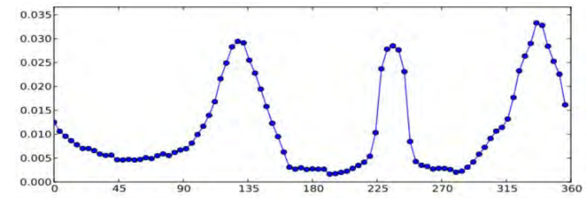
Best model view from

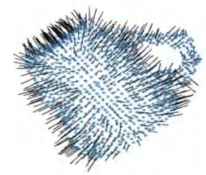
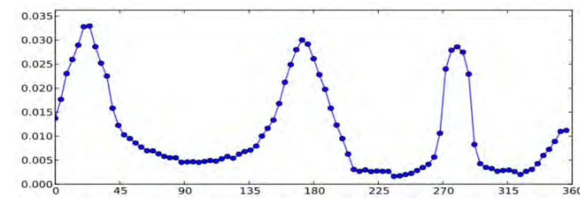
- NN search on highly probable classes from random forest

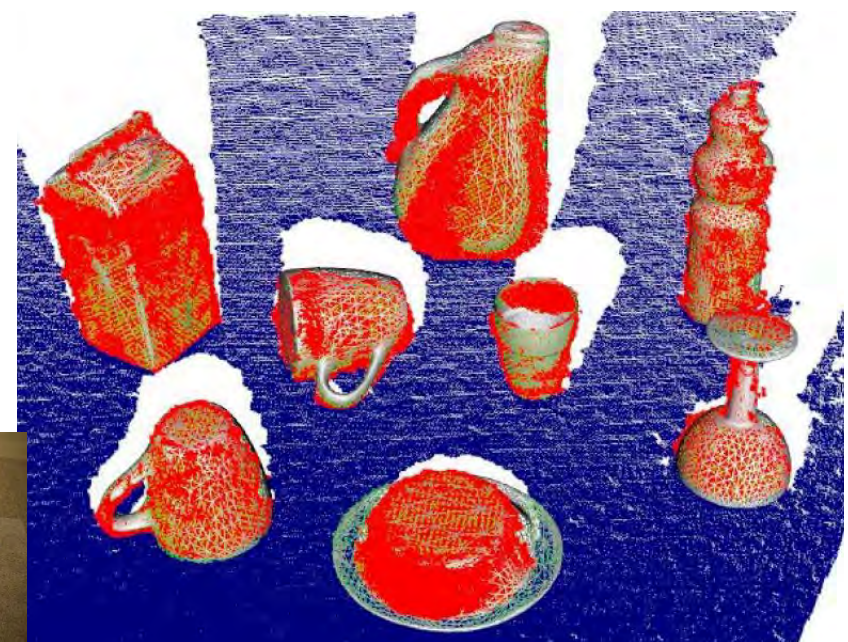
CRH to estimate roll angle

Scale from max distance of real to model view

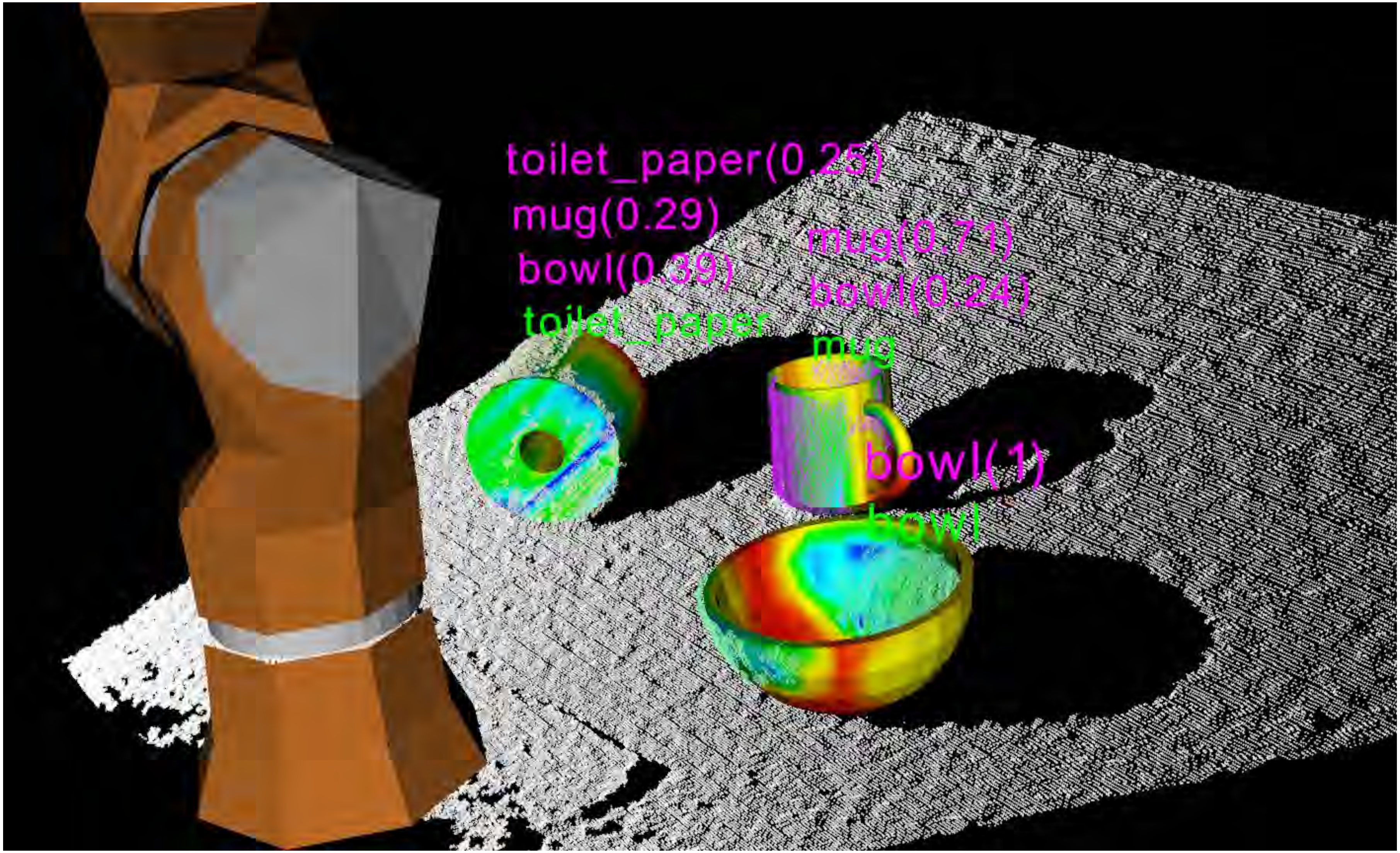
[ICCV WS 2011]

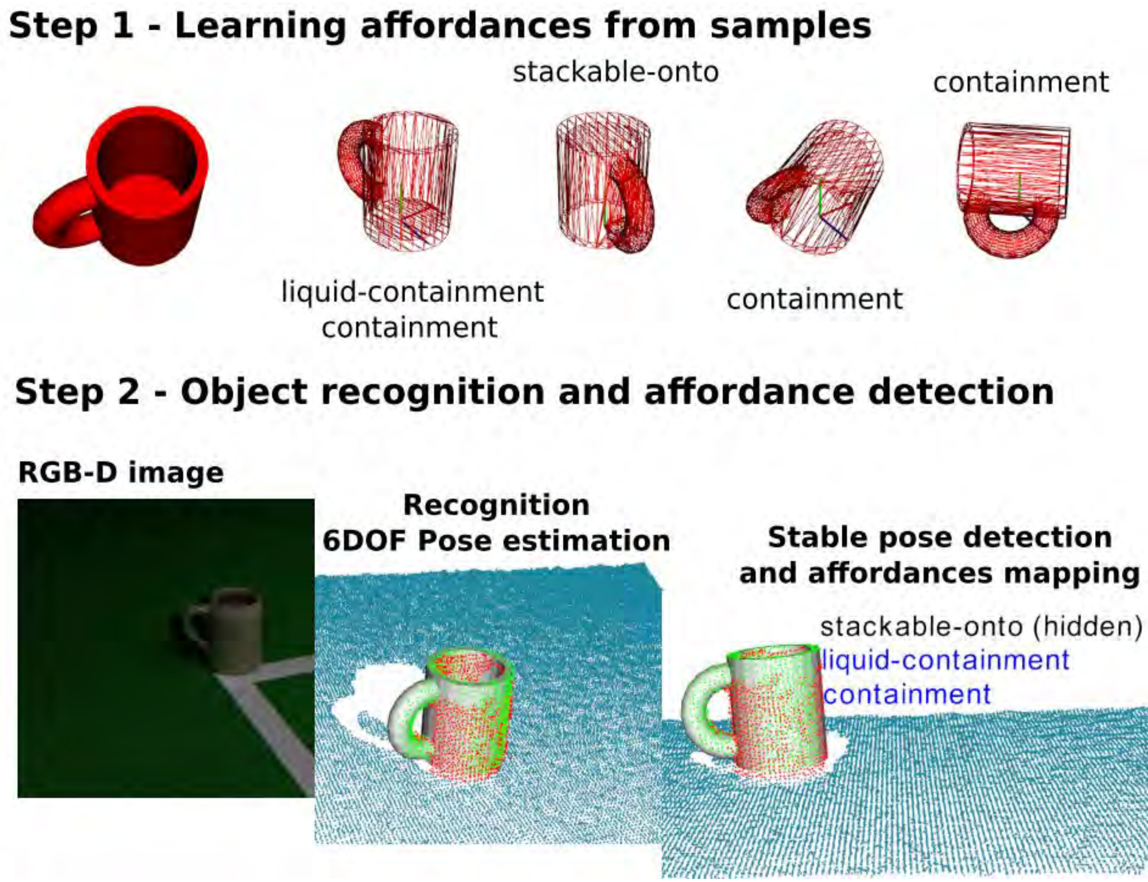


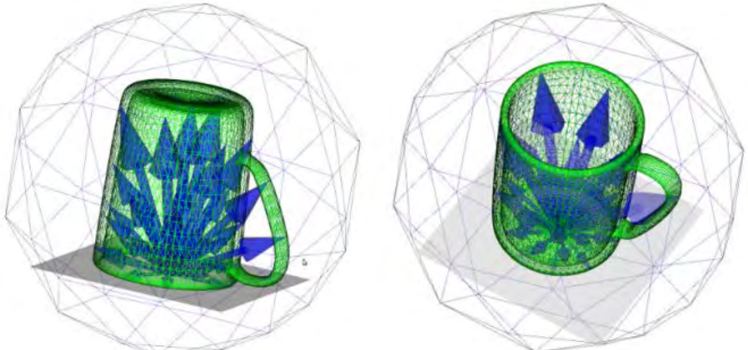

Results Categorization



Object Affordances

- Pose leads to affordances [Aldoma, ICRA 2012]
- Hidden affordance: need to change pose to use it

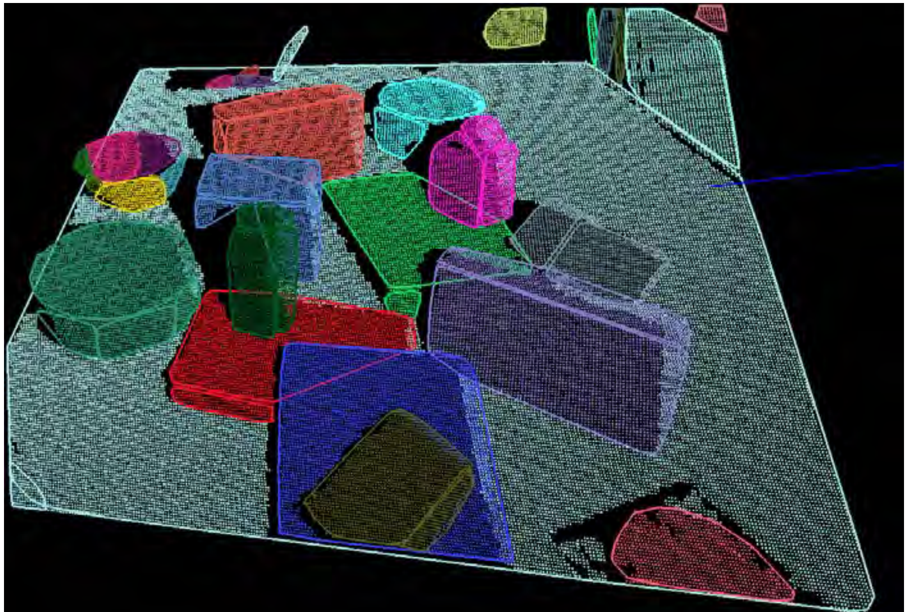




3D Categorisations and Grasping

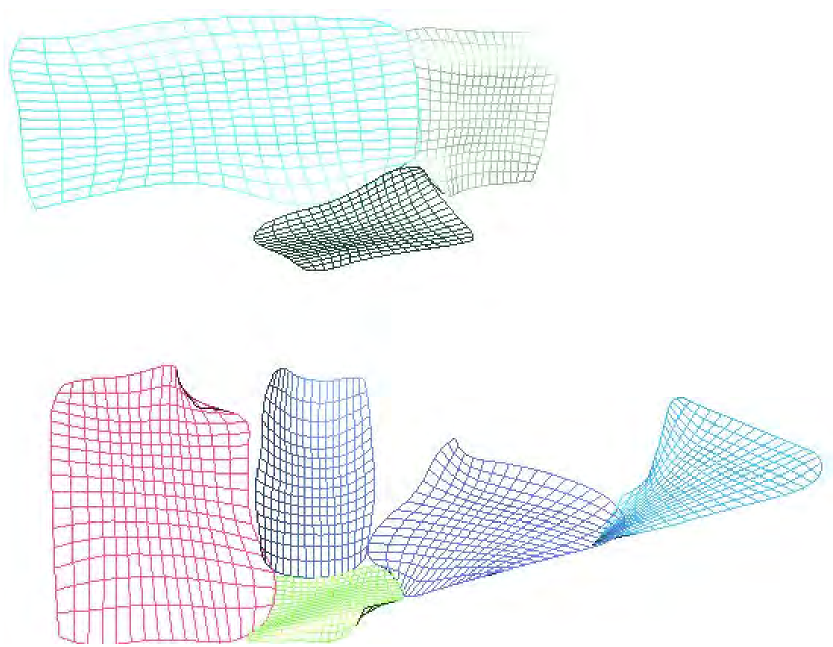
Scene Modelling with basic Shapes (Gestalts)

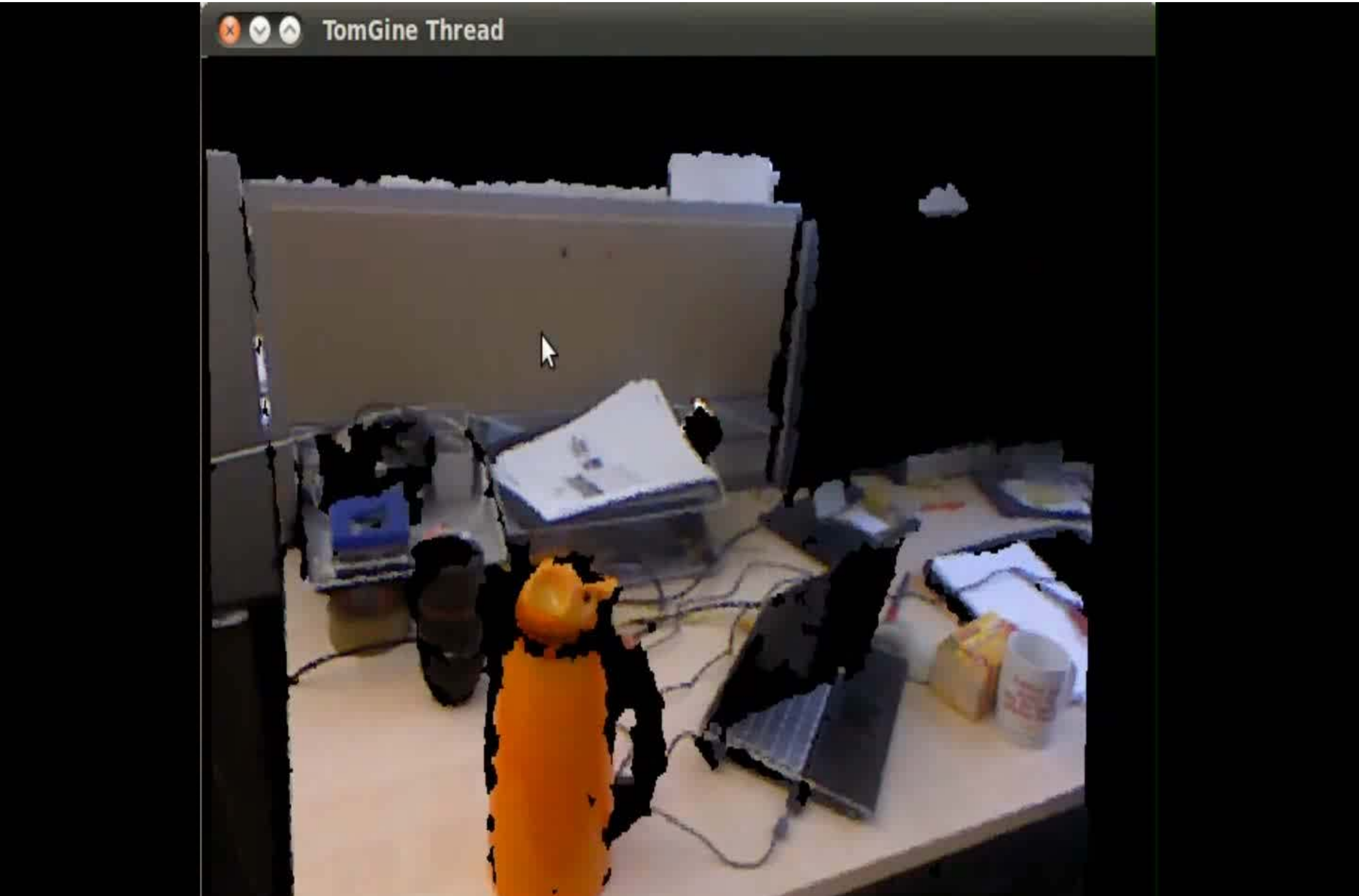
- Learning Gestalt-rules from sample scenes
- Grouping features from image and depth data
- Grouping surfaces/NURBS to regular bodies
- Used to initialise tracking model



Surface Modelling with NURBS

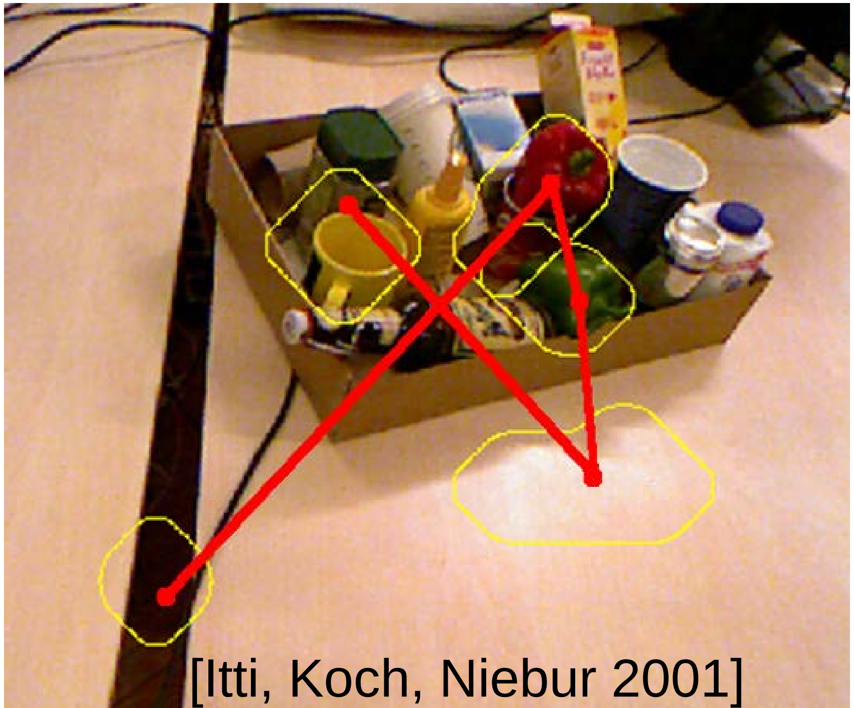
- Modelling higher order surfaces
- Fitting NURBS to segmented point-clouds
- Iterative adaption of borders to image sequences



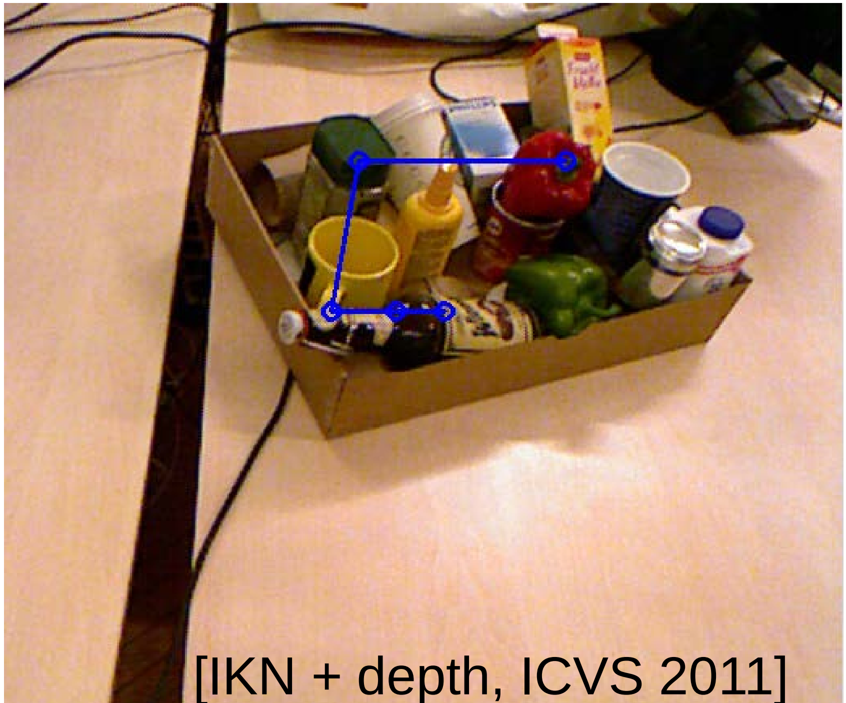

TomGine Thread

Task-based Attention

- Visual search driven by task (get me X) and partial knowledge (things on tables, in kitchens)
- Integrate 2D and 3D preattentive cues



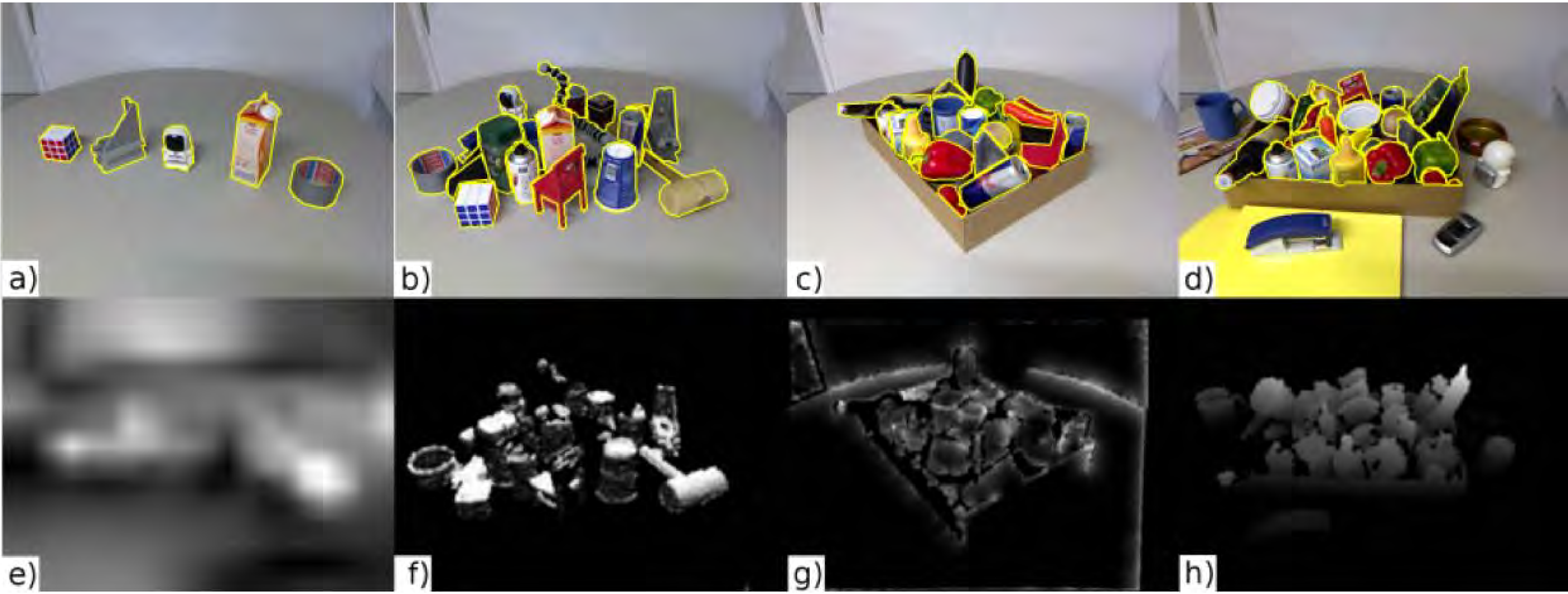
[Itti, Koch, Niebur 2001]



[IKN + depth, ICVS 2011]

Basket: Example of Attention Points

- From single objects to clutter
  - Random pose, objects close to each other
- Attention on potential object(s) to grasp
  - Combine 2D and 3D cues [ICVS 2011]

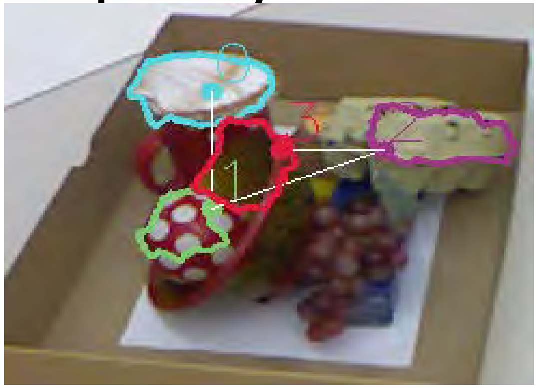
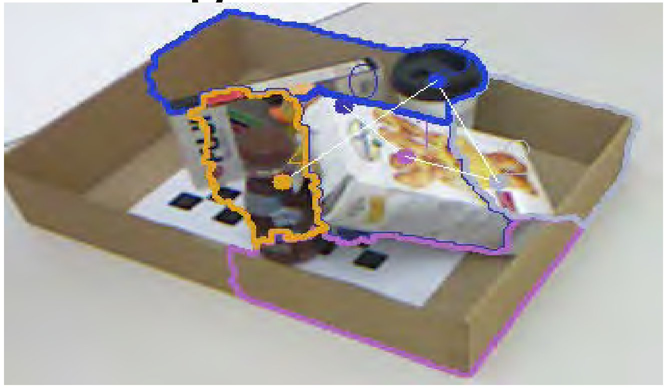
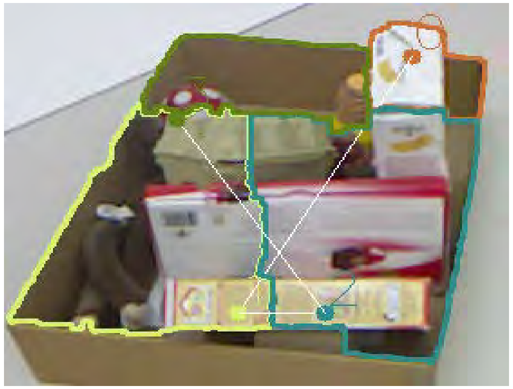


IKN [2002] | Rel. surf. ori. | Occl. edges | Surf. height

Attention-based Segmentation

Guided segmentation with attention points
[Ko et al. 2006, Mishra et al. 2009, Malik et al. 2001]

Problem: Guided segmentation highly depends on quality of calculated edges

Attention-based Segmentation

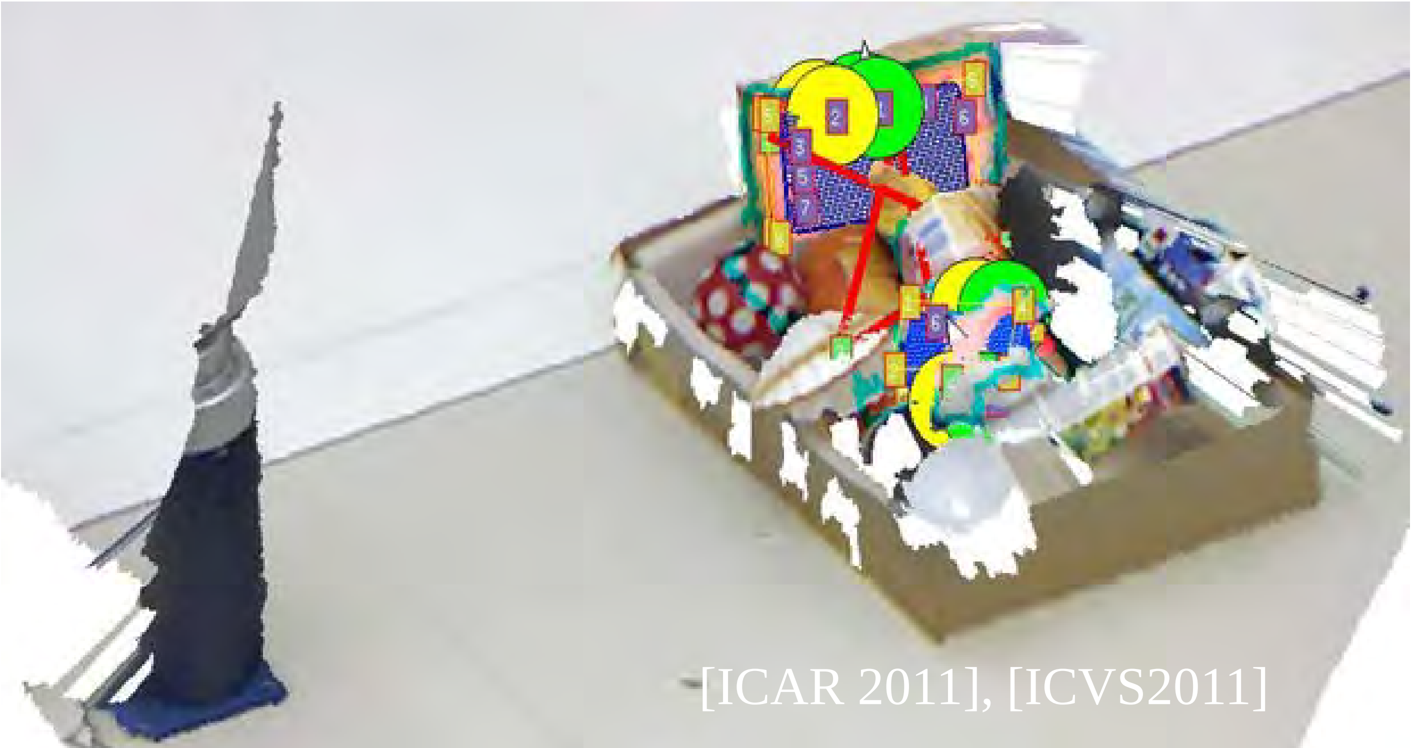
Probabilistic model integrates 2D and 3D cues

$$p(x = edge \mid s,d,c) = \frac{p(s \mid x = edge)\,p(d \mid x = edge)\,p(c \mid x = edge)\,p(s)\,p(d)\,p(c)}{p(x)}$$

s, d and c = color sobel edges, depth sobel edges and curvature (assumed to be independent)

Locally Fit Volumes for Grasp Hypotheses

- Fit superquadric in segment
  - Estimate 6 DOF Pose, scale, model primitive, deformation
  - Quality and coverage of fit measure segmentation quality
  - Parametrised grasp hypotheses



[ICAR 2011], [ICVS2011]

Empty a Box

Lessons Learned

- Natural interaction with robot?
  - Joystick and button? Wii? iPad? Kinect? Speech?
- User interface requires full technical functionality
  - Daily use at home? Soon abandoned?
- No SLAM to download to work at home
  - Too much clutter, table edges, shiny materials
  - Stereo, Laser, Kinect, …
  - Looking against windows
  - Navigation in tight spaces

Lessons Learned

- First steps to Embodied Object Classification
  - Task-based attention (humans: periphery only)
- Object Classification from 3D Web models
  - Many classes, many more instances
  - Novel objects from known classes
  - Transfer of grasp hypotheses
  - Affordances and functions
- Safe and affordable home robot
  - Technology vs. acceptance
  - Long-term usage

Thank you