Perceptual Grouping in RGB-D for Semantically Rich Visual Representations

Michael Zillich
zillich@acin.tuwien.ac.at

Vision for Robotics Group
Institute of Automation and Control
Vienna University of Technology

Hong Kong, May 31, 2014

Overview

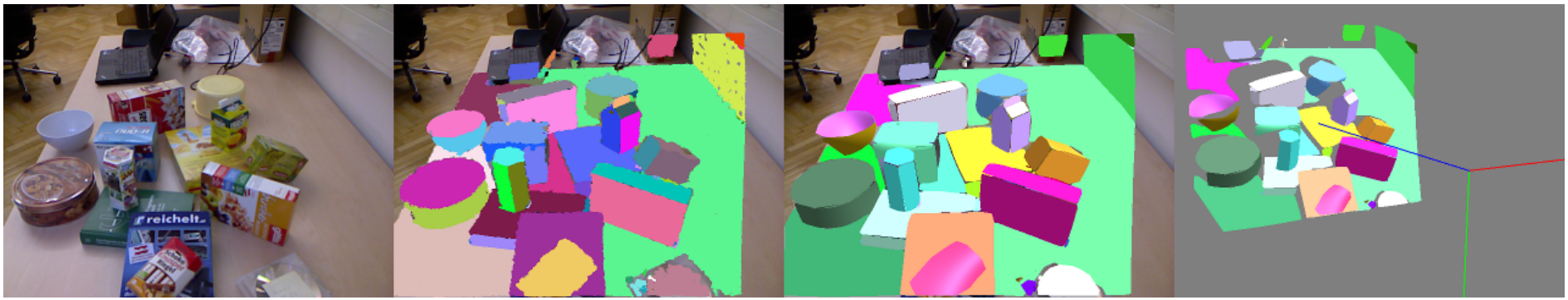
- General purpose vision
- Perceptual grouping for 3D object segmentation
- Applications
- Discussion

General purpose vision

- What does it mean to be general?
- Always purposes related to (robotics) tasks
- Less vs. “more” general
- Reflect structure of the world, independent of the tasks
- Reuse of representations for different tasks

Perceptual grouping for 3D object segmentation

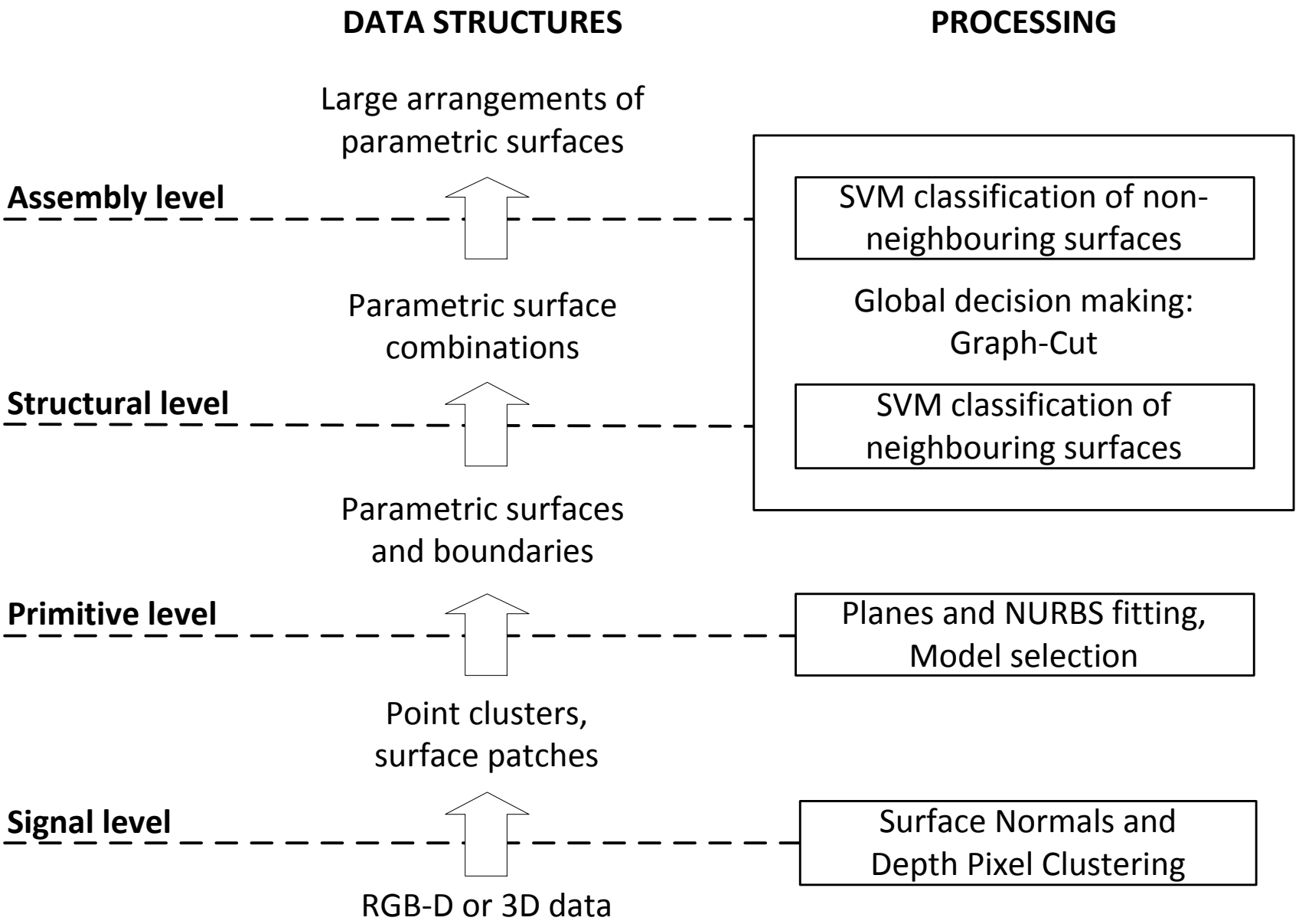
- Identify which bits of the scene could be objects
- Object hypotheses from RGBD images
- (Ückermann ea IROS 2012)
- (Mishra ea ICRA 2012)
- (Katz ea RSS 2013)
- (Hager ea IJRR 2011)



From coluoured point clouds .. .. to separated object hypotheses

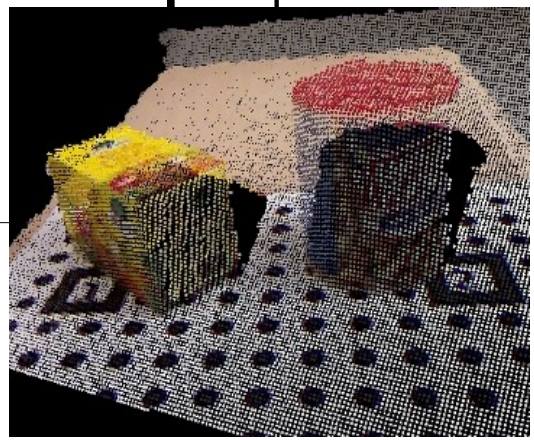
(Richtsfeld ea 2014)

Hierachy overview

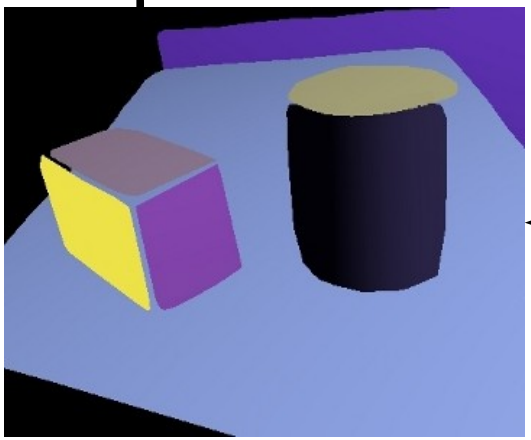


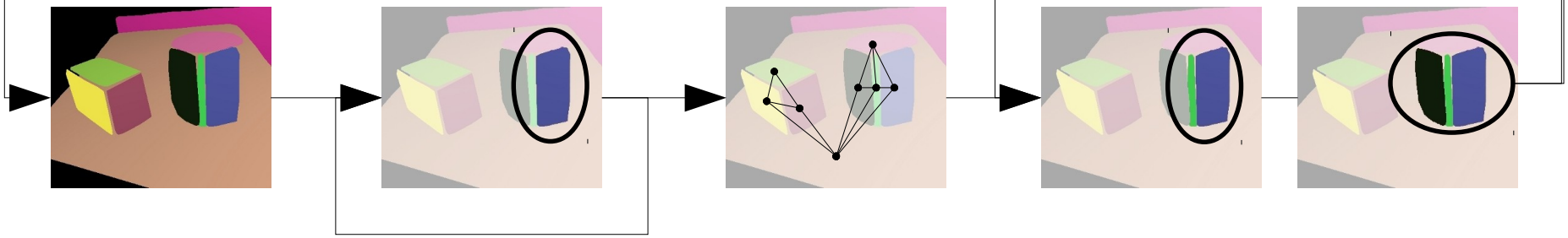
Fitting surface patches, model selection

Input: point cloud



Output: Planes / NURBS





segment patches

select patch P_i
compare models
plane ↔ NURBS

$S_N > S_P$

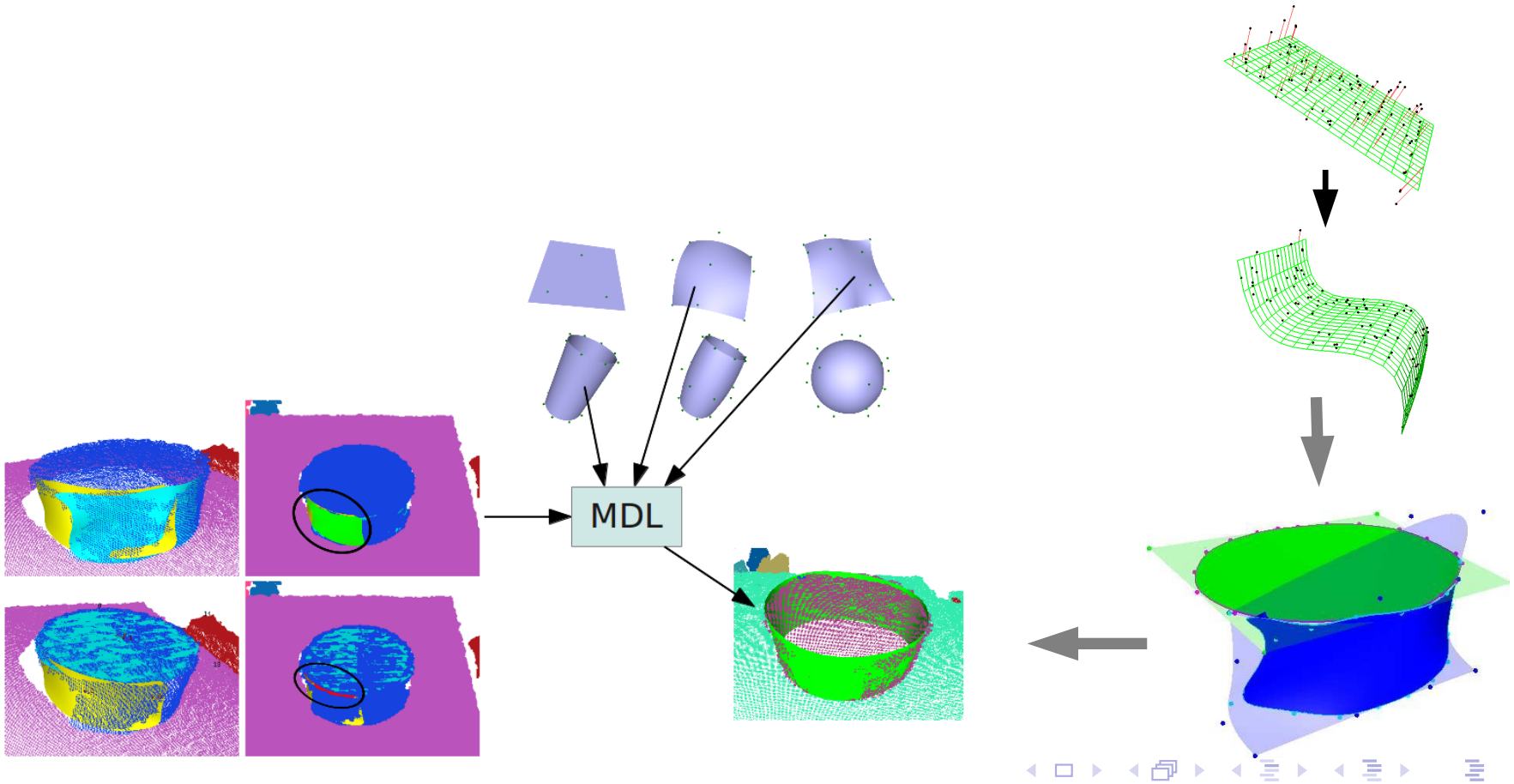
compute neighbours

Greedily fit NURBS to neighbours P_{ij} and compare models

$S_{ij} > S_i + S_j$

NURBS fitting

- Problems with noise and missing data
- Fit NURBS surface to point-cloud
- Fit B-spline curve to point cloud on parametric domain of the surface
- Multi view reconstruction using MDL



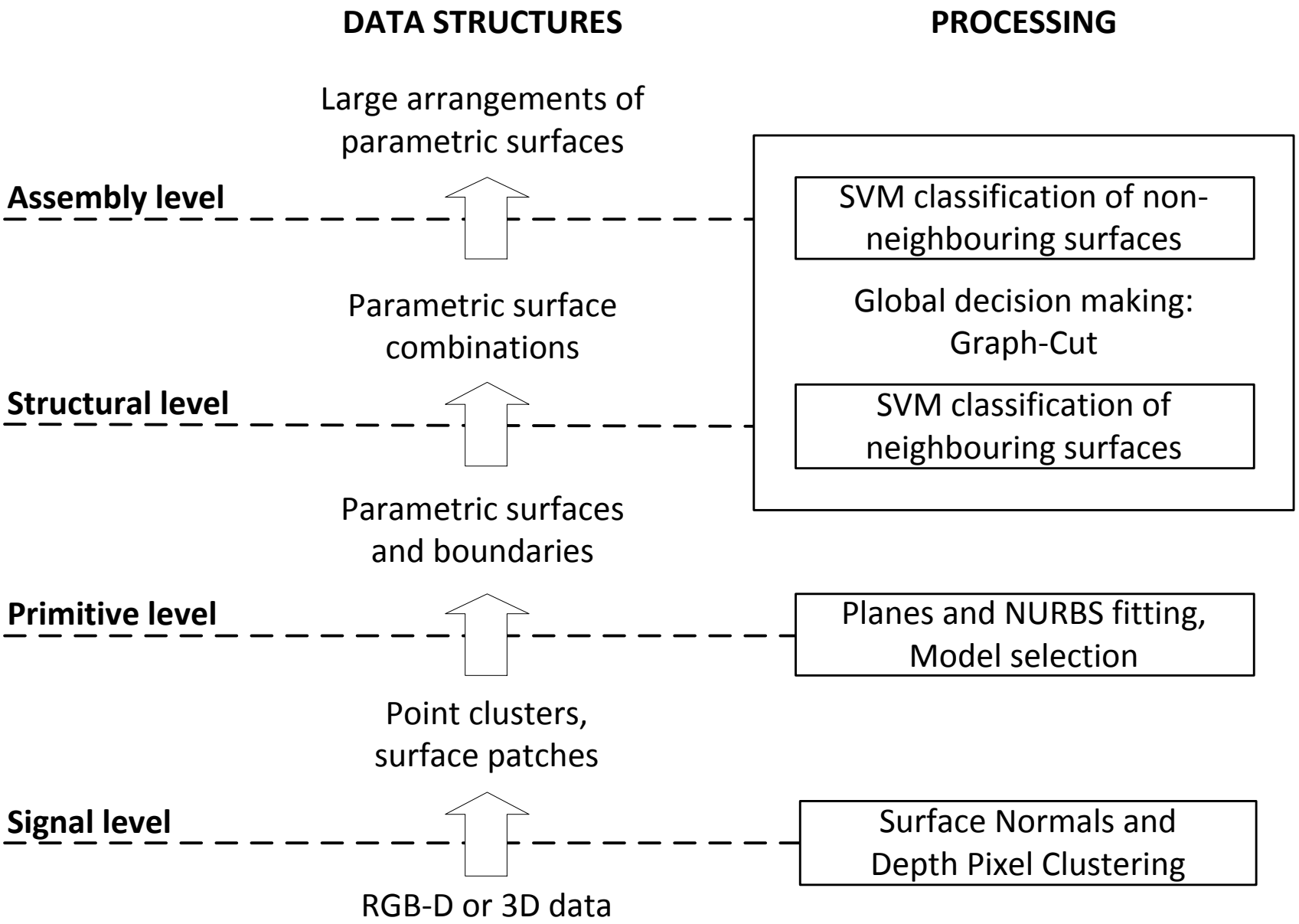
NURBS fitting

Accurate boundries in heavy clutter

Various types of surface shapes

(Mörwald ea 2013)

Hierachy overview



Gestalt principles

- Proximity
- Similarity
- Continuity
- Closure
- Symmetry
- Common region
- Element connectedness
- Common fate
- Good Gestalt

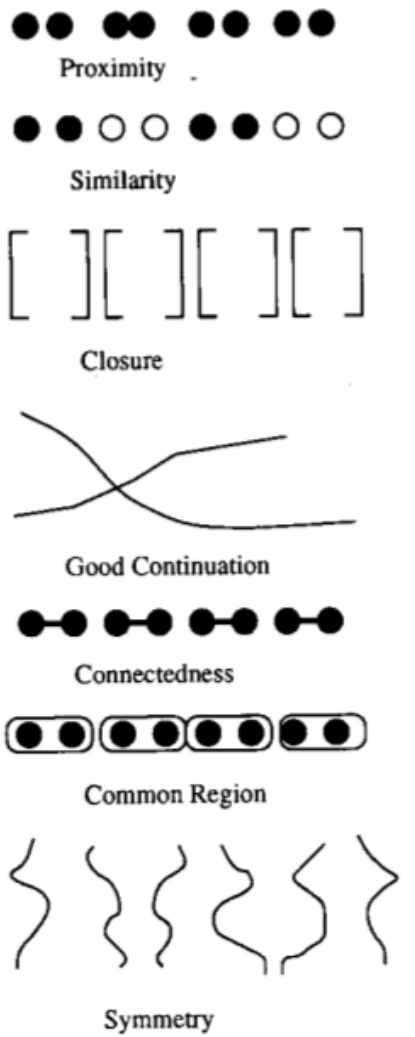
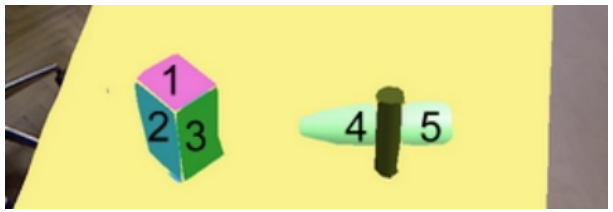


Fig. 3. Gestalt laws of grouping.

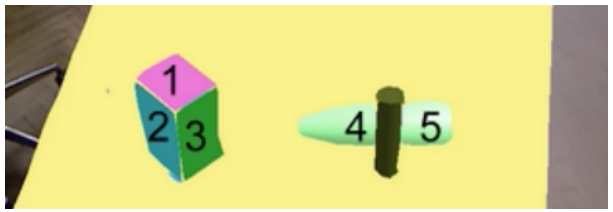
Surface relations - neighbours

- r_{co} ... similarity of patch colour
- r_{rs} ... relative patch size similarity
- r_{tr} ... similarity of patch texture quantity
- r_{ga} ... gabor filter match
- r_{fo} ... fourier filter match
- r_{co3} ... color similarity on 3D patch borders
- r_{cu3} ... mean curvature on 3D patch borders
- r_{cv3} ... curvature variance on 3D patch borders
- r_{di2} ... mean depth on 2D patch borders
- r_{vd2} ... depth variance on 2D patch borders



Surface relations - non-neighbours

- r_{co} ... similarity of patch colour
- r_{rs} ... relative patch size similarity
- r_{tr} ... similarity of patch texture quantity
- r_{ga} ... gabor filter match
- r_{fo} ... fourier filter match
- r_{md} ... minimum distance between patches
- r_{nm} ... angle between mean surface normals
- r_{nv} ... difference of variance of surface normals
- r_{ac} ... mean angle of normals of nearest contour p.
- r_{dn} ... mean distance in normal direction of nearest contour p.

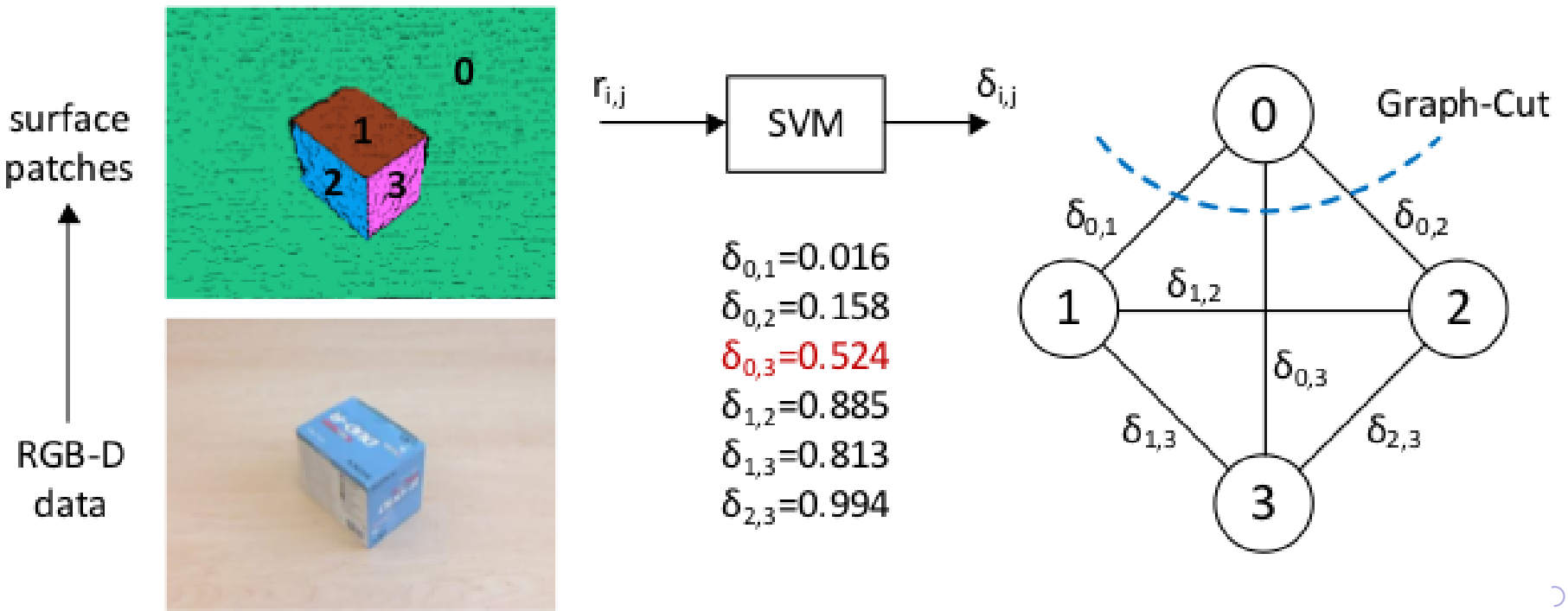


Global decision using graph cut

- Train Support Vector Machines (SVMs) on feature vectors, using annotated training data

$$R_{st} = (r_{co}, r_{rs}, r_{tr}, r_{ga}, r_{fo}, r_{co3}, r_{cu3}, r_{cv3}, r_{di2}, r_{vd2})$$
$$R_{as} = (r_{co}, r_{rs}, r_{tr}, r_{ga}, r_{fo}, r_{md}, r_{nm}, r_{nv}, r_{ac}, r_{dn})$$

- Use predicted probability of "same object" as pairwise terms for graph cut



Examples

Feature significance

F-Score

$$F(i) = \frac{(\bar{r}_i^{(+)} - \bar{r}_i)^2 + (\bar{r}_i^{(-)} - \bar{r}_i)^2}{\sigma_i^{(+)^2} + \sigma_i^{(-)^2}}$$

Balanced error rate

$$BER_{svm} = \frac{1}{2} * (\frac{fp}{tp + fp} + \frac{fn}{tn + fn})$$

	F_{score}	BER_{svm}	P	R	P^*	R^*
$r_{st} = \{r_{co}\}$	0.163	37.5%	19.89%	92.38%	89.97%	93.90%
$r_{st} = \{r_{rs}\}$	0.185	38.7%	18.11%	92.39%	87.27%	93.80%
$r_{st} = \{r_{tr}\}$	0.094	43.1%	22.99%	93.56%	90.76%	93.89%
$r_{st} = \{r_{ga}\}$	0.183	40.1%	25.01%	93.55%	90.85%	93.90%
$r_{st} = \{r_{fo}\}$	0.206	43.2%	39.40%	93.94%	90.33%	93.88%
$r_{st} = \{r_{co3}\}$	0.191	38.4%	33.33%	89.04%	91.40%	93.89%
$r_{st} = \{r_{cu3}\}$	1.418	19.6%	81.81%	94.05%	63.71%	93.61%
$r_{st} = \{r_{cv3}\}$	0.001	49.9%	5.83%	97.53%	91.29%	93.82%
$r_{st} = \{r_{di2}\}$	0.453	27.9%	27.21%	93.83%	91.29%	93.81%
$r_{st} = \{r_{vd2}\}$	0.687	27.3%	26.10%	94.01%	90.71%	93.80%
$r_{st} = \{r_{2d3}\}$	0.342	27.0%	33.35%	93.12%	90.70%	93.70%
$\mathbf{r_{st}}$		**16.7%**	**90.85%**	**93.88%**		

Segmentation results

	Mishra		Ückermann		SVM_{st}		SVM_{st+as}	
	P	R	P	R	P	R	P	R
Boxes	76.87%	75.86%	97.12%	94.72%	96.47%	97.91%	96.47%	97.91%
Stacked	70.57%	74.61%	95.61%	93.26%	86.70%	96.23%	86.72%	97.54%
Occluded	67.37%	55.81%	94.53%	74.76%	94.18%	78.23%	94.00%	91.62%
Cylindric	69.81%	87.38%	96.47%	92.50%	96.21%	97.11%	87.35%	97.71%
Mixed	62.99%	76.29%	95.27%	93.42%	91.21%	95.90%	91.21%	95.90%
Complex	61.06%	54.61%	93.14%	83.49%	87.50%	91.49%	86.78%	92.09%
OSD	66.10%	67.91%	94.91%	88.79%	90.85%	93.88%	89.95%	95.00%
Willow	77.51%	83.82%			98.11%	98.82%	98.10%	98.81%

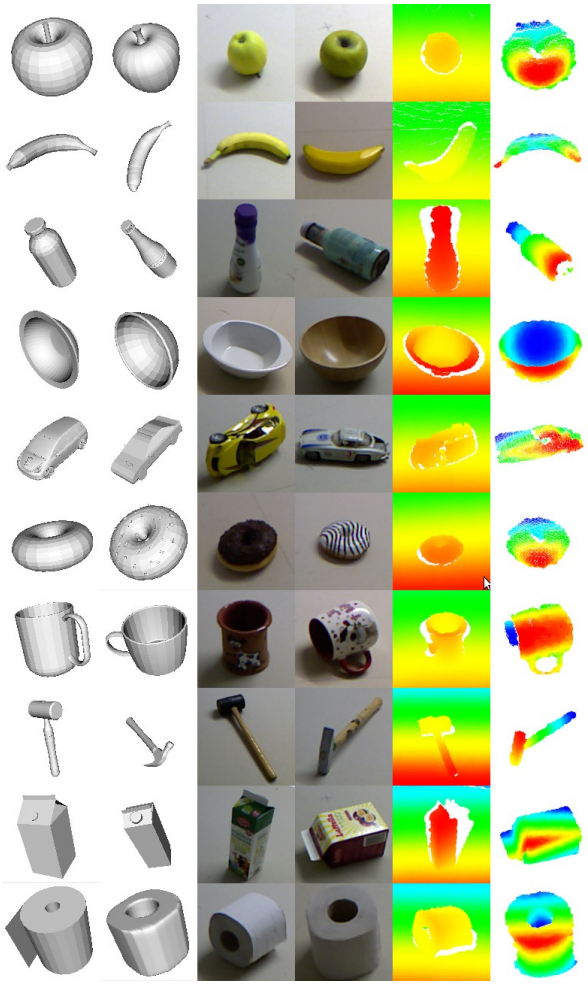
OSD: The Object Segmentation Database,
`http://www.acin.tuwien.ac.at/?id=289` (2012)
Willow: Challanges in perception database, provided by Willow Garage

Applications

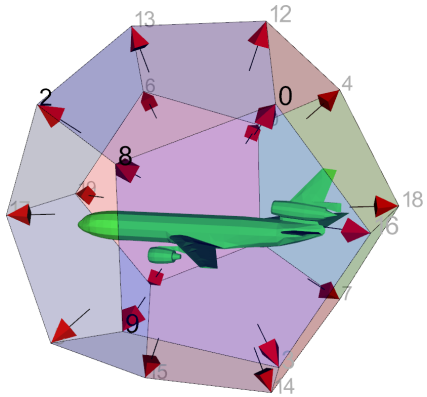
One representation - several tasks

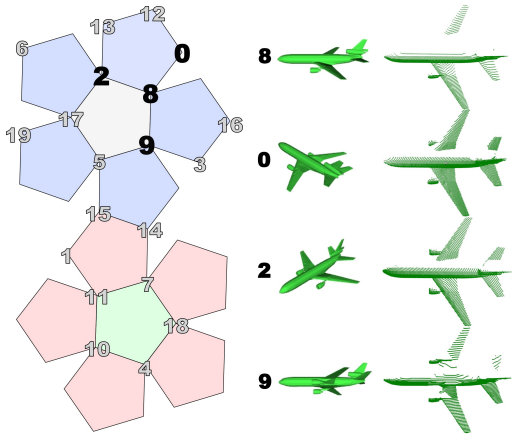
- Object classification
- Physics based tracking
- One shot learning
- (object recognition)
- (object grasping)
- (attention guided incremental segmentation)

Object classification

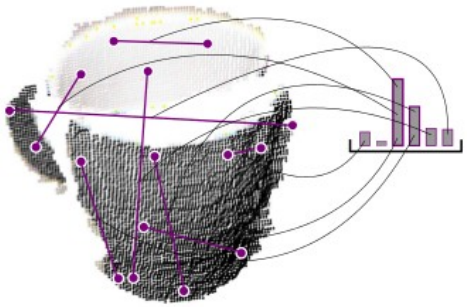


Train from 3D web models ..





.. by learning 3D descriptors of generated views ..



VFH, CVFH, SDVS, SHOT, ESF

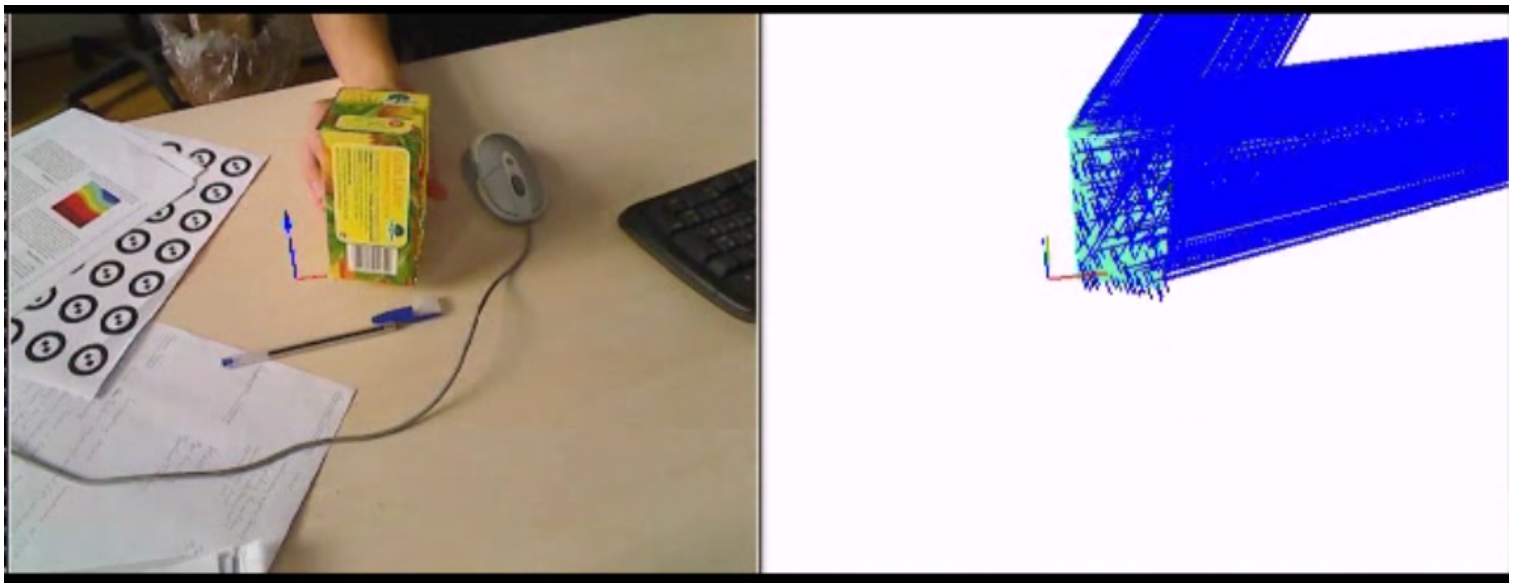
Nearest Neighbour Classifier

(Wohlkinger ea 2012)

Object classification

Physics based tracking

- Particle filter based tracking using learned shape models
- Adding texture and views for recognition on the fly

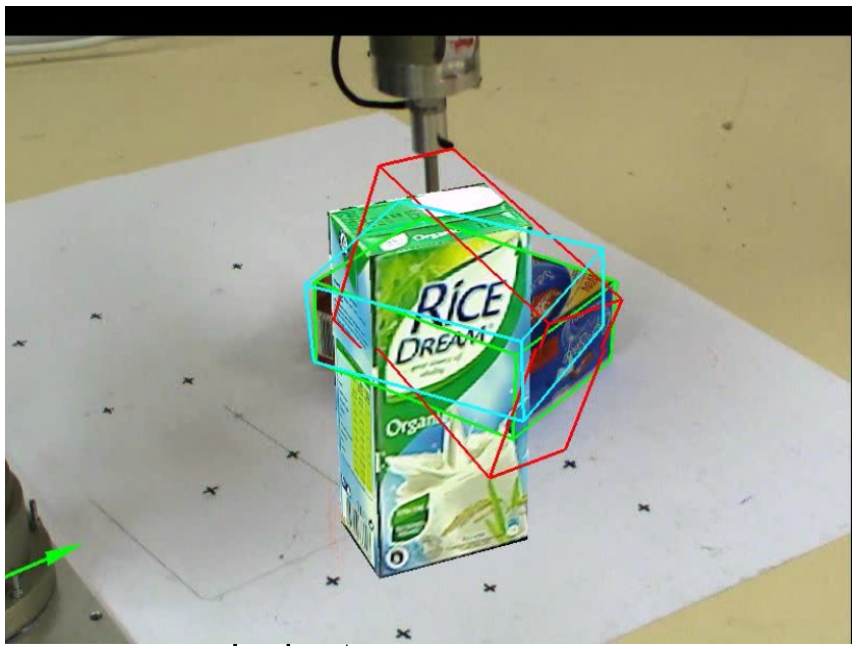


(Video)

(Mörwald ea 2013)

Physics based tracking

- Replace simplistic motion model in particle filter with actual physics model
- Physics engines are difficult to parameterise $\Rightarrow$ learn physics model, using global and local shape properties
- KDE to learn predictive model of motion given a particular interaction (Kopicki ea ICAR'09) (Birmingham Univ.) (Mörwald ea 2011)



prediction, tracking, tracking + prediction (Video)

One shot learning with language

Rich representation allows to talk about object properties, like global shape, colours and texture on specific surfaces.

H: Do you see a medkit?

R: What does a medkit look like?

H: It is a white box with a red cross on it.

R: Ok.

(Krause ea 2014)

Discussion

General purpose vision and active visual learning?

1. Generic object hypotheses from perceptual grouping
2. Hierarchical representation allows to probe various levels
3. Representations "semantically" rich enough, allow to attach physical or linguistic meaning

Thanks

Thanks to

- Andreas Richtsfeld, Johann Prankl, Thomas Mörwald, Ekaterina Potapova, Walter Wohklinger
- Marek Kopicki, Evan Krause, Matthias Scheutz

An advertisement ..

3rd Workshop on Robots in Clutter:
Perception and Interaction in Clutter

IROS 2014, Chicago, Illinois, Sept. 18, 2014
http://workshops.acin.tuwien.ac.at/clutter2014

Submissions due: July 18, 2014

References

- A. Ückermann, R. Haschke, and H. Ritter, Real-Time 3D Segmentation of Cluttered Scenes for Robot Grasping, IROS 2012.
- A. K. Mishra, A. Shrivastava, Y. Aloimonos, Segmenting Simple Objects Using RGB-D, ICRA 2012.
- D. Katz, A. Venkatraman, M. Kazemi, J. A. Bagnell, and A. Stentz, Perceiving, Learning, and Exploiting Object Affordances for Autonomous Pile Manipulation, in RSS, 2013.
- Gregory D Hager and Ben Wegbreit. Scene parsing using a prior world model. IJRR 2011.
- E. Krause, M. Zillich, T. Williams, and M. Scheutz, Learning to Recognize Novel Objects in One Shot through Human-Robot Interactions in Natural Language Dialogues in AAAI, 2014 (to appear).
- Wohlkinger, W., Vincze, M. Shape-Based Depth Image to 3D Model Matching and Classification with Inter-View Similarity. IROS 2011.

References

- Mörwald, T., Kopicki, M., Stolkin, R., Wyatt, J., Zurek, S., Zillich, M., Vincze, M. Predicting the Unobservable: Visual 3D Tracking with a Probabilistic Motion Model. ICRA 2011.
- Wohlkinger, W., Buchaca, A. A., Rusu, R., Vincze, M. 3DNet: Large-Scale Object Class Recognition from CAD Models. ICRA 2012.
- Richtsfeld, A., Mörwald, T., Prankl, J., Zillich, M., Vincze, M. Segmentation of Unknown Objects in Indoor Environments. IROS 2012.
- Mörwald, T., Richtsfeld, A., Prankl, J., Zillich, M., Vincze, M. Geometric data abstraction using B-splines for range image segmentation. ICRA 2013.
- T. Mrwald, J. Prankl, M. Zillich, and M. Vincze, Advances in real-time object tracking - Extensions for robust object tracking with a Monte Carlo particle filter. Journal of Real-Time Image Processing, December, 2013.