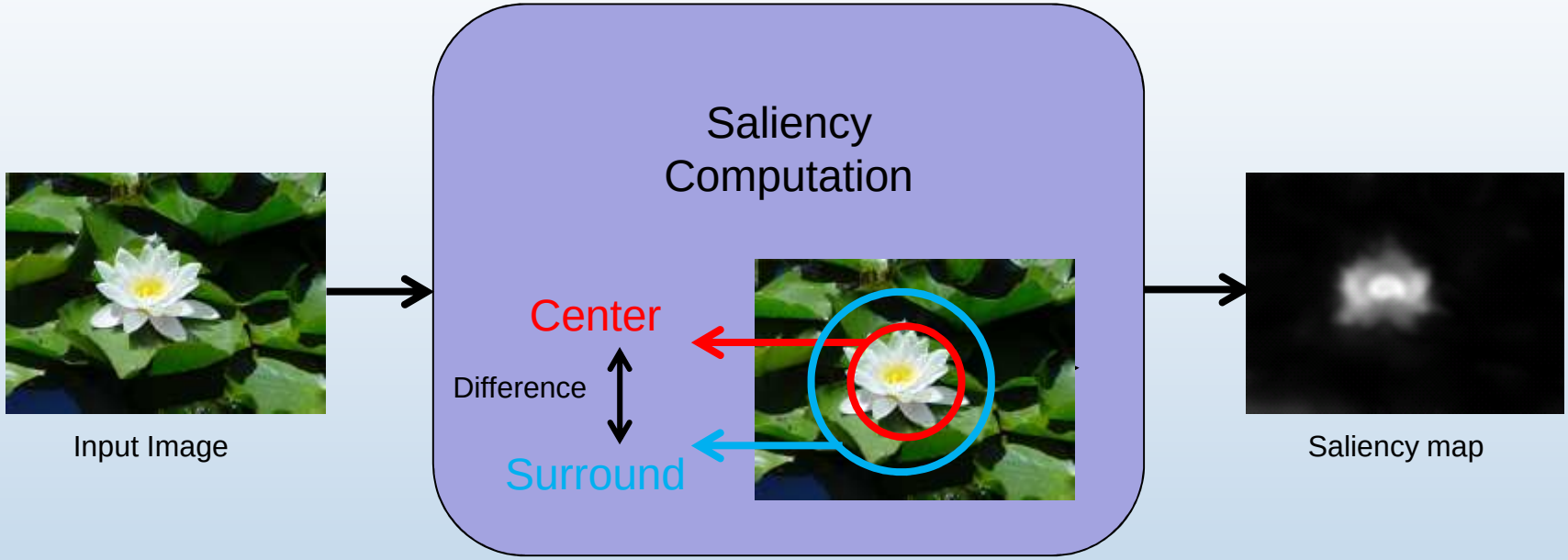
universitätbonn

IVS intelligent vision systems

An Attentional Framework for 3D Object Discovery

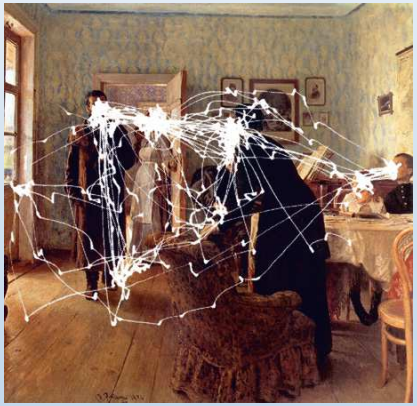
Germán Martín García and Simone Frintrop
Cognitive Vision Group
Institute of Computer Science III
University of Bonn, Germany

Saliency Computation



Attention vs Saliency

- Why visual attention is more than saliency computation:
- Attention is a process that operates over time
  - Look at one region after the other
  - „explore“ a scene by eye movements (saccades)



examine the painting freely

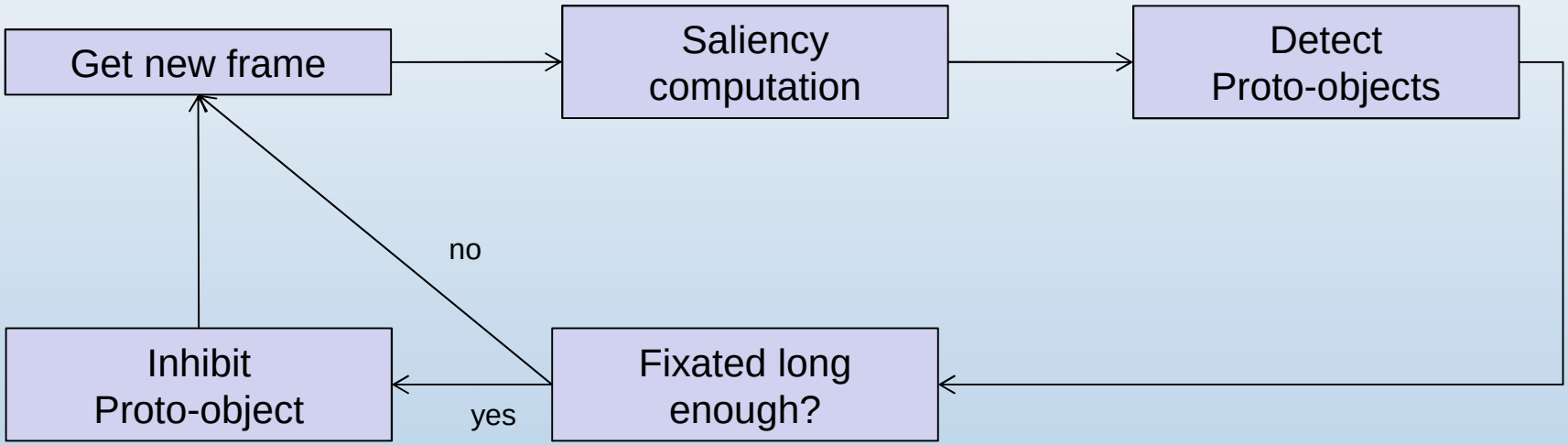
assess the ages of the characters

- The gaze is triggered by bottom-up (saliency) AND top-down cues

[Yarbus 1967; Images: Sasha Archibald, after Ilya Repin, *An Unexpected Visitor*, 1884)]

Visual attention

- Tasks to be solved by visual attention systems:
  - **Where** to direct the attention at (saliency + top-down cues)
  - **When** to switch
  - **How** to prevent returning to already visited locations

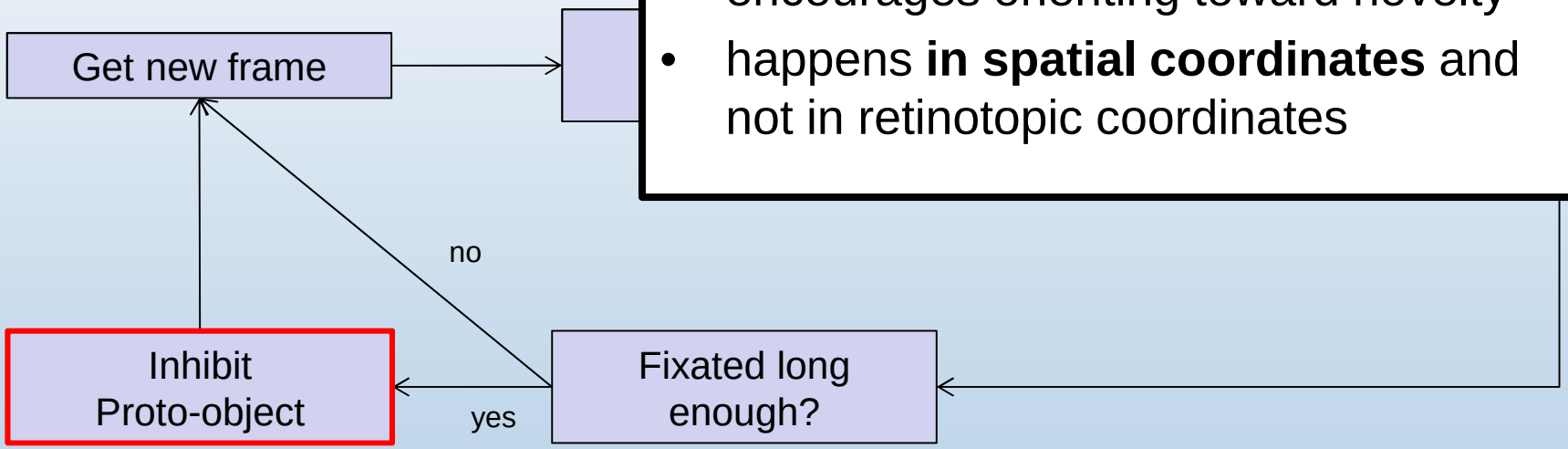


Visual attention

- Tasks to be solved b
  - **Where** to direct the a
  - **When** to switch
  - **How** to prevent retur

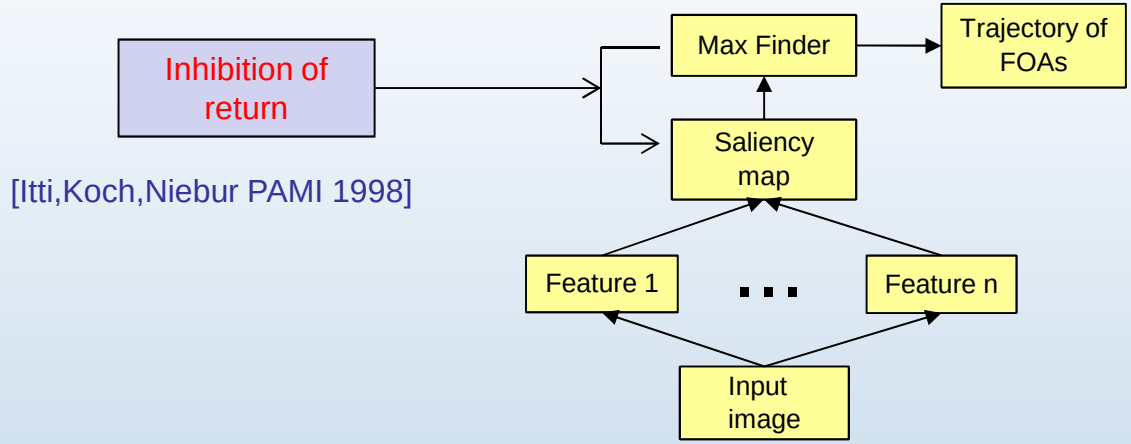
Inhibition of return (IOR) [Posner 1984]

- suppression of processing of locations and objects that had recently been the focus of attention
- encourages orienting toward novelty
- happens **in spatial coordinates** and not in retinotopic coordinates



Inhibition of Return

- Many systems simply compute the saliency map and ignore IOR
- Inhibition of return in computational attention systems:

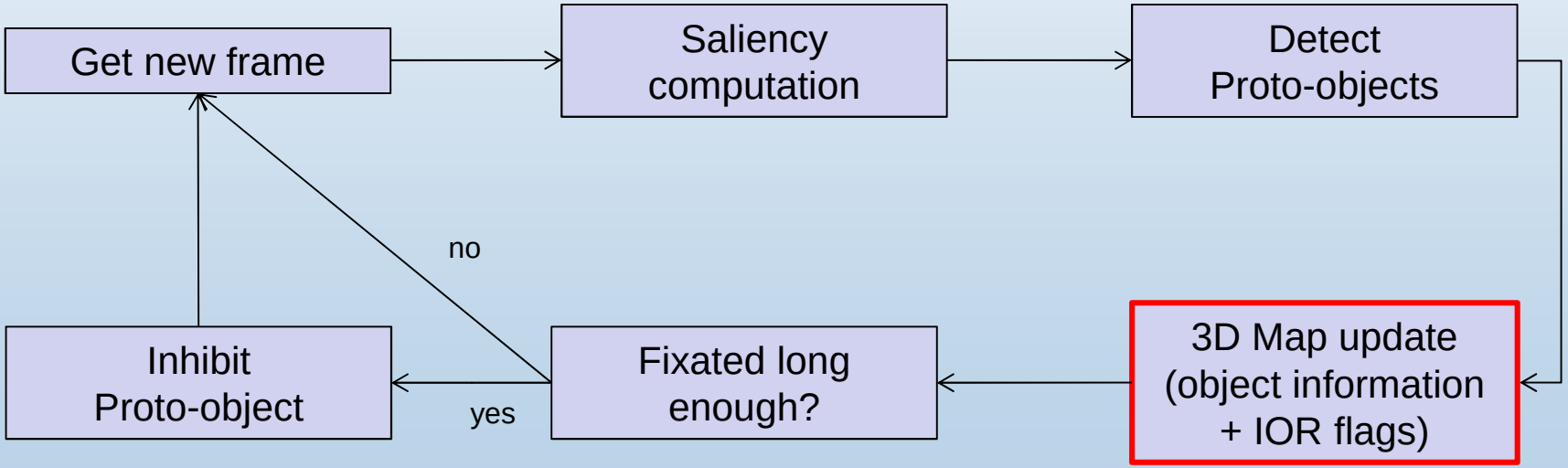


[Itti,Koch,Niebur PAMI 1998]

- If realized, usually by setting values to zero in saliency map.
- In temporal sequences, if camera positions and/or objects change location, this is not sufficient!
- It is necessary to remember which regions/objects have already been attended not in image but in spatial coordinates

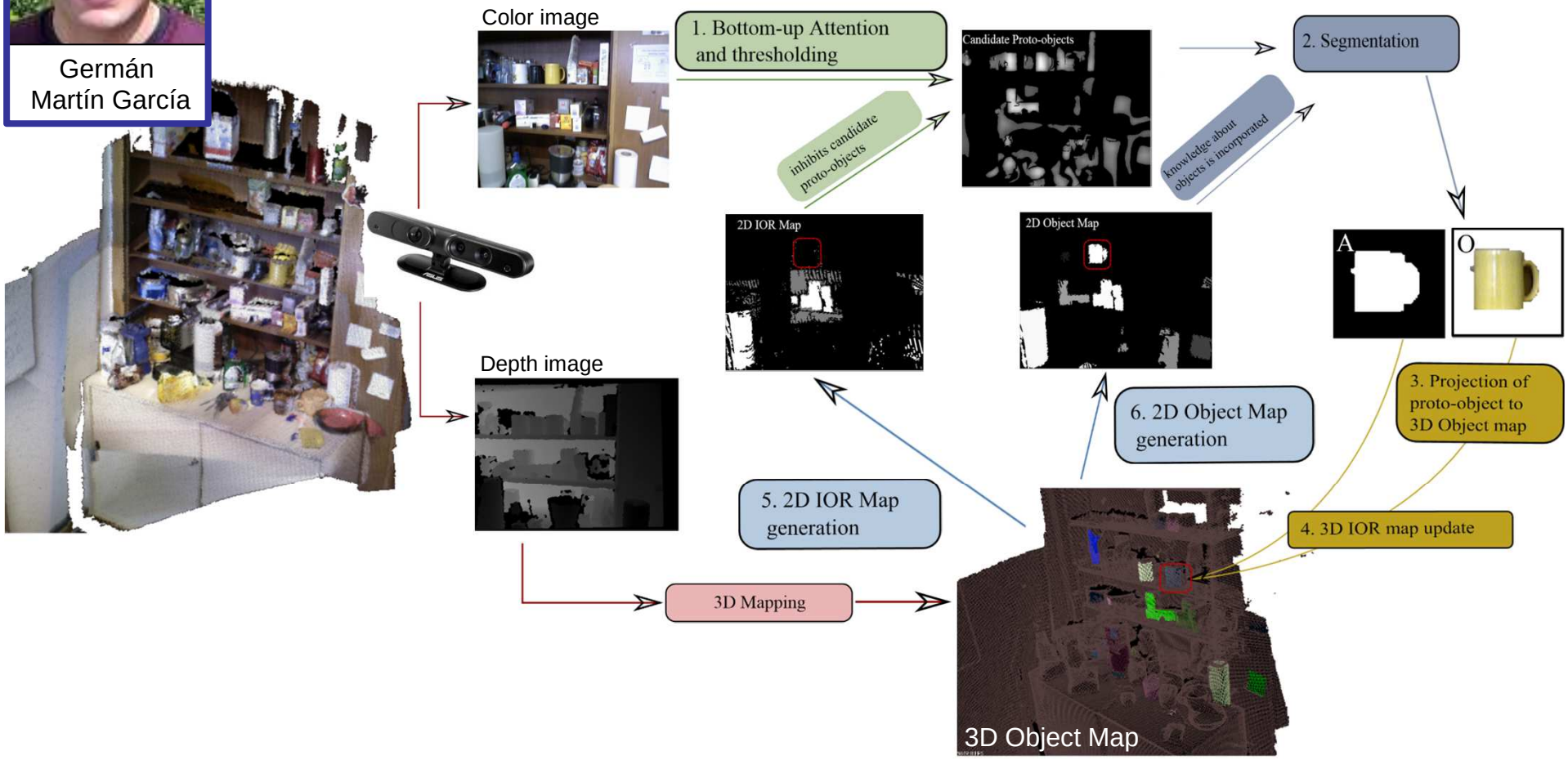
3D Object Discovery

- We propose an architecture for object discovery that operates in 3D (on RGB-D data)
- Object information is accumulated over time and 3D object models are built based on information from current and previous time steps
- 3D voxels are ‚tagged' by *inhibition of return flags* that indicate if a region was recently attended



Germán
Martín García

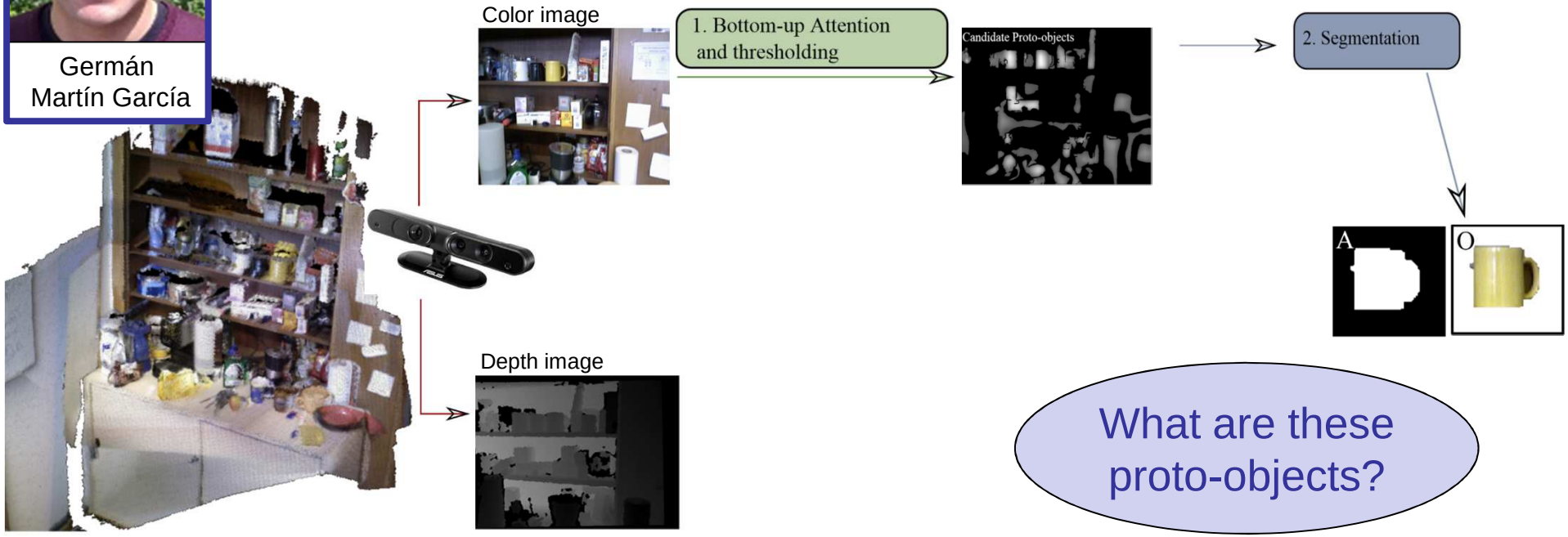
3D Object Discovery



[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

Germán
Martín García

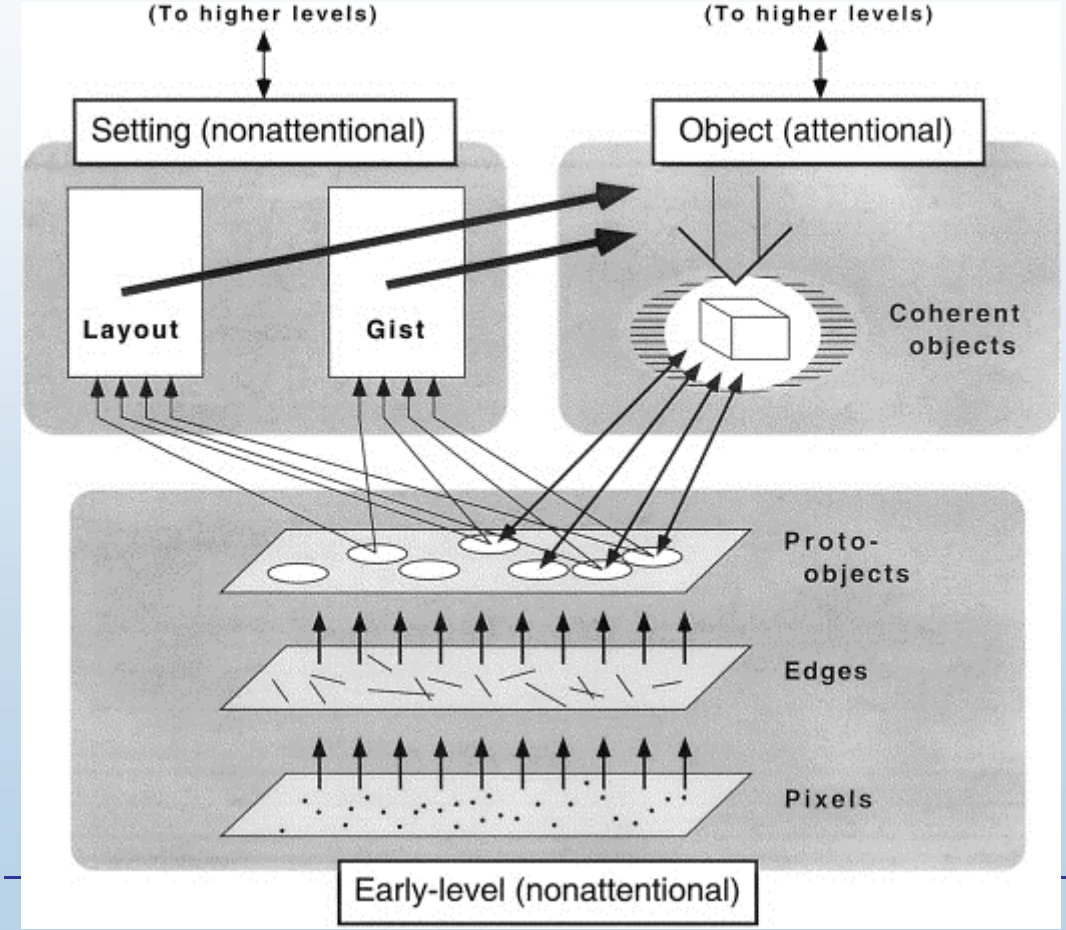
3D Object Discovery



[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

What are proto-objects?

The triadic architecture of Rensink (2000):



Proto-objects:

- Proto-objects are object hypotheses built from simple features
- Volatile structures with limited coherence in space and time
- (Collections of) proto-objects are the target of attention ("the focused attention system acts as a hand to grab proto-objects")

R. A. Rensink, **The Dynamic Representation of Scenes,** *Visual Cognition (7),* 2000

universitätbonn

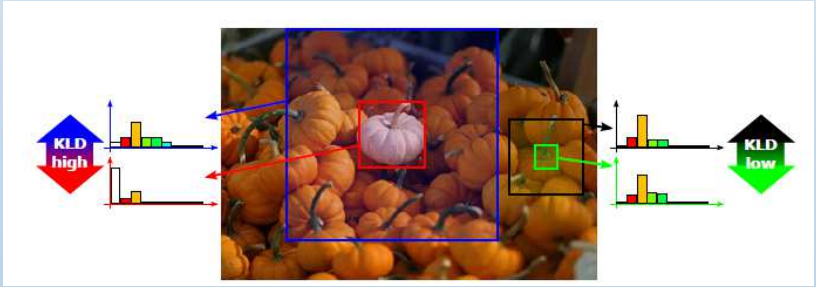
Saliency Detection


Dominik Klein

For saliency computation, use information-theoretic saliency:
Represent center and surround regions
by probability distributions
Otherwise structure similar to Itti/Koch [Itti et al, PAMI 1998]
(image pyramids, feature channels, conspicuity maps, saliency map)

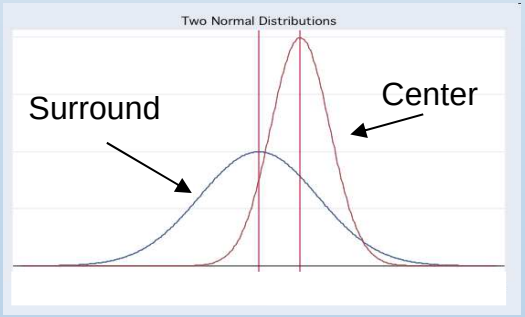
BITS [Klein/Frintrop ICCV 2011]:

Discrete distributions
compare with Kullback-Leibler Divergence



CoDi [Klein/Frintrop DAGM 2012]:

Continuous distributions (Gaussians)
compare with Wasserstein distance W_2

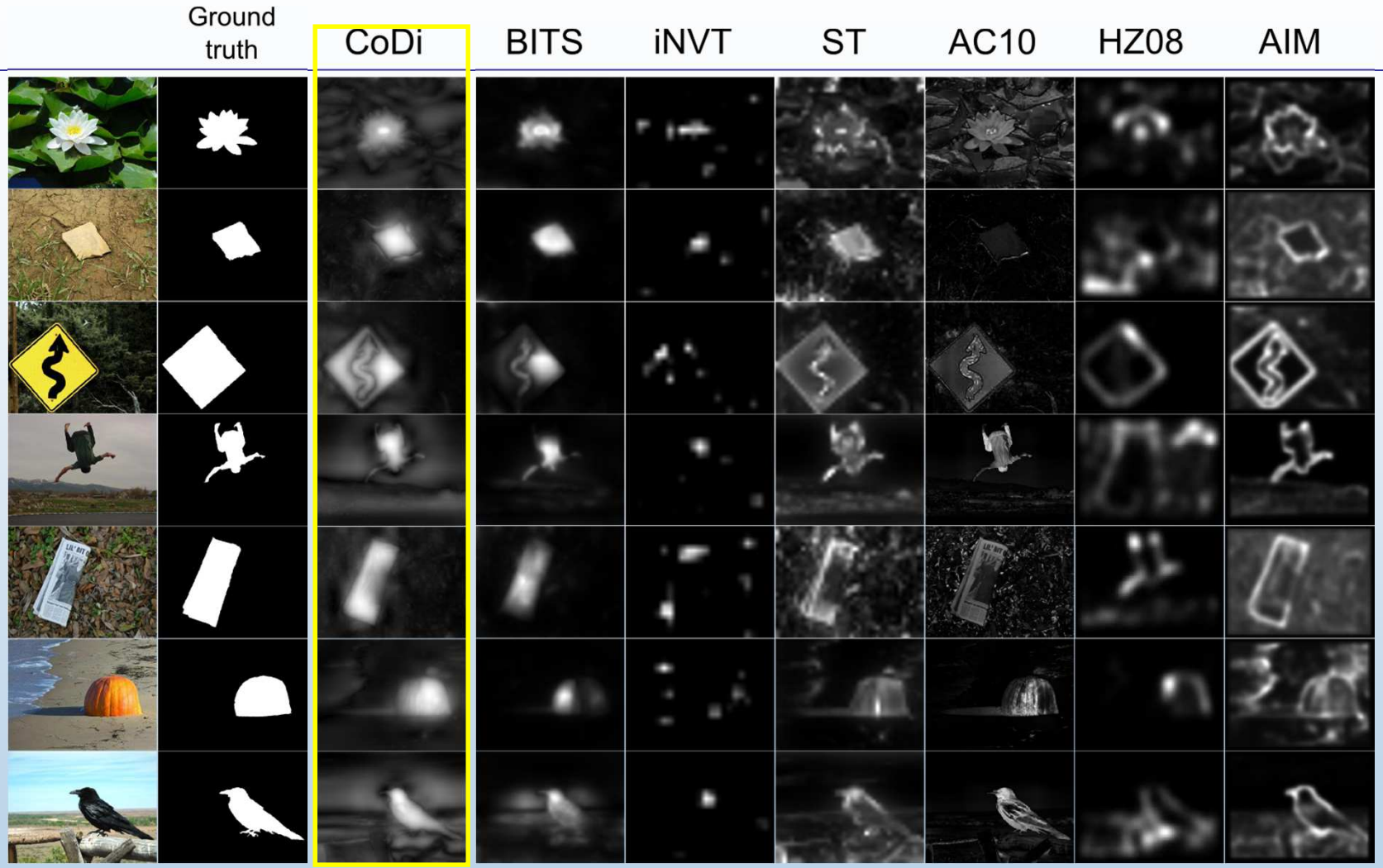


C++ code available at:
http://www.iai.uni-bonn.de/~kleind/

universitätbonn

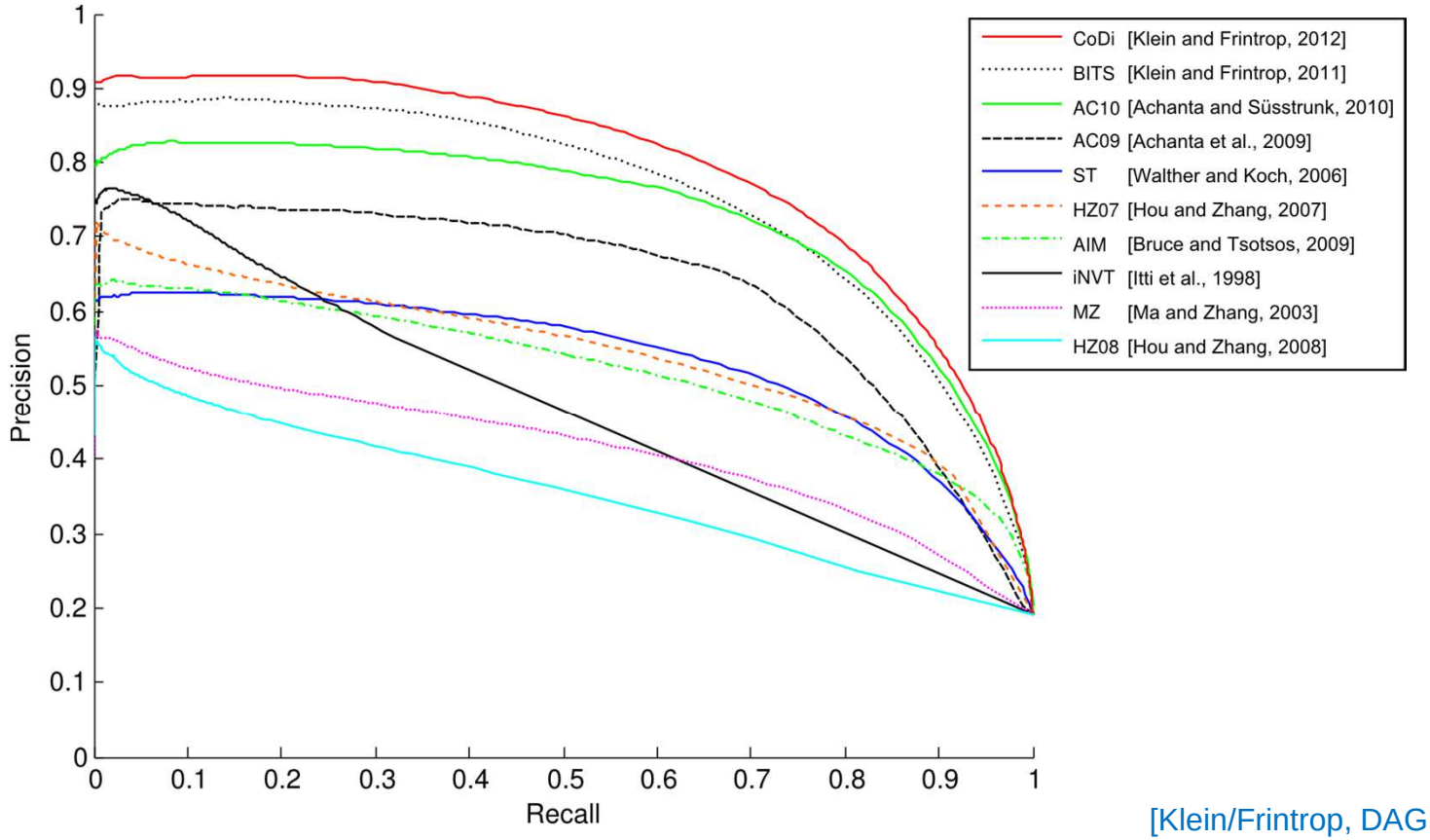
MSRA Salient Object Database

Results from: [Klein/Frintrop DAGM 2012]



BITS: [Klein/Frintrop, ICCV 2011]
iNVT: [Itti et al., PAMI 1998]
ST: [Walther/Koch, 2006]
AC10: [Achanta/Süsstrunk, 2010]
HZ08: [Hou/Zhang, 2008]
AIM: [Bruce/Tsotsos, 2009]

Saliency performance on MSRA Salient Object Database



[Klein/Frintrop, DAGM, 2012]

From saliency to proto-objects

We have now a saliency map:


To obtain proto-objects, we follow the two-stage processing strategy of the HVS:

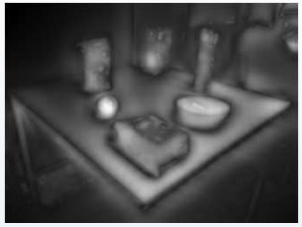
- Pre-attentive: parallel processing to find region of interest
- Attentive: serial processing, one region of interest at the time processing becomes ‚selective'.

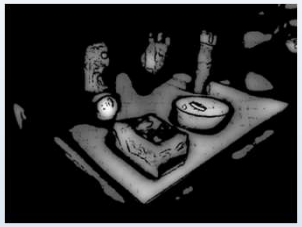
[Neisser, Cognitive Psychology, 1967]
[Treisman, *Cognitive Psychology*, 1985]

From saliency to proto-objects

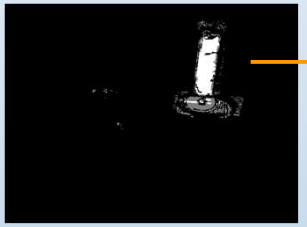
Generate proto objects:

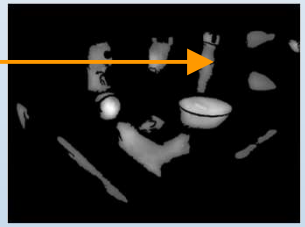
Frame t





IOR information from frames 1 to t-1



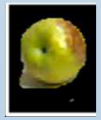


Pre-attentive (parallel) stage:

1. compute the saliency map
2. threshold the saliency map
3. find connected components (discard too big and too small objects)
4. rank by average saliency
5. Inhibit already attended regions

Attentive stage:

1. Pick one salient blob & fit a rectangle around
2. Refine shape of blob with GrabCut segmentation [Rother 2004] (OpenCV implementation) saliency values serve to initialize pixel likelihoods



[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

Proto-Objects

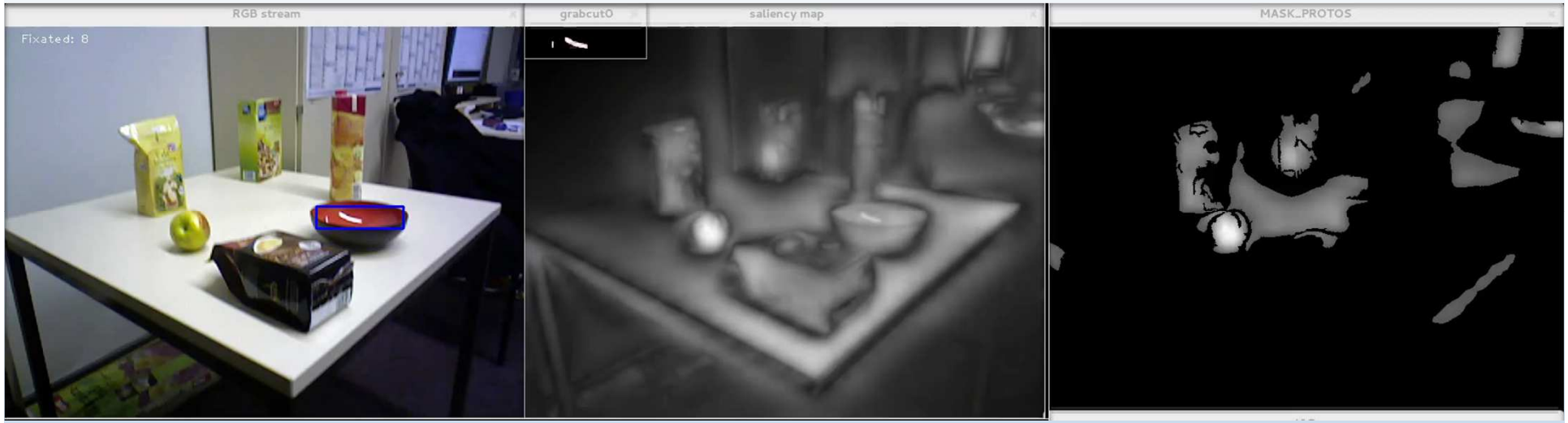
- Now, we have a segmented proto-object for the currently attended region
- Examples:

[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

universitätbonn

IVS intelligent vision systems

Salient Object Discovery in 3D



Original image | Saliency map | Object hypotheses

[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

Salient Object Discovery in 3D

That‘s was all in 2D.
Now we come to 3D…

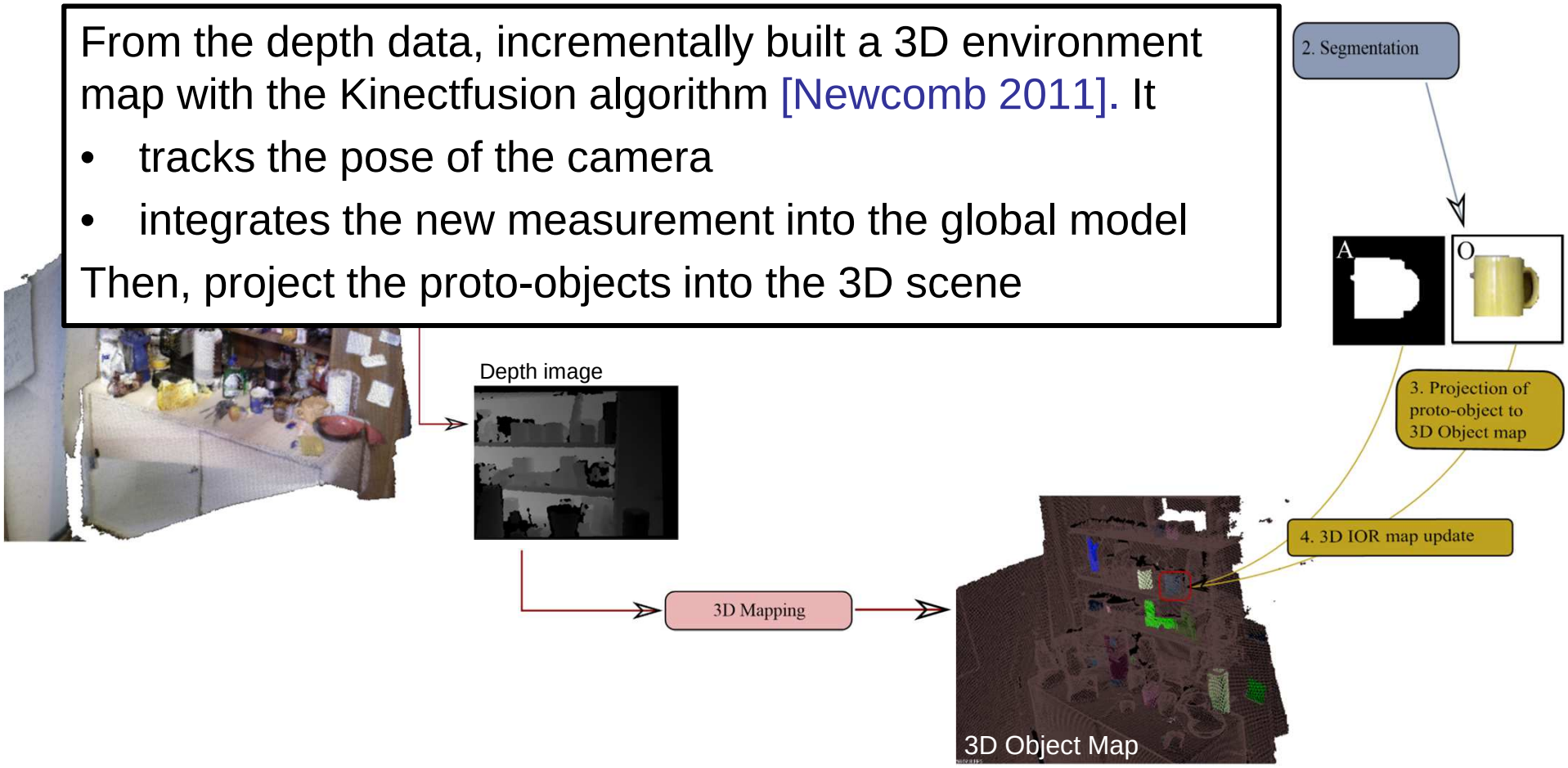
[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

Salient Object Discovery in 3D

From the depth data, incrementally built a 3D environment map with the Kinectfusion algorithm [Newcomb 2011]. It

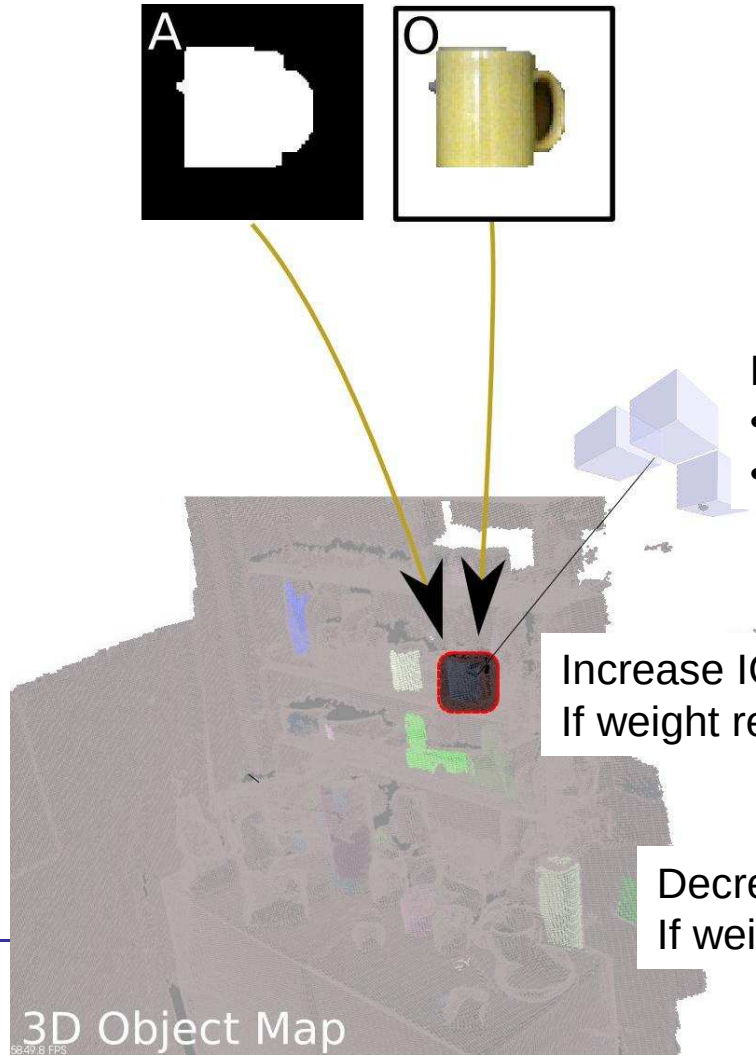
- tracks the pose of the camera
- integrates the new measurement into the global model

Then, project the proto-objects into the 3D scene



[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

Inhibition of Return in 3D



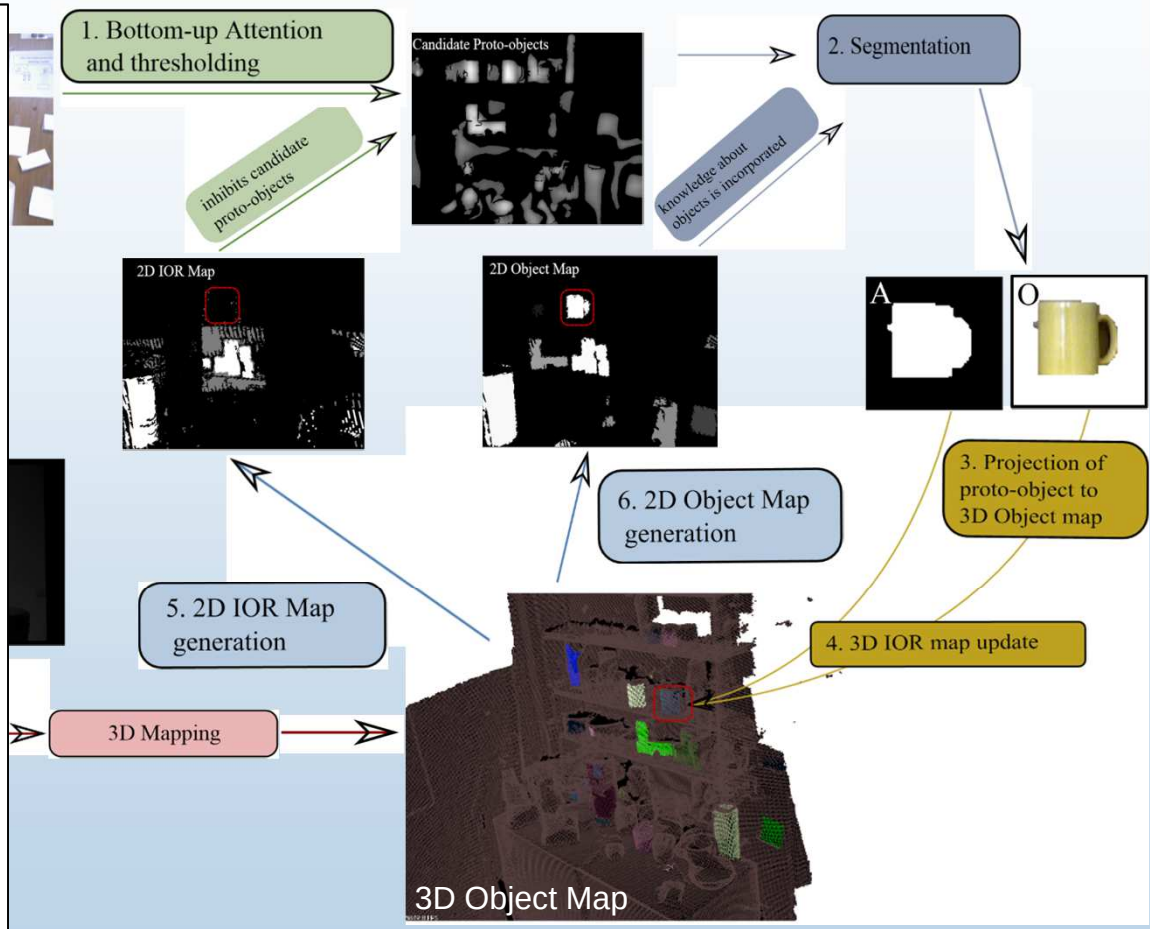
Each voxel stores
- An IOR flag (active/not active)
- An IOR weight

Increase IOR weight of attended voxels.
If weight reaches threshold, set IOR flag active

Decrease IOR weight of not attended voxels
If weight reaches 0, set IOR flag inactive

Salient Object Discovery in 3D

- Use knowledge from 3D map to improve object detection & segmentation
- raycast the object labels & inhibition flags to the current viewpoint
- constrain the segmentation step to not label pixels as foreground that belong to other objects
- Use inhibition flags to prevent selecting already attended blobs

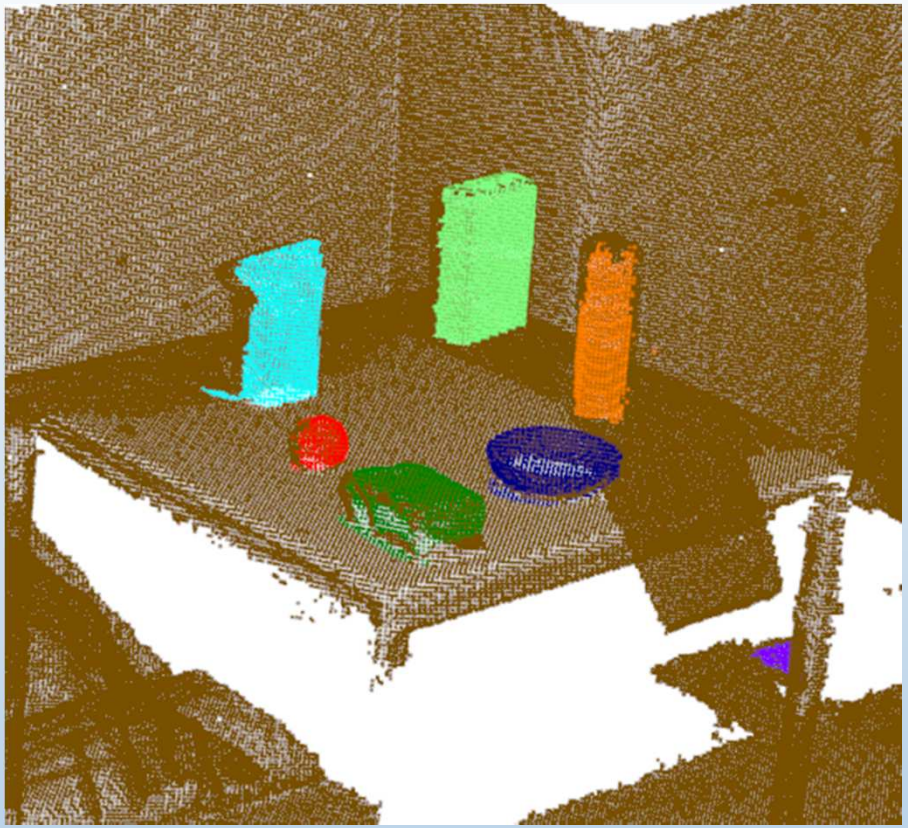


[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

Results: Table top sequence

Ground truth

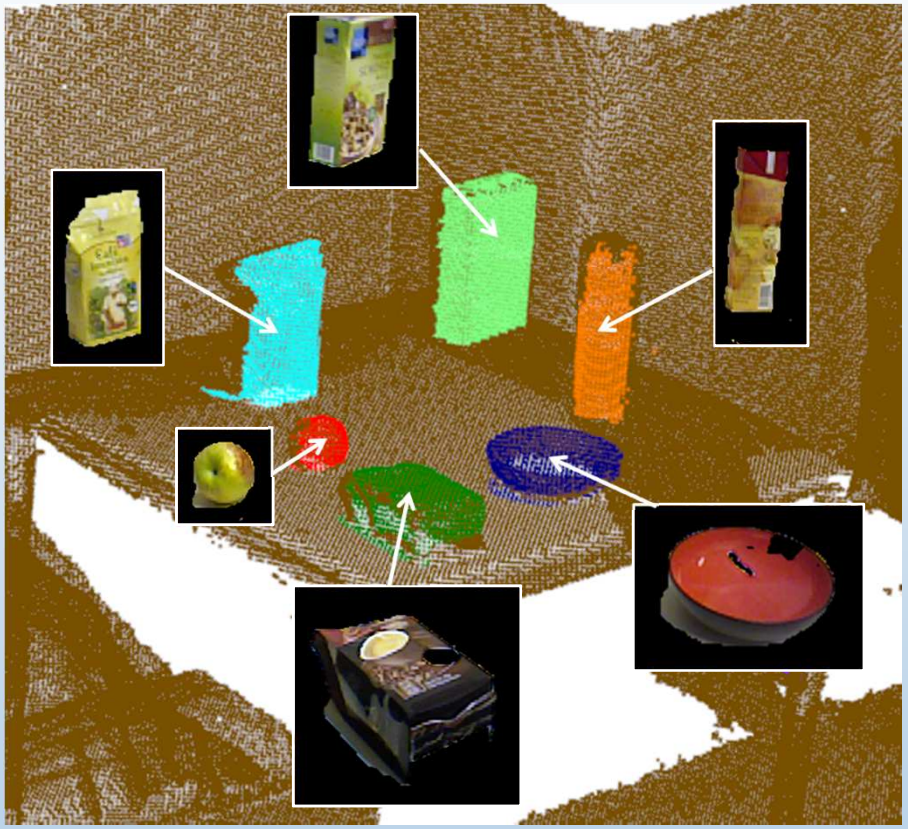
Detected Objects



Results: Table top sequence

Ground truth

Detected Objects



Salient Object Discovery in 3D

[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]

Results: Table top sequence

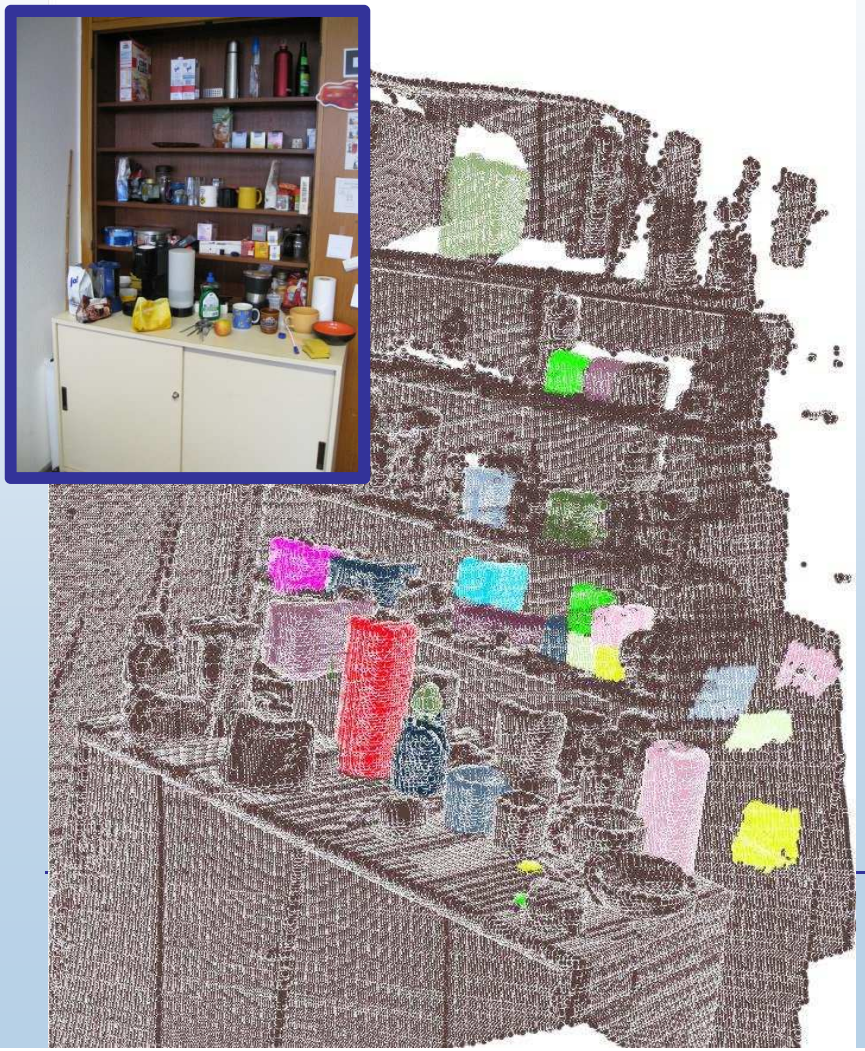
Object	Precision	Recall
Juice	97	31
Bowl	92	56
Cereals	98	61
Apple	98	45
Coffee 1	90	47
Coffee 2	96	49

universitätbonn

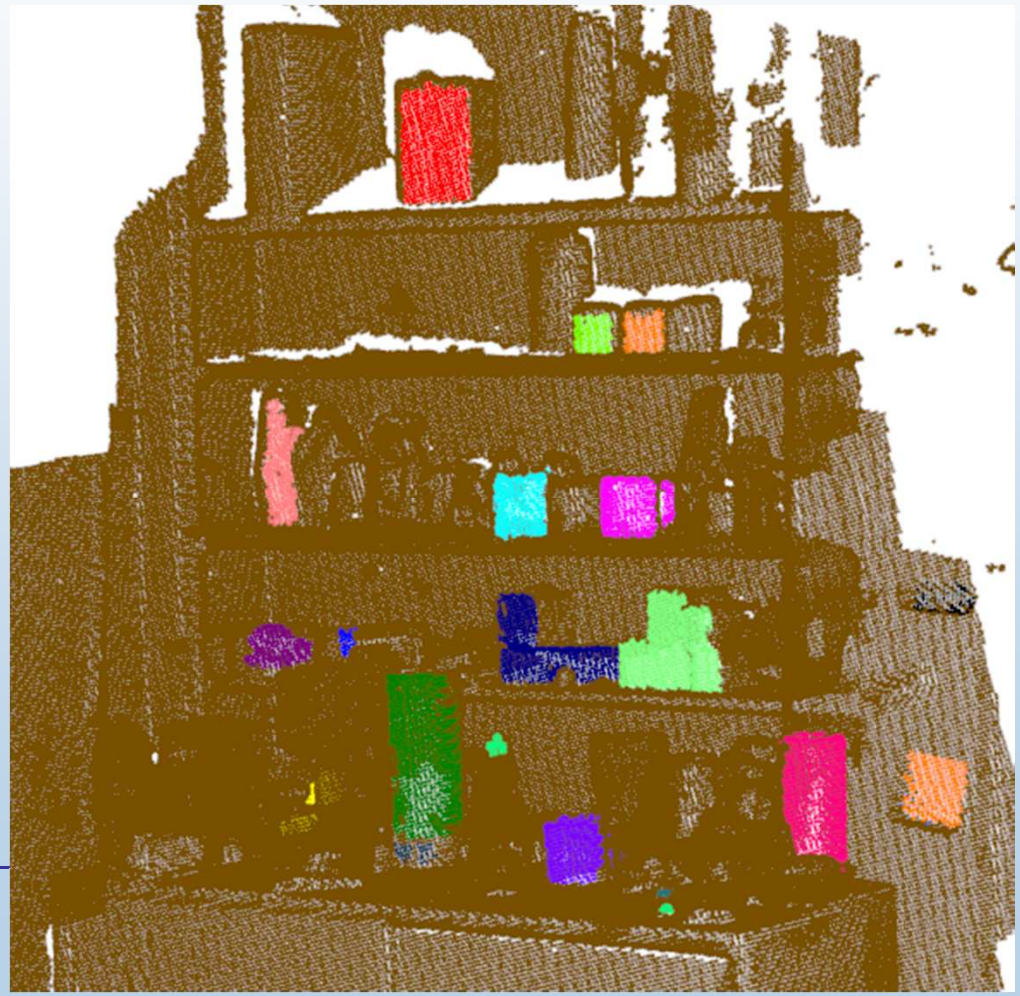
IVS intelligent vision systems

Results: Coffee machine sequence

Ground truth



Detected Objects

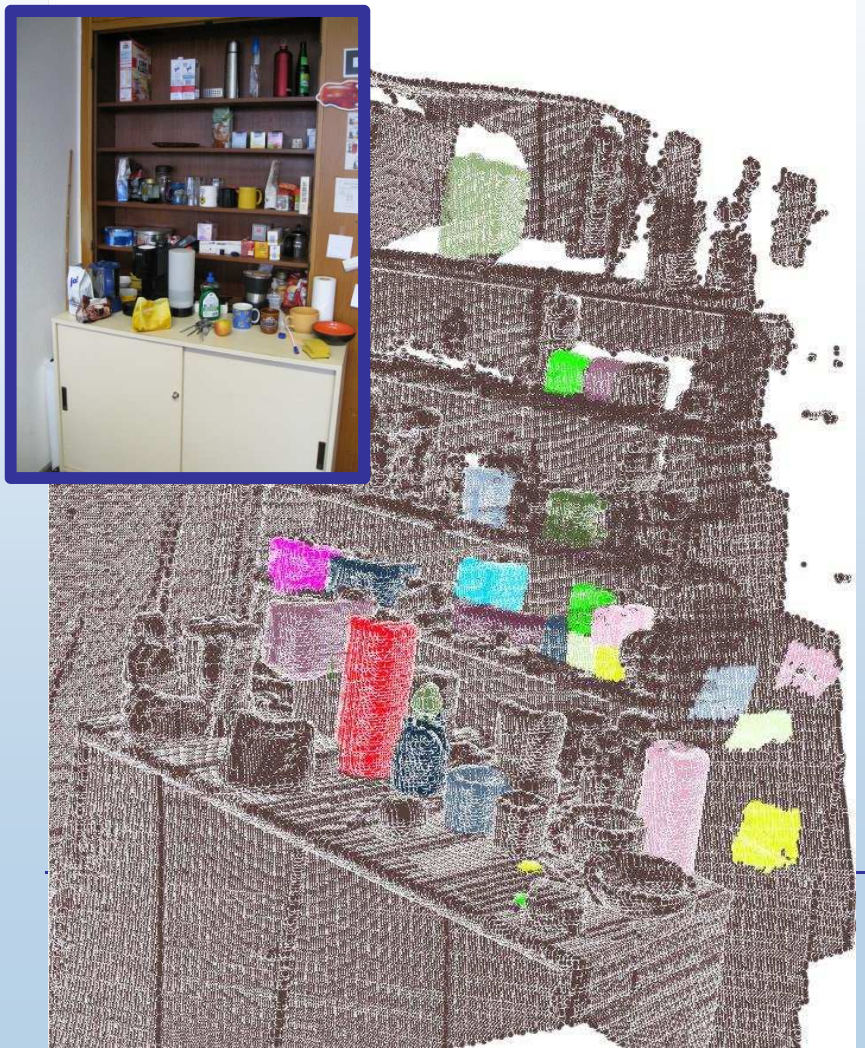


universitätbonn

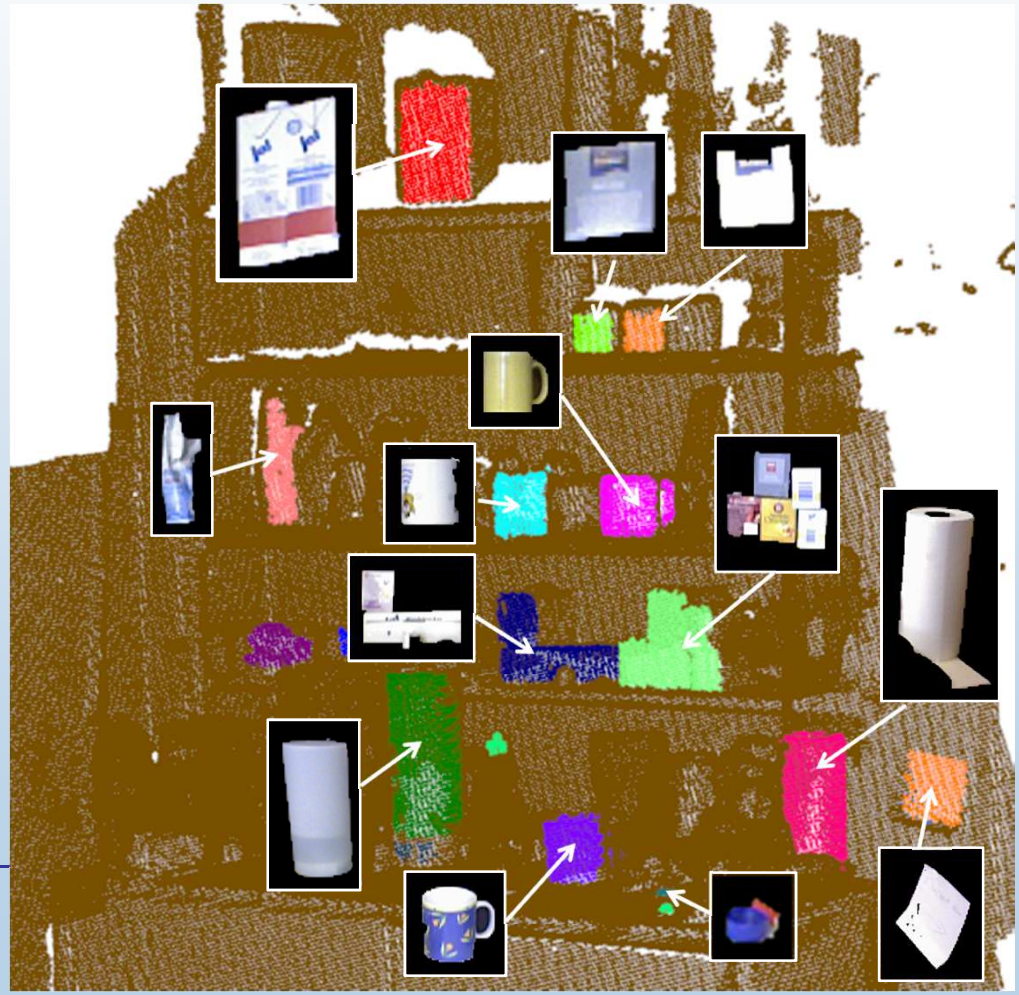
IVS intelligent vision systems

Results: Coffee machine sequence

Ground truth



Detected Objects



Results: Coffee machine sequence

Object	Precision	Recall
Note	69	43
Y. cup	90	40
W. cup	93	40
Box	99	30
Paper	99	40
Sugar	62	61
Cup	90	36
Water	99	37
Milk	98	39

3D Object Discovery

Take home message:
Real-world attention requires inhibition of return in spatial coordinates

Thank you for your attention!

[G. Martín-García & S. Frintrop, Proc. of the annual meeting of Cognitive Sciences (CogSci), 2013 (to appear)]
[G. Martín-García & S. Frintrop, German Journal of Artificial Intelligence, 2013 (to appear)]