AUTOMATION & CONTROL INSTITUTE
INSTITUT FÜR AUTOMATISIERUNGS- & REGELUNGSTECHNIK

TU WIEN

From holistic scene understanding to semantic visual perception:
A vision system for mobile robots

Kai Zhou, Karthik Mahesh Varadarajan, Andreas Richtsfeld, Michael Zillich and Markus Vincze

9-May-2011, Shanghai, China

Outline

- Motivation
- Related work
- Methodology
- Experiments
- Conclusion

Motivation

- Robots can or are designed to performe many tasks today, e.g., house cleaning, nursing, dancing, instrument playing etc.

Sliced noodle

Motivation

- Robots can or are designed to performe many tasks today, e.g., house cleaning, nursing, dancing, instrument playing etc.

Japanese Ramen

First sub-task

- The fisrt sub-task for all these cool thing is:
  - Object detection

Related Work

- How robot detects objects?

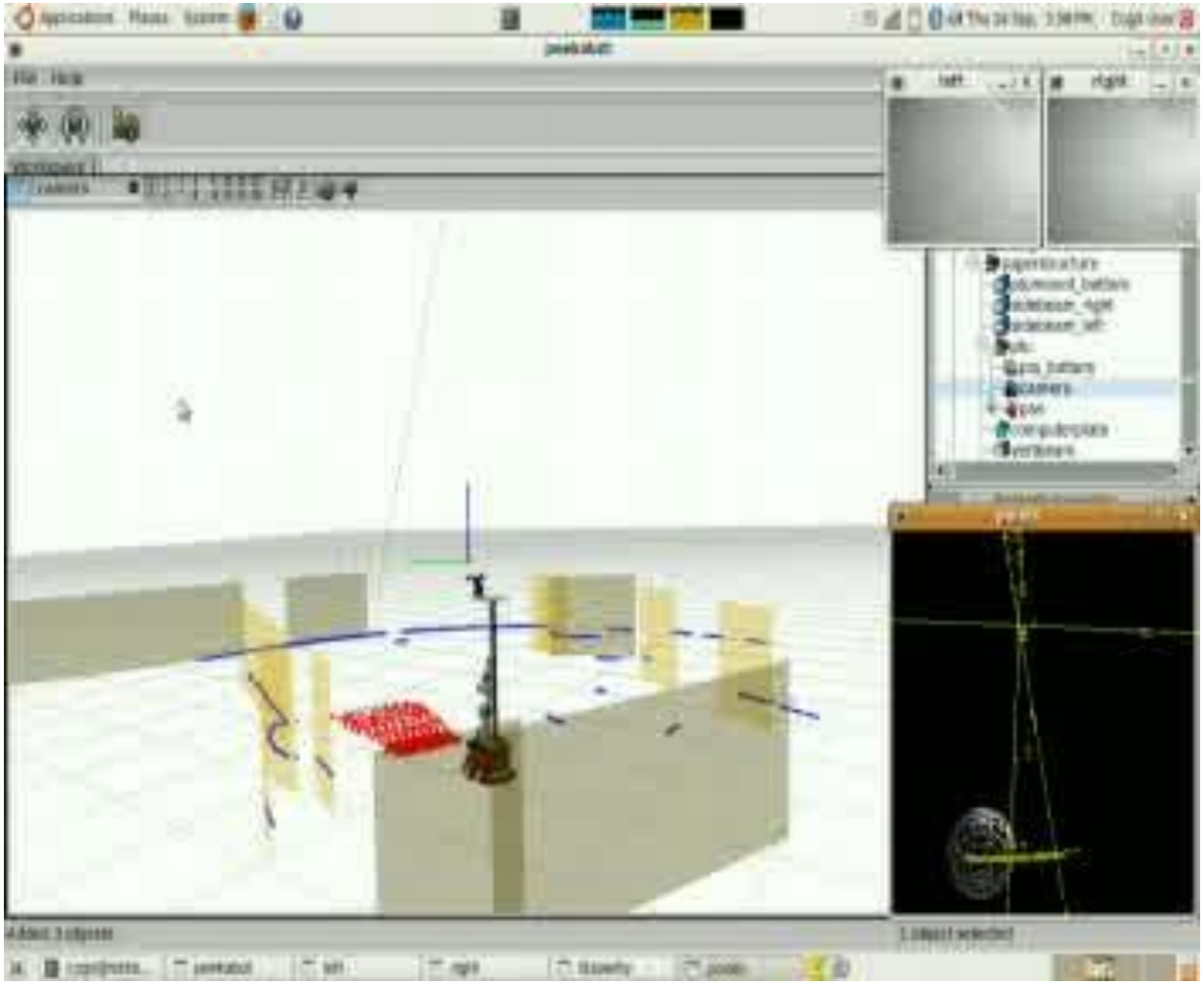
-- Segmenting irregular regions which we think the robot should be interested in

  - Colored blob
  - object motion
  - saliency
  - spatial analysis
  - Mixture of the models

Why detect supporting surface?

- Researchers from psychology, computer vision and robotics have provided evidences that the supporting surface detection is
  - the way to build up the hierarchical structure of the scene
  - dominative feature in the man-made environments

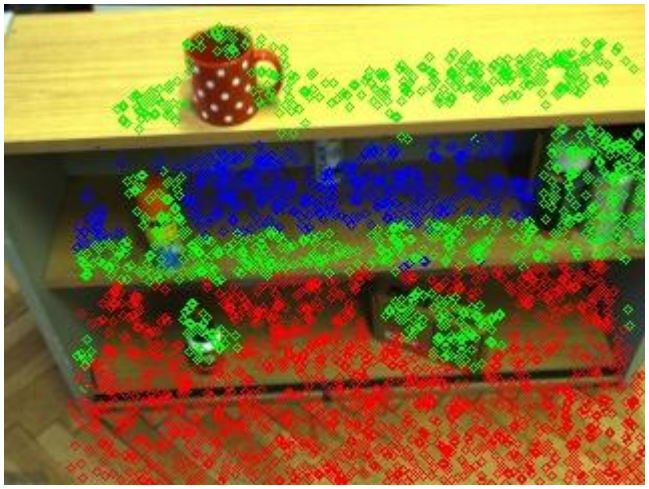
Robotic visual search with supporting plane detection



Problem solved?

No!

Connected component RANSAC[1] for single plane estimation, considers the spatial coherence exhibited by inlier points. This novel fitness evaluation criterion can also be used for multi-plane estimation in a proper way.



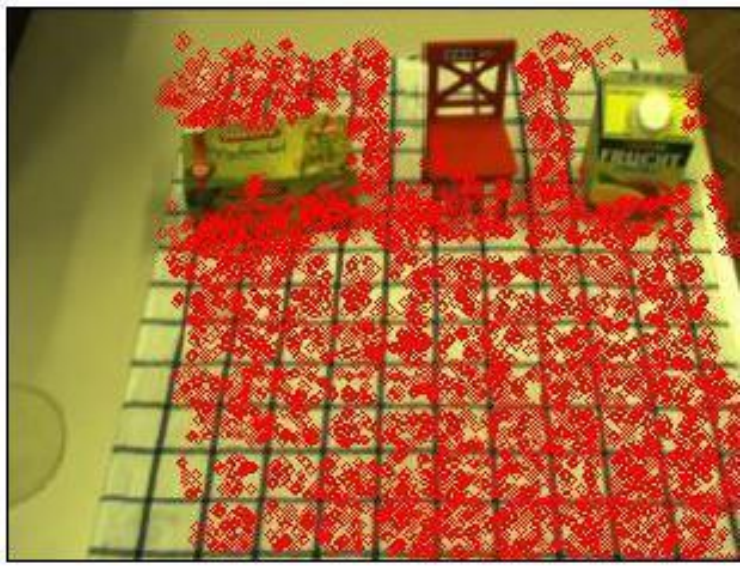
RANSAC neglects the connectivity of inliers and finds a veritical „plane“ (green cycles)

CC-RANSAC will only collect the connected inliers with 8-neighbor topology inherited from image grid

Sub-tasks

- The fisrt sub-task is:
  - Object detection
- The second sub-task is:
  - Supporting surface detection

RANSAC



Methodology

- Trackling these two relative sub-tasks in a unified probabilistic framework

[Hoiem, etal.] "Putting objects in perspective"

Best paper reward of CVPR 06, cited by 220

$$p(O)\, p(O \mid S)\, p(S)$$

Our problem is different

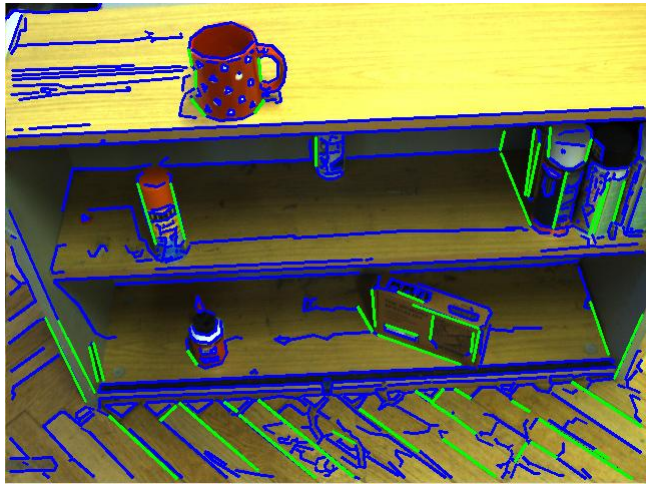
$$p(O)\, p(O \mid S)\, p(S)$$

- O -- without probabilistic representaion
- S -- from the supporting surface detection

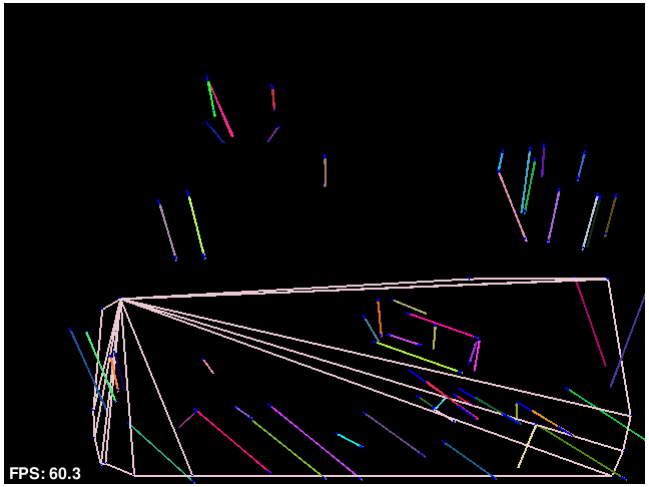
For Object detection: p*(O)*

Stereo line detection: $$p(f) = \frac{\theta_{2Dl}}{\pi/2} \cdot \frac{\theta_{2Dr}}{\pi/2} \cdot \frac{\theta_{3Dz}}{\pi/2}$$

Lines which are parallel with the epipolar line or point away from the viewpoint usually have bad 3D reconstruction.



Scene with the matched lines



3D Scene of the matched lines and dominant plane

For supporting surface estimation: p*(S)*

- CC-RANSAC [Gallo, etal.]: Generate hypotheses which have 8-neighbor connection in image grid, evaluate these hypotheses and select the best one as RANSAC

- Collect all the inliers, compute the average normal vector of all inliers, measure the angle θ_{CS} between average normal and normal of estimated plane

$$Con(S) = (1 - \frac{\theta_{CS}}{\pi/2}) \cdot \frac{k}{N}$$

P(O|S)

- This probability is determined by the distance and angle between detected stereo lines and planes.
- We assign a higher confidence value to lines which are parallel or perpendicular with the estimated plane, as well as lines which are geometrically close to the plane.

Coherent probability framwork

Coherent probability framework contains three parts:

- Confidence value of estimated plane indicates the reliability of plane estimation
- Significance value of object detection denotes the importance or correctness of detection
- Relation factor is determined by geometrical context of detected objects and planes

By maximization this joint probabolity, we are able to reduce false object detection and refine 3D plane estimation simultaneously.

$$p(O,E,S) = \prod_{i=1}^{N} p(S) \prod_{j=1}^{M} p(O|S)p(E|O)$$

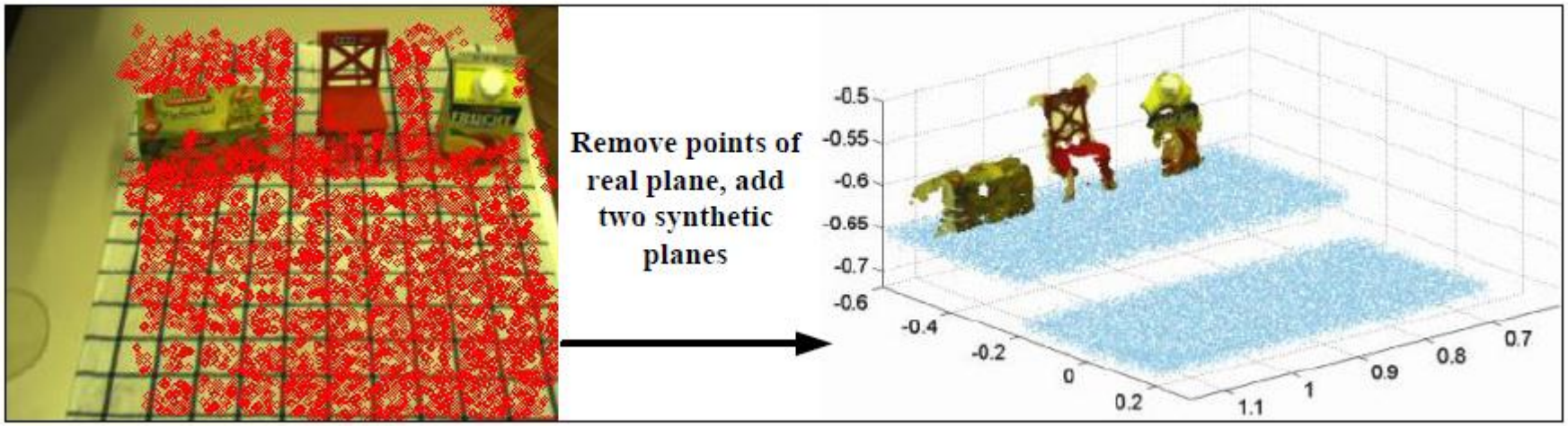
$$= \prod_{i=1}^{N} p(s_i,h_i) \prod_{j=1}^{M} p(m_j,t_j|S)p(e_j|m_j,t_j)$$

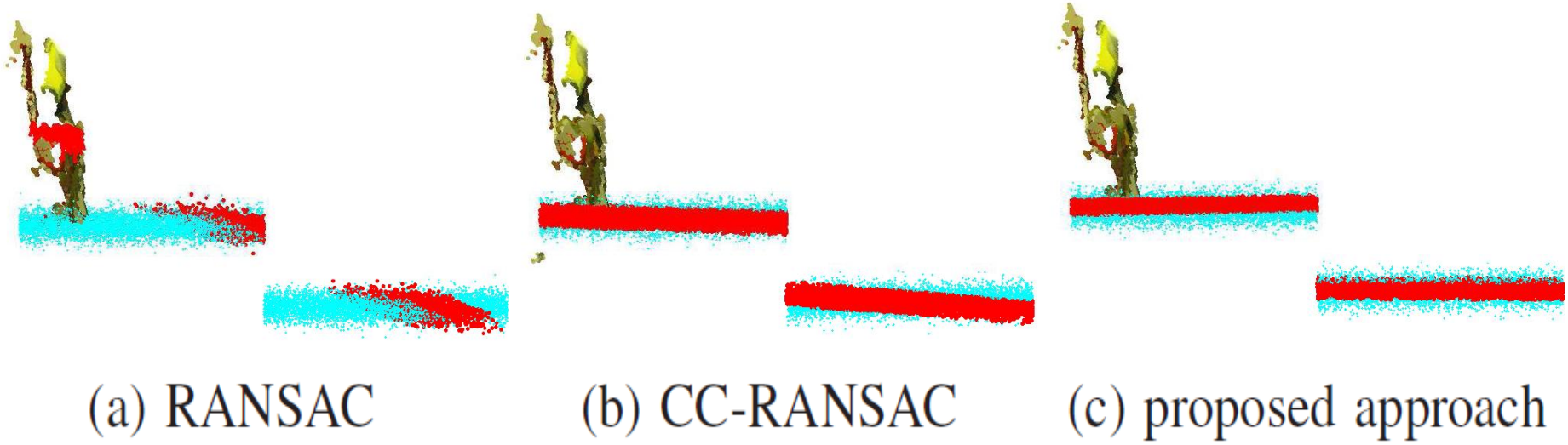
confidence value of estimated planes

Relation of estimated planes and deteted objects

significance value returned by stereo line detector

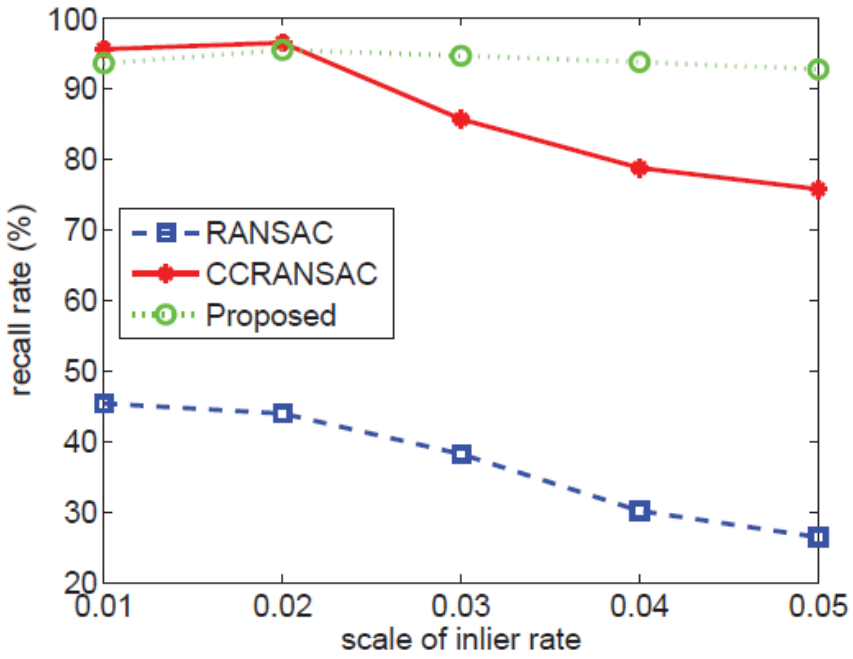
Experiments



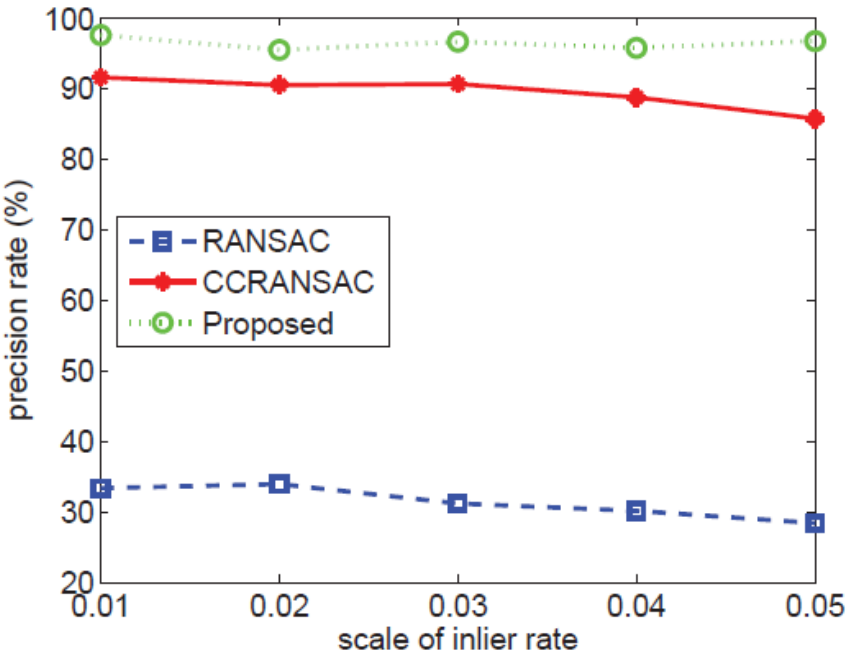


(a) RANSAC (b) CC-RANSAC (c) proposed approach

Qualitative comparison



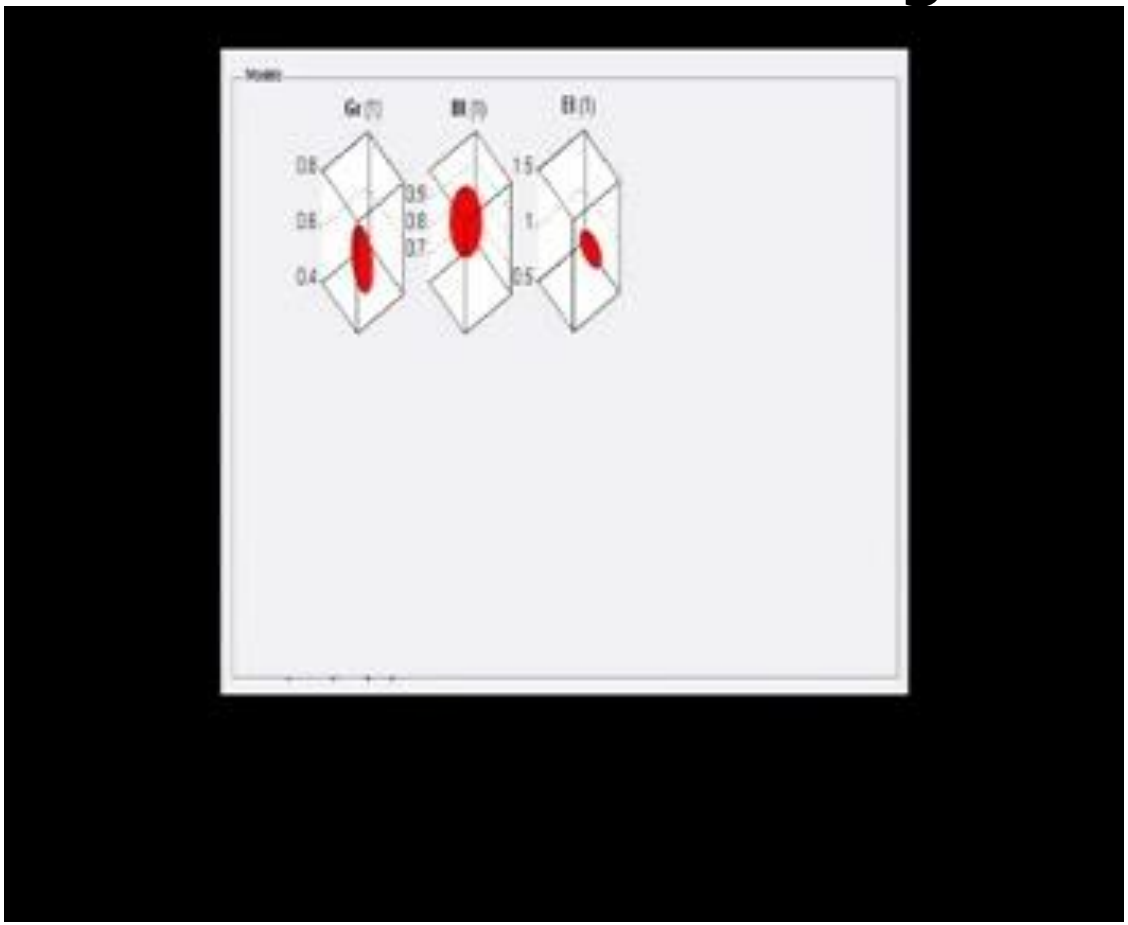
(a) Recall



(b) Precision

Real robotic applications

- Continuous interactive learning robot



Thank you for your attention.