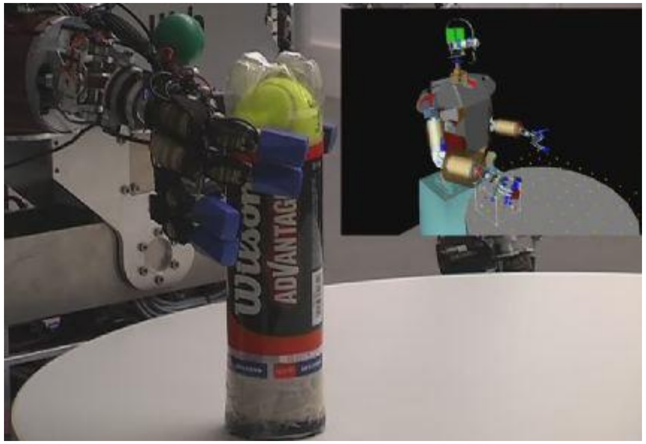
# Active Perception: Active vision and active touch

KIT-Department of Informatics, Institute for Anthropomatics and Robotics, High Performance Humanoid Technologies (H2T)



# Outline of the lecture

- Active Perception
  - Active vision
    - Motivation and definition
    - Discovery, segmentation and grasping unknown objects
    - Active visual search
  - Active touch
    - Haptic exploration of unknown objects

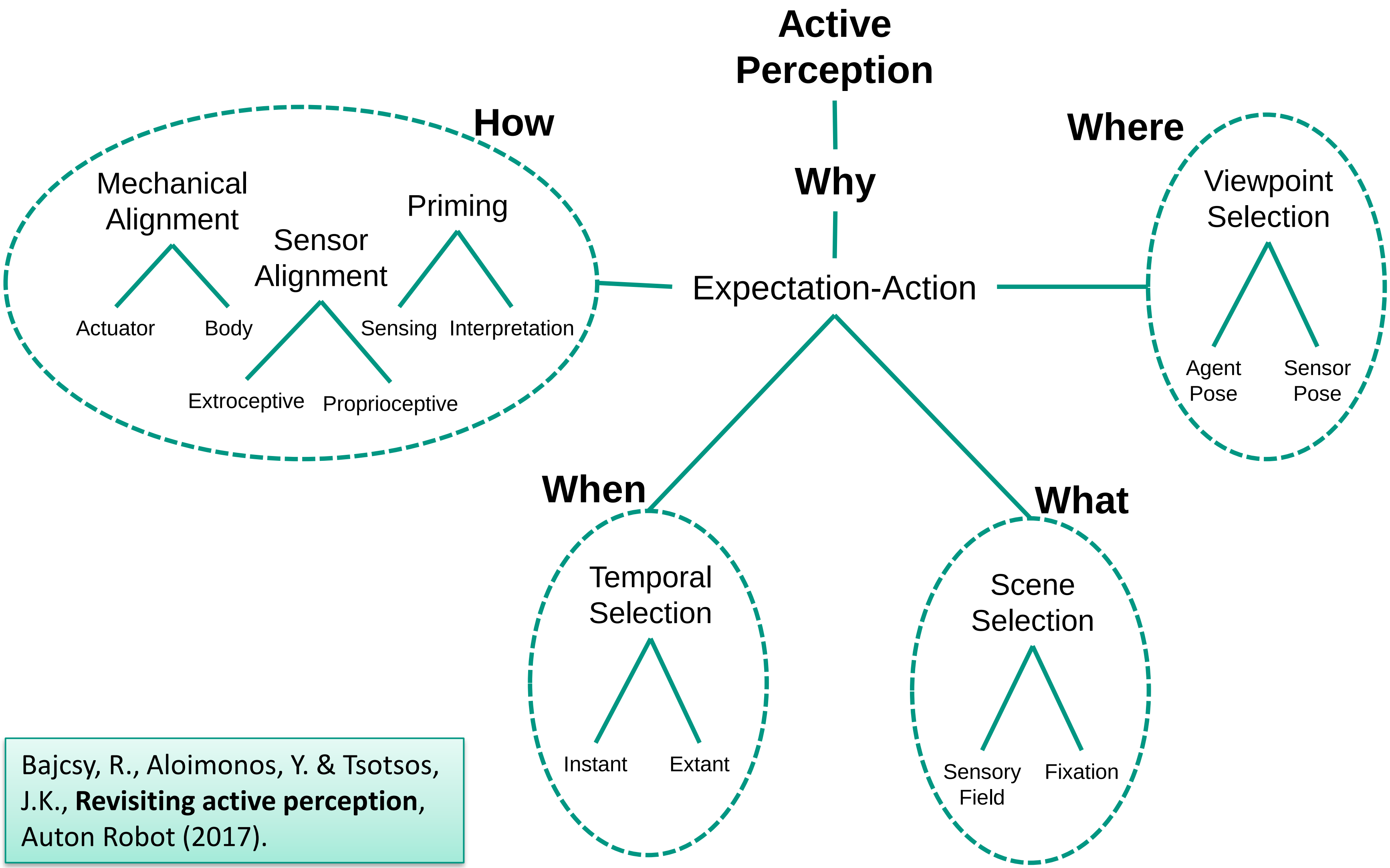
  - Visuo-haptic exploration
  - (Active hearing)

# Active Perception: Definitions

- "Active sensing is the problem of intelligent control strategies applied to the data acquisition process which will **depend on the current state** of data interpretation including recognition," Bajcsy (1988)
- "An observer is called active when engaged in some kind of activity whose purpose is to **control the geometric parameters of the sensory apparatus**. The purpose of the activity is to manipulate the constraints underlying the observed phenomena in order to improve the quality of the perceptual results," Aloimonos et al. (1988)
- "An agent is an active perceiver if it knows **why** it wishes to sense, and then chooses **what** to perceive, and determines **how**, **when** and **where** to achieve that perception," Bajcsy et al. (2017)

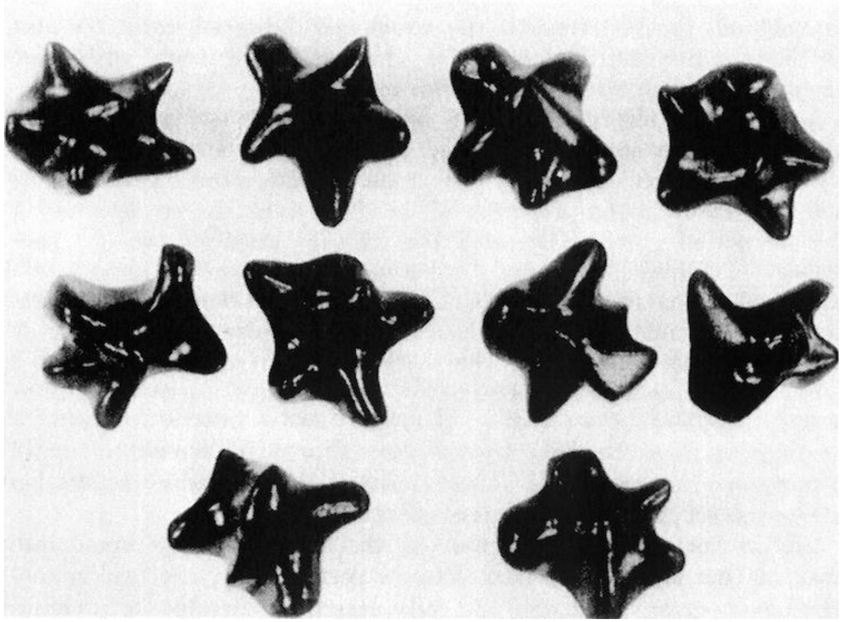

- Bajcsy, R., **Active perception,** Proceedings of the IEEE (1988), 76(8), 966–1005
- Aloimonos, J. et al., **Active vision**, Int Journal of Computer Vision (1988), 1(4), 333–356.
- Bajcsy, R., Aloimonos, Y. & Tsotsos, J.K., **Revisiting active perception**, Auton Robot (2017).

# Active Perception: Five Questions




Bajcsy, R., Aloimonos, Y. & Tsotsos, J.K., **Revisiting active perception**, Auton Robot (2017).

# Active Perception: Biological Motivation



| Presentation | Recognition Rate | Comparable to |
|---|---|---|
| Single image | 49% | Classic Computer Vision |
| Rotating object | 72% | Active Vision |
| Object in hand | 99% | Active Perception |

J. J. Gibson, “The senses considered as perceptual systems”, Boston, Houghton Mifflin, 1966.

# Classical CV vs. Active Vision vs. Active Perception

| | Classical CV | Active Vision | Active Perception |
|---|---|---|---|
| Image processing | ✓ | ✓ | ✓ |
| Viewpoint selection | - | ✓ | ✓ |
| Multi-modal sensory input | - | - | ✓ |
| Changing agent's state | - | ✓ | ✓ |
| Changing the environment | - | - | ✓ |

# (Inter-)active Perception

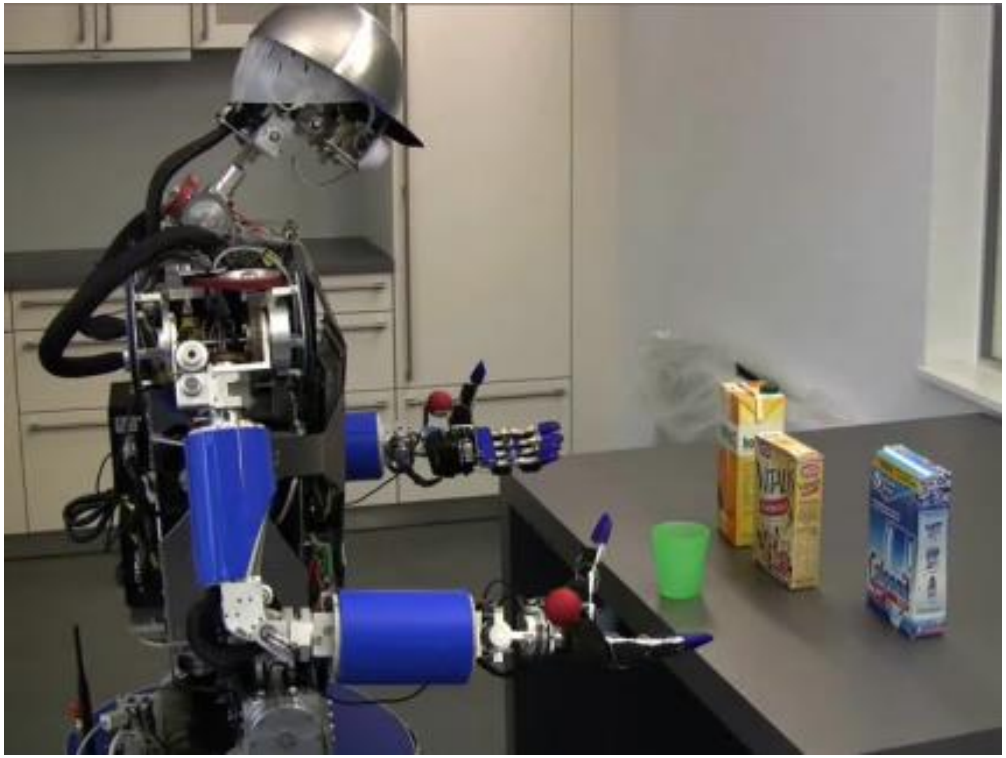
- Forceful interaction with the environment
  - Creates a novel sensory signal

- Exploit the regularity in $S \times A \times t$
  - $S$: Multi-model sensory input
  - $A$: Executed action(s)
  - $t$: Time

- Benefits
  - Generation of new sensory input
  - Using the regularity in $S \times A \times t$ to predict, update world state
  - Prior knowledge makes interpretation easier
  - Learn the regularity $S \times A \times t$

Bohg, Jeannette, et al., **Interactive perception: Leveraging action in perception and perception in action**, arXiv:1604.03670, 2016.

# Perception



# Perception

- Perception modes:
  - Vision
  - Audio
  - Tactile, force, pressure
  - Laser, infrared, sonar, …
  - Internal state (temperature, force, voltage, …)
  - …

- This chapter: visual and tactile perception

# Why visual Perception?

- Recognition (of known objects)
- Localization (i.e. determine spatial relationship between objects, and between the robot and the environment)

  → Get an internal representation of the world state for planning and acting

- Observation (of motion, actions, relations over time)

  → Learn: trajectories, possible actions, probabilities of events…

- Discovery (of new things)

  → Learn: visual appearance of new, unknown objects

# Visual Perception

- **Recognition, localization, observation**: see computer vision lectures
  - Robotik 3
  - Inhaltsbasierte Bild- und Videoanalyse
  - Computer Vision für Mensch-Maschine-Schnittstellen
  - …
- **Discovery and learning**: this chapter

# Discovery and learning of unknown objects

- Given: A humanoid robot with limited real world knowledge
- Will frequently have to cope with unknown objects
- Possible goals:
  - learn about them
  - be able to recognize them when seen again
  - grasp/manipulate them

# Discovery and learning of unknown objects

- Goal: Learn the visual appearance of an unknown object for future recognition

- Necessary steps:
  - discover the new object
  - segment it from the background
  - learn its visual appearance for recognition

# Discovery and learning of unknown objects


- **Task:** Discovery and segmentation are difficult in a cluttered environment, may be impossible by vision only
- **Reason:** difficult / impossible to define the concept of “objectness” in full generality, especially when restricted to a purely perceptional level

- Additional information for segmentation can be provided by physical interaction with the object

  → **(Inter)Active Perception**

# Overview

Create initial object hypotheses

Push

Re-localize, determine transformation

No motion

Object moved

Verify and improve hypothesis

Push again

# Overview

**Create initial object hypotheses**

Push

Re-localize, determine transformation

No motion

Object moved

Verify and improve hypothesis

Push again

# Create initial object hypotheses

- Generate initial object hypotheses based on camera images
- Three heuristics:
  - Planes, cylinders and spheres amongst SIFT features (RANSAC)
    → textured, artificial objects
  - Unicolored regions of promising size (color MSERs)
    → single-colored objects
  - Visually salient regions (DoG filter)
    → objects that are neither textured nor unicolored

# Create initial object hypotheses

- Generate hypotheses using all three heuristics
- Dense stereo matching to get 3D position of all pixels of the image
- Each hypothesis is represented by the set of RGBD points in its image region

- Hypothesis selection for verification by pushing
  - Filter out those which are lower than their local neighborhood
  - Select reachable hypothesis

Each group of points of the same color represents an object hypothesis

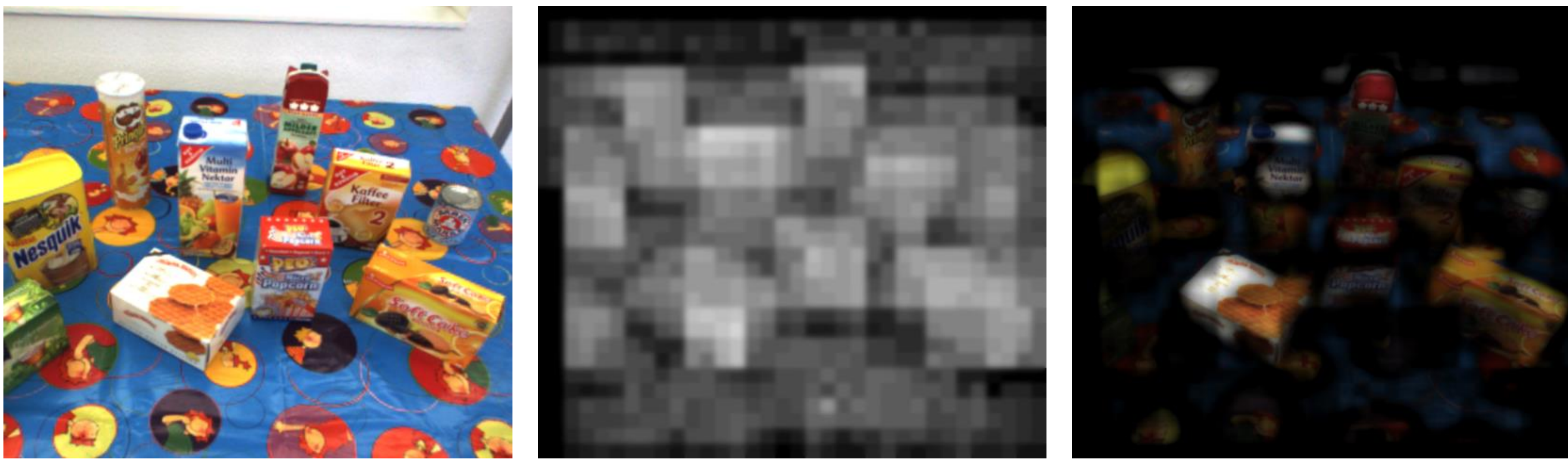
# Create initial object hypotheses

- Use all 3 criteria → get a lot of hypotheses

- Only „pushable“ hypotheses desired

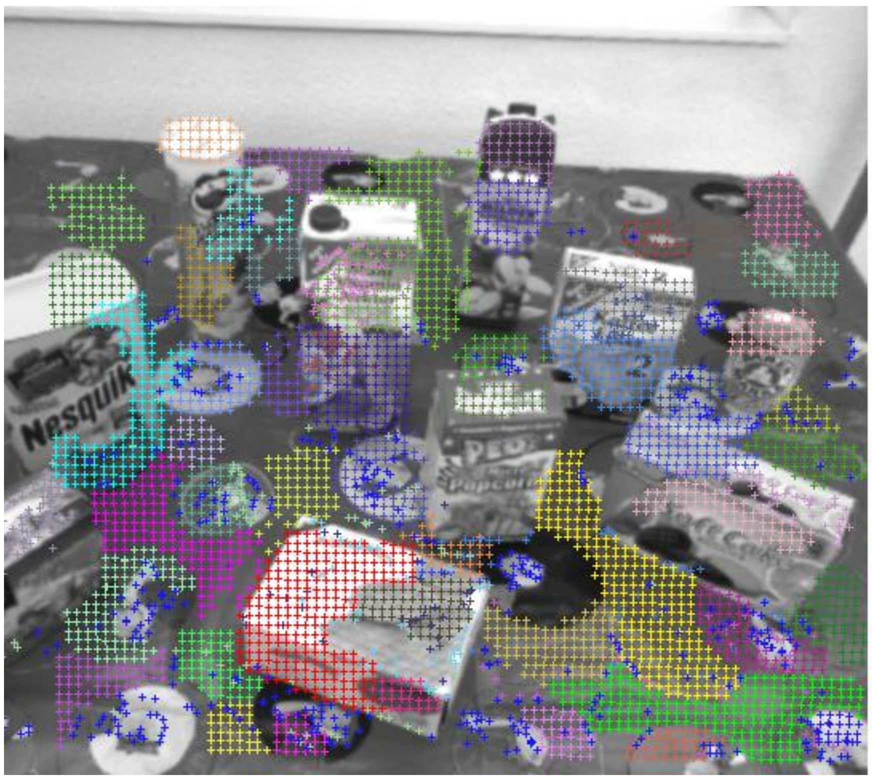

# Generation of Object Hypotheses

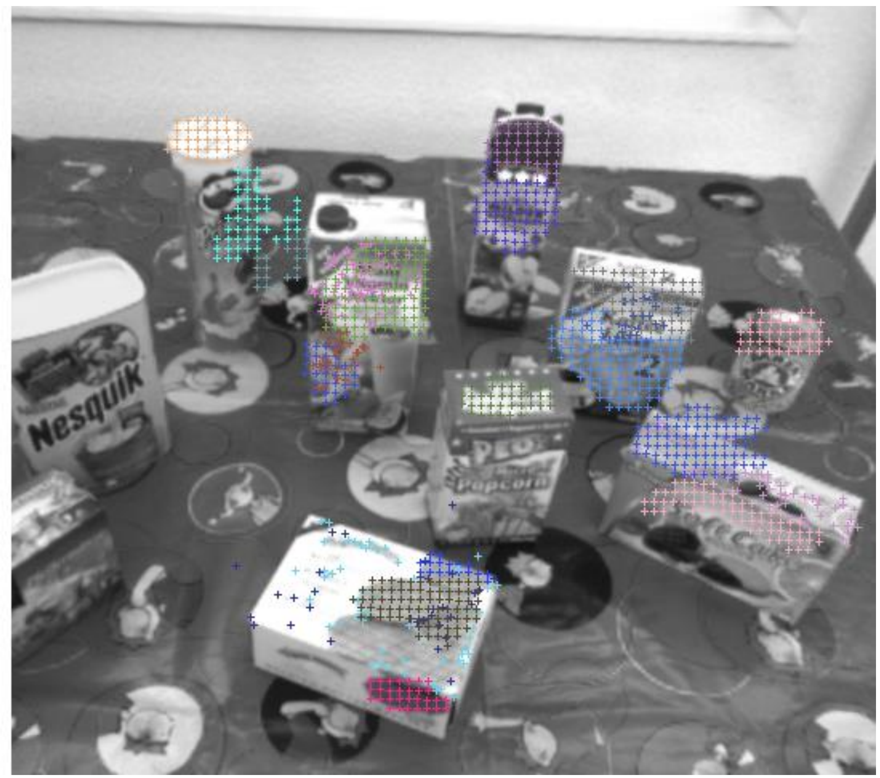
- Additional criterion: image region should correspond to high part of the scene
- Calculate proximity to local maxima of image parts based on the 3D point cloud



# Generation of Object Hypotheses

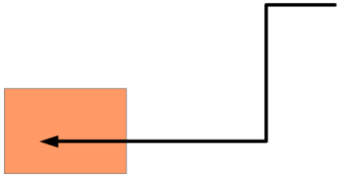
- Use proximity to local maxima to filter object hypotheses before choosing one for pushing





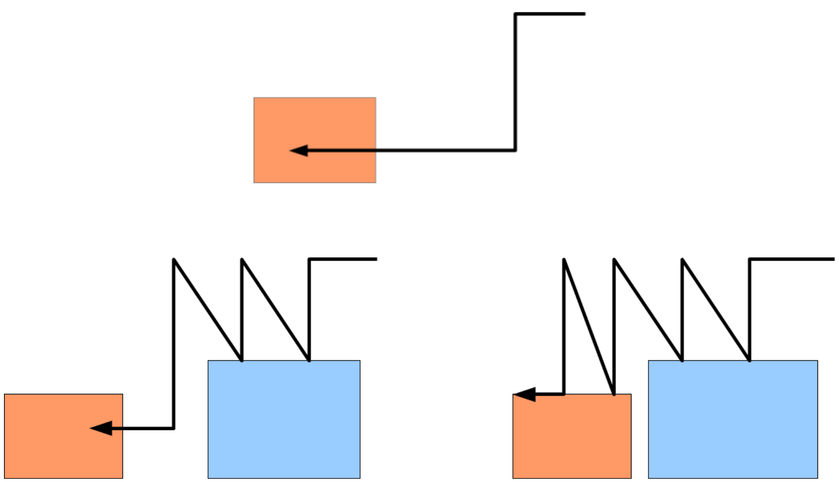
# Object pushing

- Pushing: move the object sufficiently for segmentation, but:
  - Keep object in field of view
  - Do not change visual appearance too much
- Push object over a fixed distance towards a central point in front of the robot



# Object pushing

- Minimize risk of collisions with other objects:
  - Approach the object from the top
  - Move the hand down beside the object,  then push it
  - Raise the hand, move it back out of sight
- Detect collisions using force-torque-sensor in the wrist
- Adapt approaching path reactively



# Overview: Interactive Segmentation



Create initial object hypotheses

Push

Re-localize, determine transformation

No motion

Object moved

Verify and improve hypothesis

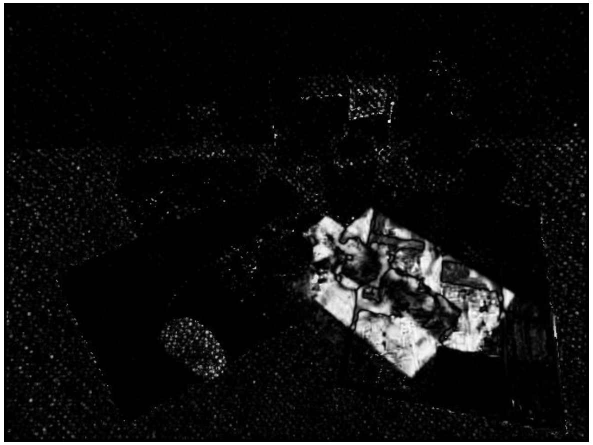
Push again

# Motion estimation

- After pushing: Re-localize object hypotheses
  - Textured objects: match SIFT features
  - More general alternative: use point cloud matching
- For each hypothesis: estimate motion
  - If it didn‘t move, ignore it
  - If it moved: objectness verified
- Verified object: Segment it to learn a visual object descriptor

# Motion estimation (I)

Check all hypotheses from changed image regions



# Motion estimation (II)

Check all hypotheses from changed image regions

↓

First: ICP starting at old position

- Re-localize object and estimate motion by point cloud matching
- Object hypotheses represented by sets of RGBD points
- Iterative Closest Point (ICP) for matching, using a distance in cartesian and color space

# Motion estimation (III) - ICP

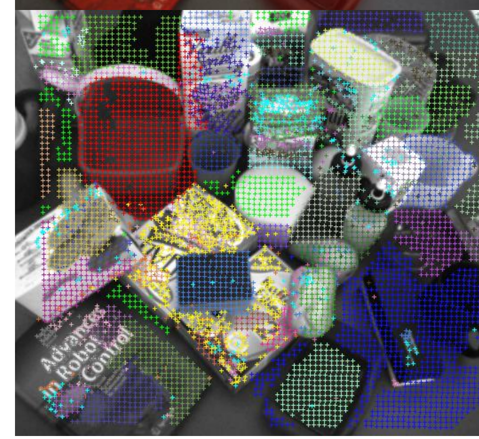
- Match two point clouds ***A*** and ***B*** using ICP:
  1. For each ***a*** ϵ ***A*** find closest point in ***B***
  2. Calculate transformation ***T*** that minimizes the mean squared distance of the correspondences
  3. Apply ***T*** to all ***a*** ϵ ***A***
- Iterate until convergence

# Motion estimation (III) - ICP

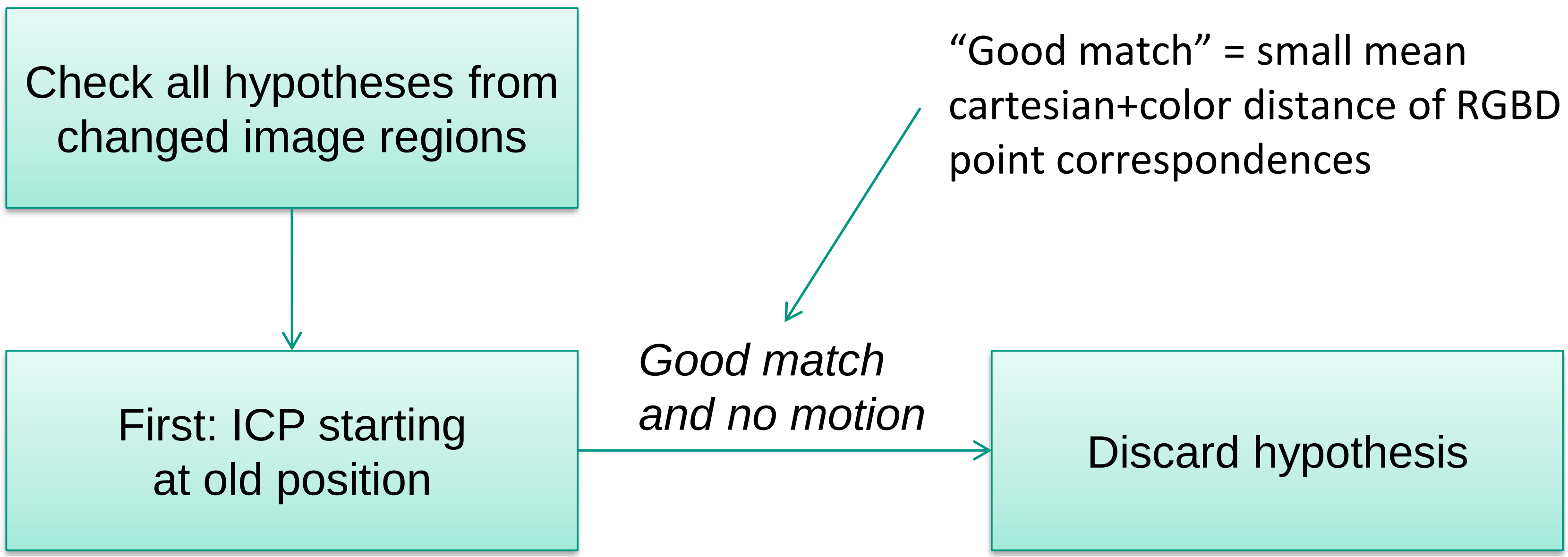
- Match two point clouds ***A*** and ***B*** using ICP:
  1. For each ***a*** ϵ ***A*** find closest point in ***B***
  2. Calculate transformation ***T*** that minimizes the mean squared distance of the correspondences
  3. Apply ***T*** to all ***a*** ϵ ***A***
- Iterate until convergence

- Problems:
  - Find small object in complex scene
  - Object only partially covered, false points included
  - 3D shapes ambiguous, e.g. many planes in most scenes

- Use weighted cartesian+color distance in step 1 of ICP
  - Removes most shape ambiguities
  - Gives more reliable point correspondences

# Motion estimation (IV)



# Motion estimation (V)

Check all hypotheses from changed image regions

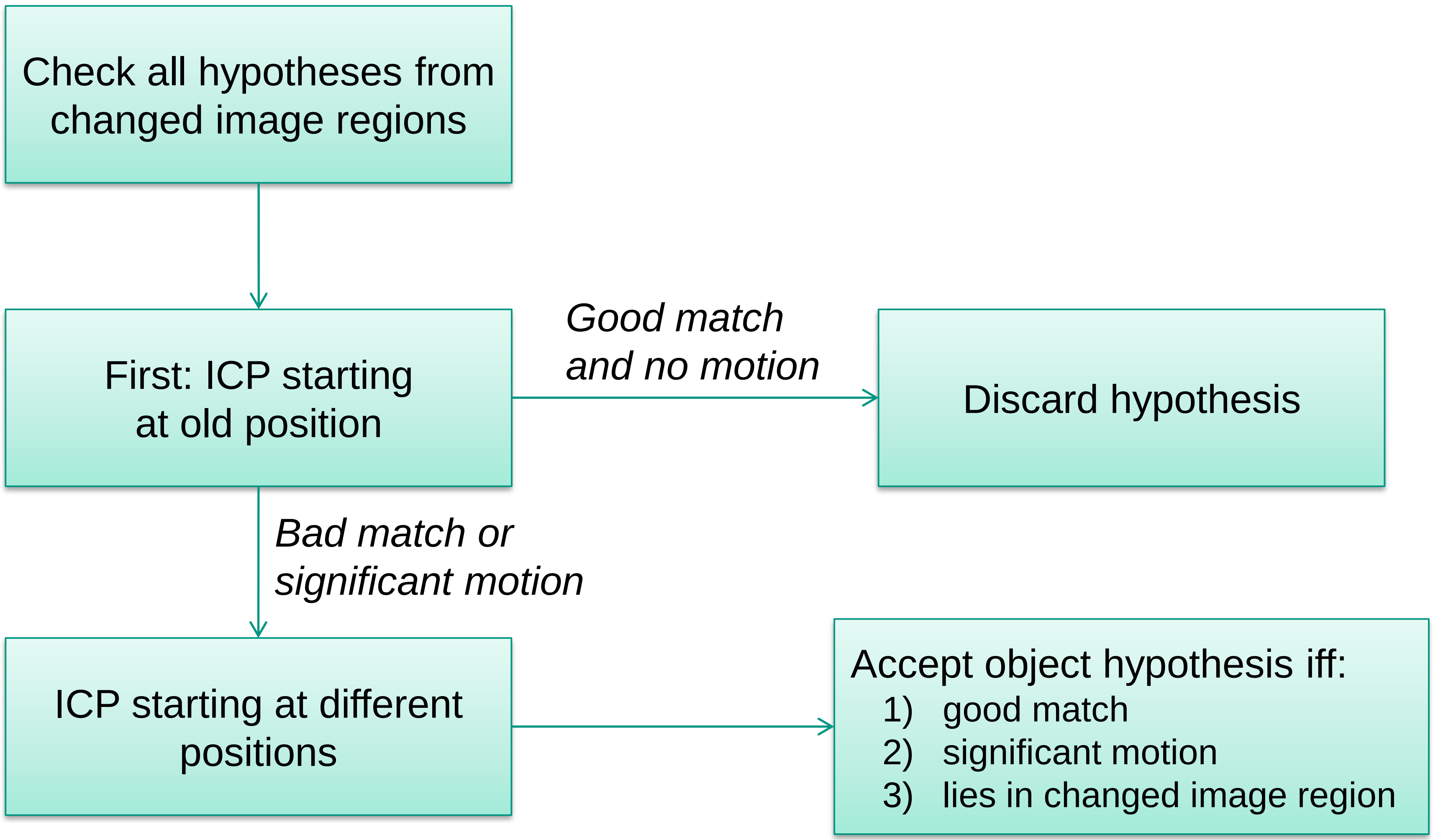
First: ICP starting at old position

*Good match and no motion* → Discard hypothesis

*Bad match or significant motion* → ICP starting at different positions

- ICP finds a local optimum
  → initiate it in different regions with color histogram similar to hypothesis

# Motion estimation (VI)



# Overview

Create initial object hypotheses

Push

Re-localize, determine transformation

No motion

Object moved

**Verify and improve hypothesis**

Push again

# Hypothesis correction and extension

- Discard points that
  - don’t accord with the overall motion or
  - come to lie in an unchanged image region

i.e. the point from the initial hypothesis has no good position+color match in the new point cloud after the transformation of the whole object has been applied to it

- Add new candidate points that lie in a changed image region close to the hypothesis

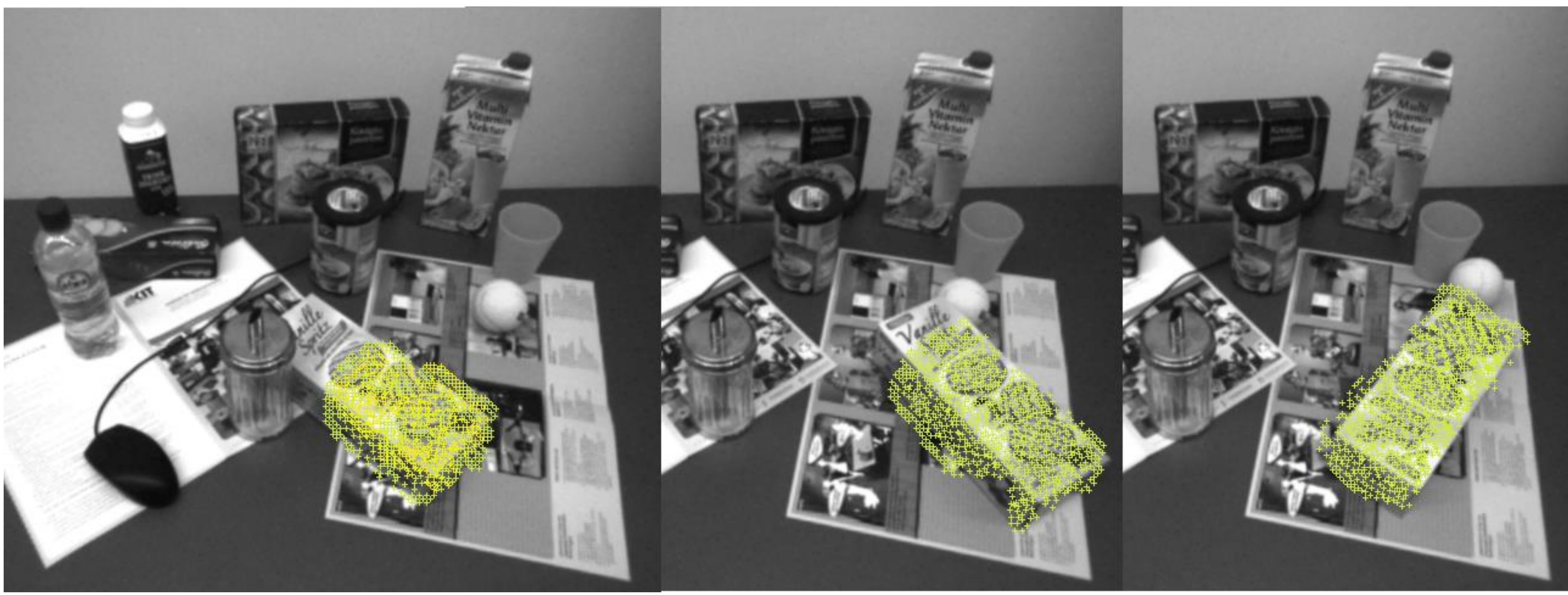
- Improve hypothesis over several pushes

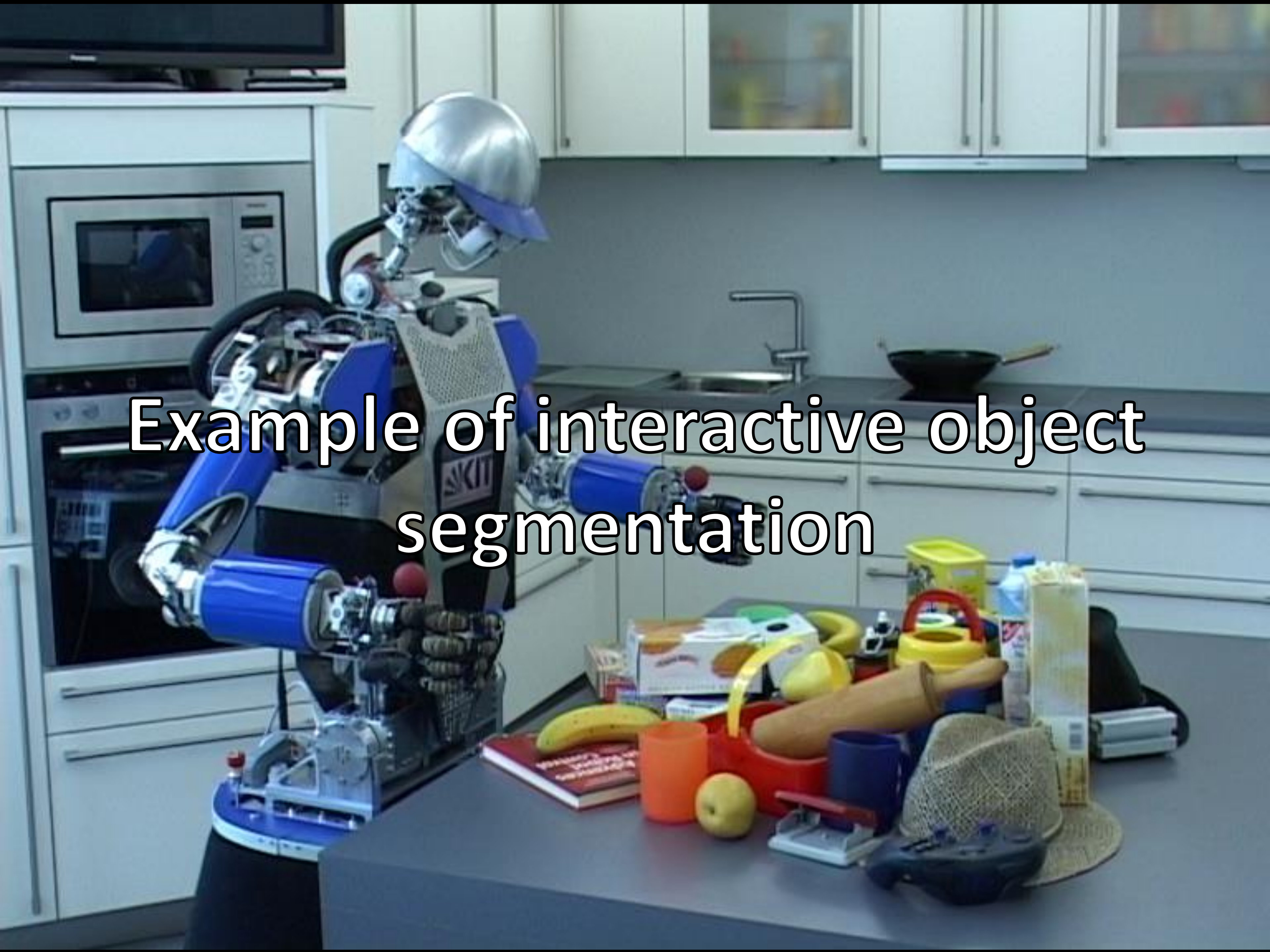
# Hypothesis correction and extension

- Push the object 2-3 times

  $\Rightarrow$ complete segmentation
- More pushes reveal different sides

  $\Rightarrow$ generate a multi-view descriptor




Example of interactive object segmentation
KIT

Left robot camera image

# Initial object hypotheses


Left camera
Nesquik
Advances in Robot Control
x2
Initial hypotheses
Outside view
Planned push through the center of the object hypothesis

# Old camera image

# New camera image

# Changed image regions

# Confirmed object hypotheses


Crosses are confirmed points, dots newly added candidates


Left camera
New hypotheses
x2
Outside view
Old hypotheses
Nesquik
Advances in Robot Control
Popcorn

# Old camera image

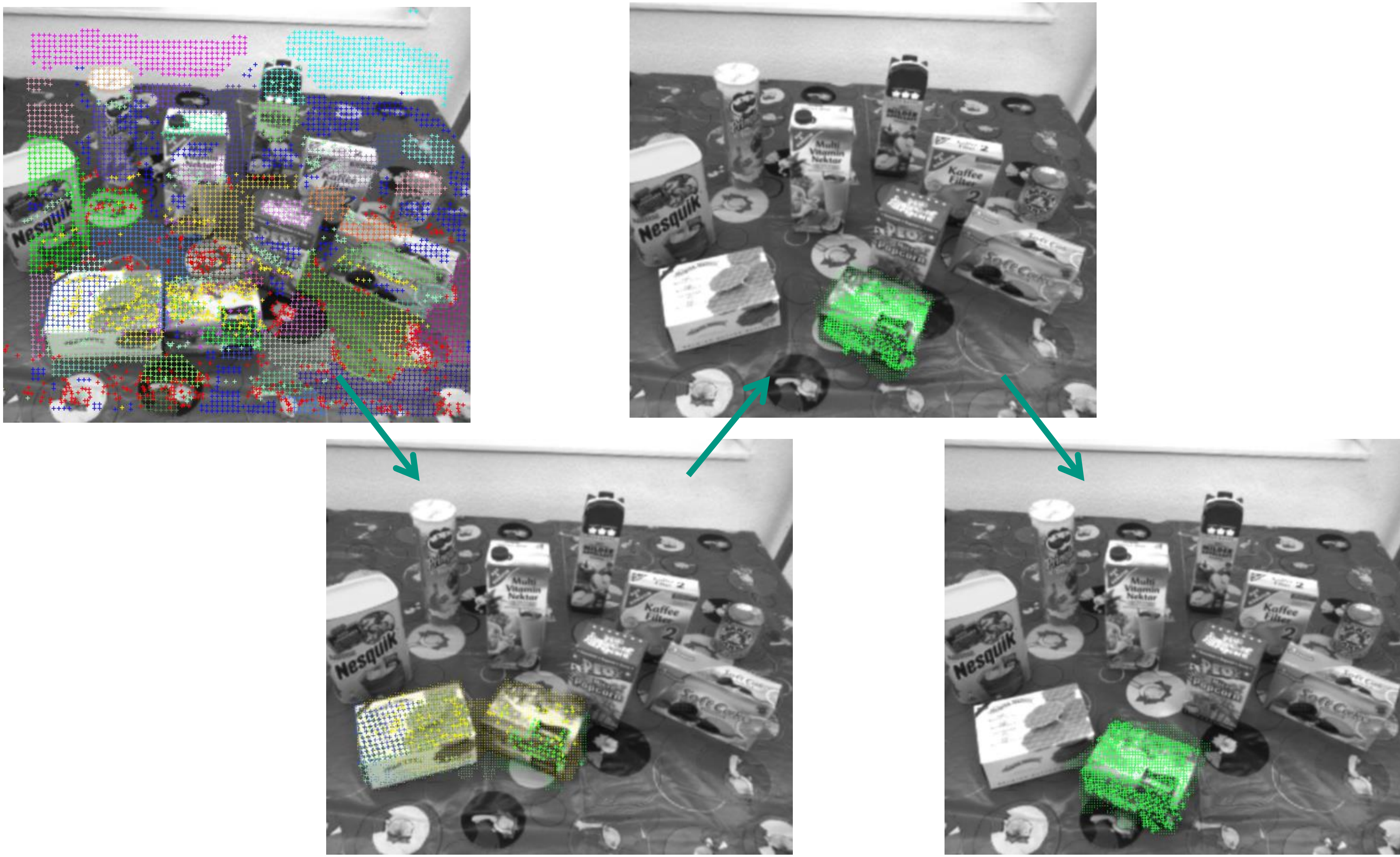
# New camera image

# Changed image regions

# Confirmed object hypothesis

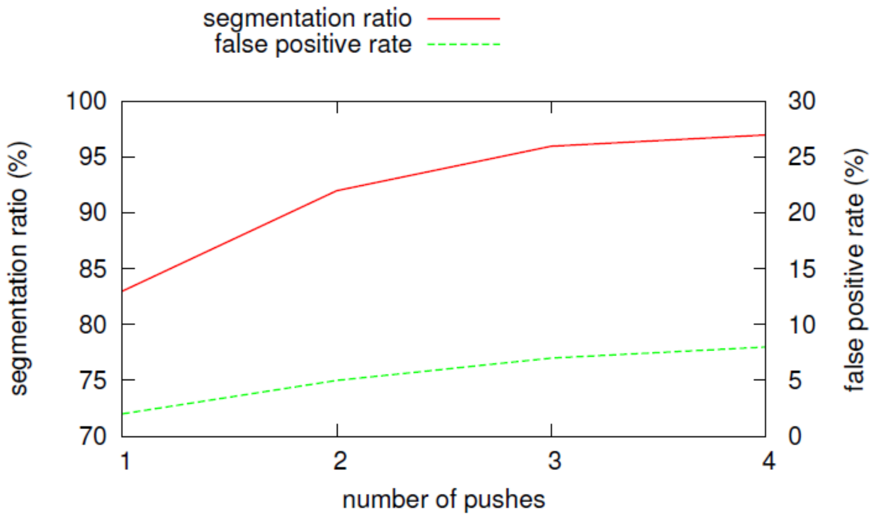

Crosses are confirmed points, dots newly added candidates

# Object segmentation example



# Object Learning for Recognition: Results

- Segmentations usually correct and complete




- Proof of concept: Simple object descriptor created based on the segmentation
  - ⇒ solid recognition results

| similar point of view | different point of view | partly occluded | false positive rate |
|---|---|---|---|
| 98.5 % | 70.6 % | 67.2 % | 3.8 % |

Object recognition rates

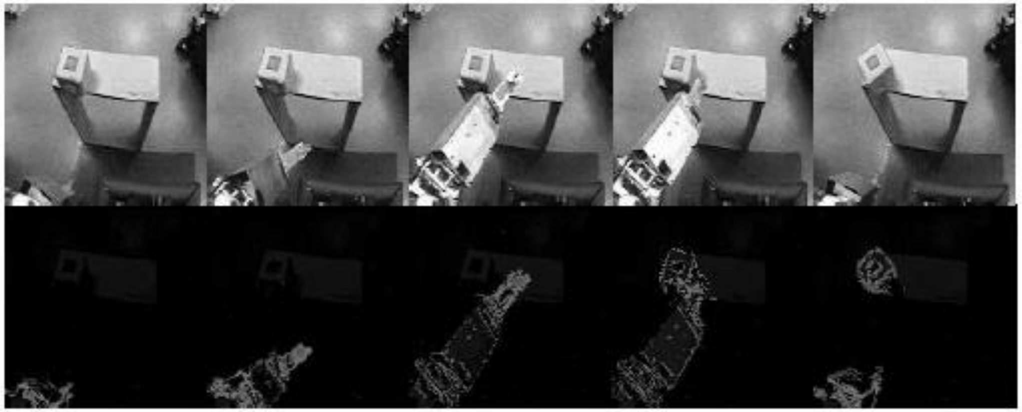
# Publications

- D. Schiebener, A. Ude and T. Asfour, *Physical Interaction for Segmentation of Unknown Textured and Non-textured Rigid Objects*, IEEE International Conference on Robotics and Automation (ICRA), 2014
- D. Schiebener, J. Morimoto, T. Asfour and A. Ude, *Integrating visual perception and manipulation for autonomous learning of object representations*, Adaptive Behavior, 2013
- A. Ude, D. Schiebener, N. Sugimoto and J. Morimoto, *Integrating surface-based hypotheses and manipulation for autonomous segmentation and learning of object representations*, IEEE International Conference on Robotics and Automation (ICRA), 2012
- D. Schiebener, A. Ude, J. Morimoto, T. Asfour and R. Dillmann, *Segmentation and learning of unknown objects through physical interaction*, IEEE/RAS International Conference on Humanoid Robots (Humanoids), 2011

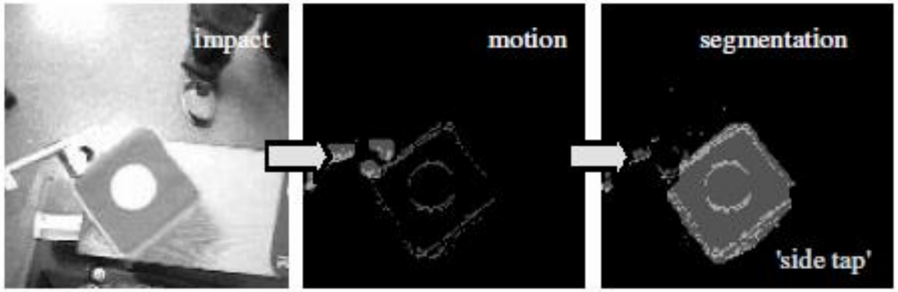
Red: relevant for the exam

# Related work

- G. Metta and P. Fitzpatrick, *Grounding vision through experimental manipulation*, Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences, 2003



- Approach object with arm
- Track arm, observe optical flow
- Sudden spread of motion next to the hand indicates collision with object



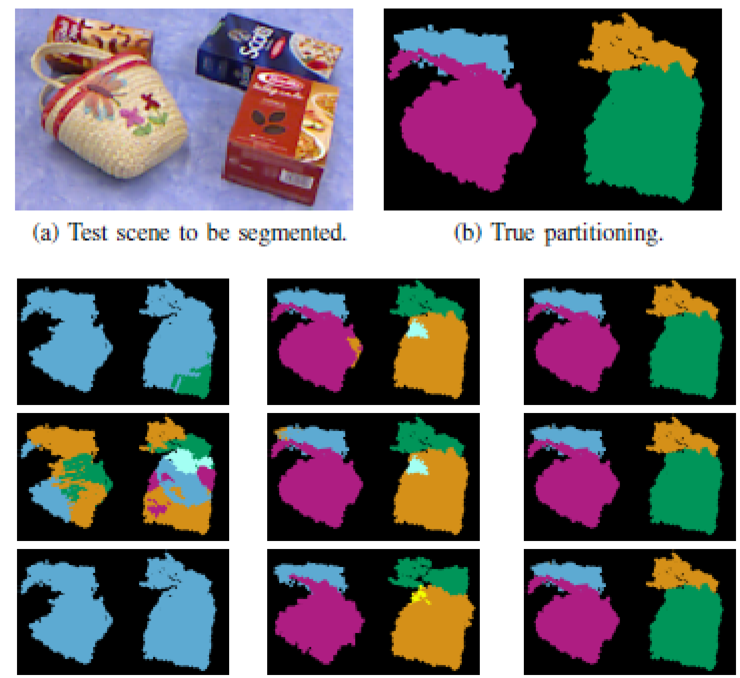
# Related work

- D. Katz and O. Brock, *Manipulating articulated objects with interactive perception*, IEEE Int. Conf. Robotics and Automation (ICRA), 2008
- Touch a 2D articulated object
- Observe motion of SIFT-features
- Group them by relative motion
- Deduce position of joints

# Related work

- H. van Hoof, O. Kroemer and J. Peters, *Probabilistic Interactive Segmentation for Anthropomorphic Robots in Cluttered Environments*, IEEE/RAS International Conference on Humanoid Robots (Humanoids), 2013

- Over-segment scene into regions
- Interaction to cause motion
- Use observed motion of regions to update probabilistic partitioning of the whole scene into objects



(a) Test scene to be segmented.

(b) True partitioning.

Prior samples (0 actions)

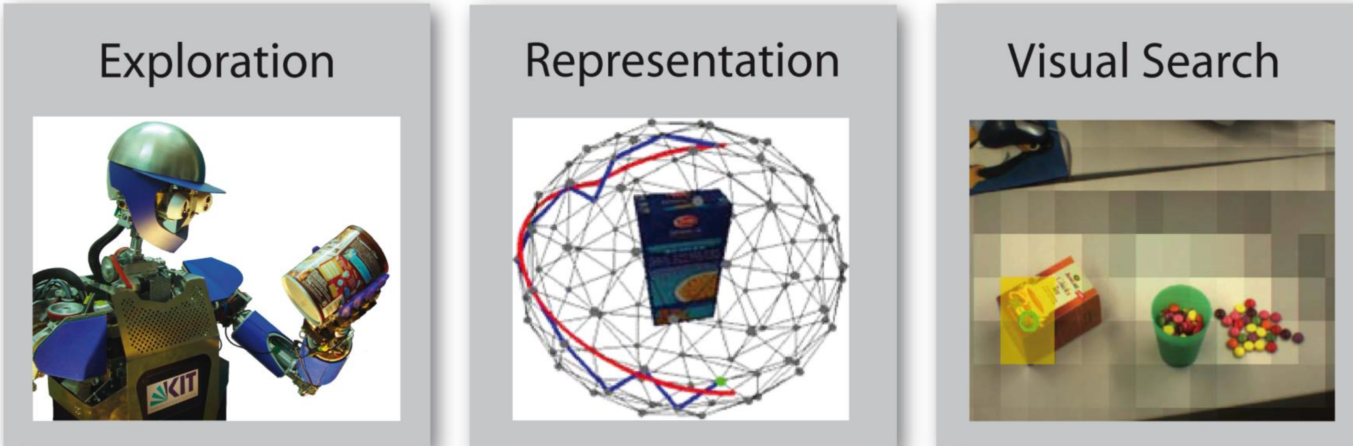
Posterior samples (5 actions)

Posterior samples (15 actions)

# More Related work


- E. S. Kuzmic and A. Ude, *Object segmentation and learning through feature grouping and manipulation*, IEEE-RAS Int. Conf. Humanoid Robots (Humanoids), 2010
- W. H. Li and L. Kleeman, *Segmentation and modeling of visually symmetric objects by robot actions*, Int. Journal of Robotics Research, 2011
- L. Chang, J. Smith and D. Fox, *Interactive singulation of objects from a pile*, IEEE International Conference on Robotics and Automation (ICRA), 2012
- M. Gupta and G. Sukhatme, *Using manipulation primitives for brick sorting in clutter*, IEEE International Conference on Robotics and Automation (ICRA), 2012
- K. Hausman et al., *Tracking-based Interactive Segmentation of Textureless Objects*, IEEE International Conference on Robotics and Automation (ICRA), 2013

# Active Visual Object Exploration and Search

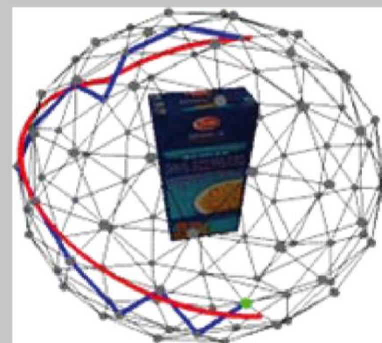



- *Welke, K., Issac, J., Schiebener, D., Asfour, T. and Dillmann, R. Autonomous Acquisition of Visual Multi-View Object Representations for Object Recognition on a Humanoid Robot. IEEE International Conference on Robotics and Automation (ICRA 2010)*
- *Welke, K., Asfour, T. and Dillmann, R. Active Multi-View Object Search on a Humanoid Head. 2009 Proceedings of the IEEE International Conference on Robotics and Automation, pp. 417-423, (ICRA 2009)*
- *Ude, A., Omrcen, D. and Cheng, G. Making object learning and recognition an active process In International Journal of Humanoid Robotics, 5(2), pp. 267-286, 2008*

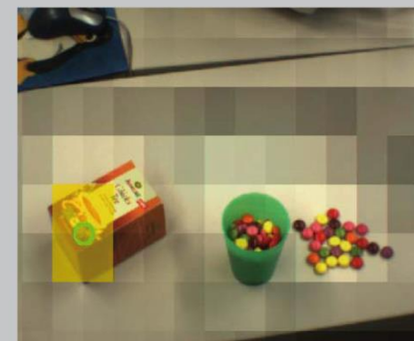
# Active Visual Object Exploration and Search

Exploration

Representation

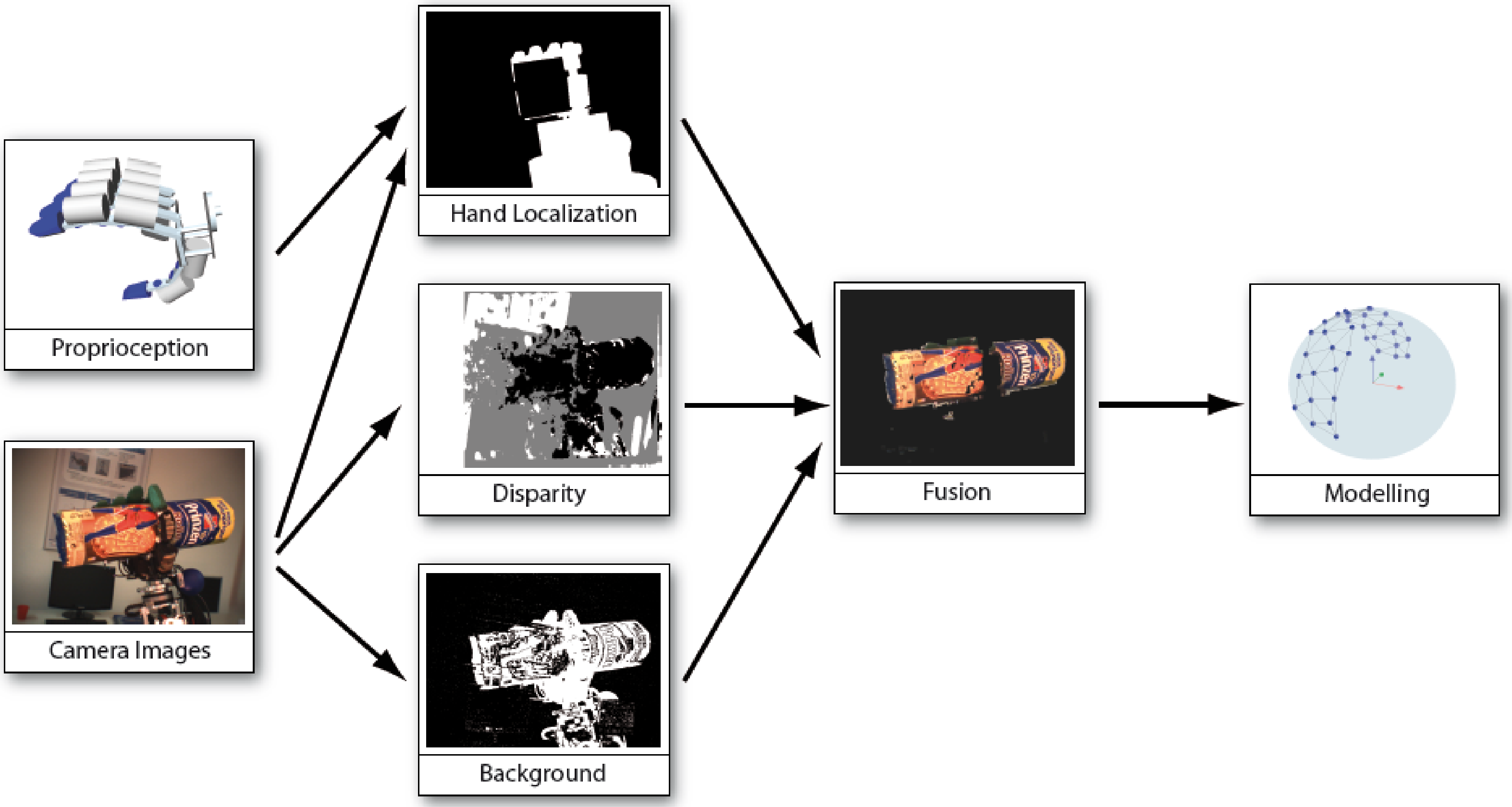


Visual Search



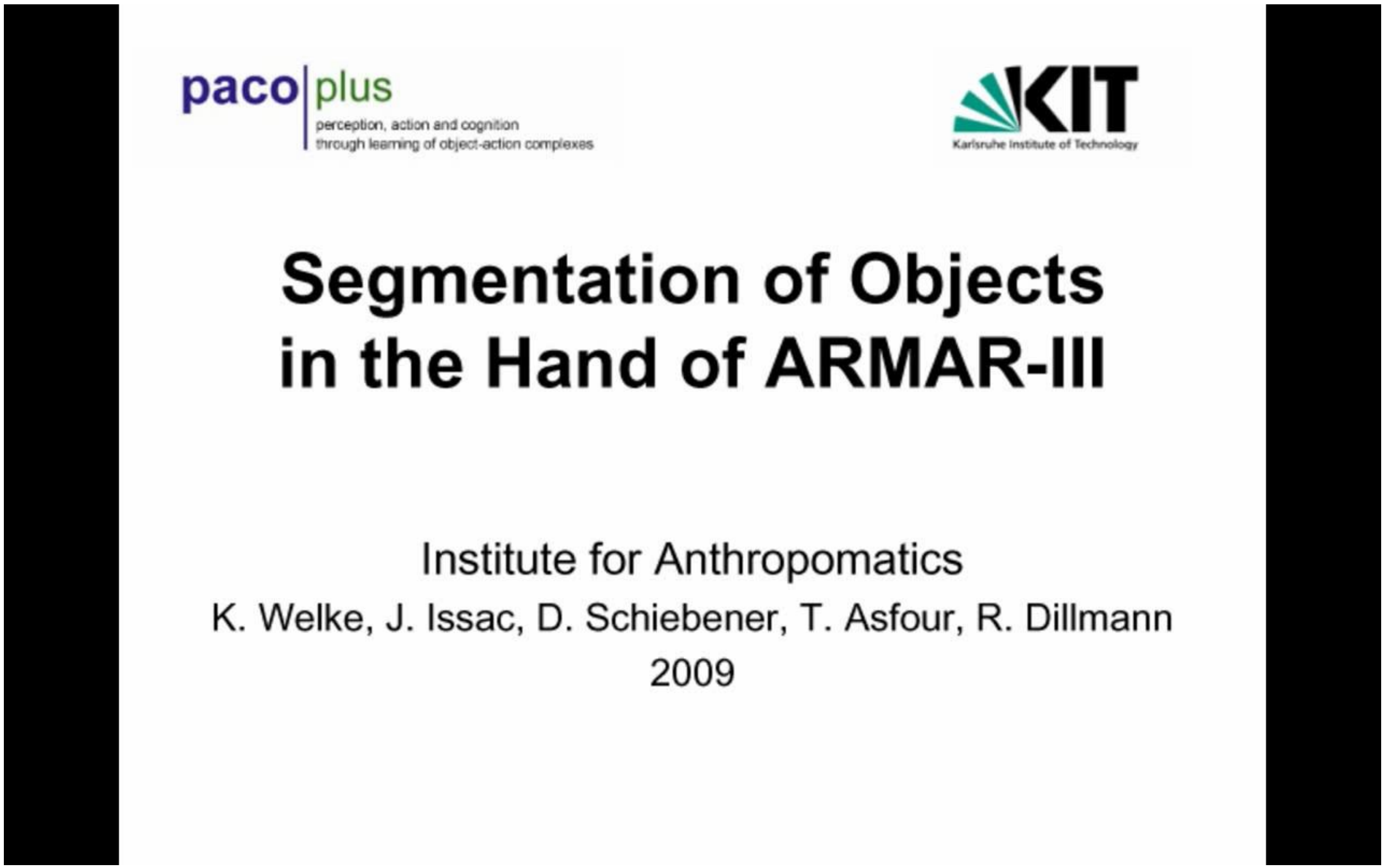
- Generation of visual representations through exploration
- Application of generated representations in recognition tasks

# Visual exploration of unknown objects

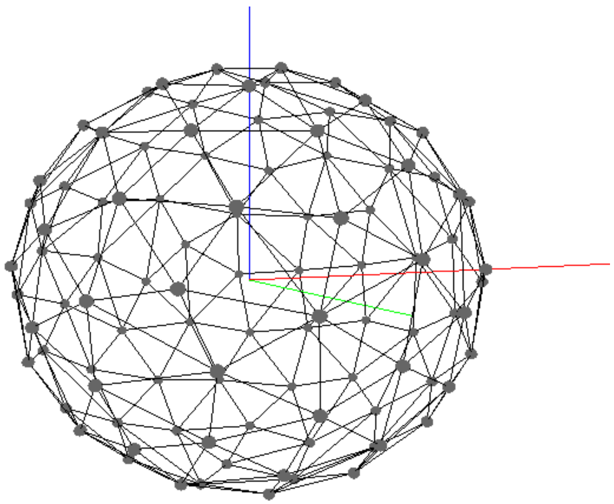


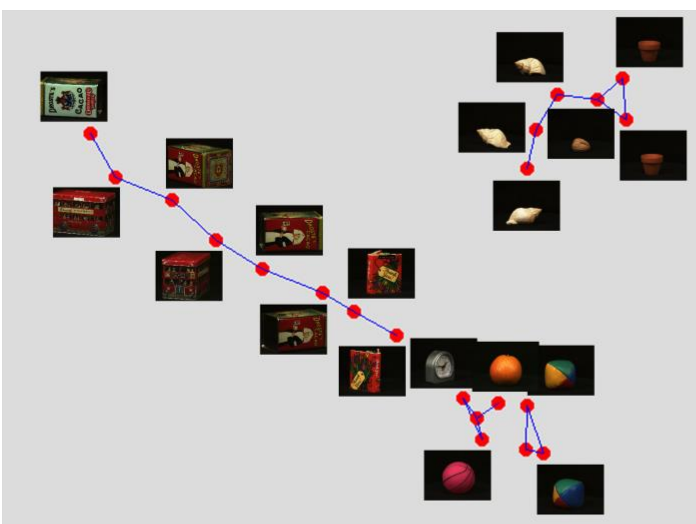
# Visual exploration of unknown objects

- Background-foreground and hand-object segmentation
- Generation of different views through manipulation



# Representation

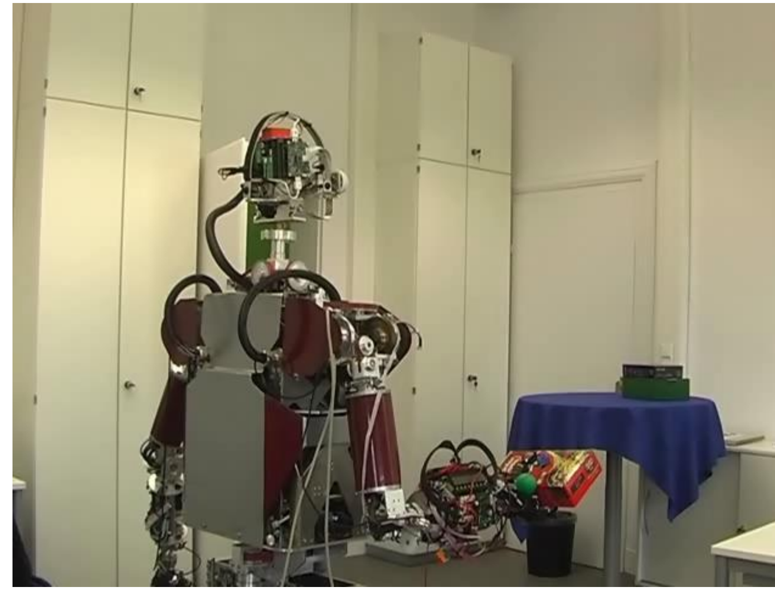




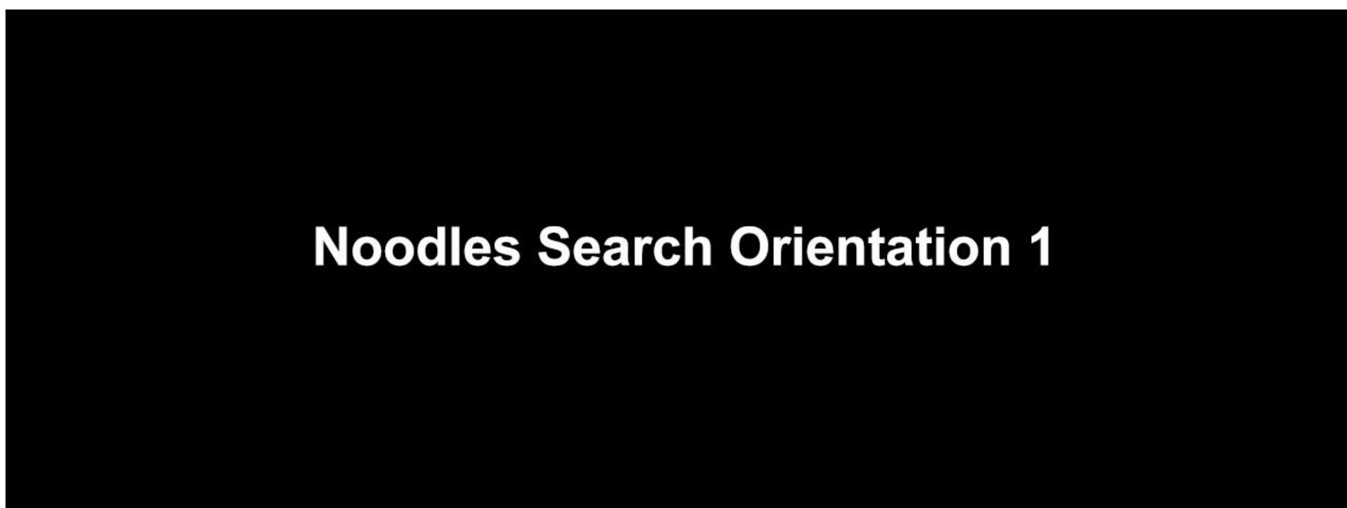
- Aspect Graph
  - Multi-view appearance-based representation
  - Each node corresponds to one view
  - Edges describe neighbor relations
- Feature Pool
  - Compact representation of views with prototypes
  - Grouping based on visual similarity

# Active visual search

- Active Search
  - Object search using perspective and foveal camera of Karlsruhe Humanoid Head
- Scene memory
  - Integration of object hypotheses in an ego-centric representation

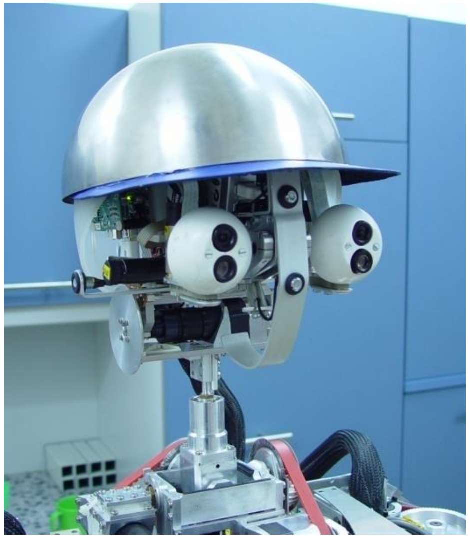


ICRA 2010
Humanoids 2009
ICRA 2009



# Active Visual Search

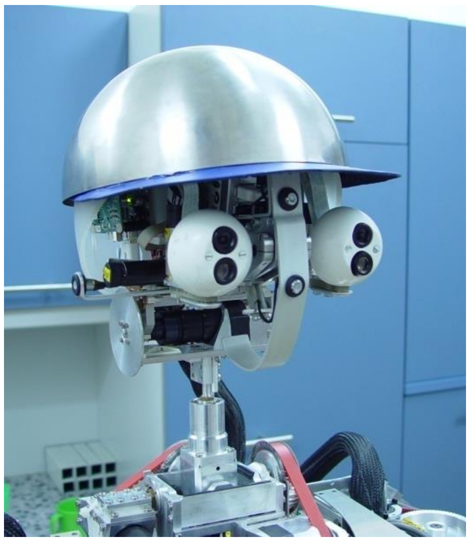
# Active Visual Search



peripheral view

foveal view

# Active Visual Search



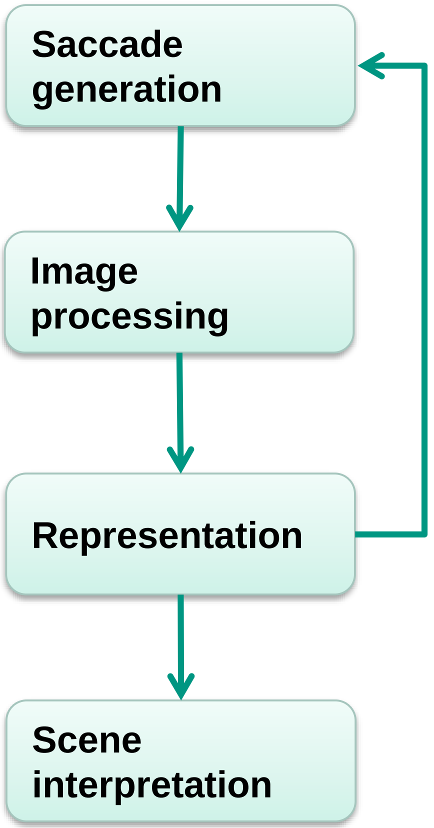
peripheral view        foveal view

- Tasks

Saccade generation → Image processing → Representation → Scene interpretation

# Active Visual Search and Representation

- Active visual search
  - Search for known target object
  - Generation of saccadic eye movements
  - Object detection and recognition
- Representation
  - Transsaccadic memory
  - Perception as continuous process

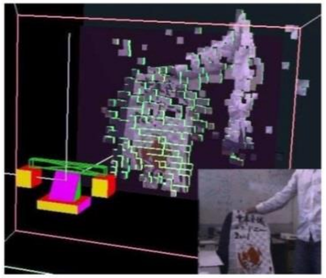



Kai Welke "Memory-Based Active Visual Search for Humanoid Robots", phd thesis, KIT, 2011

# Related Work

- Foveal Vision
  - Search and pursuit using signatures
    *[Ude et al., 2003]*
  - Search based on depth information
    *[Bjorkman and Kragic, 2004]*
  - Bottom-up saliency and weights
    *[Rasolzadeh et al., 2010]*
  - Saliency based on color *[Orabona et al., 2005]*
- Representations
  - Occupancy Grid (3D)
    *[Dankers et al., 2009]*
  - Sensory Egosphere (2D)
    *[Figueira et al., 2009]*

*[Ude et al., 2003]*



*[Dankers et al., 2009]*

*[Figueira et al., 2009]*

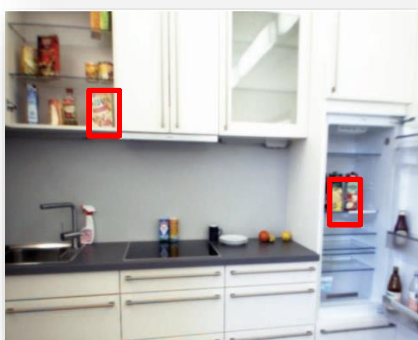
No integration of active visual search and representation.

# Active Visual Search

- Complexity of visual search
  - General visual search problem: NP-complete

- Approach
  - Knowledge of the target object model: linear complexity
  - Decomposition of the problem:

Coarse Analysis → Validation



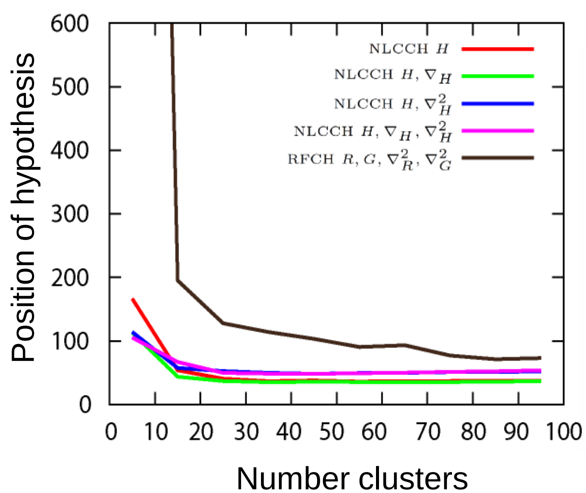
Peripheral view


Foveal view

# Object search in the peripheral view

- **Goal:** Restriction of the search space
- **Approach**
  - Coarse analysis of the scene in peripheral view
  - Detection of object candidates

- **Methods**
  - Color Cooccurrence Histograms (CCH)
  - Search window for object candidate detection



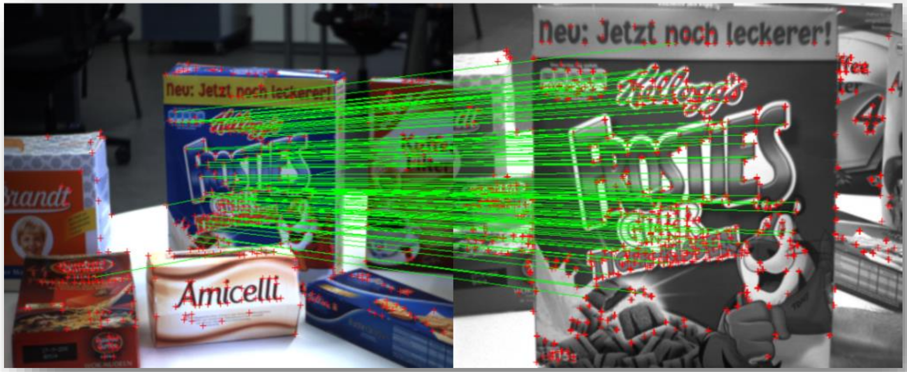

# Object recognition in the foveal view

Karlsruhe Institute of Technology

- **Goal:** Validation of object candidates
  - Foveal view allows for detailed analysis
  - Elimination of false positive object candidates


## ➡ Object recognition

  - Texture-based recognition based on Harris-SIFT features *[Azad et al., 2008]*
  - Calculation of feature correspondences with object model
  - Classification of object candidates

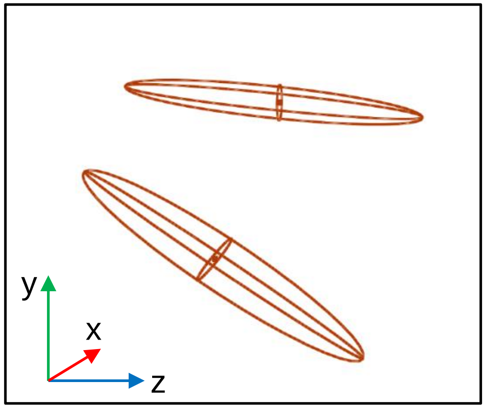


# Saccade generation

- **Goal**
  - Minimal number of saccades until object recognition
  - ➡ Gaze direction with maximum probability of recognition
- **Approach**
  - Saliency based on the Bayesian Strategy *[Torralba, 2003]*

$$p(O=1,X|F) = \underbrace{\frac{1}{p(F)}}_{\text{Bottom-up}} \cdot \underbrace{p(F|O=1,X)}_{\text{Object model}} \cdot \underbrace{p(X|O=1)}_{\text{scene priors based on the spatial location}} \cdot \underbrace{p(O=1)}_{\text{prior of object existence}}$$

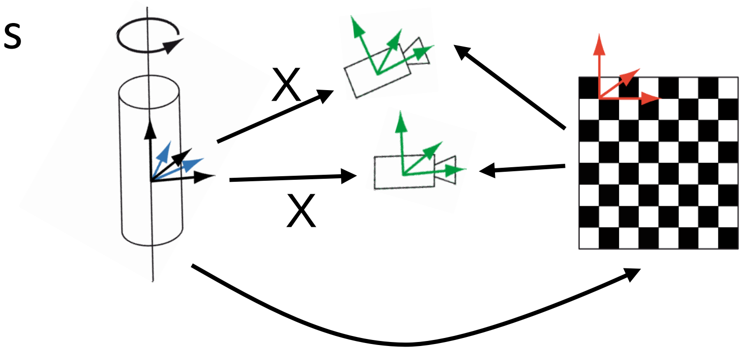
- **Representation of saliency**
  - Landmark-based map of candidates
    - Localization uncertainty
    - Probability of existence
  - Approximates $p(O=1,X|F)$



Localization unceraint for 2 candidates

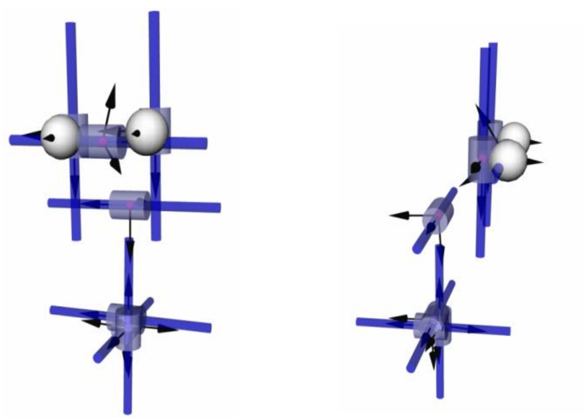
## Execution of saccades

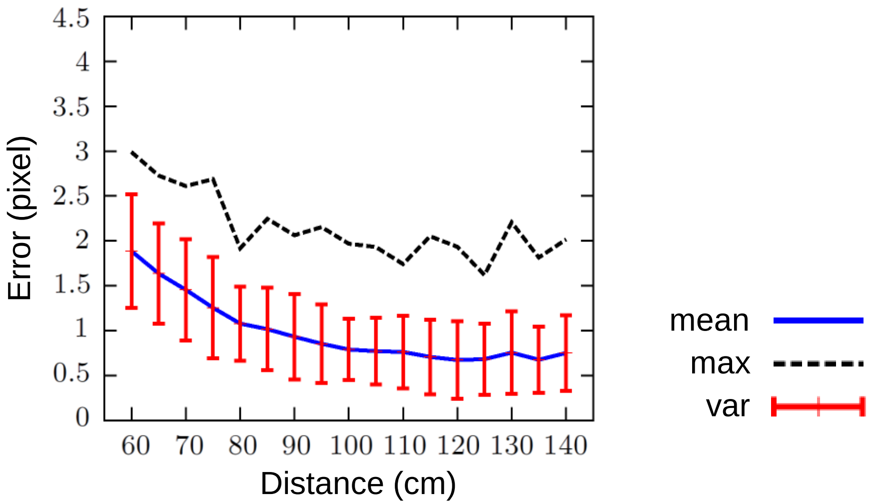
- **Kinematic model for saccade execution**
  - Pose of the camera coordinate systems unknown
  - Inaccuracies in CAD model




**➡ Kinematic Calibration**

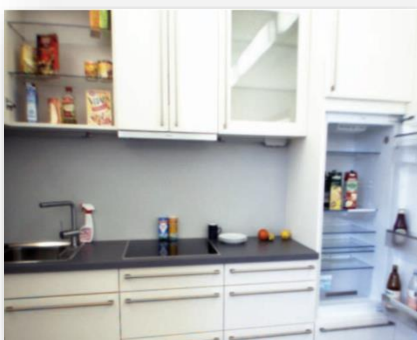
- Visually-based
- Calibration of all joints



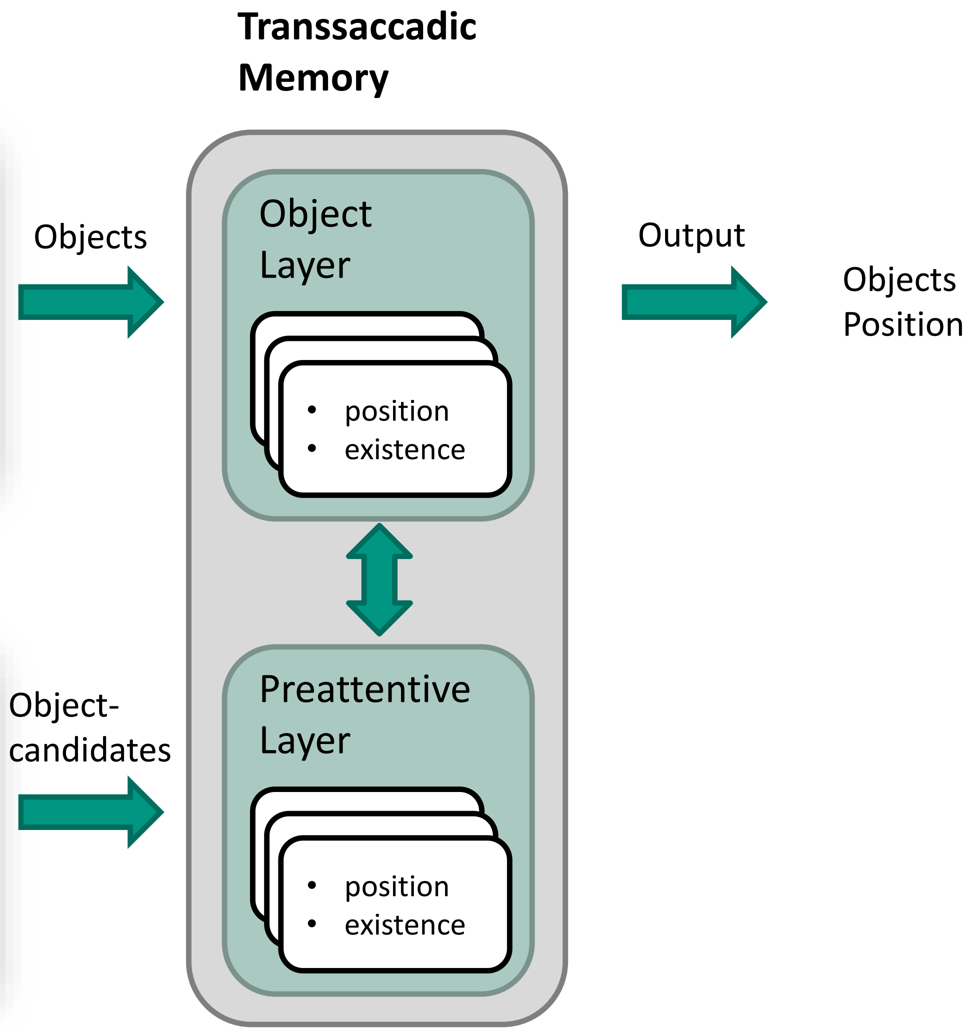


# Transsaccadic Memory

**Foveal camera image**



**Peripheral camera image**



# Transsaccadic Memory – Update

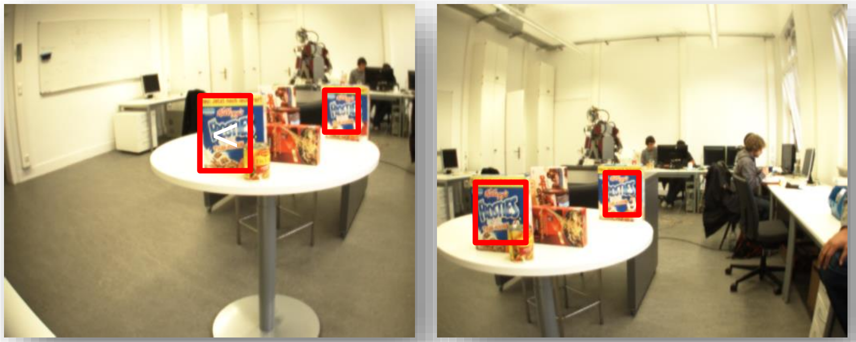
- Update of the Preattentive Layer

- Update of the Object Layer

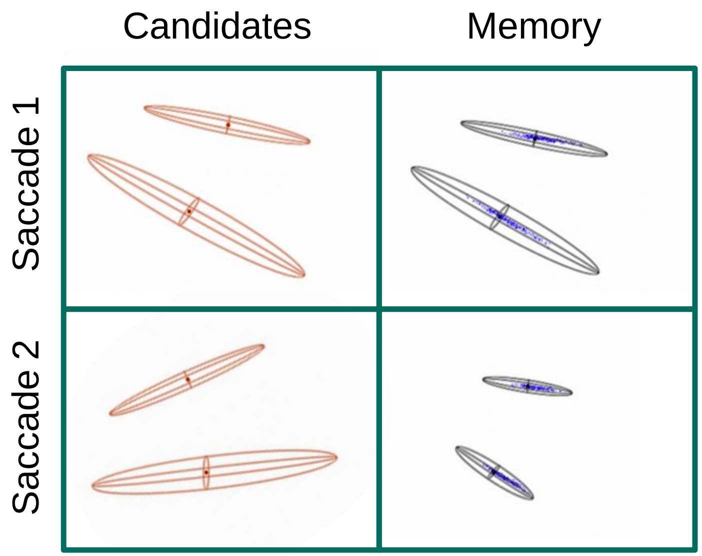
- Consistency of scene and memory

# Update of the Preattentive Layer

- **Problem**
  - Object candidates observed from different viewing directions
  - Correspondence problem

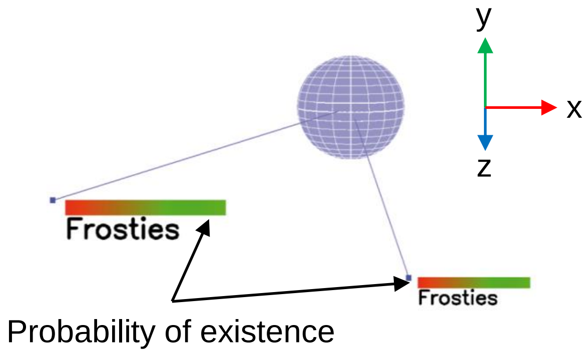


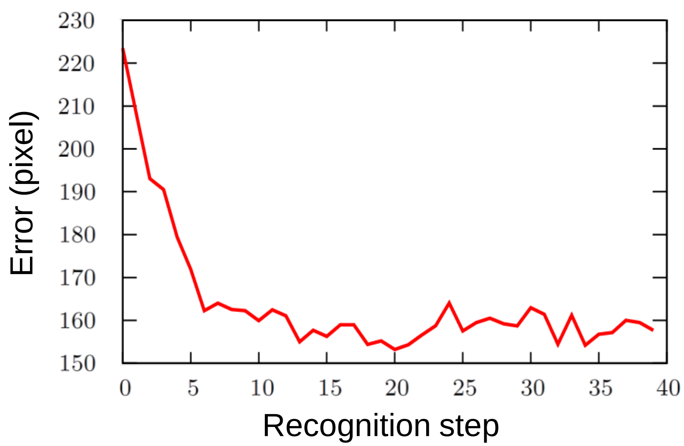
- **Estimation of correspondences and update**
  - Probabilistic model
    - Uncertainty of execution and calibration
    - Inference using Rao-Blackwell Particle Filters
  - Update of position and existence using Bayes and Kalman Filters
  - Correspondence using maximum a posteriori estimate



# Update of the Object Layer

- **Prerequisite**
  - Object candidate fixated in foveal cameras
  - ➡ Correspondence solved
- **Update of object existence**
  - Match probability
  - Update using Bayes Filter
- **Update of object position**
  - Closed loop
  - 2D position error in left and right camera



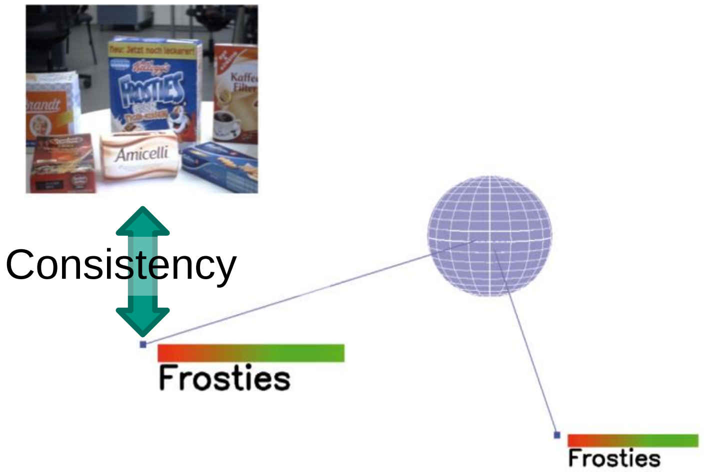



# Memory and Saccade Generation (I)

- **Requirement**

  Consistency of scene and memory
  - For each object instance a corresponding representation exists in memory
  - For each representation in memory a corresponding object instance exists

- **Approach**
  - Consistency is assured using foveal validation



# Memory and Saccade Generation (II)

- **Consequences for Saccade Generation**
  - Account for consistency of Object Layer
  - Gaze directions towards inconsistent memory entities

- Inconsistency $I$ depends on
  - Validation using foveal object recognition $V$
  - Change of the world $C$

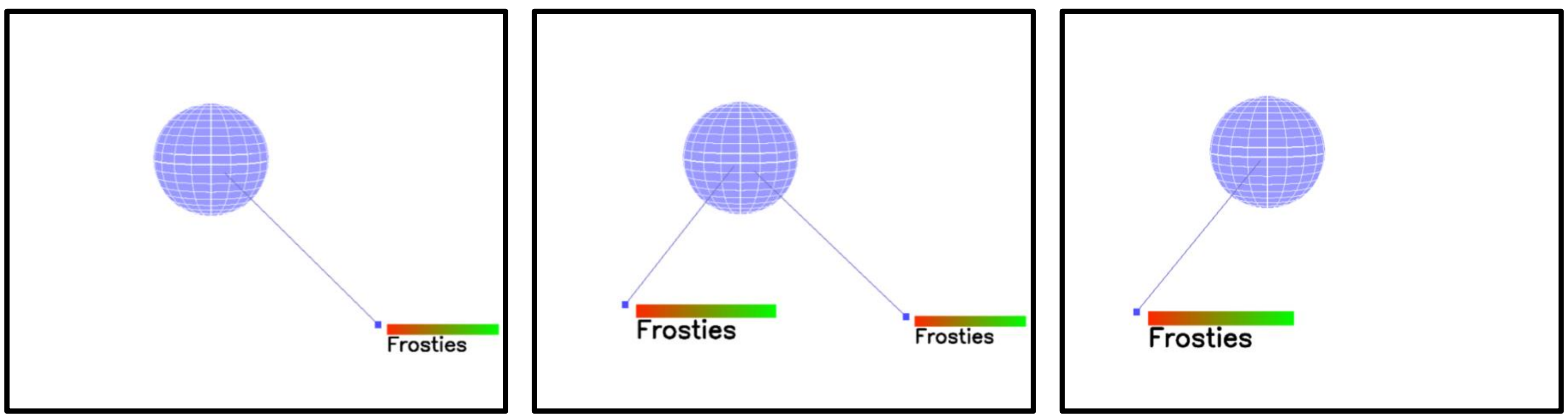
- **Active Saliency**

$$
\begin{aligned}
s_a &= p(O=1, X, I=1|Z) \\
&= \underbrace{p(O=1, X|F)}_{\text{Bayesian Strategy}}\underbrace{p(I=1|C,V)}_{\text{Inconsistency}}
\end{aligned}
$$

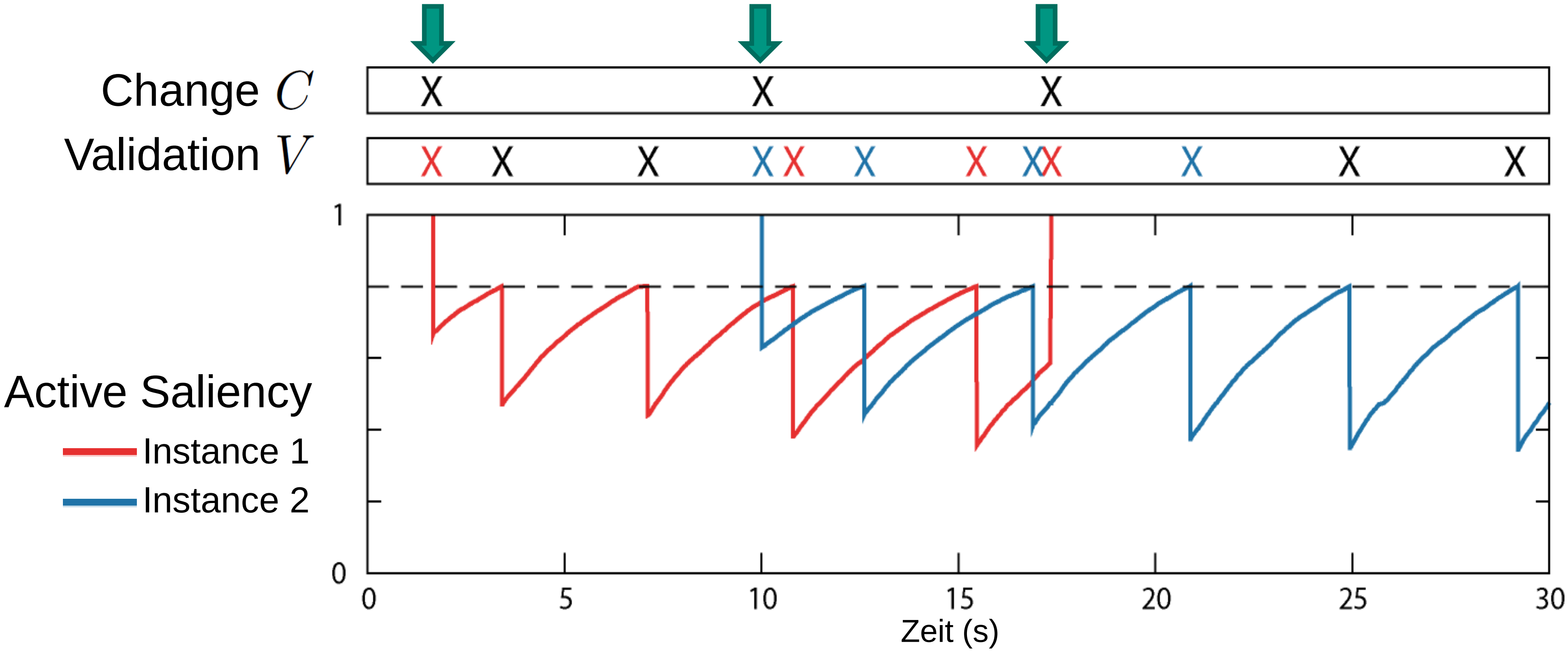
# Active Saliency: Example

- Changing scene with two object instances

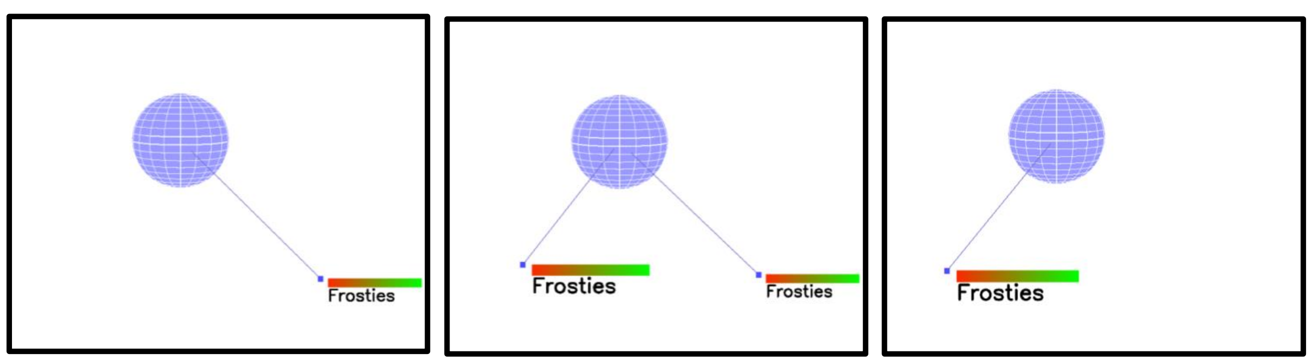
- **Approach:** Consistency of Object Layer



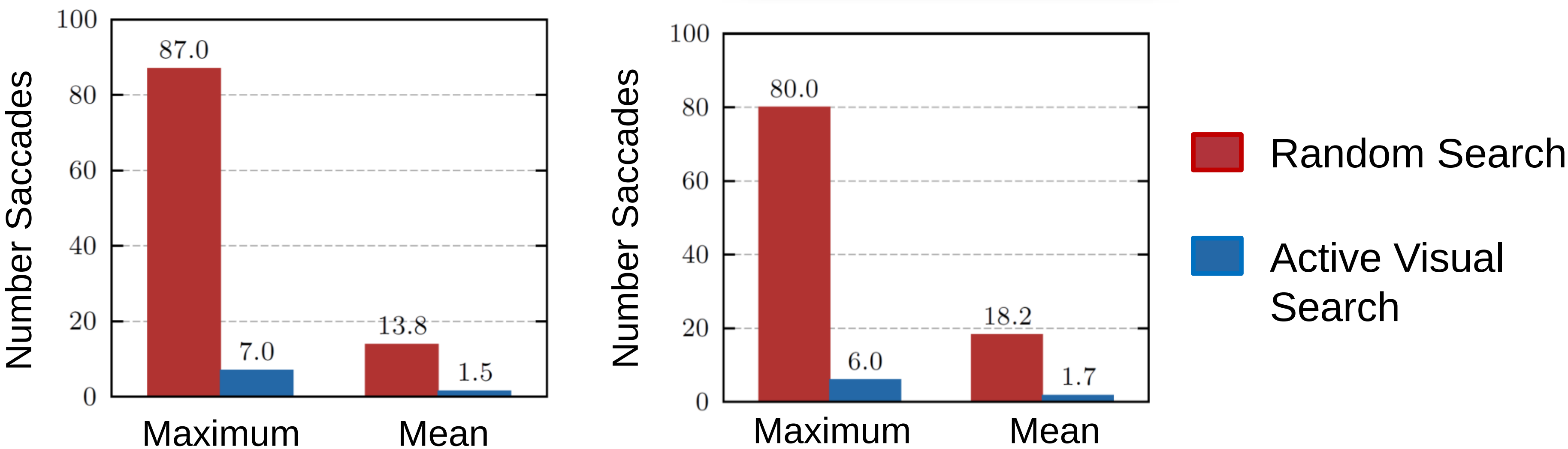
# Active Saliency: Example




Object Layer

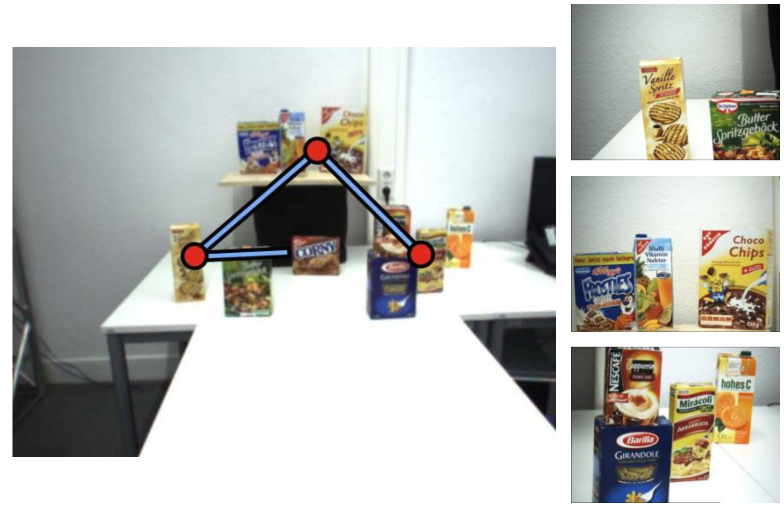


# Active Visual Search: 10 objects in 20 scenes

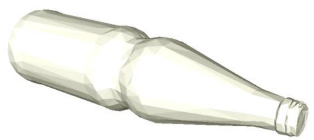


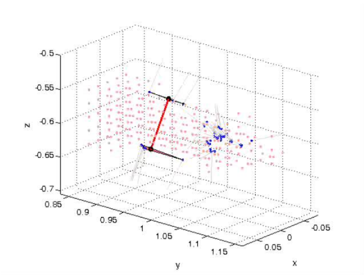
# Active scene exploration

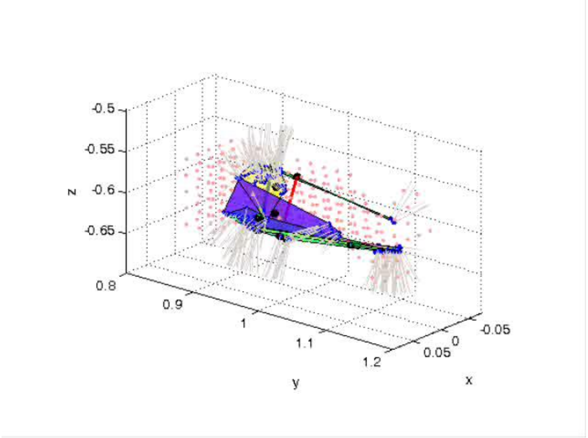
- **Active visual search**
  (Welke et al., 2009; 2011)
  - Analyze scene exploiting active foveal camera system
  - Build consistent scene representation
  - Continuous perception in changing environments



# Haptic exploration



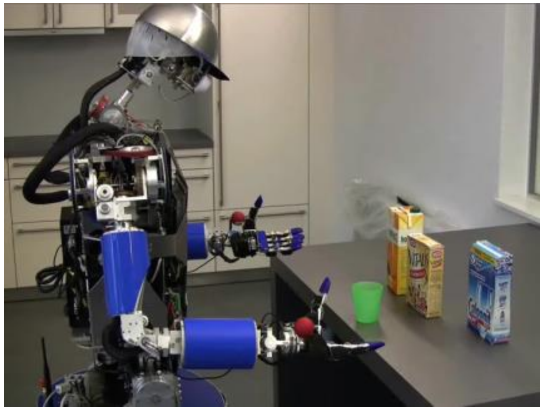



- *Bierbaum, A., Rambow, M., Asfour, T., Dillmann, R. Grasp Affordances from Multi-Fingered Tactile Exploration using Dynamic Potential Fields. In IEEE/RAS International Conference on Humanoid Robots, 2009.*
- *Bierbaum, A., Schill, J., Asfour, T., Dillmann, R. Force Position Control for a Pneumatic Anthropomorphic Hand. In IEEE/RAS International Conference on Humanoid Robots, 2009*
- *Bierbaum, A., Asfour, T., Dillmann, R. Dynamic Potential Fields for Dexterous Tactile Exploration. In Workshop on Human-Centered Robotics Systems (HCRS), 2009.*

Red: relevant for the exam

# Motivation



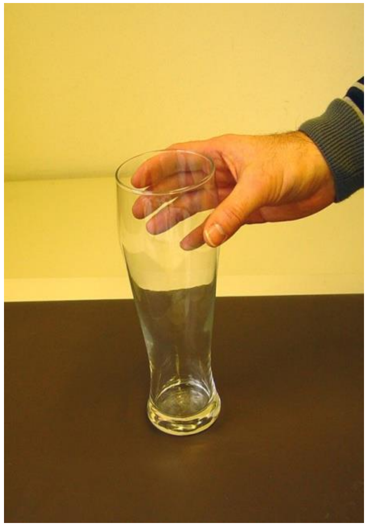
Humanoid robots in human-centered environments

- Manipulation of unknown objects
- Enhance and augment object information from visual sense

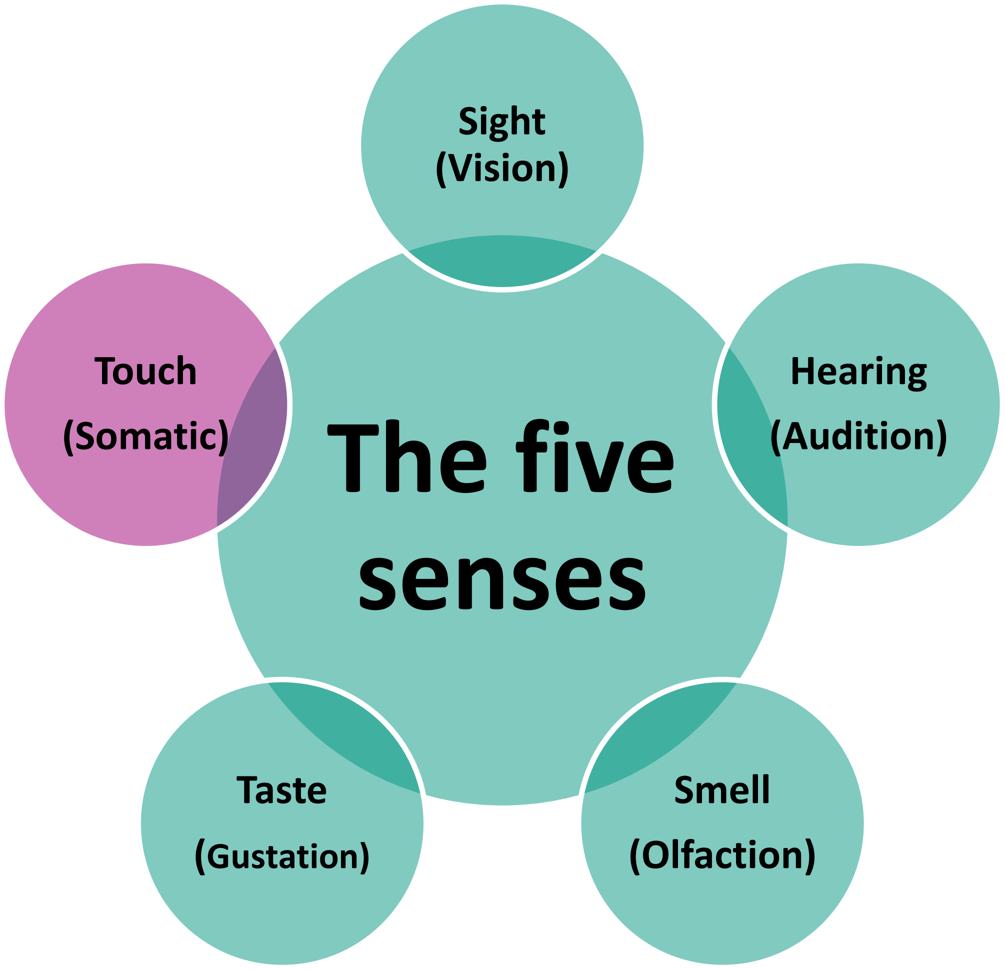
→ Haptically explore unknown objects

- Active touch information from haptic exploration enables human to sense object shape

  → Hints for classification, recognition and manipulation

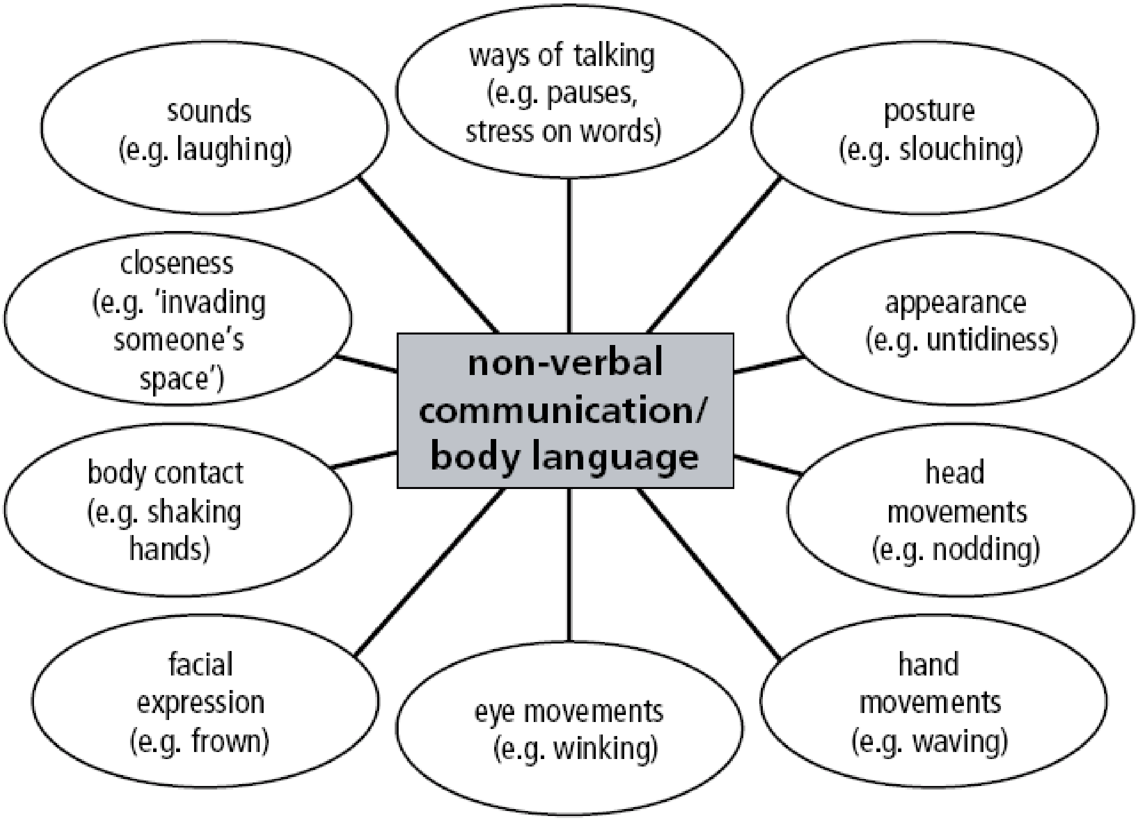


# Human senses



# Was is Haptics?

- The sense of touch
- Any form of nonverbal communication involving touch



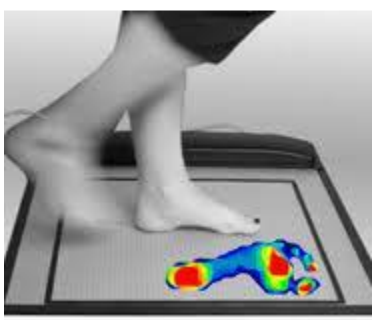

# What is Haptics

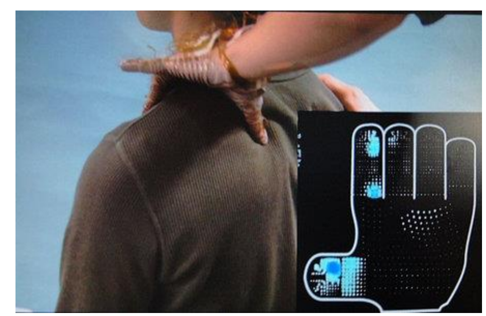
- We touch intending to
  - do a task
  - probe an object
  - poke to elicit a reaction
  - fidget to relieve tension
  - communicate a message
  - verify that an action is completed
  - enjoy aesthetic pleasure or comfort
  - connect physically or emotionally to living things
  - …

# What is there to sense?

Pick your body part:

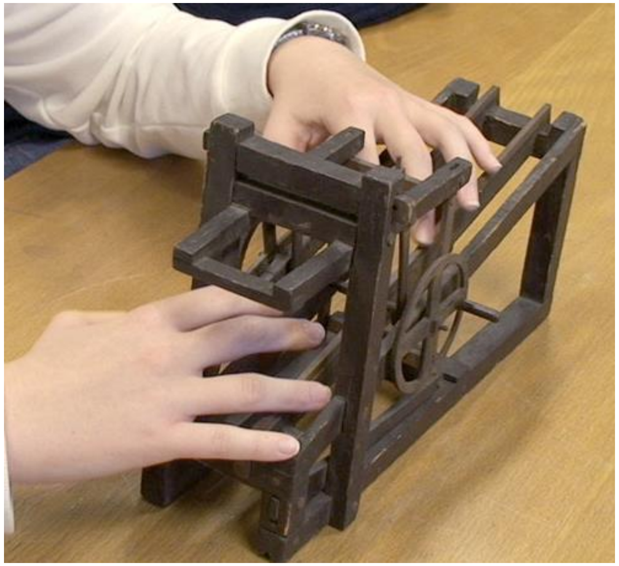
Contact, position, velocity, acceleration, applied force, pressure (squeeze, press), type of grasp, temperature, … pain, …





# Was is Haptics ?

- The sensibility of the individual to the world adjacent to his body by use of his body (Gibson)
- People can rapidly and accurately identify three-dimensional objects by touch
- The sense of touch is natural for humans to feel surface roughness, object softness, lightness or heaviness, etc
- Loss of the sense of touch is a catastrophic deficit that can impair skilled actions such as holding objects or using tools and walking



Gibson, J.J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin.

# Haptic perception

- Tactile / Cutaneous:
  - temperature, pressure, vibration, slip, pain
  - Sensation arising from stimulus to the skin

- Proprioception / kinesthesia:
  - Limb position/location, motion, force
  - End organs located in muscles, tendons, and joints
  - Stimulated by body movement

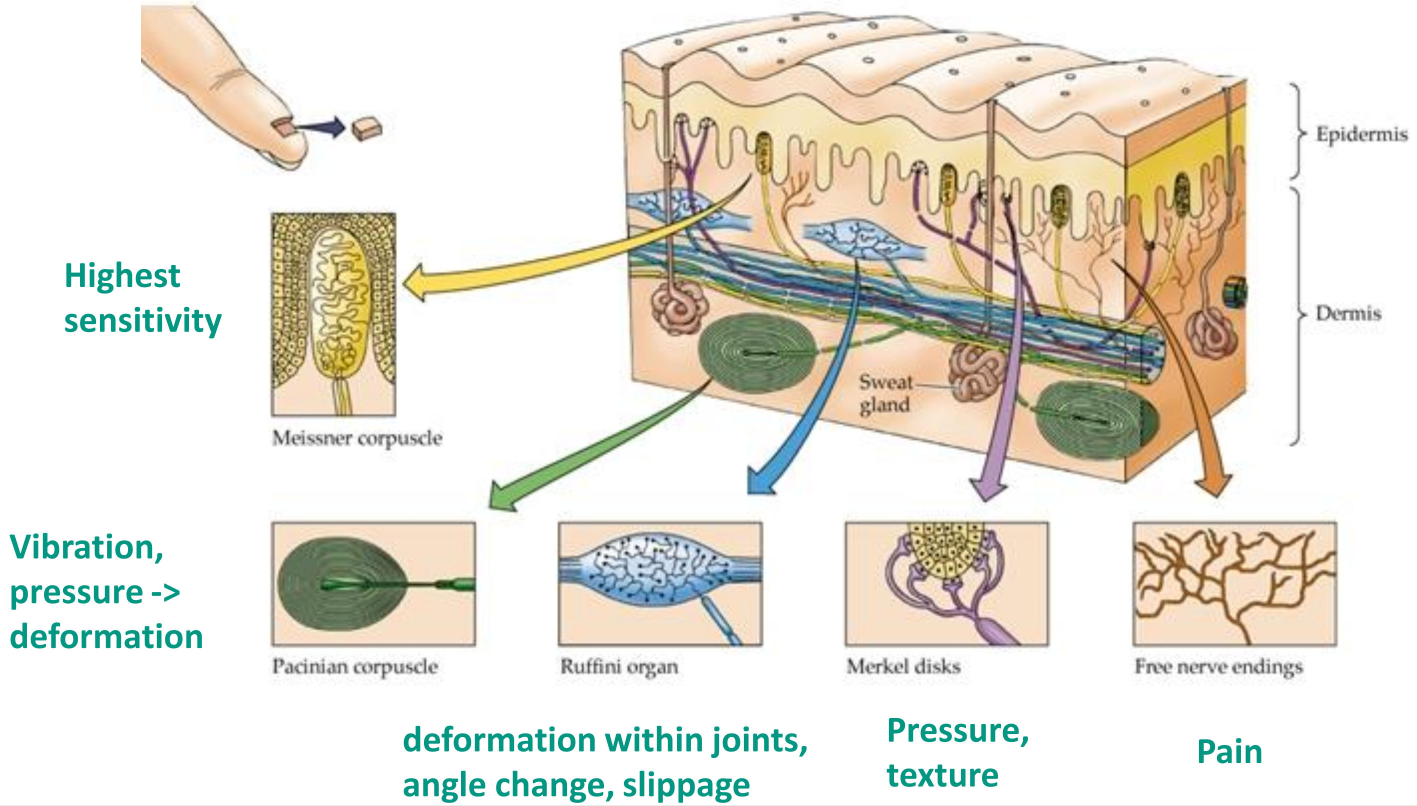
**Haptics = Tactile + Proprioception**

# Haptic perception

- The process of recognizing objects by touch
- It involves a combination of somatosensory perception of patterns on the skin surface (e.g., edges, curvature, and texture) and proprioception of hand position and conformation
- The somatosensory system is a complex sensory system. It is made up of a number of different receptors, including thermoreceptors, nociceptors, mechanoreceptors and chemoreceptors.
- It also comprises essential processing centres, or sensory modalities, such as proprioception, touch, temperature, and nociception. The sensory receptors cover the skin and epithelia, skeletal muscles, bones and joints, internal organs, and the cardiovascular system.

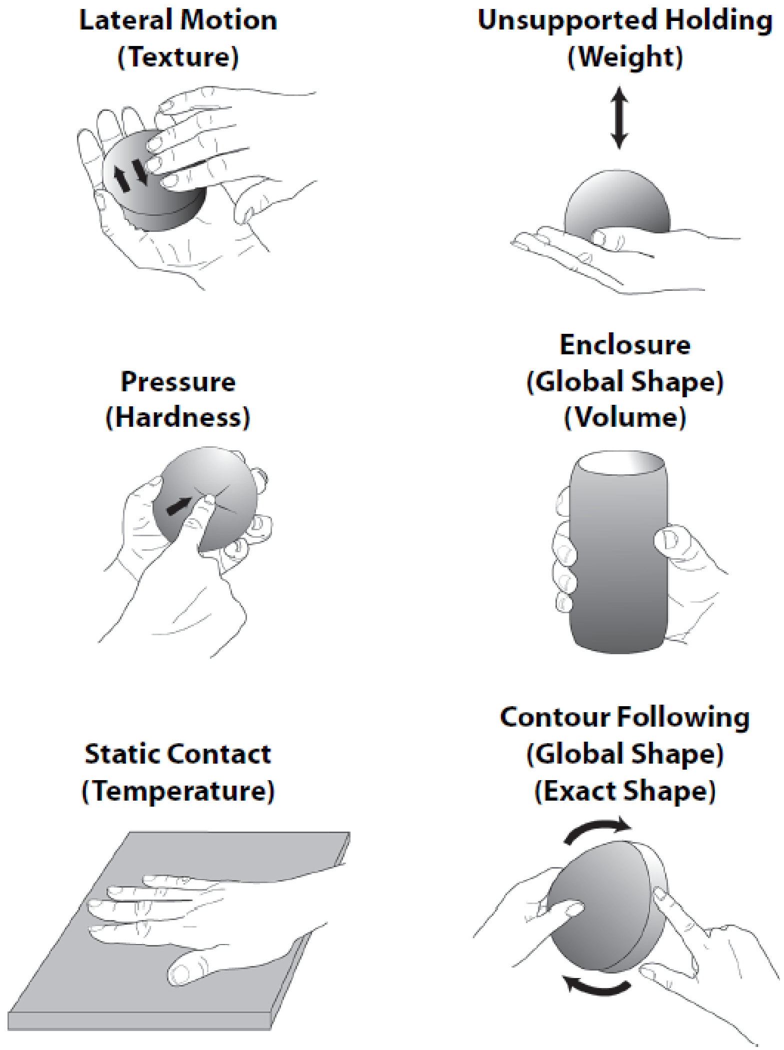
# Human skin

- A mechanoreceptor is a sensory receptor that responds to mechanical pressure or distortion. There are four main types in glabrous skin: Pacinian corpuscles, Meissner's corpuscles, , Merkel's discs, and Ruffini endings



# Haptic exploration can only be active

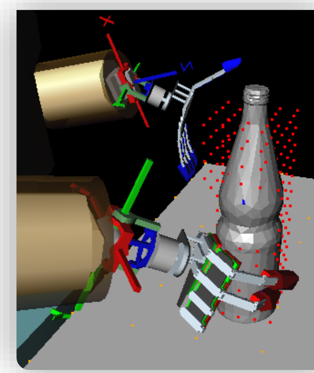
- Six manual “exploratory procedures” and their associated object properties (in parentheses)




From “Hand Movements: A Window Into Haptic Object Recognition,” by S. J. Lederman and R. L . Klatzky, 1987, *Cognitive Psychology*, *19*, p. 346. Copyright 1987 by Elsevier

# State of the art

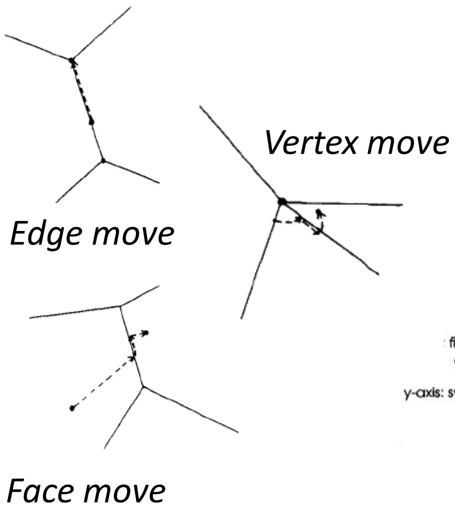
- Haptic exploration
  - Potential field based  [Bierbaum e al., 2009]
  - Haptic object recognition [Allen et al., 1989]
  - Active contact exploration [Roberts et al., 1990]
  - Active exploration and recognition of convex objects [Caselli et al., 1996]
  - Repetitive grasping [Takamuku, 2008]
  - Haptic exploration with slippage [Okamura et al., 1997]
  - Recognition of internal states [Chitta et al., 2010]
  - Multimodal  scene exploration [Bohg, 2010]

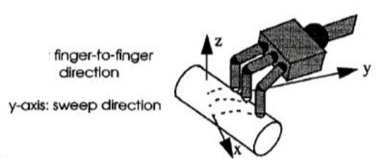


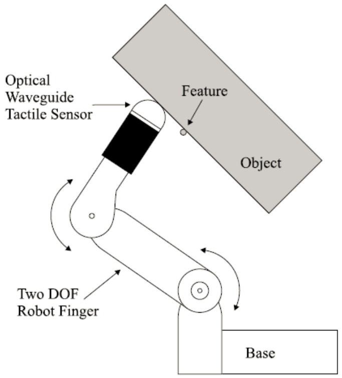
## State of the art

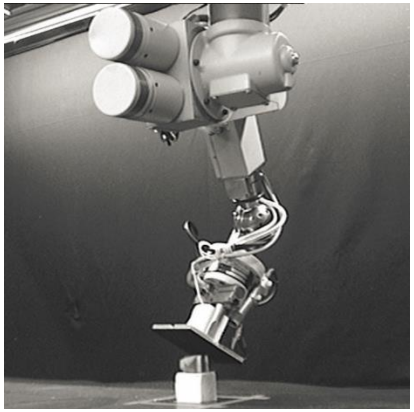
# Tactile acquisition of local object features

- Strategies for punctiform end-effectors [Roberts 1990, Caselli 1996]
- Contour following algorithms [Chen 1995, Charlebois 1997]
- Identification of local object features [Okamura 1999, 2003]
- Haptic aspect graph [Kinoshita 1992, Schopfer 2007]



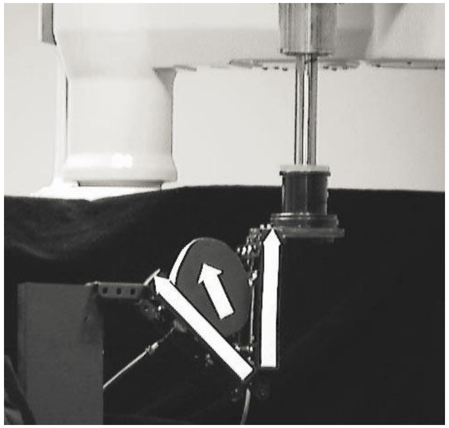



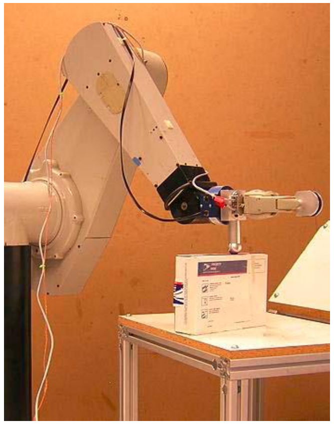



## Related Work

# Recognition and pose estimation of objects

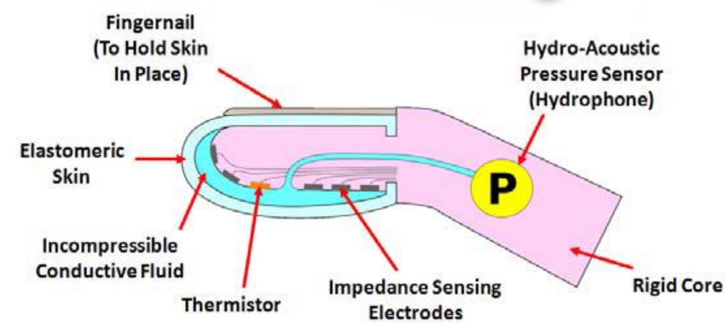
- Contour and motion of 2D objects [Moll & Erdmann 2003]
- Pose estimation using polygonal models [Petrovskaja 2006]
- Learning global object features by enclosing [Takamuku 2008]

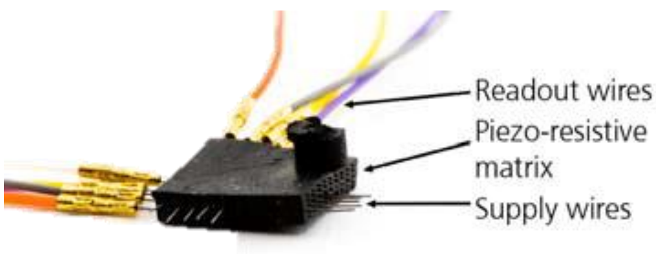


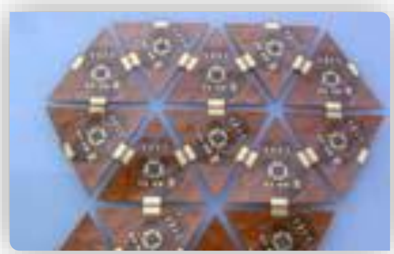


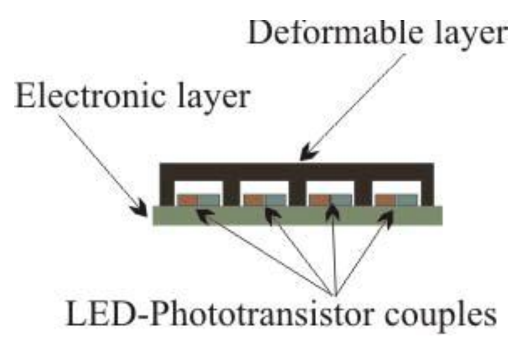
# Tactile sensors

- Resistiv
  - FSR
  - Weiss Robotics
- BioTAC [Wettels, 2007]
- Piezo-Resistiv [Strohmayr, 2009]
- Optical [Pirozzi, 2009]
- Capazitiv



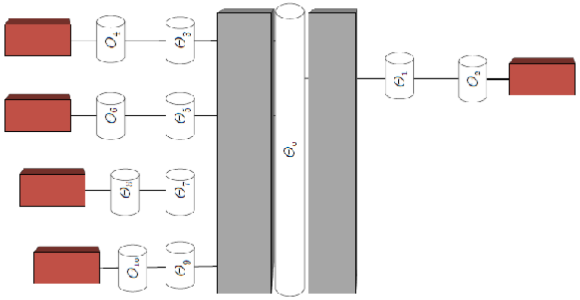



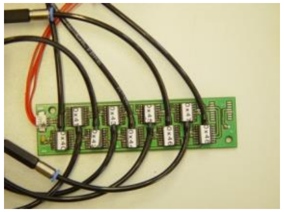



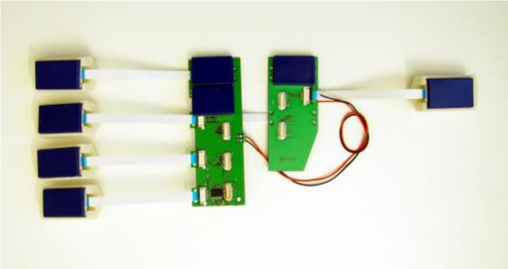


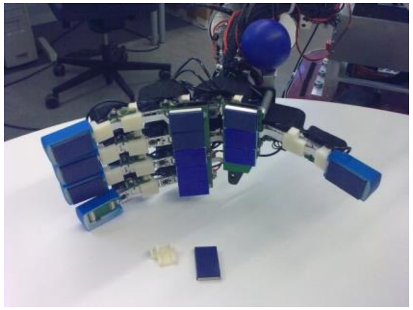
# The ARMAR-IIIb hand

- Five finger hand
- Carbon and aluminium structure
- 8 independent Degrees of Freedom
  - 2 DoF for index, middle and thumb
  - 1 DoF in the palm
  - Coupled pinkie and ring finger
- Pneumatically actuated fluid actors
- Sensors
  - Absolute joint angles
  - Pressure sensors → joint torques
  - Tactile sensors





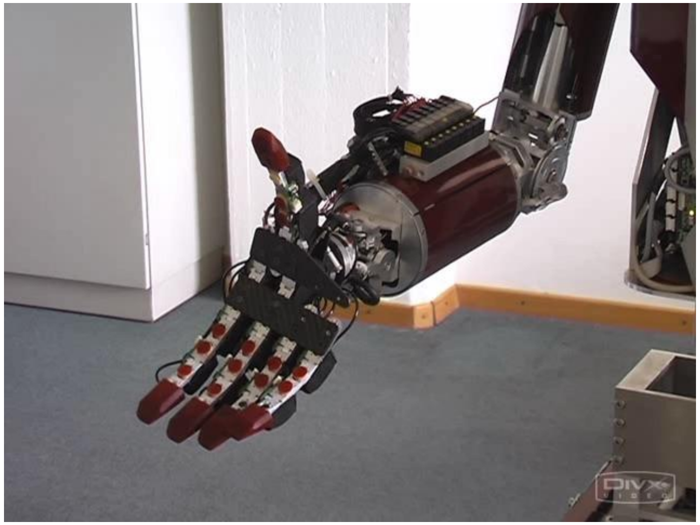


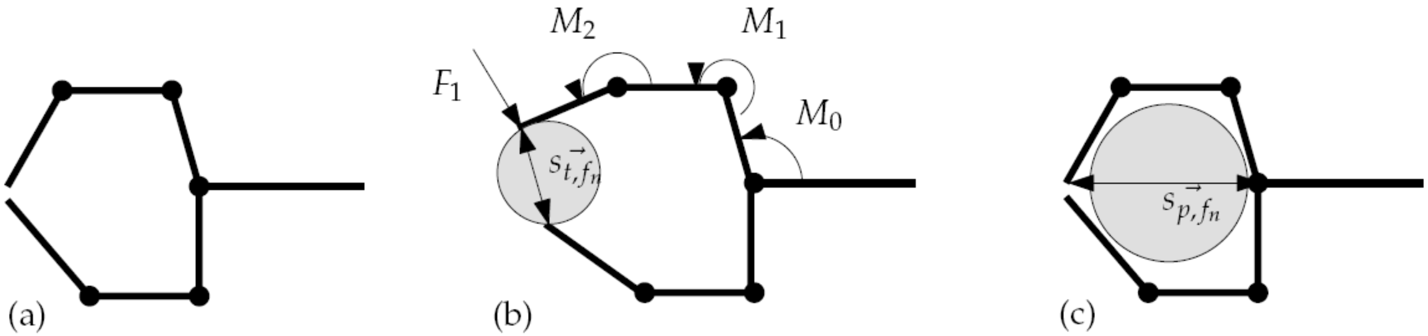


# Hand: Available skills

- Direct Kinematics
- Inverse Kinematics
- Position/force control

- Detection of contact and "objectness"
- Assessment of object deformability





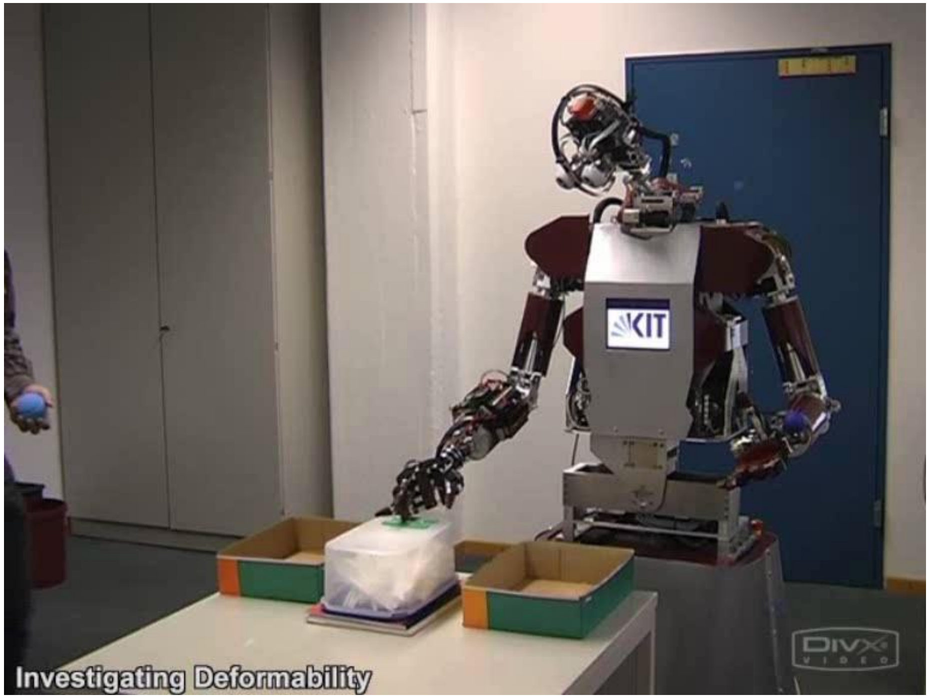

Failed grasp "no object"

Precision grasps: Distance between fingertips

Power grasps: Distance between fingertips and palm

# Visually-guided haptic exploration

Fusion of tactile, proprioceptive and visual sensor data with a five-fingered hand



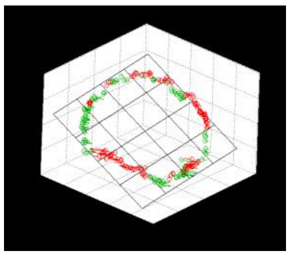
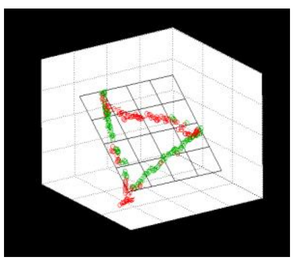
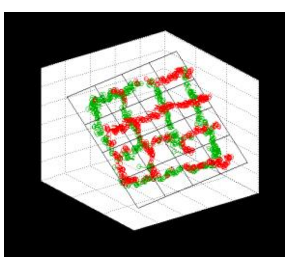
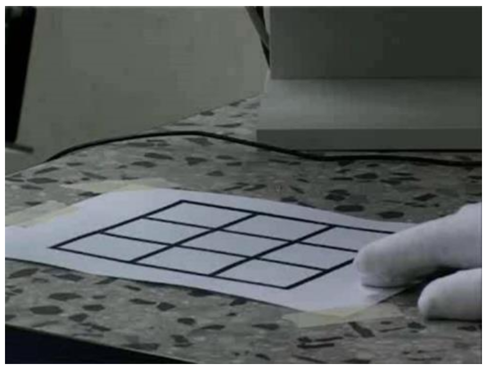

Verification of object size

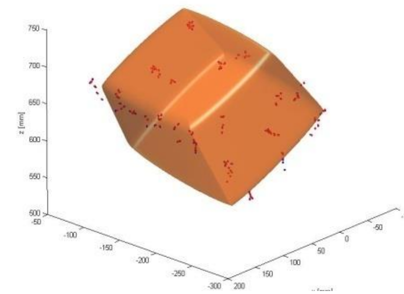
Verification of object deformability

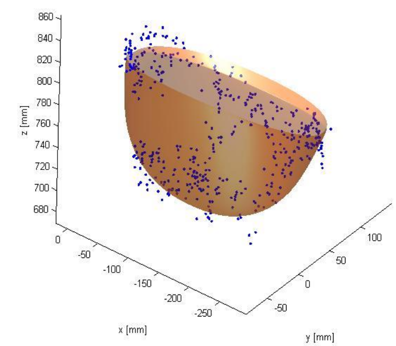
# Exploration with dataglove

**Contour Following (2D)**
of z-plane reference shapes, σ < 6.1mm

**Active Touch Exploration (3D)**
of 3D objects and superquadric fitting results.

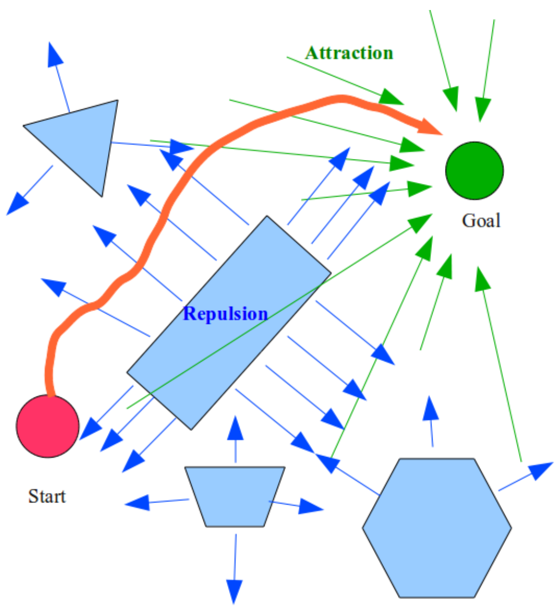


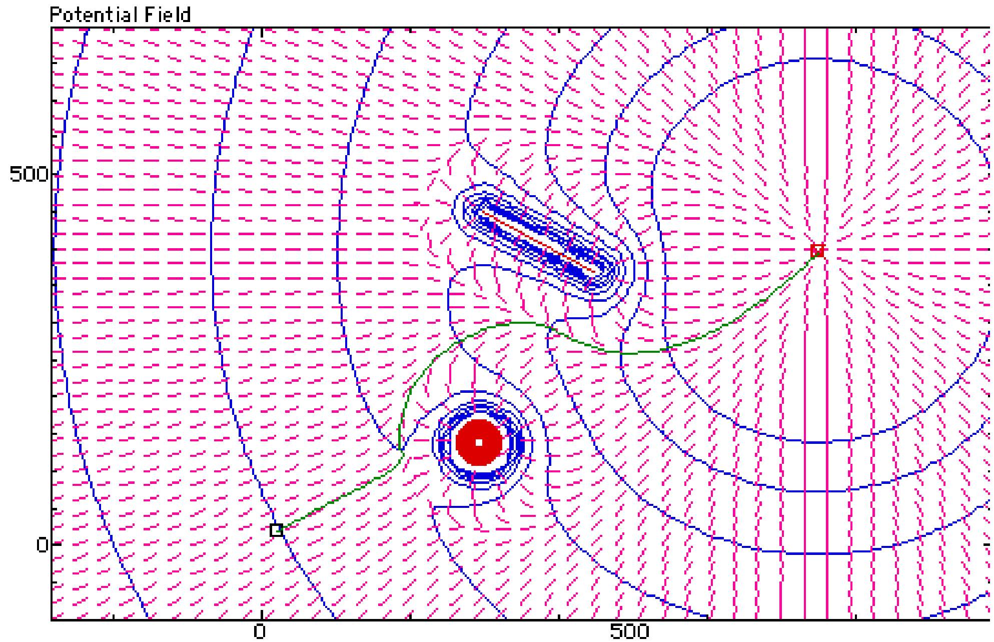



# Potential Field Based Exploration

Method originally developed for

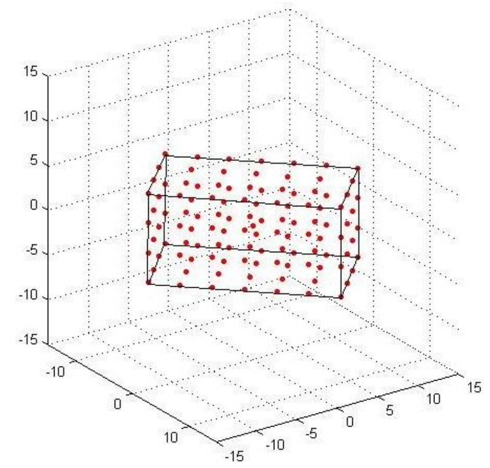
- Motion planning [Kathib 1986]
- Mobile robot SLAM, e.g. [Prestes 2002]



# Exploration using dynamic potential fields

- Field gradient direction in operational space
  - Unknown regions → attractive $\Phi_a < 0$
  - Known regions → repellent $\Phi_r > 0$
- Dynamic adaptation of potential field configuration from tactile response
- Superposition of individual potential sources

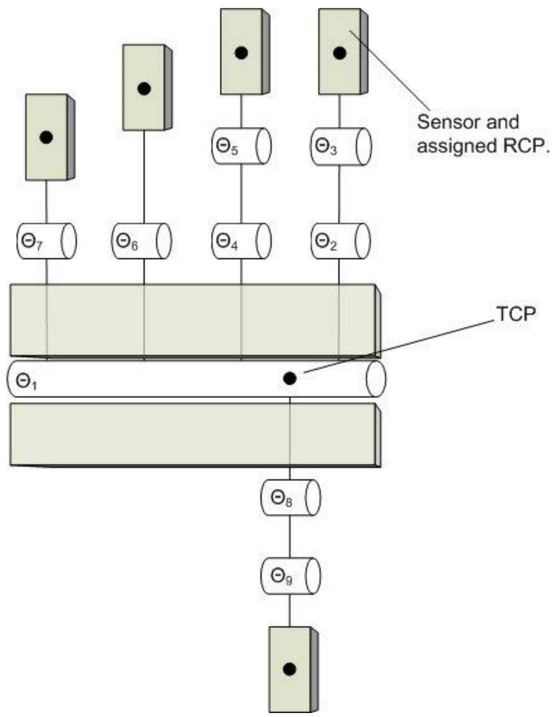



$$\Phi(x) = \sum_i \Phi_{r,i}(x) + \sum_j \Phi_{a,j}(x)$$

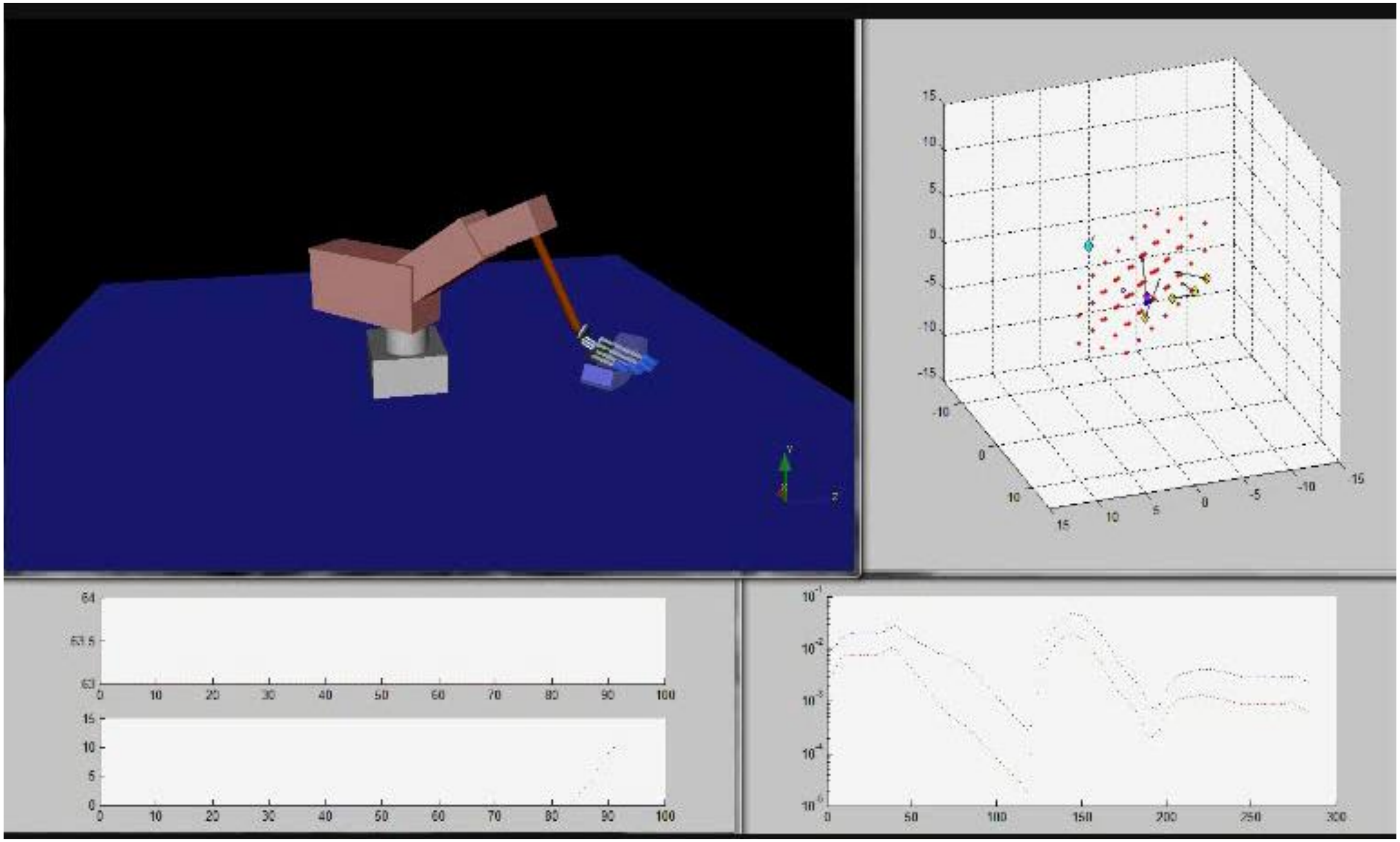
- Field initialization from pose and extension estimation of target object, e.g. by computer vision.

# Exploration using dynamic potential fields

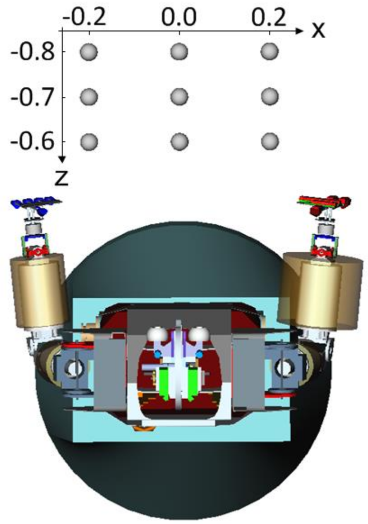
- Generation of trajectories for multi-point end-effectors (Robot Control Points, RCPs) using real-time gradient calculation [Khatib 1986]
- Harmonic potential functions to minimize number of local minima
- Reconfiguration strategy for resolving structural local minima of the hand
- Real-time inverse kinematics using Virtual Model Control (VMC) [Pratt 1996]
- Result: Oriented 3D point set with irregular density



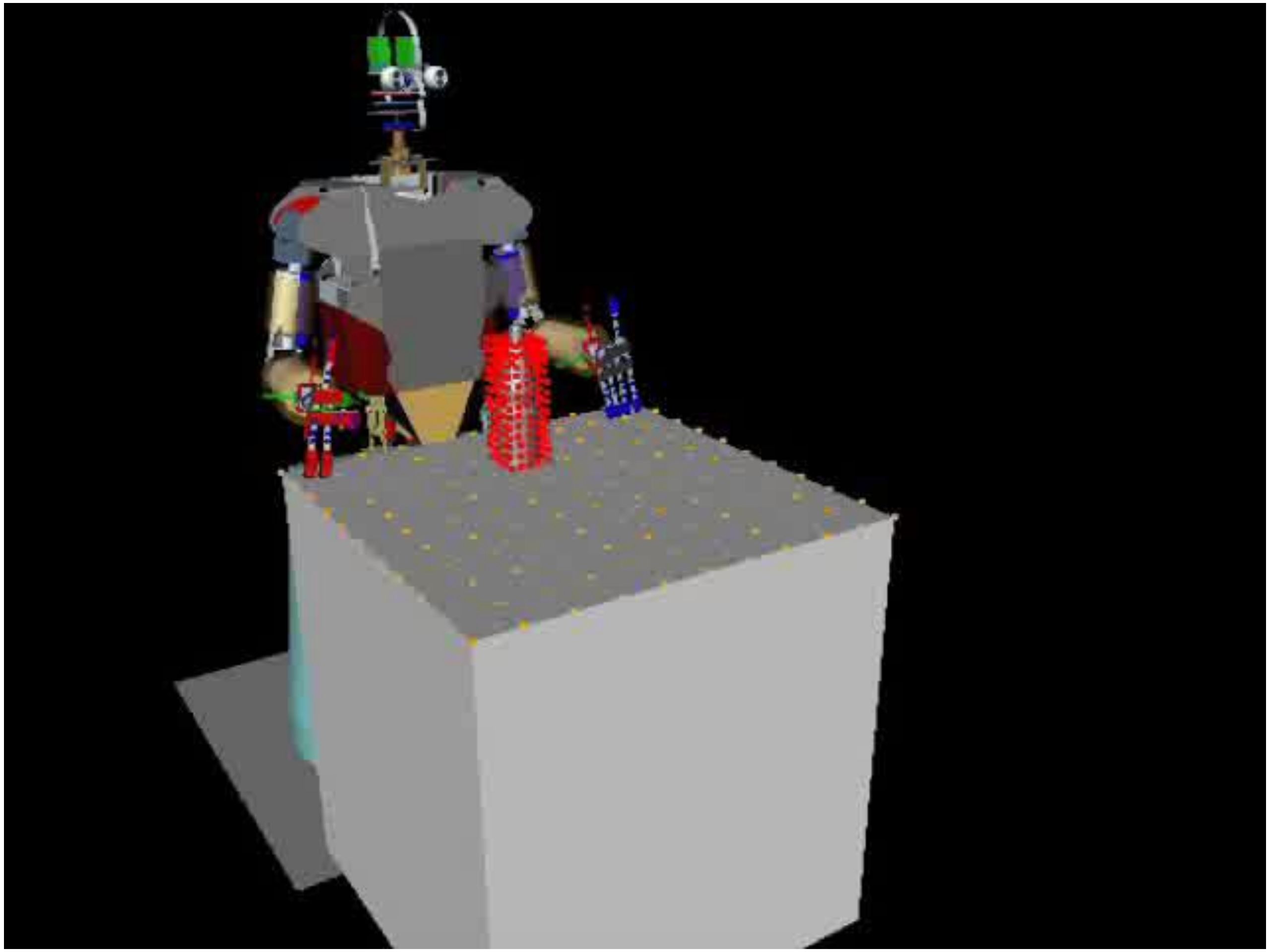
# Haptic Exploration using a Movemaster RM-501 Manipulator



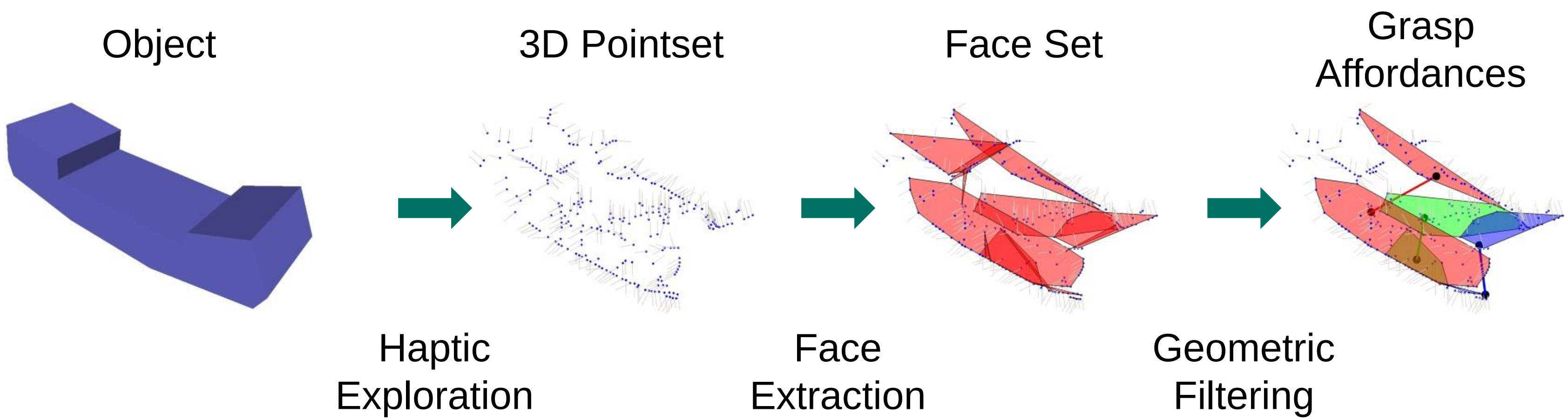
# Haptic Exploration (ARMAR-III)




Object Positioning



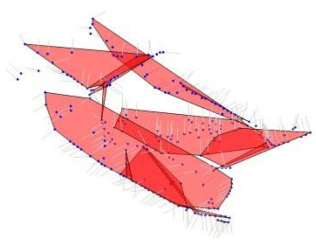
# Extracting Grasp Hypotheses



# Geometric Filtering and Grasp Computation

- Generate all face pairings and compute grasp affordance quality [Pertin-Troccaz 88].

- Grasp affordance quality $s(f_1, f_2)$ for each face pairing from 4-stage filter pipeline
  - Parallelism
  - Minimum face size
  - Mutual visibility (intersection of projection)
  - Face distance

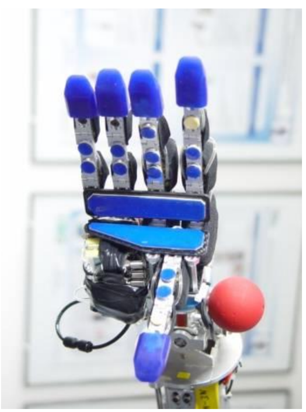


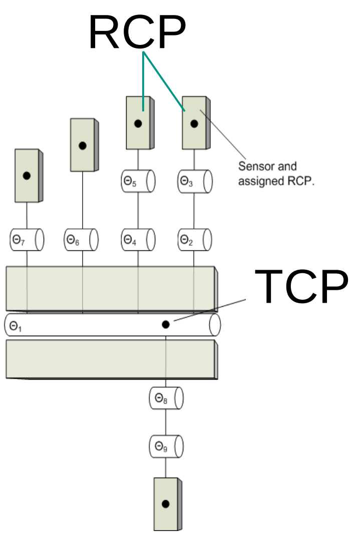
$$s(f_1, f_2) = \prod_{i=1}^{4} o_i(f1, f2)$$

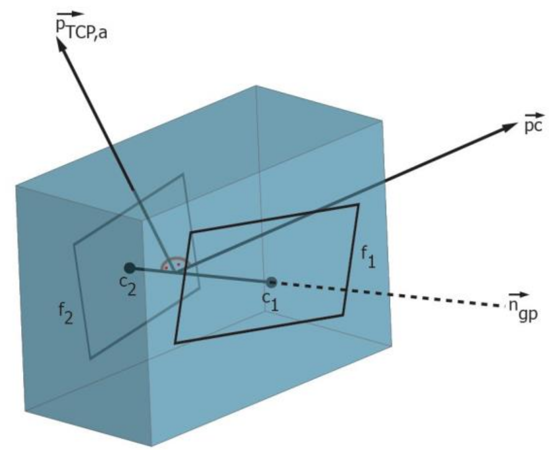
# Geometric Filtering and Grasp Computation

- Calculation of grasp parameters for highest quality grasps
  - Approach direction of TCP *(Tool Center Point)*
  - Hand orientation
  - Target configuration of RCPs *(Robot Control Points)*

$$s(f_1, f_2) = \prod_{i=1}^{4} o_i(f1, f2)$$

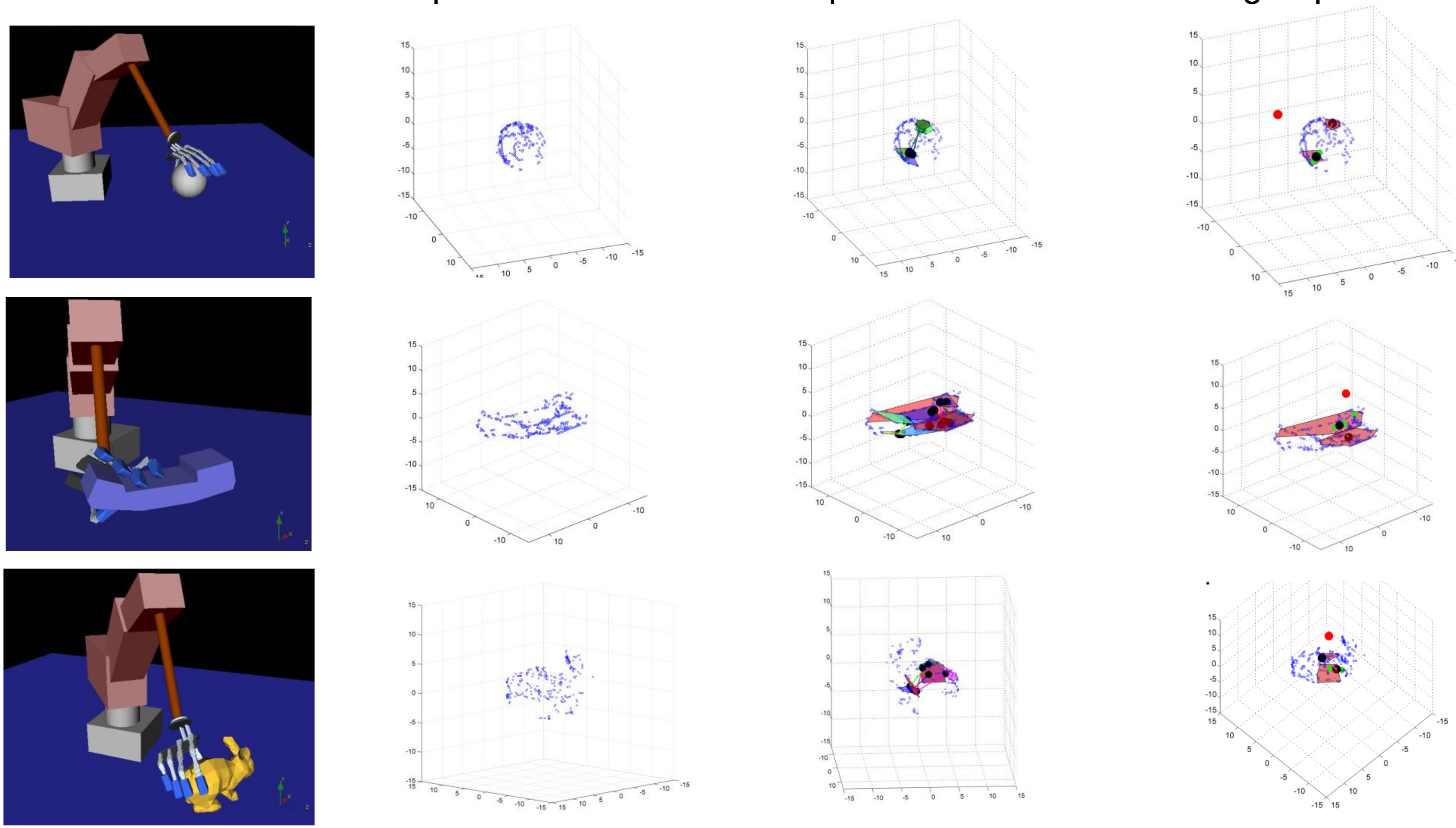




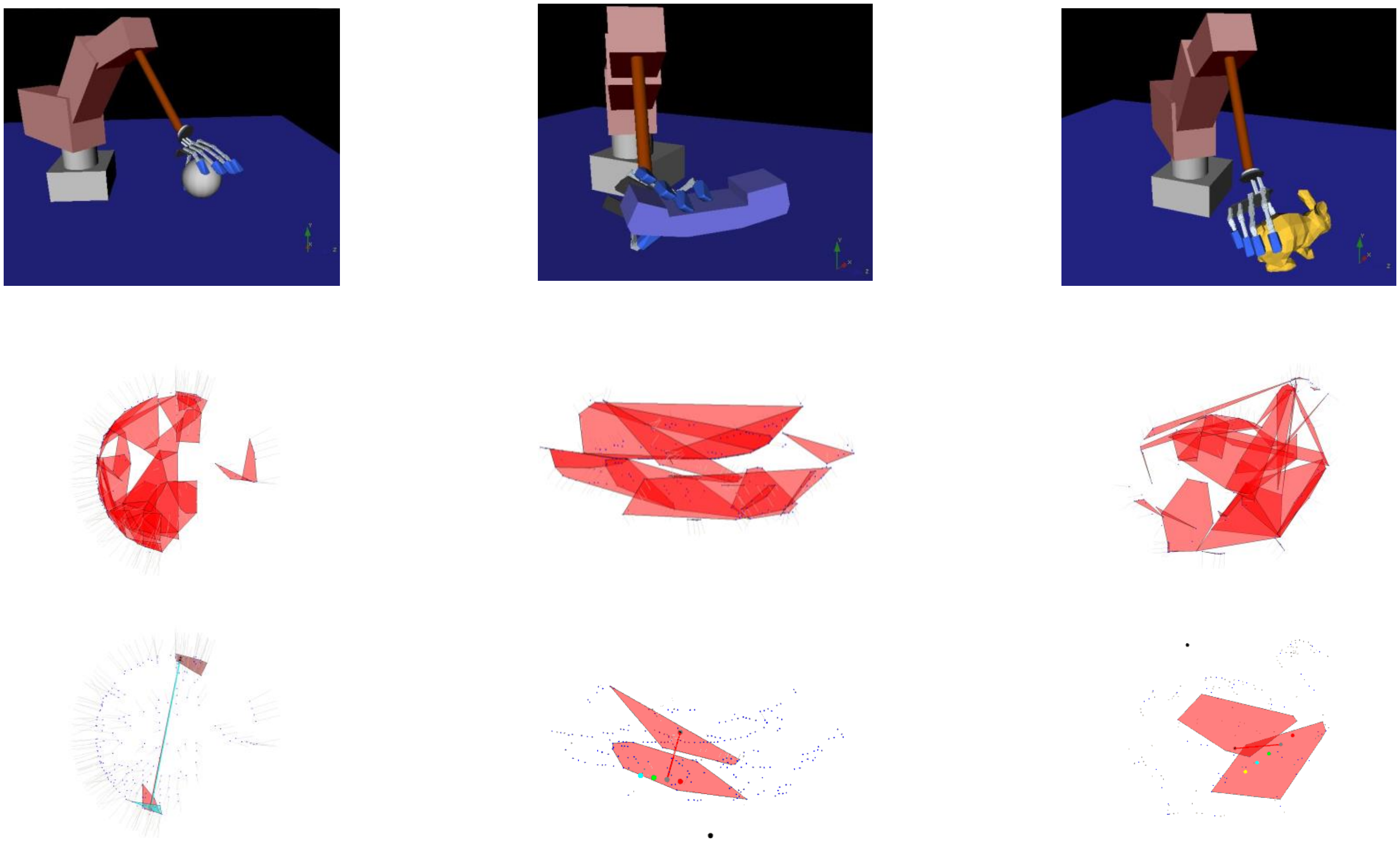



# Generation of grasp hypotheses

3D point clouds | Grasp candidate | Best grasp



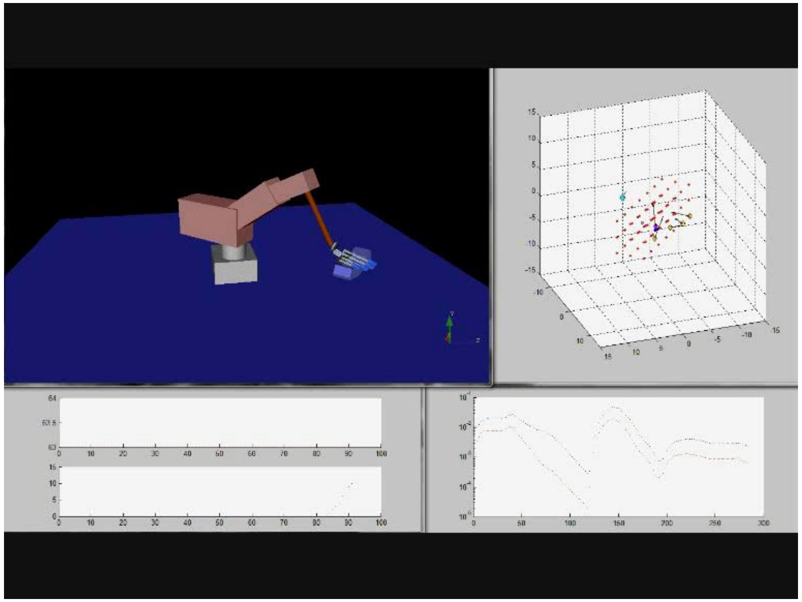
# Examples

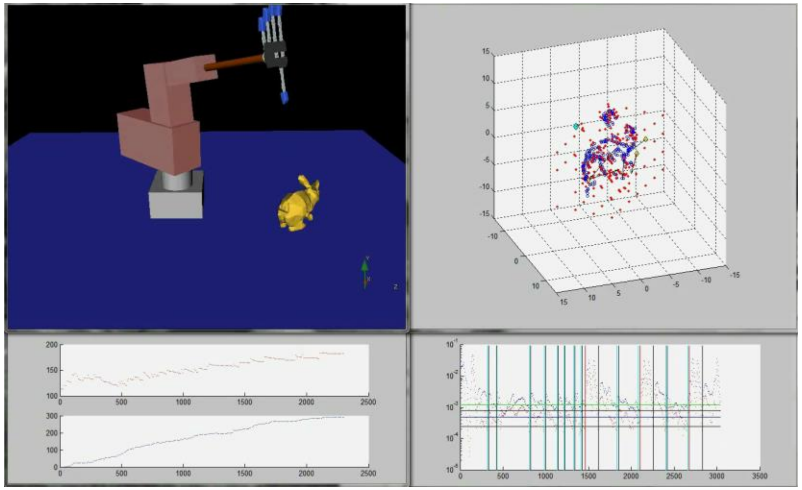


# Tactile Object Exploration

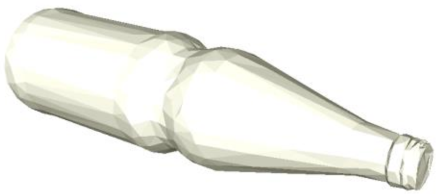
- Potential field approach to guide the robot hand along the object surface
- Oriented 3D point cloud from contact data
- Extract faces from 3D point cloud in a geometric feature filter pipeline
    - Parallelism
    - Minimum face size
    - Face distance
    - Mutual visibility

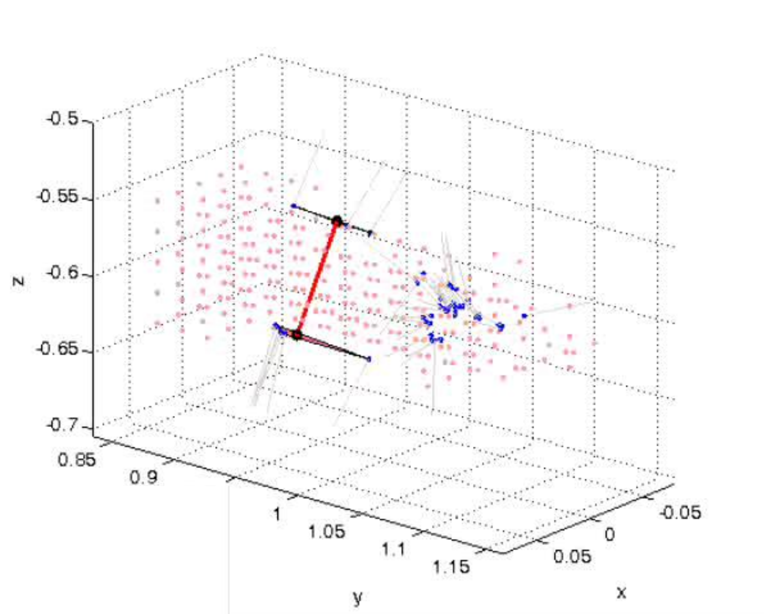
→ Association between objects and actions (grasps) → Symbolic grasps (grasp affordances)

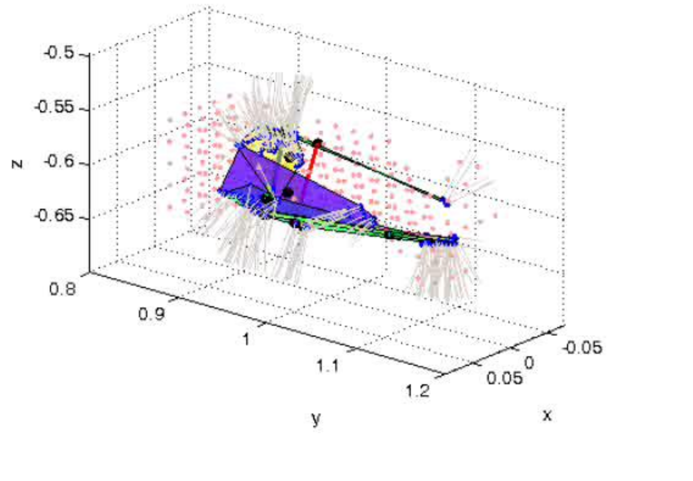




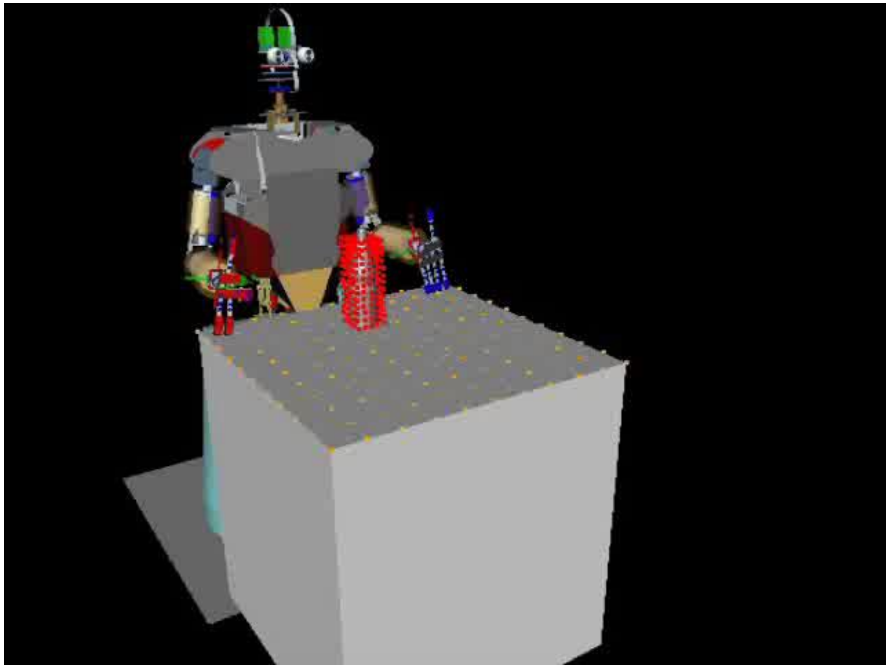
# Examples: Bottle

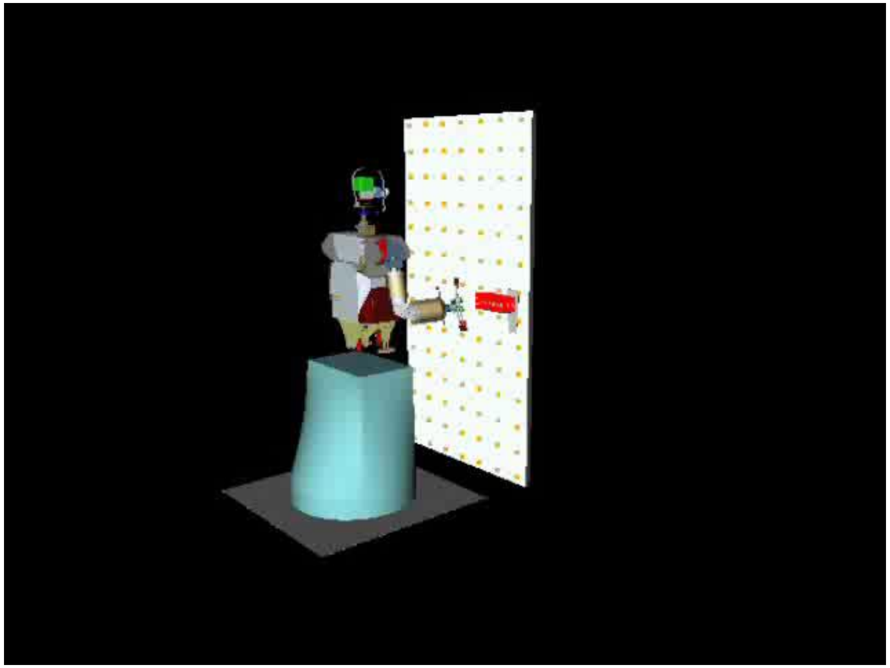




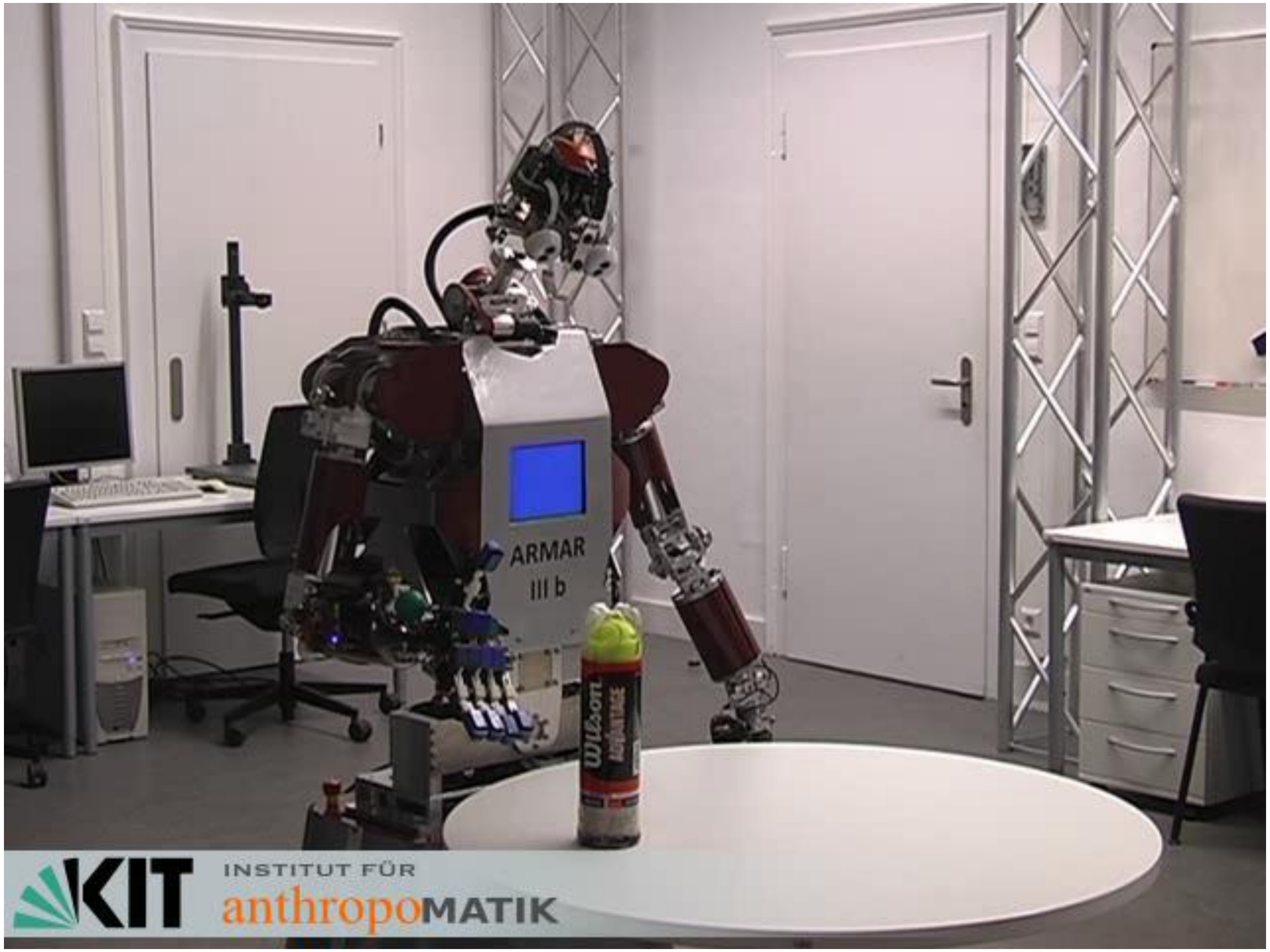
# Visually guided exploration on ARMAR





- Exploration in simulation
  - Physics extension for Open Inventor/VRML modeling of complex mechanical systems
  - Modeling of virtual sensors
  - Virtual Model Control (VMC) - based inverse kinematics

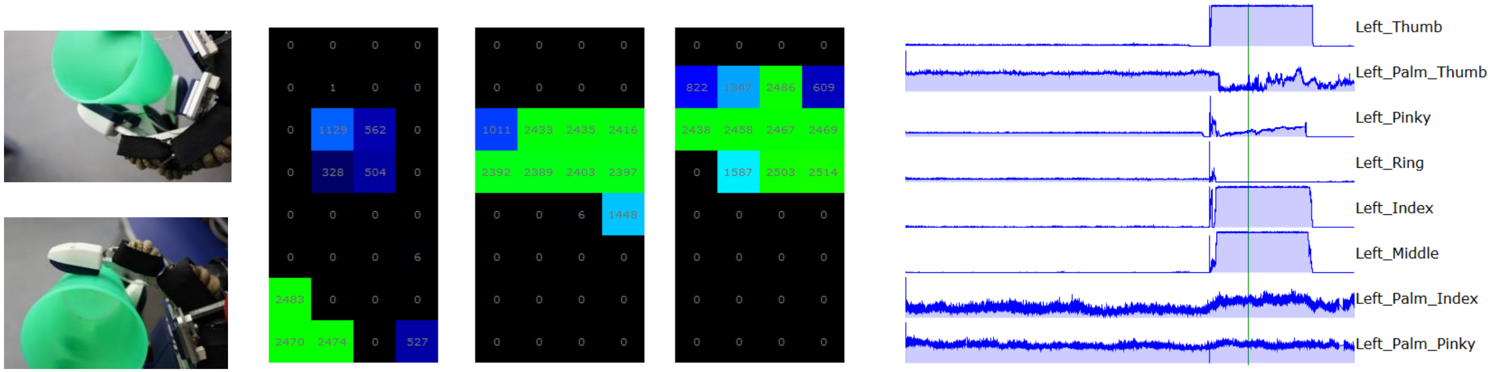
# Haptic Exploration using ARMAR-III



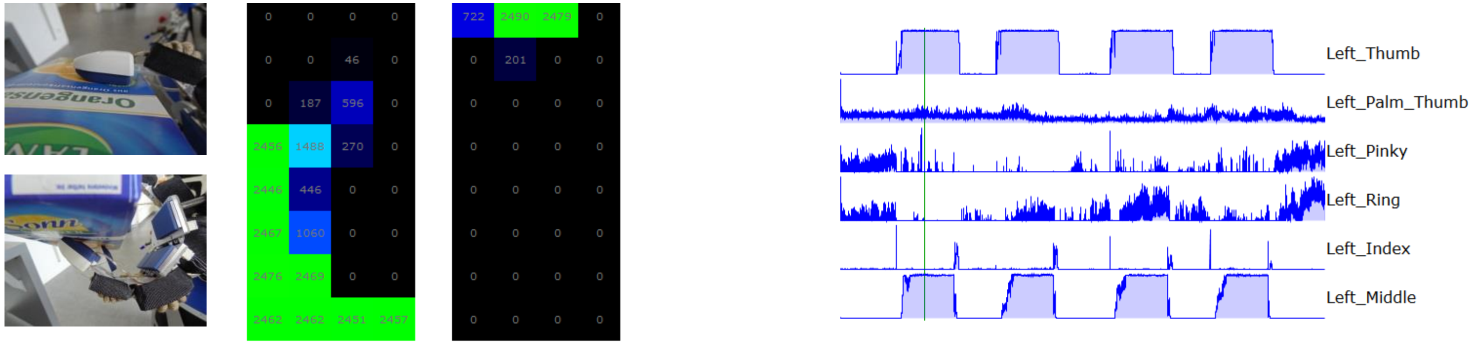
# New Approach for Haptic exploration

# Tactile-based contact detection

Contact with 3 fingers




Contact with 2 fingers



# Haptic representations based on information gain

- Haptic object exploration for extracting object features
  - Surfaces
  - Edges and corners
  - Grasp hypotheses

- Idea: Define the extraction of desired object features as information gain

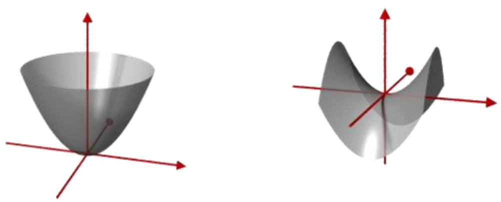
*Maximize information gain:* $\dfrac{\Delta\, Information}{\Delta t}$

# Haptic Exploration – Algorithm

- Requirements on exploration algorithm
  - Efficient object shape exploration
  - Avoid local minima
  - Stick to object surface

H: Heuristic
P: Sample Point on Local Object Model
C: Contact Point

- Algorithm:
  - Approximate local object shape as paraboloid
  - Information gain for a point P:

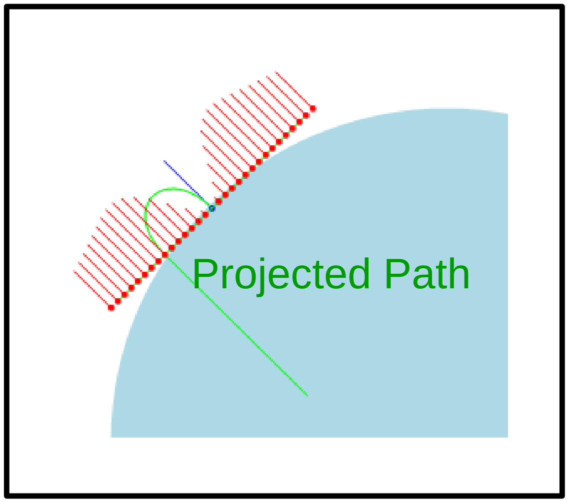


$$I(P) = \frac{\min_{C} H(P,C)}{len\big(Path(P)\big)} \qquad H(P,C) = 1 - e^{-(P-C)^2} \cdot \frac{P_N \cdot C_N + 1}{2}$$

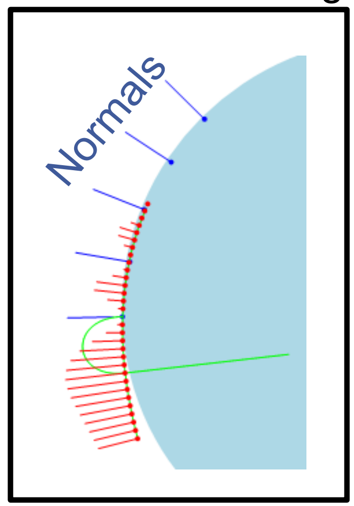
  - Select the point on the paraboloid with the highest information gain as the next exploration target

# Haptic Exploration – Results

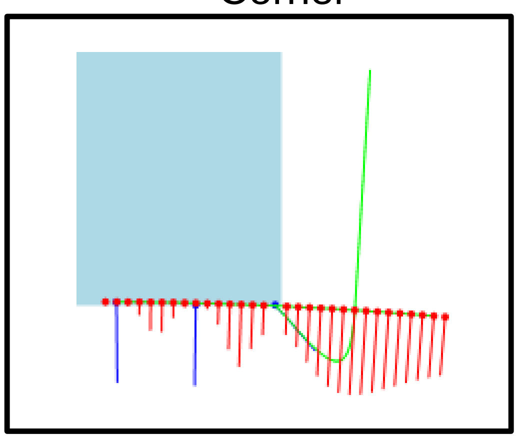
Initial contact



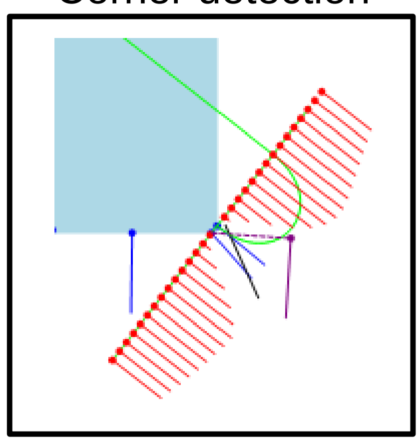

Surface following



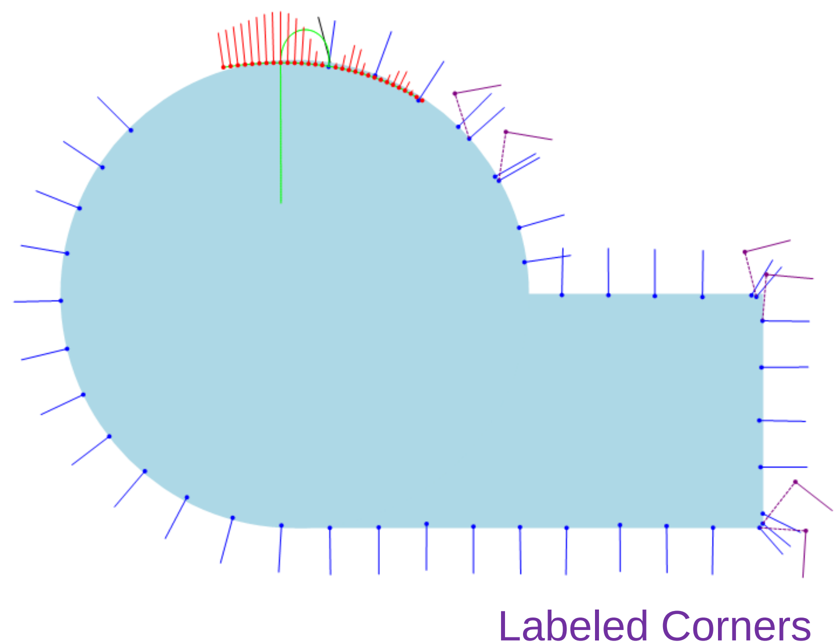

Corner



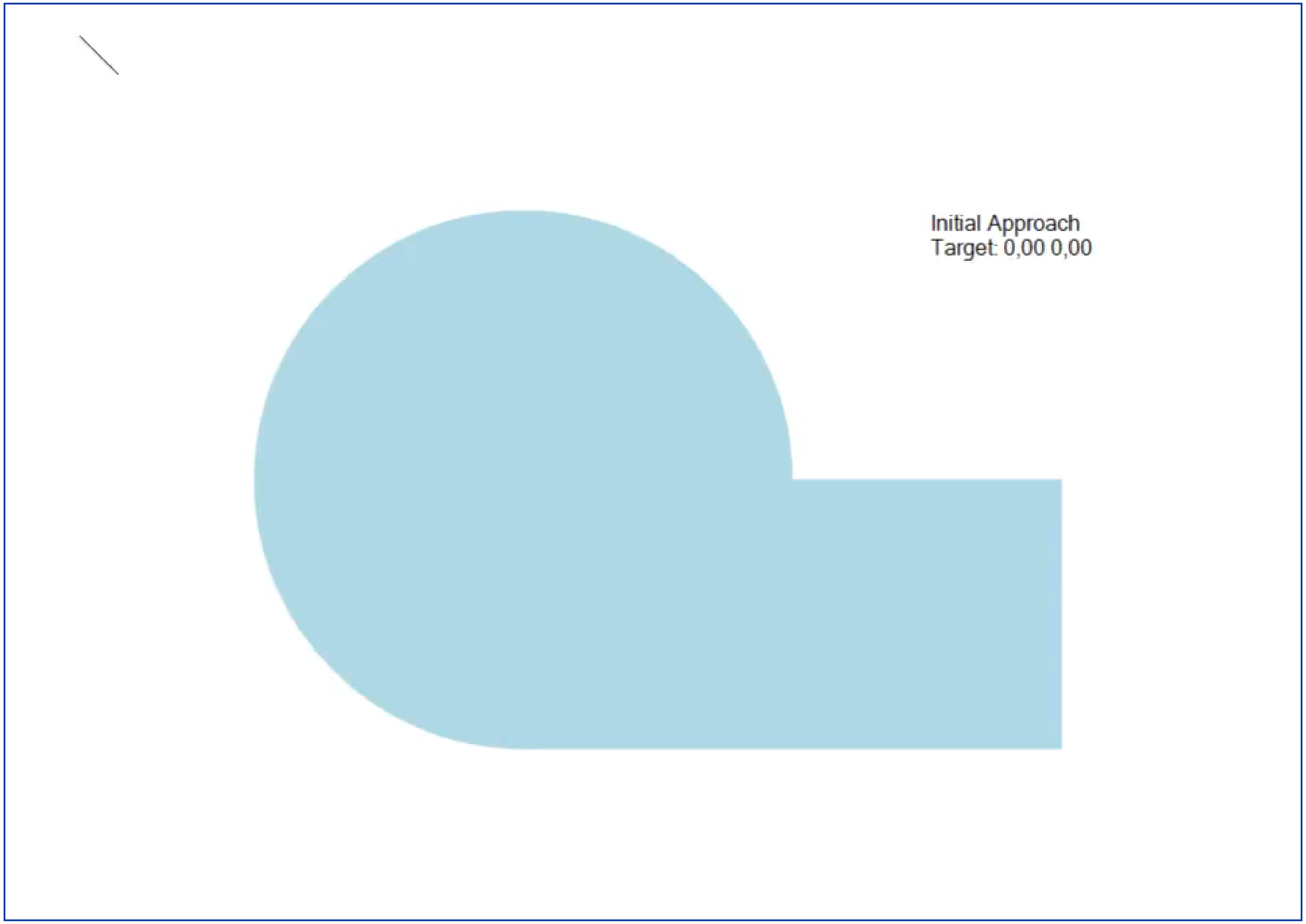
Corner detection



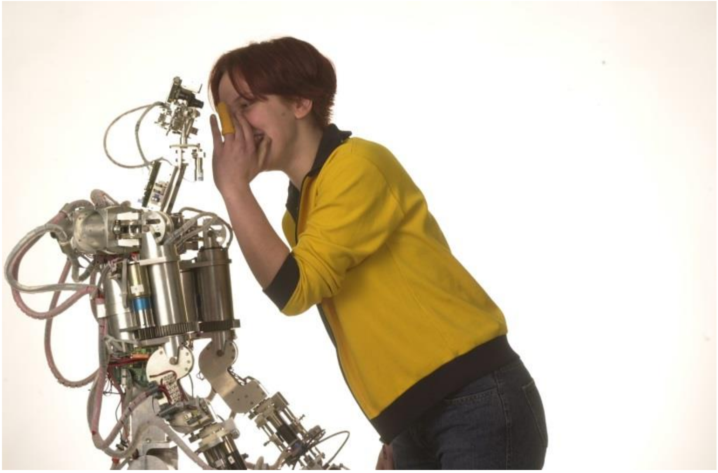
Heuristic:
Estimated information gain



# Haptic Exploration – Results



# Active Hearing



# Research questions

- Hearing capability with a pair of ears in humans and animals as well as with a pair of microphones for machines, is a rudimental function for perception and communication
- To make robots, either physical or virtual, be in symbiosis with people, they must be endowed with the ability to localize, separate and process sounds under noisy environments or from a mixture of sounds
- Research topics
  - Generic design of binaural sensors
  - Active binaural sound source localization
  - Voice detection
  - Binaural speaker recognition
  - Ego-noise cancellation

- **How to move the body to improve the quality of the perceived sound?**

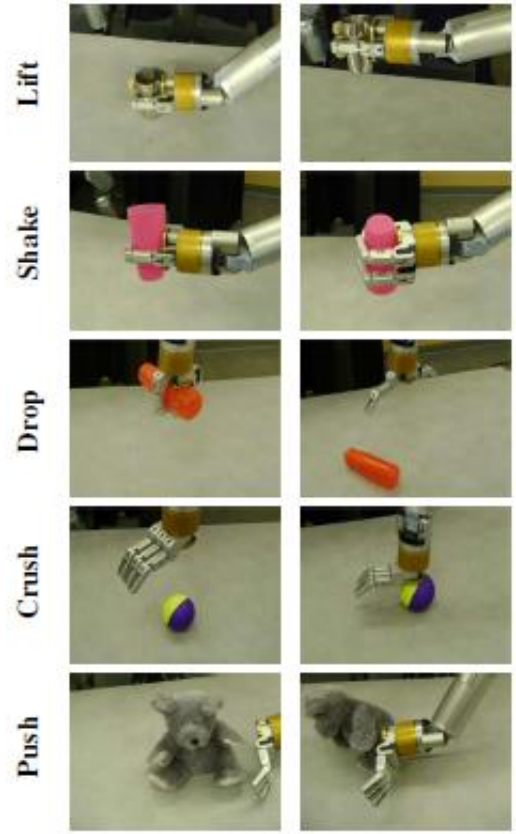
# Learning object categories using audio and video

- Interacts with objects to learn object categories
- The robot captures audio and video as it performs six different exploratory behaviors
- A separate object categorization is formed for each behavior and sensory modality combination
- The resulting 12 categorizations are unified into a single one

**Object Categorization in the Sink: Learning Behavior–Grounded Object Categories with Water**
Shane Griffith, Vladimir Sukhoy, Todd Wegter, and Alexander Stoytchev

# Learning object categories using audio and video

- The robot initially explores the objects by applying five exploratory behaviors (lift, shake, drop, crush and push) on them while recording the proprioceptive and auditory sensory feedback
- A graph-based recognition model is trained by extracting features from the estimated similarity relations, allowing the robot to recognize the category memberships of a novel object based on the object's similarity to the set of familiar objects




*Before* and *after* snapshots of the five behaviors used by the robot.

**Object Category Recognition by a Humanoid Robot Using Behavior-Grounded Relational Learning**
**Jivko Sinapov and Alexander Stoytchev**