Karlsruher Institut für Technologie

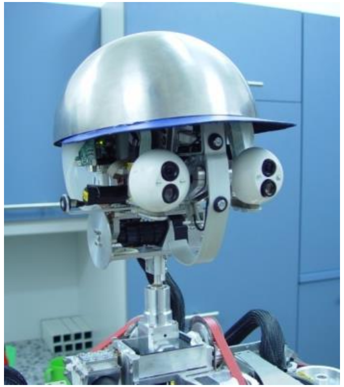
Hochperformante Humanoide Technologien

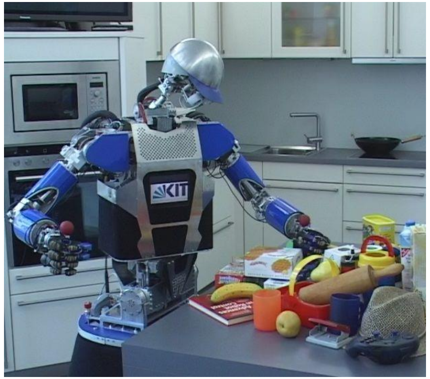
# Robotics III: Sensors and Perception in Robotics

## Chapter 07: Active Perception

Tamim Asfour

http://www.humanoids.kit.edu



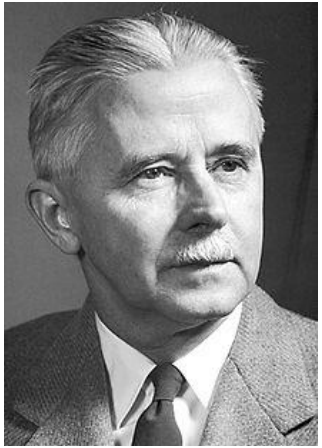


# Outline of This Lecture

- Introduction to Active Perception
  - Definition of Active Perception
  - The Human Eye
  - Human Visual Attention
- Active Visual Perception
  - Gaze Control & Stabilization
  - Object Discovery and Segmentation
- Active Haptic Perception
  - Tactile Exploration
  - Visuo-Haptic Grasping
- Active Hearing

# Vision as dynamic process

*"**Vision** itself is a **dynamic process**. There is little in the world that stands still, at least not as imaged in our retinas, for our eyes are always moving. The visual system is almost exclusively organized to **detect change and motion**."*



Keffer Hartline - Nobel Lecture, December 12, 1967
https://en.wikipedia.org/wiki/Haldan_Keffer_Hartline

# Active Perception: Definitions

Karlsruher Institut für Technologie

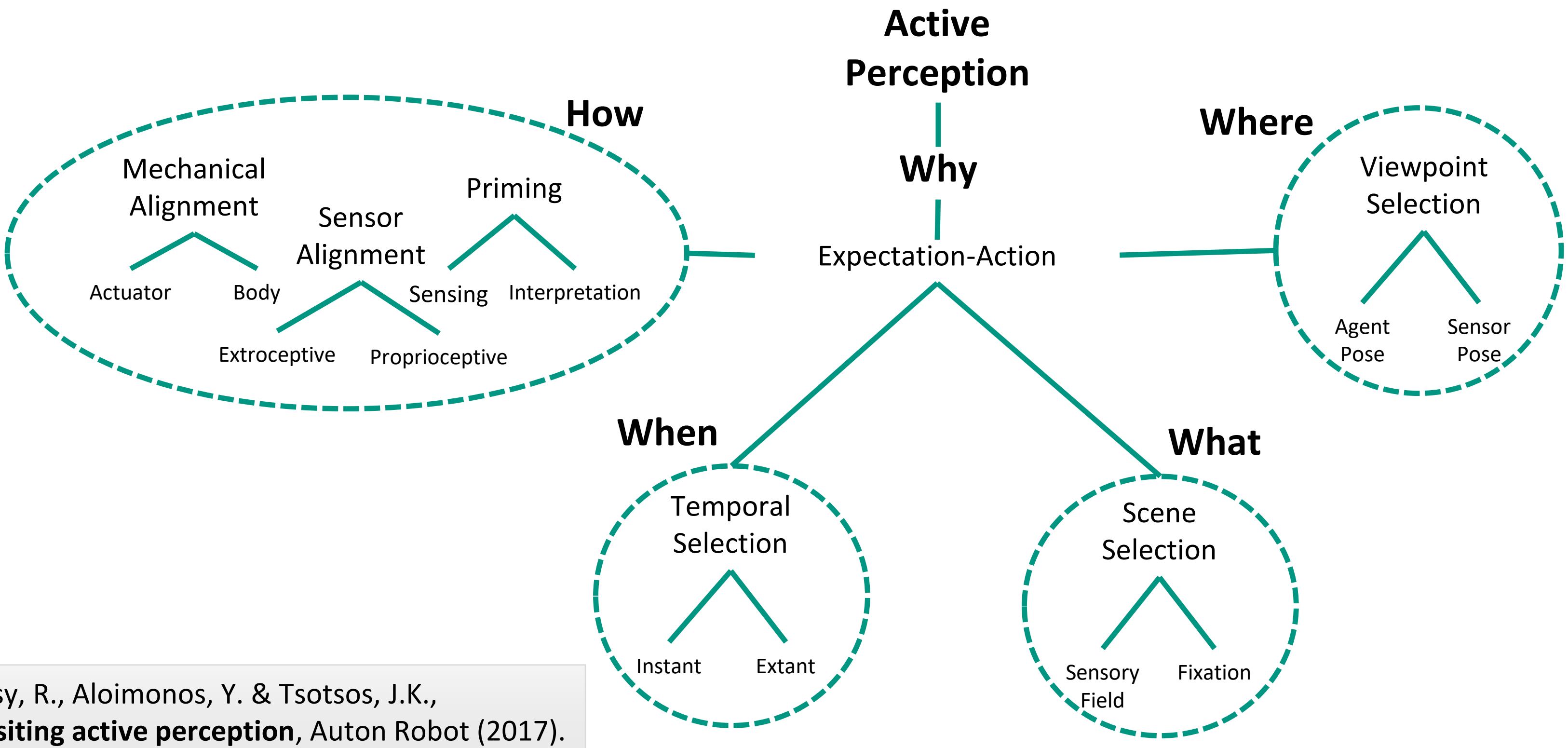
- "Active sensing is the problem of intelligent control strategies applied to the data acquisition process which will **depend on the current state** of data interpretation including recognition," Bajcsy (1988)

- "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)

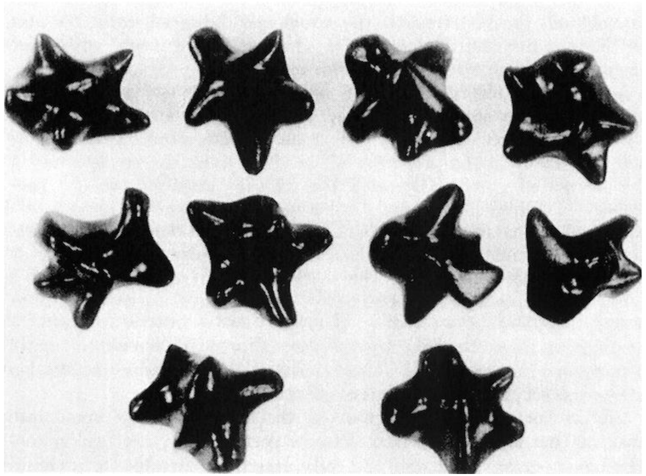

- Bajcsy, R., **Active perception,** Proceedings of the IEEE (1988), 76(8), 966–1005
- Aloimonos, J. et al., **Active vision**, Int Journal of Computer Vision (1988), 1(4), 333–356.
- Bajcsy, R., Aloimonos, Y. & Tsotsos, J.K., **Revisiting active perception**, Auton Robot (2017).

# 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.

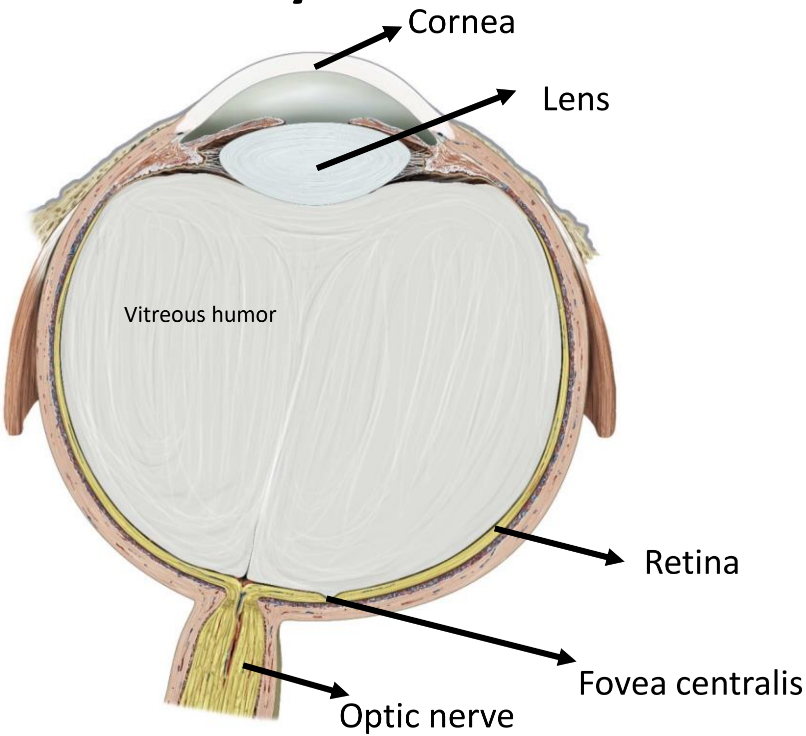
# Classical CV vs. Active Vision vs. Active Perception

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

# Outline of This Lecture

- Introduction to Active Perception
  - Definition of Active Perception
  - The Human Eye
  - Human Visual Attention
- Active Visual Perception
  - Gaze Control & Stabilization
  - Object Discovery and Segmentation
- Active Haptic Perception
  - Tactile Exploration
  - Visuo-Haptic Grasping
- Active Hearing

# The Human Eye




**Prometheus : LernAtlas der Anatomie: Kopf und Neuroanatomie**
Michael Schünke ; Erik Schulte ; Udo Schumacher; Markus Voll ; Karl Wesker

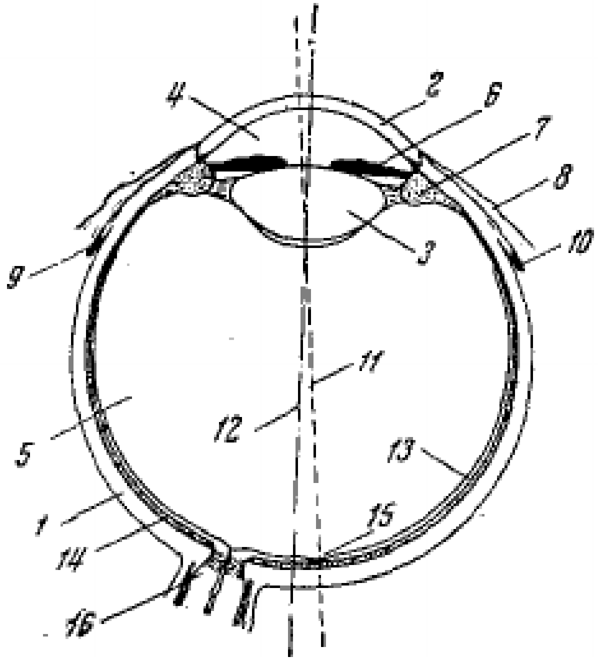
- Light enters the eye through the cornea, through the pupil and then through the lens
- Photons fall then on the photoreceptor cells of the retina
- Photoreceptor cells, i.e. cones and rods, are light sensitive and convert light into electrical signals (signal transduction)
- Electrical signals are transmitted to the brain by the optic nerve

# Anatomy of the Human Eye

Karlsruher Institut für Technologie

- „Sensor“ reacting to light and pressure;
- Perception of light, color and depth
- Captures visual stimuli which are then carried to the brain for visual perception
- Principle of a pinhole camera

Visual axis (11)
Optical axis (12)
Retina (13)
Fovea centralis (15)
Optic nerve (16)




Alfred L. Yarbus, **Eye Movements and Vision,** (1967)

# Peripheral and Foveal Vision

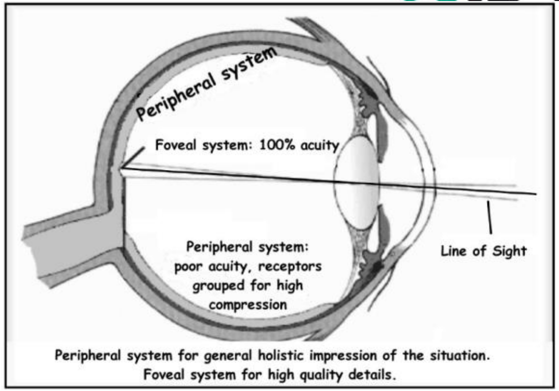
## Foveal vision (Fovea centralis)

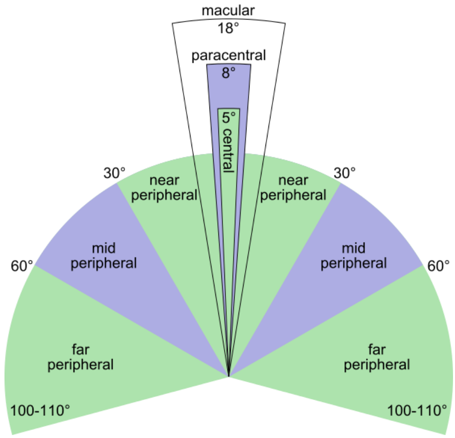
- Among mammals the area with highest visual acuity
- Located near the visual axis
- Color impression  (high density of cones)

## Peripheral vision

- Low resolution
- Poor visual acuity
- Gray-scale vision (high density of rods)

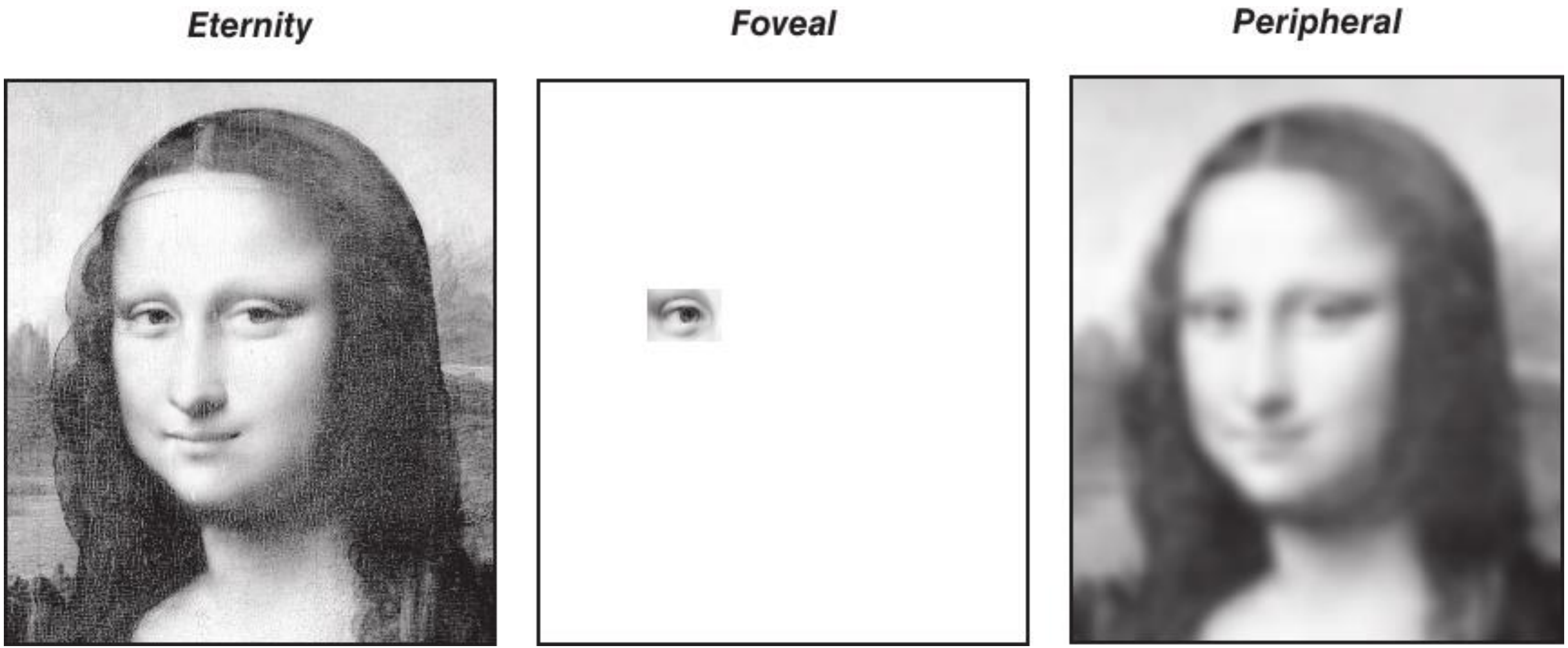
- Allows to monitor the scene
- Determine salient regions to shift the gaze




https://en.wikipedia.org/wiki/Peripheral_vision

# Example: Foveal vs. Peripheral Vision

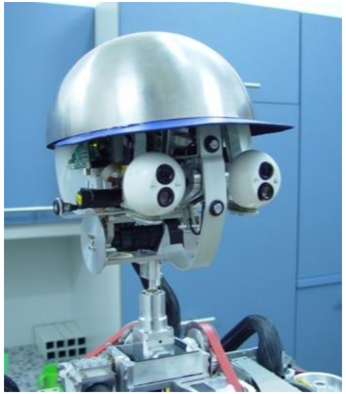



- Eye movements are required

J. Lauwereyns, ***Brain and the gaze: On the active boundaries of vision***, 2012

# Peripheral Vision and Foveal Vision in Robotics

- Active stereo head with two cameras per eye
- Foveated vision realized using different camera lenses
- 4x PointGrey Dragon Fly 2 @ 640 x 480



foveal view

peripheral view

# Outline of This Lecture

- Introduction to Active Perception
  - Definition of Active Perception
  - The Human Eye
  - Human Visual Attention
- Active Visual Perception
  - Gaze Control & Stabilization
  - Object Discovery and Segmentation
- Active Haptic Perception
  - Tactile Exploration
  - Visuo-Haptic Grasping
- Active Hearing

# Visual Attention

Human visual attention can be independent of the eye movement:

- overt attention:  visual attention **associated with** eye movement
- covert attention: attention shift **not associated with** of eye movement

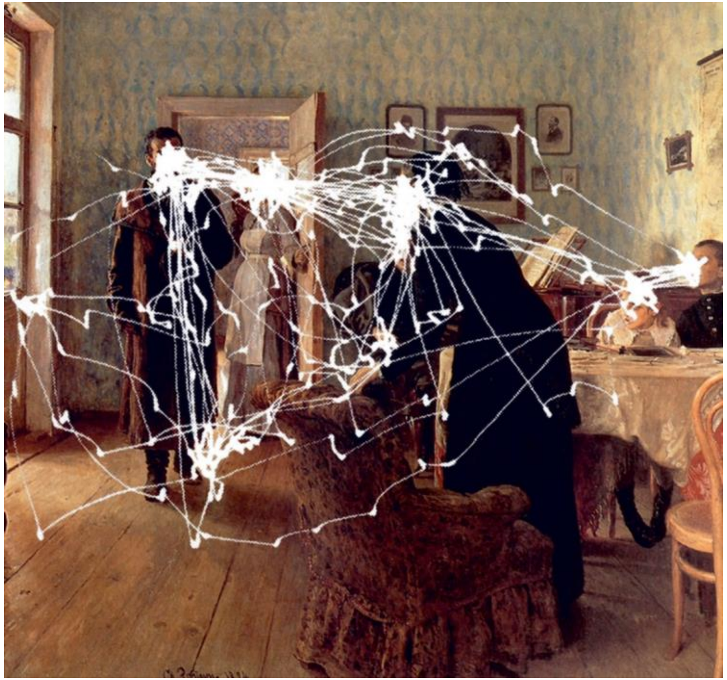
Attention can be bottom-up or top-down

- bottom-up: **stimulus driven** attention. Pre-attention stage, automatic involuntary attention behavior with high speed
- top-down: **task driven** attention. It is related to knowledge experience and goals

- Fixation of regions is the result of three mechanisms: bottom-up, top-down, or both mechanisms simultaneously

# Eye Movements during Perception of Complex Objects

- Eye movements of a subject while perceiving Ilya Repin's picture “An Unexpected Visitor”
- Patterns of eye movements are similar (but not identical) for
  - different people
  - a single individual over the course of multiple days
- How viewing behavior changes over extended periods of time?
  - 3 minutes recordings



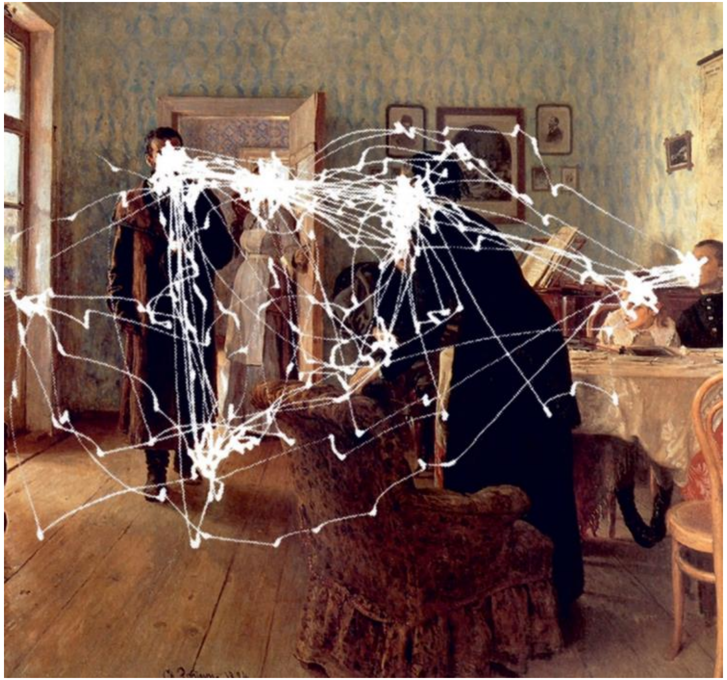
Alfred L. Yarbus, **Eye Movements and Vision,** (1967)

Overlay by Sasha Archibald http://www.cabinetmagazine.org/issues/30/archibald.php

# Eye Movements during Perception of Complex Objects

- 3 minutes recordings
- One participant viewing the same image seven times, each with a different set of Task.
- Different pattern of eye movements depending on the task

- **Task 1:** Free examination of the picture



Overlay by Sasha Archibald http://www.cabinetmagazine.org/issues/30/archibald.php

# Eye Movements during Perception of Complex Objects

- 3 minutes recordings
- One participant viewing the same image seven times, each with a different set of Task
- Different pattern of eye movements depending on the task

- **Task 3:** give the ages of the people

Overlay by Sasha Archibald http://www.cabinetmagazine.org/issues/30/archibald.php

# Eye Movements during Perception of Complex Objects

- 3 minutes recordings
- One participant viewing the same image seven times, each with a different set of Task
- Different pattern of eye movements depending on the task

- **Task 5:** remember the clothes worn by the people

Overlay by Sasha Archibald http://www.cabinetmagazine.org/issues/30/archibald.php

# ACTIVE VISUAL PERCEPTION

# Active vs. Static Vision

## Static Vision

- Passive sensor
- Limited field-of-view
- Occlusion, ambiguity
- *Classic* computer vision

## Active Vision

- Change the camera viewpoint and
- Update camera parameters
- Makes ill-posed problems tractable

- Robotic vision: the camera can be actively controlled

# Outline of This Lecture

- Introduction to Active Perception
  - Definition of Active Perception
  - The Human Eye
  - Human Visual Attention
- Active Visual Perception
  - Gaze Control & Stabilization
  - Object Discovery and Segmentation
- Active Haptic Perception
  - Tactile Exploration
  - Visuo-Haptic Grasping
- Active Hearing

# GAZE STABILIZATION

# Motivation

- Stable camera images are key for visual processing
  - Relevant objects are not visible
  - High perceptual blur

# Gaze Control

## Gaze = Head + Eye movements

**Type of Eye movements**

- Maintaining / Stabilizing gaze
  - Vestibulo-Ocular Reflex (VOR)
  - Optokinectic nystagmus / Optokinectic reflex (OKR)
  - Fixation
- Switching Gaze
  - VOR cancellation
  - Saccade movements
    - Rapid eye movements (ballistic)
    - Can be generated on command or involuntarily (nystagmus)
  - Smooth pursuit
  - Vergence

# Example: Smooth pursuit

EMBalance project: http://www.embalance.eu


https://www.youtube.com/watch?v=w9l-IZbX1NI

# Smooth pursuit on ARMAR-III

# Gaze Stabilization

**Goal**: compute compensatory eye and head movements, i.e., $\dot{\boldsymbol{q}}_{\boldsymbol{eye}}\ \dot{\boldsymbol{q}}_{\boldsymbol{head}}$, to stabilize the current view

## Sensory cues

- Visual feedback
- Vestibular information
- Proprioception

## Gaze Stabilization Methods

- Vestibulo-ocular reflex (VOR)
- Vestibulocollic reflex
- Optokinetic reflex (OKR)
- Inverse kinematics (IK)

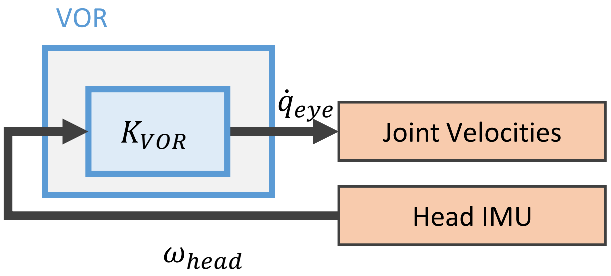
# Vestibulo-ocular Reflex (VOR)

**Idea:** The VOR stabilizes the gaze by producing **eye** movements **counteracting** head movements

- Inertial Measurement Unit (IMU) located in the head to mimic the human vestibular system
- Reflex is triggered by a measurement of the head rotational velocity $\omega_{head} = [\omega_{yaw} \quad \omega_{pitch}]^T$

**Control output:** $\dot{q}_{eye} = -k_{vor} \cdot [\omega_{yaw} \quad \omega_{pitch}]^T$

Gain $k_{vor}$ should be close to 1 to fully compensate head rotations



# Optokinetic Reflex (OKR)

**Idea:** The OKR stabilizes the gaze by producing **eye** movements **cancelling the retinal slip** (the perceived optical flow)

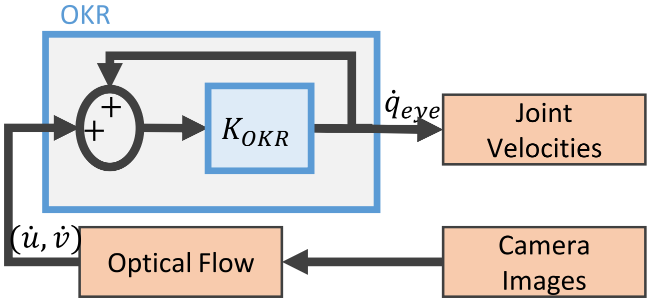
- Optical flow $(\dot{u}, \dot{v})$ in the image is computed using feature tracking
- Translate the optical flow to rad/s using the angle of view of the cameras

**Control output:** $\dot{q}_{eye} = k_{okr} \cdot [\dot{u} \quad \dot{v}]^T$

Gain $k_{OKR}$ should be close to 1

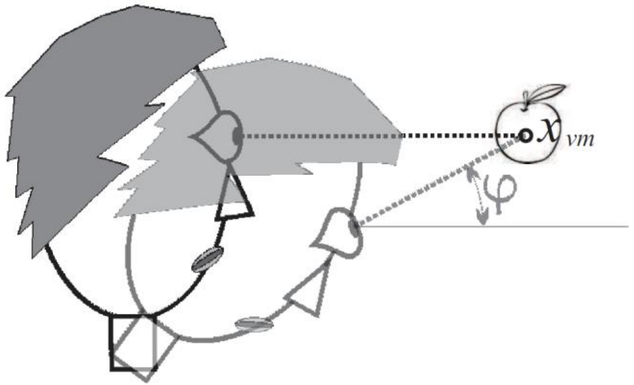
More efficient implementation using current eye velocities:

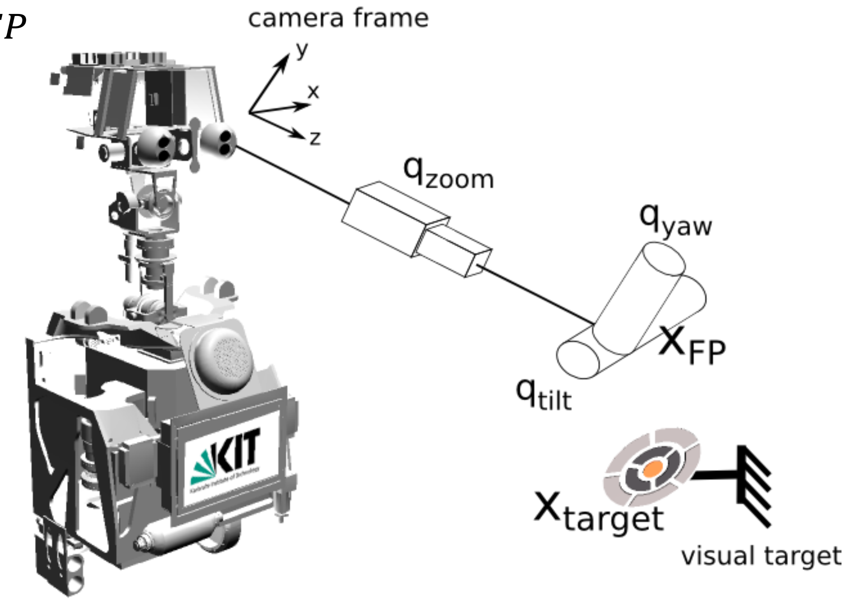
$$\dot{q}_{eye} = k_{okr} \cdot [\dot{u} + \dot{q}_{yaw} \quad \dot{v} + \dot{q}_{pitch}]^T$$



# Inverse Kinematics Control - Virtual Linkage Model

- Formulate the gaze stabilization problem as control problem in task space
- Fixation point $x_{FP}$ is the intersection of the lines of sight of each eye
- Virtual linkage between the eye and $x_{FP}$

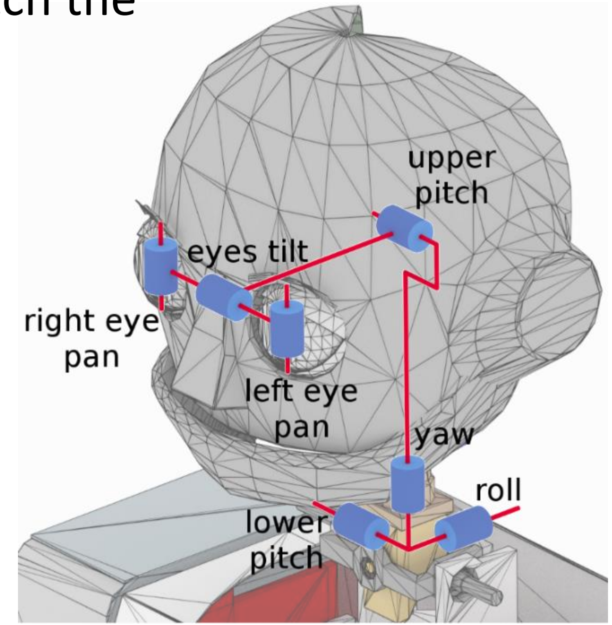


# Inverse Kinematics Control (IK)

**Idea:** Control the current fixation point $x_{FP}$ to match the input view target $x_{target}$

- Compensate for the motion induced by the body's own movements
- Intercept the motor commands and apply them to an internal robot model
- Predict new gaze target and correct gaze by computing head & eye movements




Kinematic model of the Karlsruhe Humanoid Active Head

# Inverse Kinematics Control (IK)

- Let $FK$ denote forward kinematic relationships, i.e. an internal robot model (purely kinematic)

$$x_{FP} = FK\left(q_{body}, q_{head}, q_{virt}\right)$$

$$\dot{x}_{FP} = [J_{body} \quad J_{head} \quad J_{virt}][\dot{q}_{body} \quad \dot{q}_{head} \quad \dot{q}_{virt}]^T$$

- Assume that $x_{target}$ is centered in the image
- Then the gaze is stabilized by $\dot{x}_{FP} = 0$ and thus

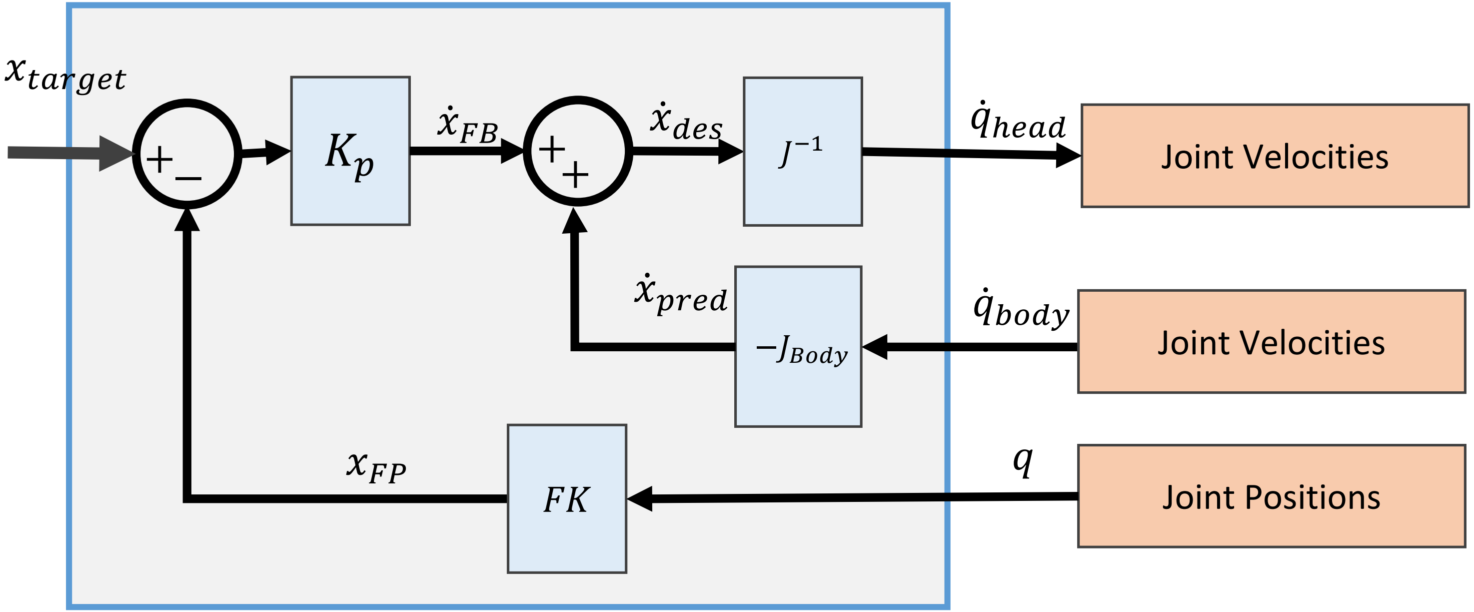
$$[J_{head} \quad J_{virt}][\dot{q}_{head} \quad \dot{q}_{virt}]^T = -J_{body}\,\dot{q}_{body}$$

- Use Moore-Penrose pseudoinverse $J^\dagger$ to compute velocities

$$\begin{bmatrix} \dot{q}_{head} \\ \dot{q}_{virt} \end{bmatrix} = -J^\dagger \left(J_{body}\,\dot{q}_{body}\right)$$

# Inverse Kinematics Control (IK)

Inverse kinematics control (IK)



# Advantages & Limitations

## VOR

### Advantages

- High sample rate
- Requires little computation and is easy to implement

### Limitations

- Can only compensate perturbations due to robot motions
- Requires an IMU
- Control only the eye joints

## OKR

### Advantages

- Versatile: The only source of feedback that can stabilize the image in a dynamic environment (i.e., with unpredictably moving objects)

### Limitations

- Input of the OKR is usually noisy and available at a lower frequency (e.g., 30 Hz for standard cameras).
- Low sample rate due to inherent drawback of image processing (i.e., less reactive & accurate)
- Control commands only for the eye joints

## IK

### Advantages

- Controls both head and eye joints

### Limitations

- Requires a target point
- Can only measure and thus stabilize self-induced perturbations
- Depends on the accuracy of the robot's kinematic model

# Experimental Evaluation

## Three evaluation scenarios

1. *Self Robot*

   Apply a sinusoidal motion to the torso joint (self-induced)

2. *External Robot*

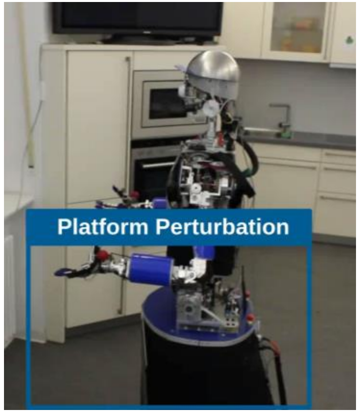
   Apply an external perturbation by rotating the platform

3. *External Target*

   Move a visual target (chessboard) on a TV screen

1.


2.

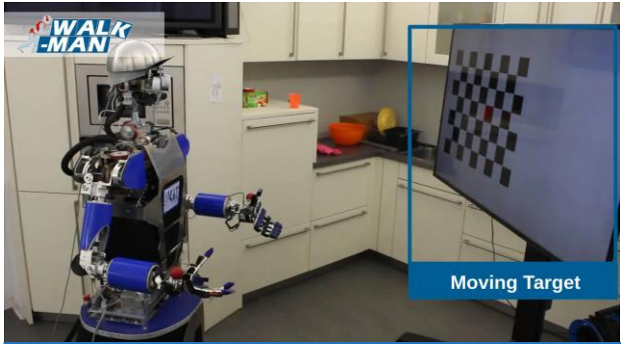



3.

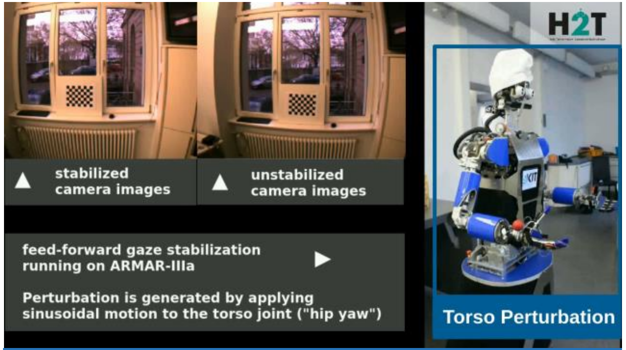
# Individual gaze stabilization modalities

Three stabilization methods:

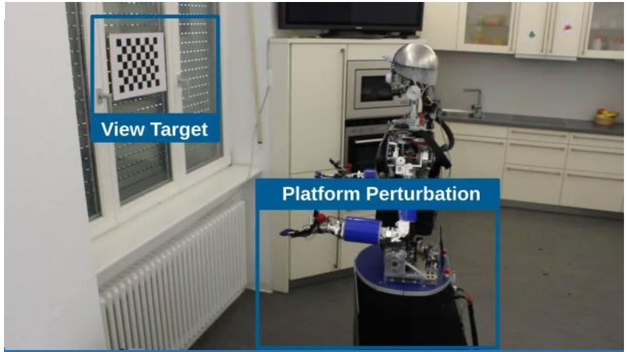
1. **Inverse Kinematics** method (IK)
2. **Vestibulo-ocular Reflex** (VOR)
3. **Optokinetic Reflex** (OKR)




OKR

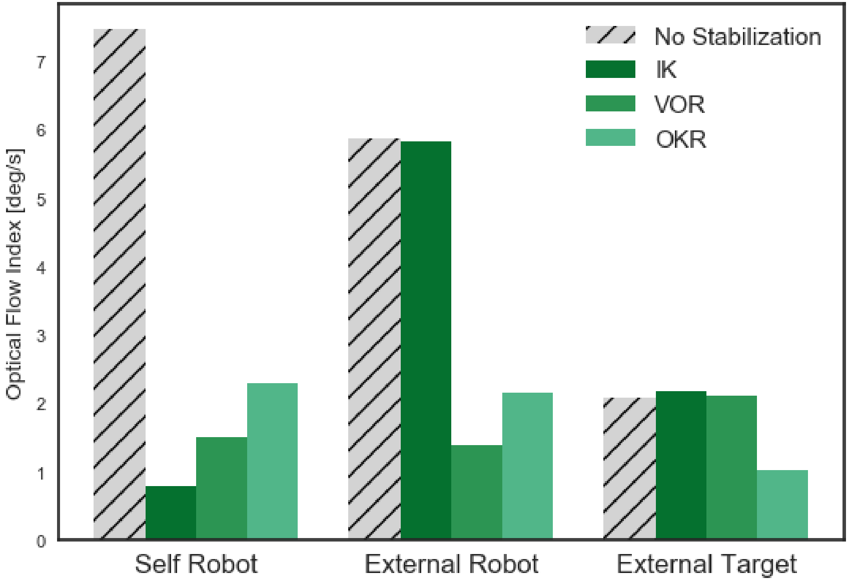



IK




VOR

# Results




- Stabilization methods are complementary, depending on the type of perturbation
- Ideal gaze stabilization should combine the three sources of information in order to be both versatile and efficient.

# Discussion and Combination

- Each individual modality works best for different tasks
- OKR and VOR are counter-acting (VOR compensates perturbations due to head motion and OKR stabilizes the motion perceived in the image)

## Combination

- Simple summation of the commands would lead to over-compensation
- Averaging the commands tends to under-compensation
- Can we find a way to combine the methods to increase **versatility?**

# The Reafference Principle

Concept proposed by Erich von Holst and Horst Mittelstaedt 1950

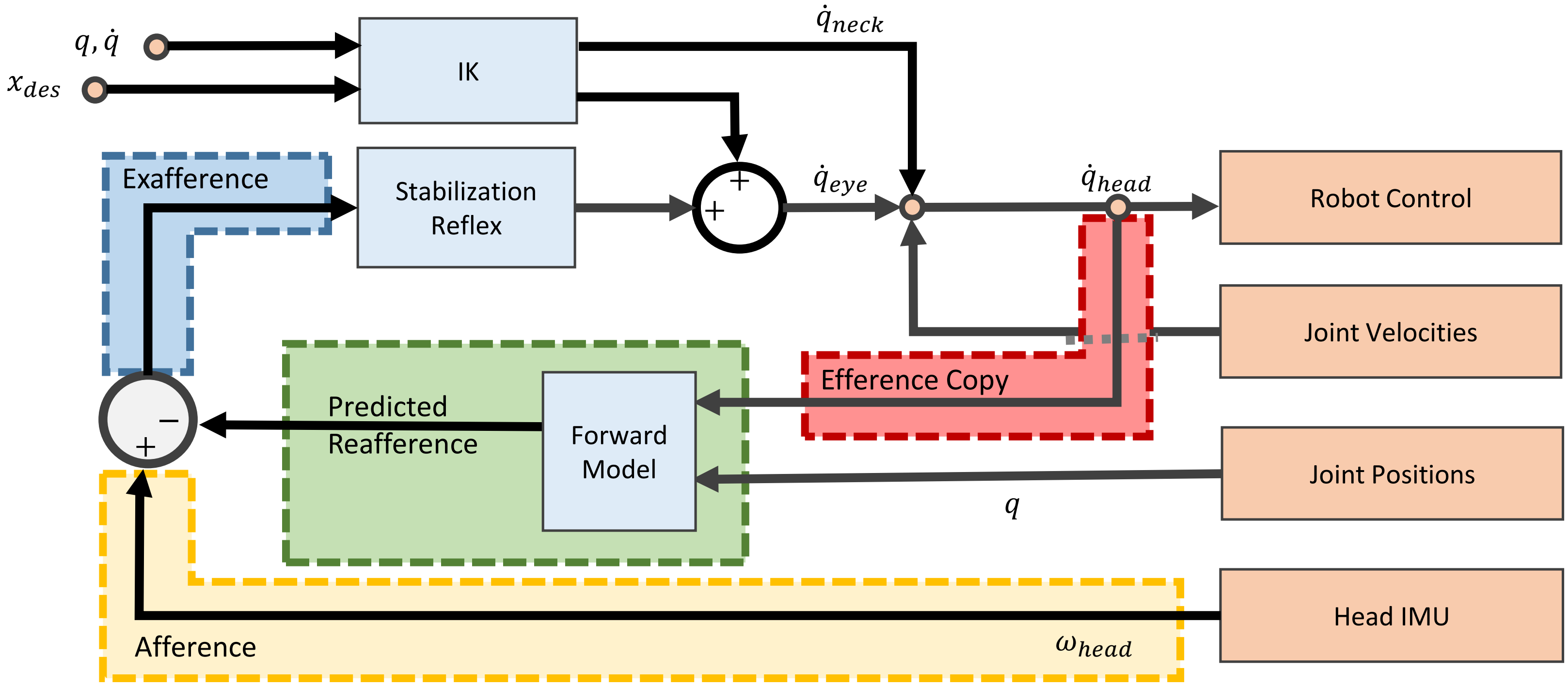
## Basic Idea

- Use **copies** of the motor commands (***efference* copy**) to **predict** the expected sensory outcome of self-induced motions (**predicted *reafference***)
- These *reafferences* are then subtracted from the actual sensor measurements and thus isolating the sensory consequences of externally induced perturbations (called *exafference*)

## Notes

- Efference copies play an important role in grasping, speech production, …
- Self-produced tickling motion is less "*tickly*"

# Gaze Stabilization based on the Reafference Principle




T. Habra, M. Grotz, D. Sippel, T. Asfour, and R. Ronsse, "Multimodal gaze stabilization of a humanoid robot based on reafferences. *International Conference on Humanoid Robotics (Humanoids)*, 2017

# Head rotational velocity prediction

- Combination requires to predict the head rotational velocity

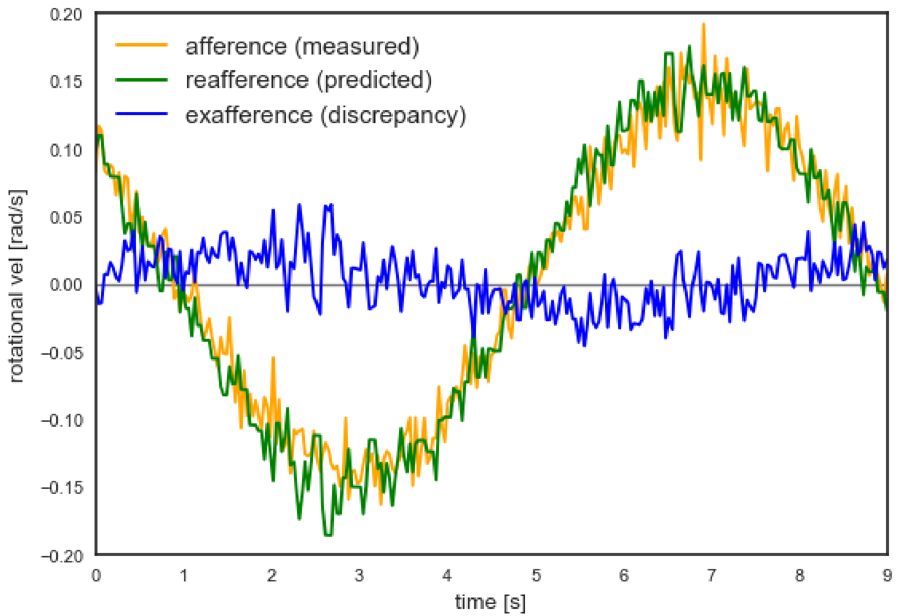
Let

- $R_{imu}$ denote the location of the IMU in the robot model
- $\omega_{imu}$ the IMU‘s rotational velocity as function of the joint positions $q$ and velocities $\dot{q}$

Then the reafference for the gyroscope velocities is given by

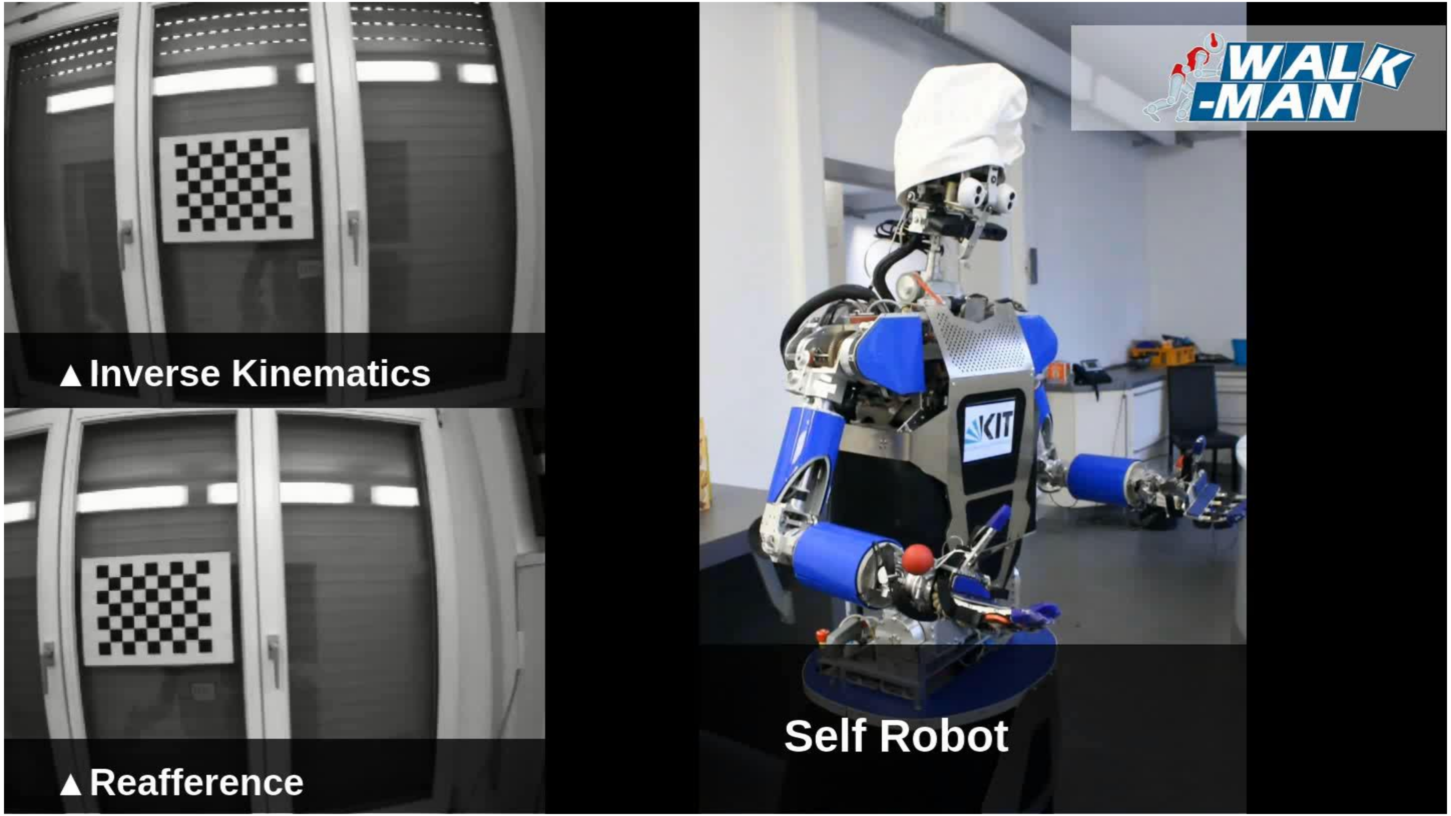
$$\widehat{\omega}_{head} = R_{imu}(q) \cdot \omega_{imu}(q, \dot{q})$$

# Exafference ARMAR-III




- Torso joint was subjected to a sinusoidal perturbation
- Diagram shows the rotational velocity of the head

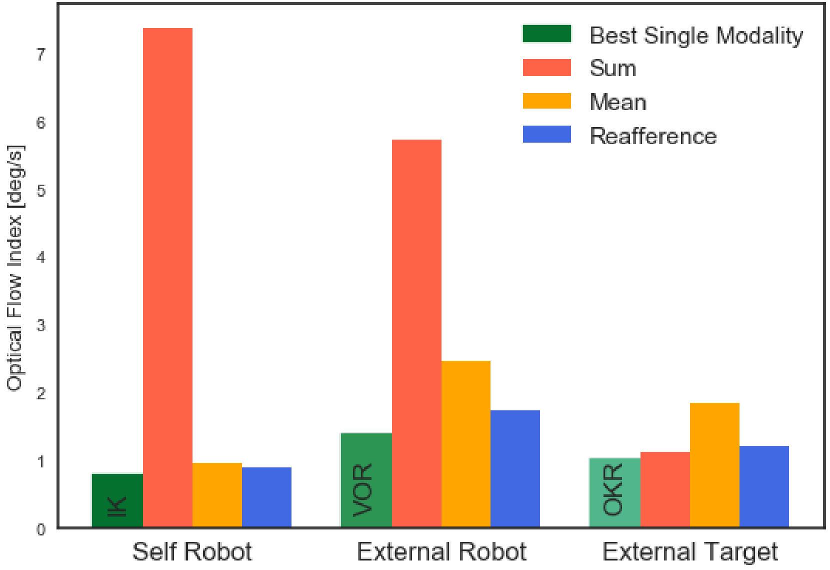
# Combination based on Reafference: Results



# Conclusion

More **versatile** and as robust as the best single modality

**(reafference-based stabilization perfoms as the best individual modality)**




- **Combining** the gaze stabilization modalities with the **reafference principle** enhances the **versatility** of gaze stabilization.
- Reflexes are only invoked if there is an actual perturbation

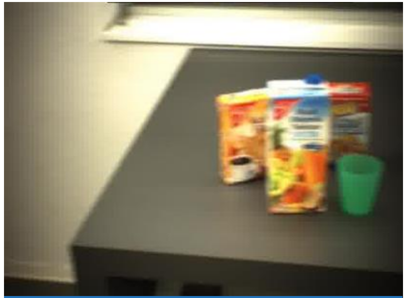
# Applications: Object Localization While Moving

## Object localization

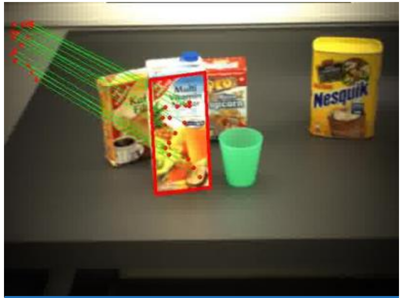
- Textured objects: SIFT features
- Single colored objects: appearance-based matching

## Perturbation

- Torso joint is subjected to sinusoidal perturbation while localizing objects

foveal camera unstabilized




foveal camera stabilized

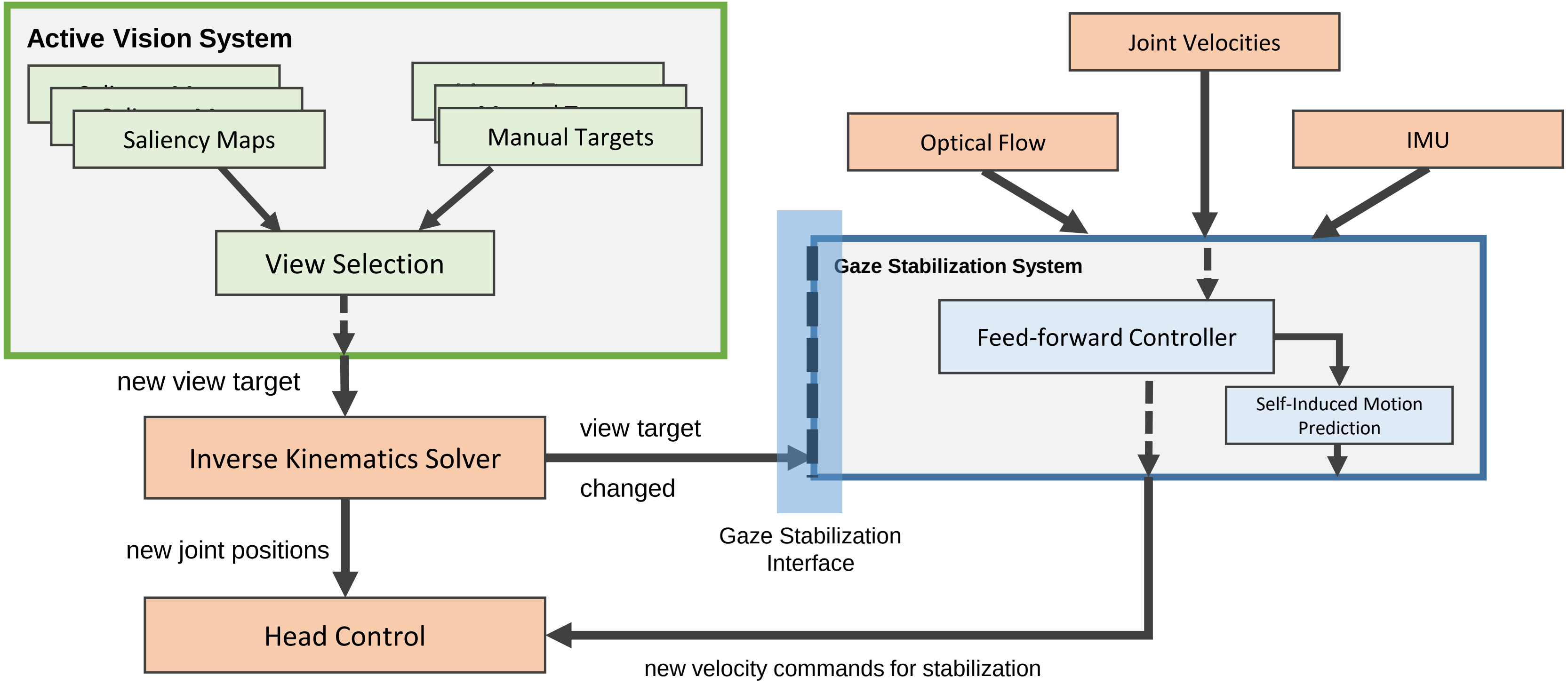
# Applications: Object Localization While Moving

# Application: Grasping while Moving Experiment

- Gaze stabilization decreases the optical flow by 50%

- Results in a better object localization result (localization uncertainty)

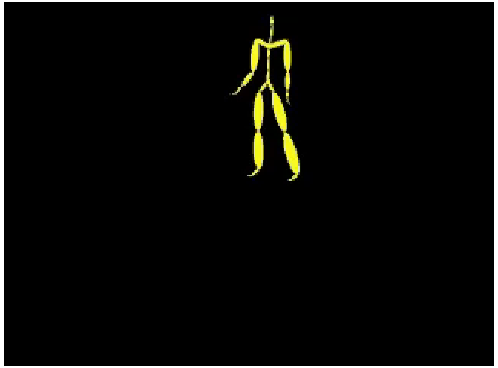
# System Architecture

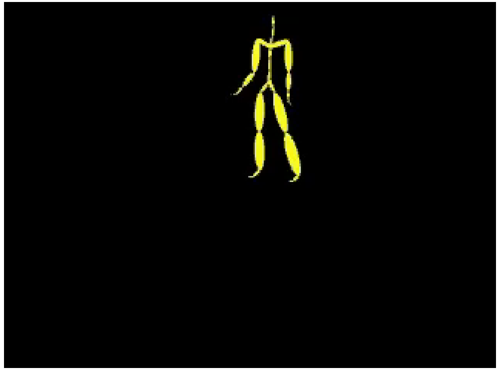


# Outline of This Lecture

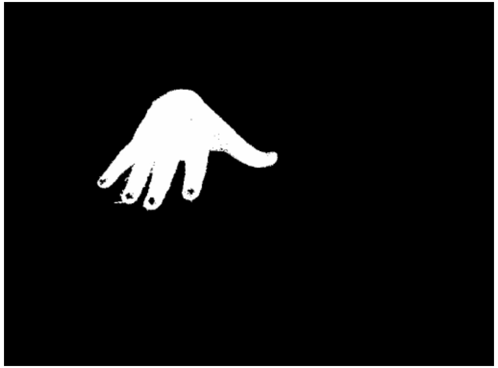
- Introduction to Active Perception
  - Definition of Active Perception
  - The Human Eye
  - Human Visual Attention
- Active Visual Perception
  - Gaze Control & Stabilization
  - Object Discovery and Segmentation **(by pushing)**
- Active Haptic Perception
  - Tactile Exploration
  - Visuo-Haptic Grasping
- Active Hearing

# The Role of Action





CMU Graphics Lab Motion Capture Database (http://mocap.cs.cmu.edu)



Antonis Argyros, FORTH

# Object-Action Complexes

- Research in humans has shown that perception and action are **tightly coupled, intertwined, equivalent!**

- PACO-PLUS: Objects and Actions are inseparably intertwined
  → Object-Action Complexes (OACs)
    - Actions define Objects
    - Objects suggest Action

Krüger, N., Geib, C., Piater, J., Petrick, R., Steedman, M., Wörgötter, F., Ude, A., Asfour, T., Kraft, D., Omrčen, D., Agostini, A. and Dillmann, R., *Object-Action Complexes: Grounded Abstractions of Sensorimotor Processes*, Robotics and Autonomous Systems, vol. 59, no. , pp. 740-757, 2011

# Why Visual Perception?

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

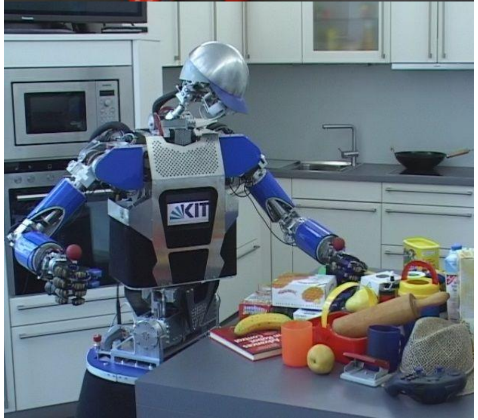
# Visual Perception

- **Recognition, localization, observation**: see also computer vision lectures
  - Robotics-III
  - “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:
  - be able to recognize them when seen again
  - grasp/manipulate them
  - learn about them



# Discovery and learning of unknown objects

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

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

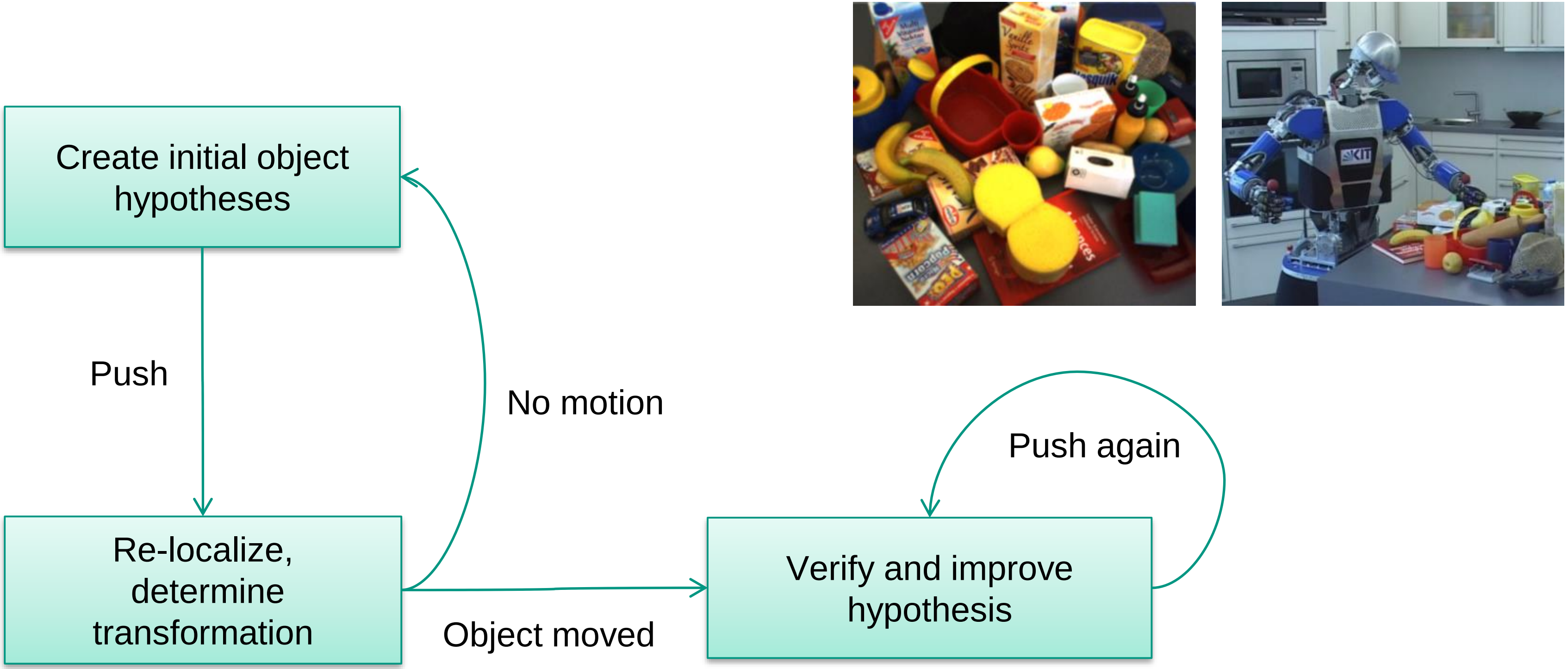
# Discovery and learning of unknown objects

- **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

No motion

Re-localize, determine transformation

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 objects
  - Unicolored regions of promising size (color MSERs (Maximally stable extremal regions) → single-colored objects
  - Visually salient regions (Difference of Gaussians filter)

  → objects that are neither textured nor unicolored

# Create initial object hypotheses

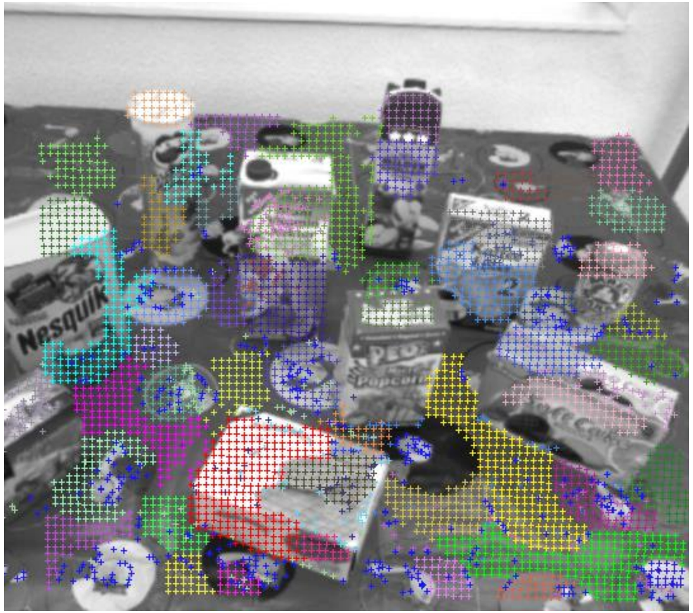
- Generate hypotheses using all three heuristics
- Dense stereo matching to get 3D position of all pixels of the image
- Each hypothesis is represented by the set of RGBD points in its image region

- 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

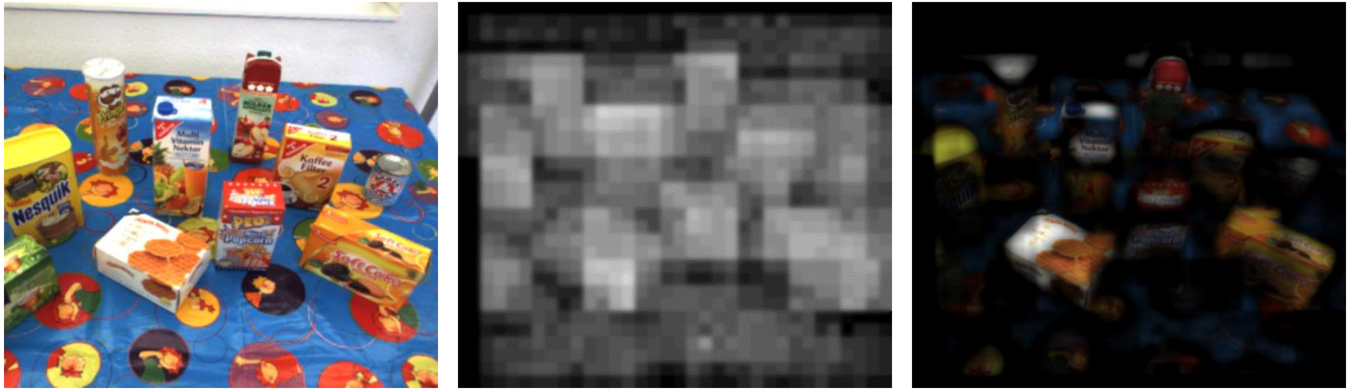
# Create initial object hypotheses

- Use all 3 criteria → lot of hypotheses
- Only „pushable“ hypotheses desired



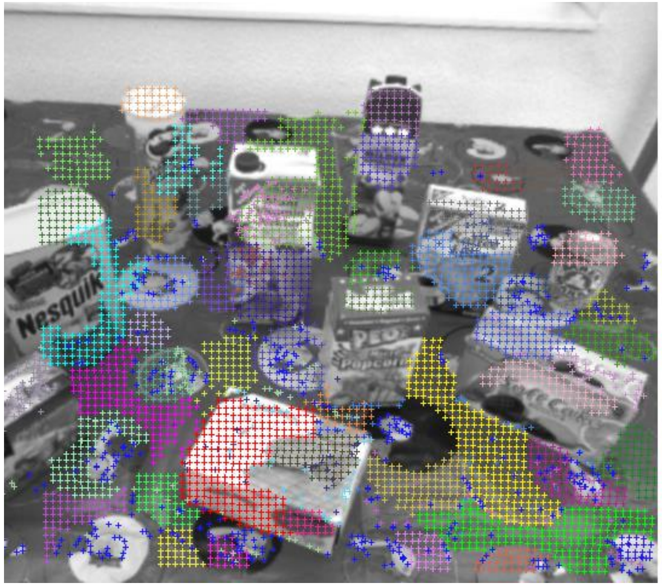
# Generation of Object Hypotheses

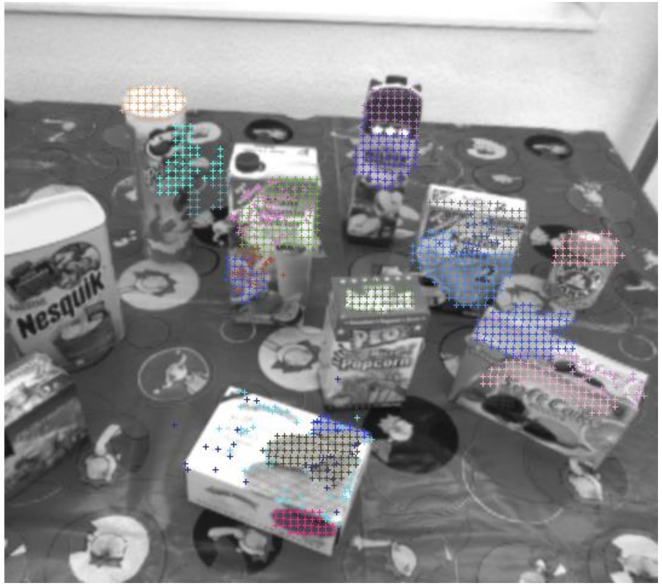
- Additional criterion:
  - image region should correspond to high part of the scene
  - Calculate proximity to local maxima of image parts based on the 3D point cloud



# Generation of Object Hypotheses

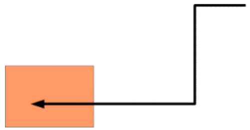
- Use proximity to local maxima to filter object hypotheses before choosing one for pushing





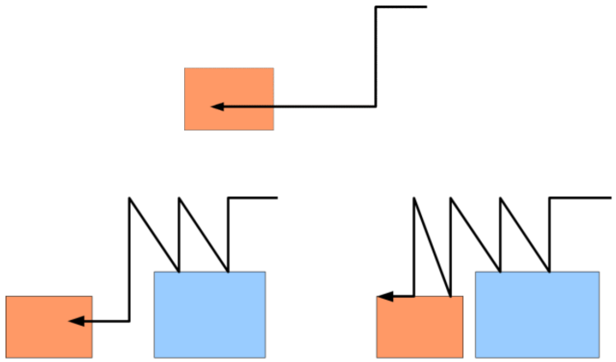
# Object pushing

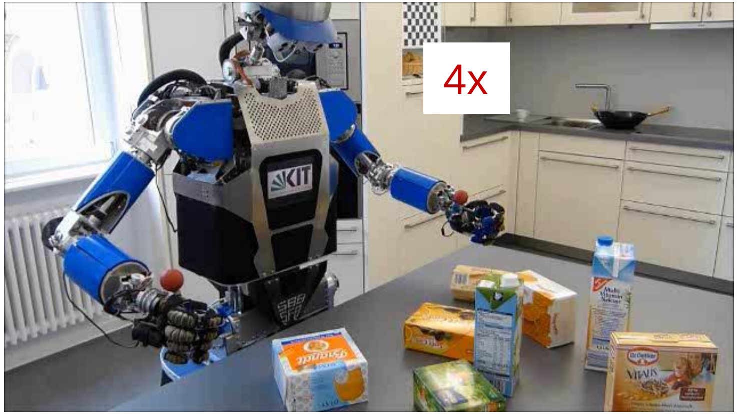
- Pushing: move the object sufficiently for segmentation, but:
  - Keep object in field of view
  - Do not change visual appearance too much
- Push object over a fixed distance towards a central point in front of the robot



# Object pushing – Optimization strategies

- 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

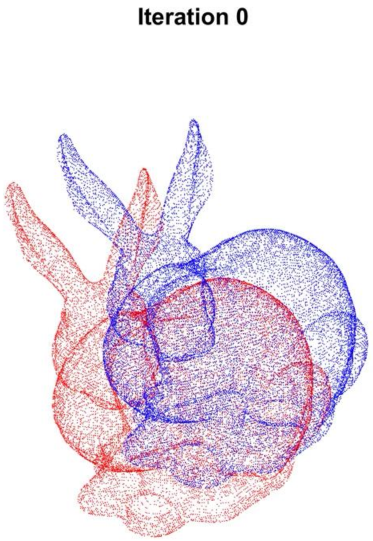
↓

First: ICP starting at old position

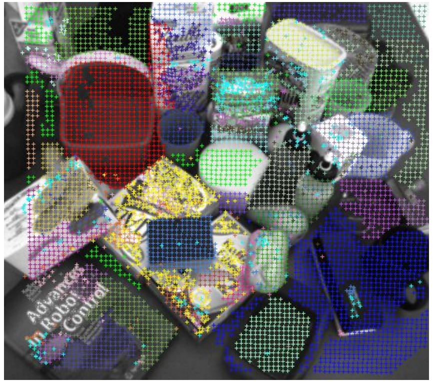
- Re-localize object and estimate motion by point cloud matching
- Object hypotheses represented by sets of RGBD points
- Iterative Closest Point (ICP) for matching, using a distance in cartesian and color space

# 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







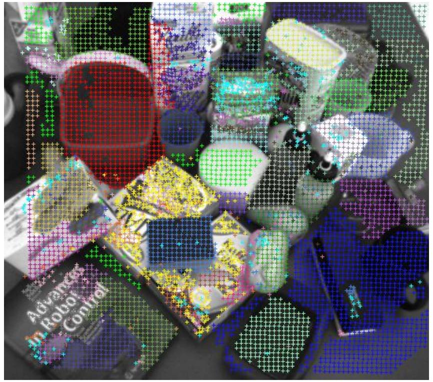
# Motion estimation (III) - ICP



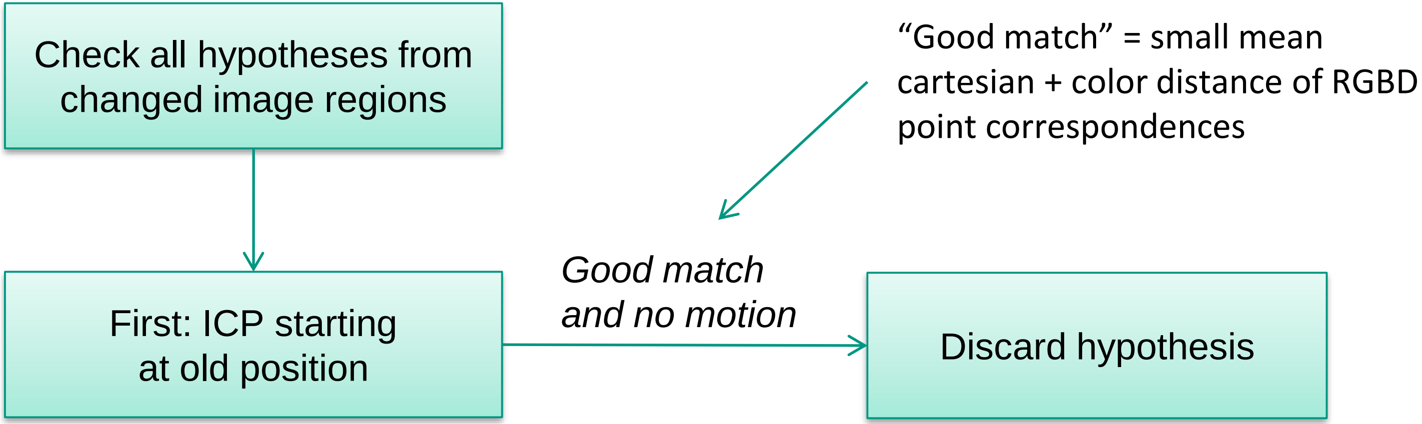
- Match two point clouds ***A*** and ***B*** using ICP:
  1. For each ***a*** ∊ ***A*** find closest point in ***B***
  2. Calculate transformation ***T*** that minimizes the mean squared distance of the correspondences
  3. Apply ***T*** to all ***a*** ∊ ***A***
- Iterate until convergence

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

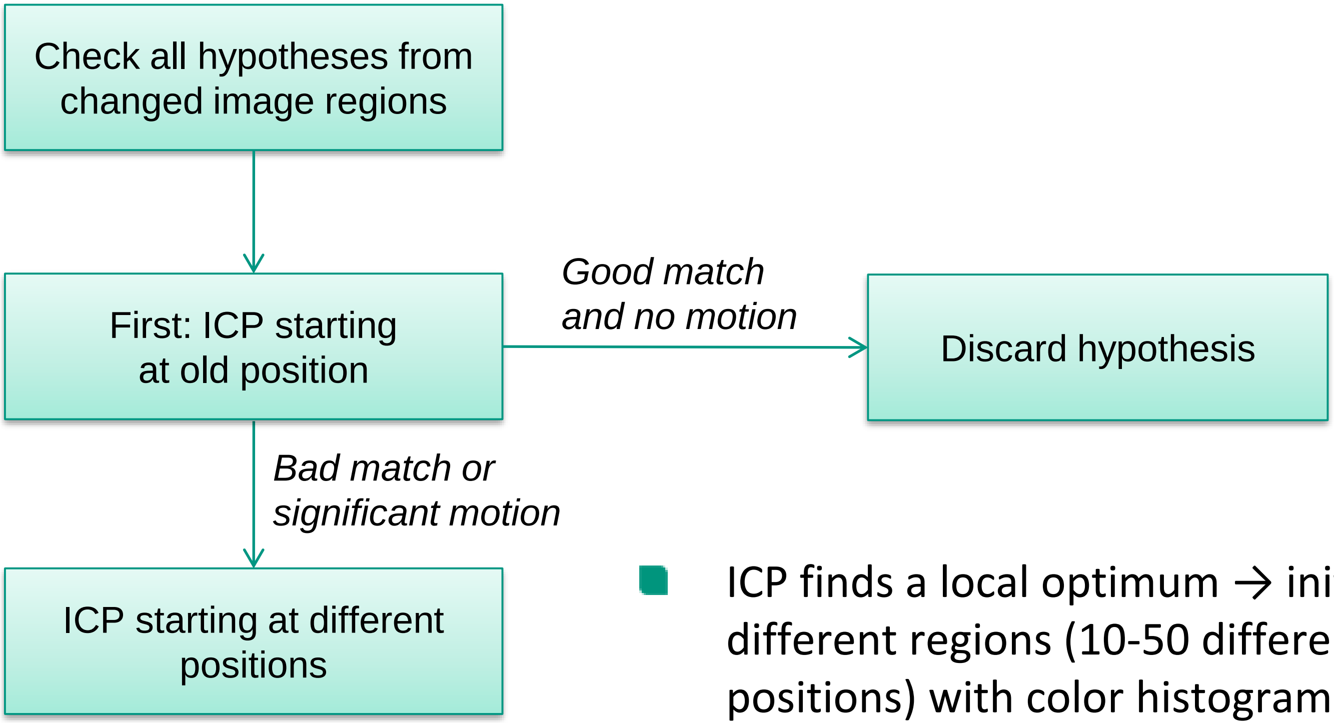
- Use weighted cartesian + color distance in step 1 of ICP
  - Removes most shape ambiguities
  - Gives more reliable point correspondences

# Motion estimation (IV)

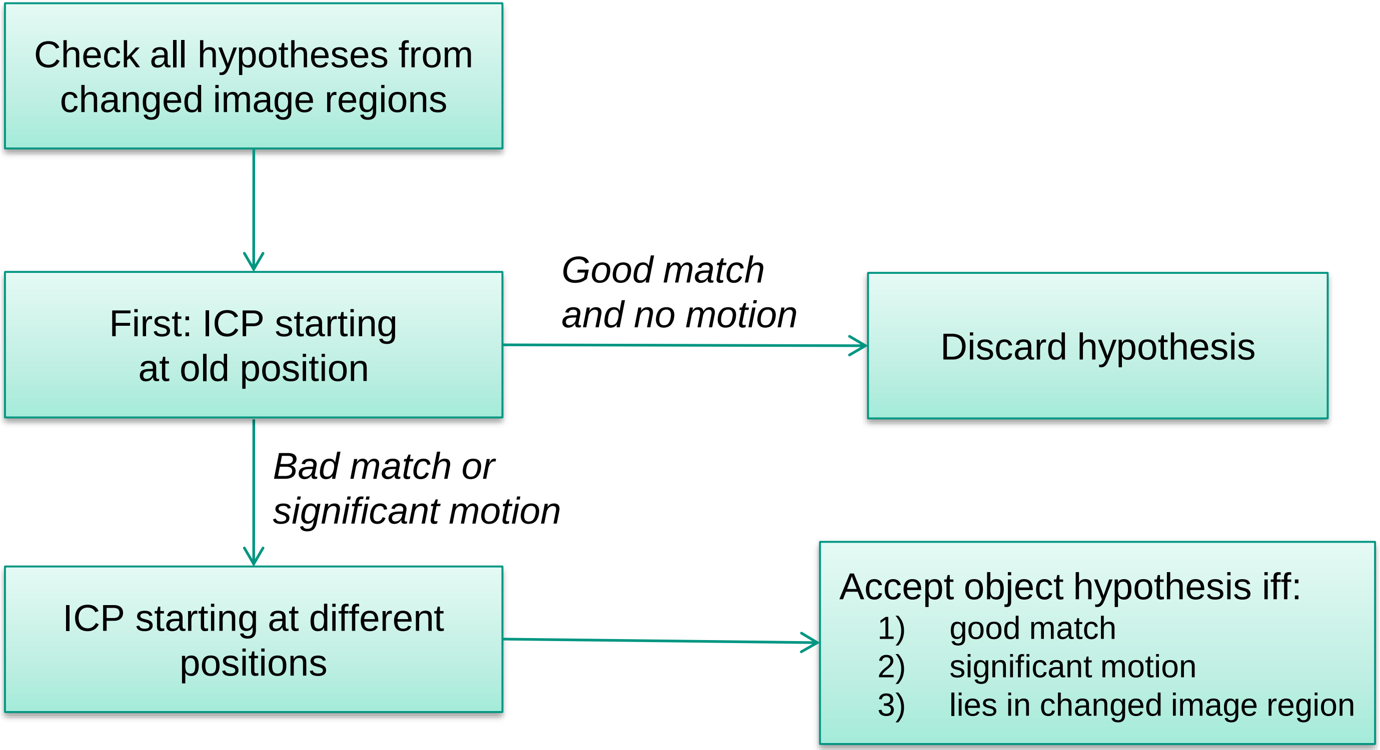


# Motion estimation (V)




- ICP finds a local optimum → initiate it in different regions (10-50 different start positions) with color histogram similar to hypothesis

# Motion estimation (V)



# Overview

Create initial object hypotheses

Push

Re-localize, determine transformation

No motion

Object moved

**Verify and improve hypothesis**

Push again

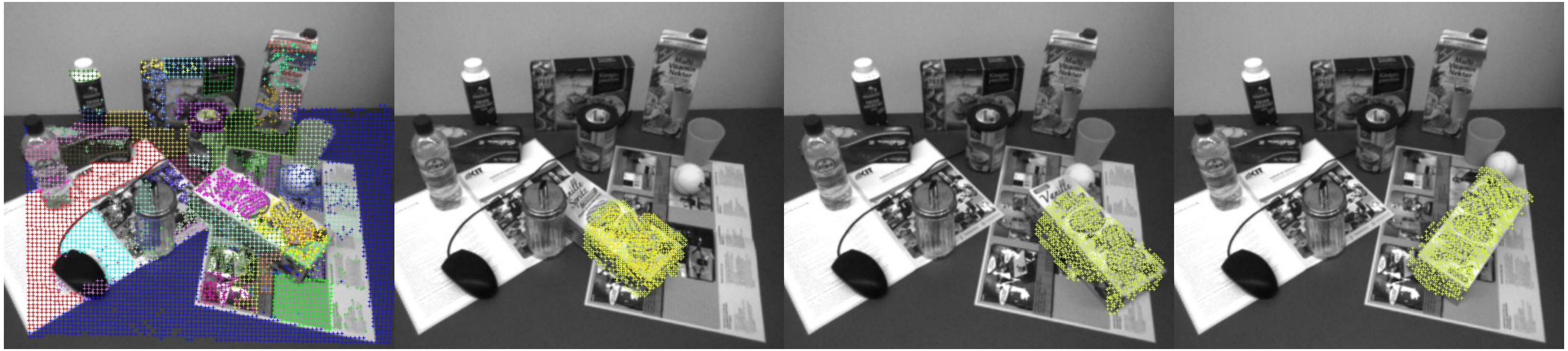
# Hypothesis correction and extension

- 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

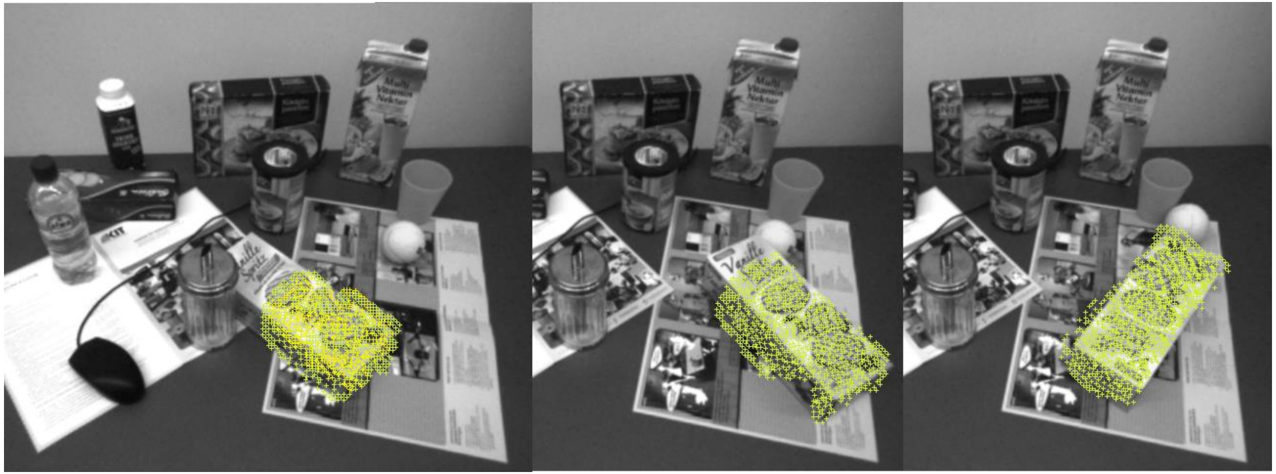
- Add new candidate points that lie in a changed image region close to the hypothesis

- Improve hypothesis over several pushes



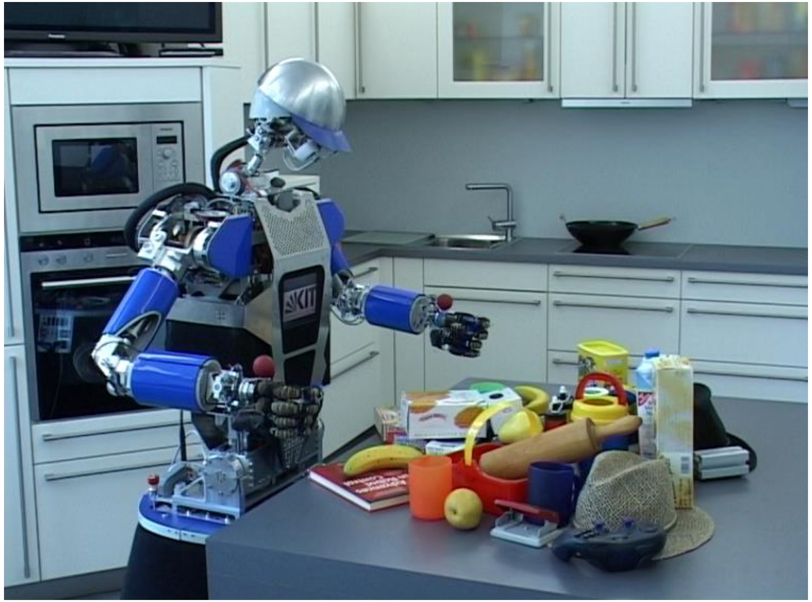
# Hypothesis correction and extension

- Push the object 2-3 times
  - ⇒ complete segmentation
- More pushes reveal different sides
  - ⇒ generate a multi-view descriptor



# Interactive Object Segmentation Example

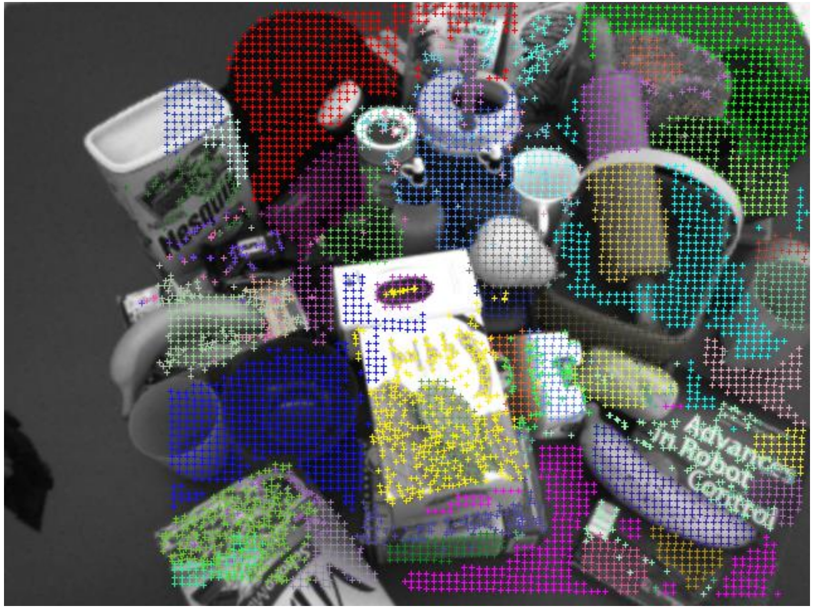




External view


Robot’s view (left camera image)

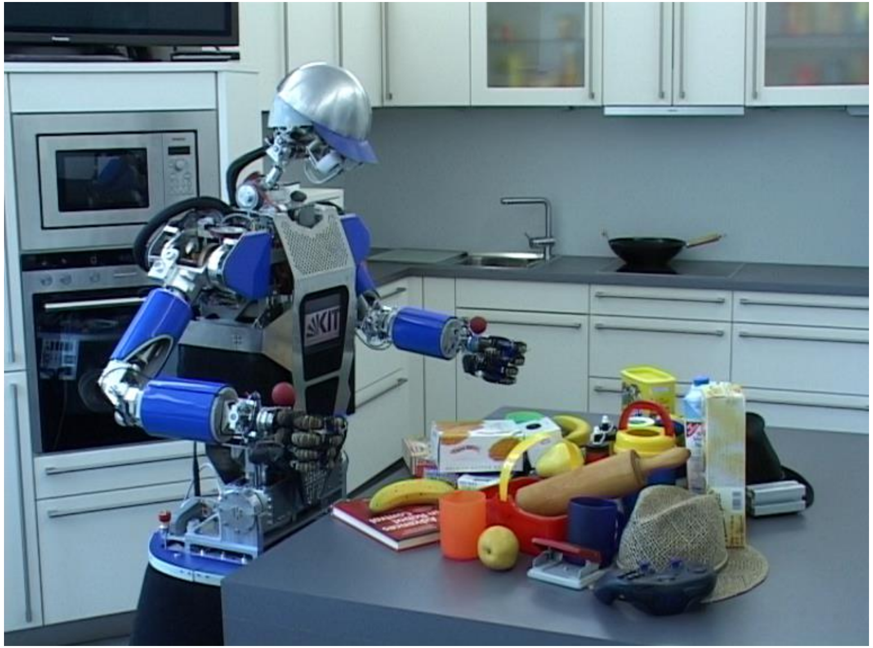
# Object Hypotheses




Initial object hypotheses


Planned push

# Push Execution



External view


Robot’s view (left camera image)

# Visual Observation: Before and After Push


Old camera image


New camera image

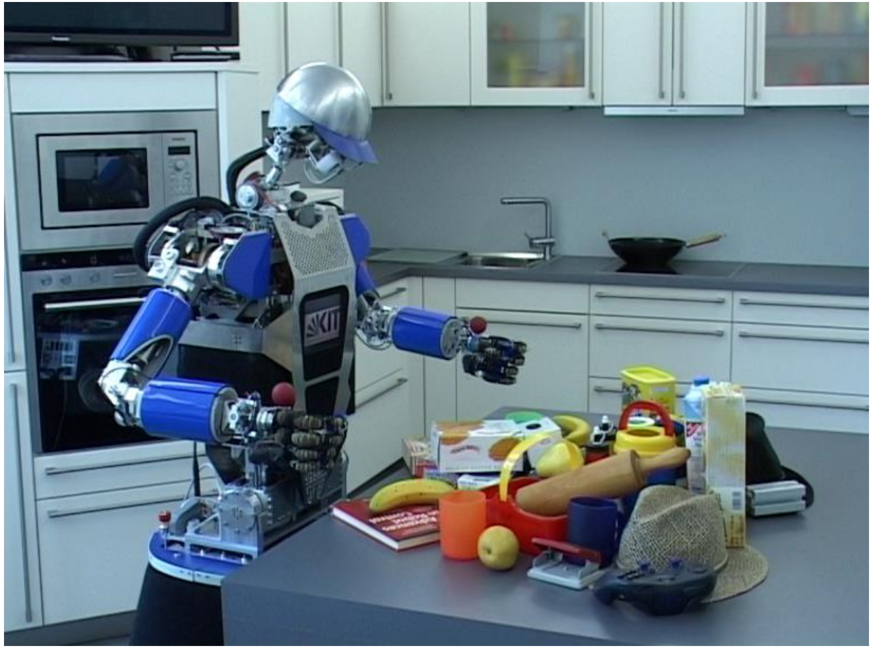
# Confirmed Object Hypotheses

Changed image regions



Crosses are confirmed points, dots newly added candidates

# Second Push Execution



External view

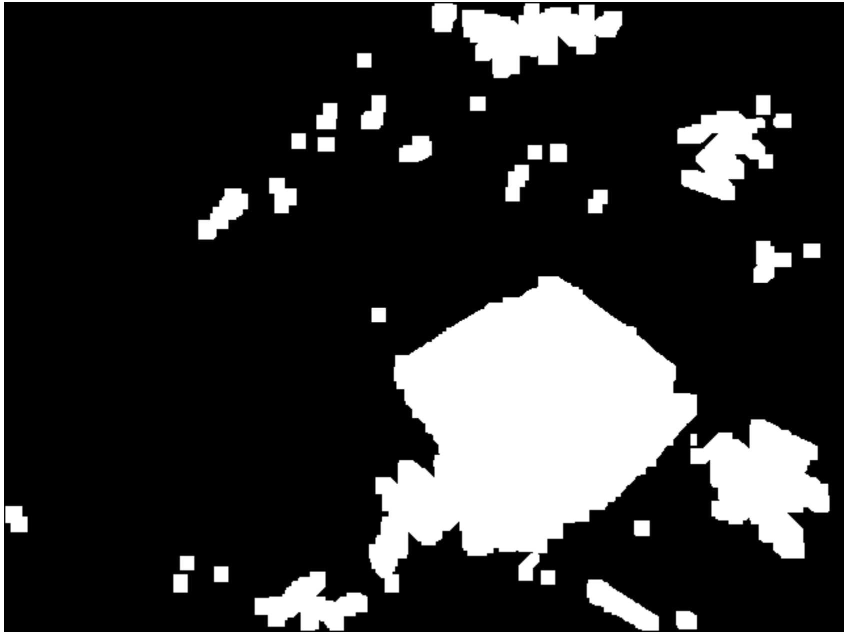

Robot’s view (left camera image)

# Visual Observation: Before and After Push


Old camera image


New camera image

# Confirmed Object Hypotheses

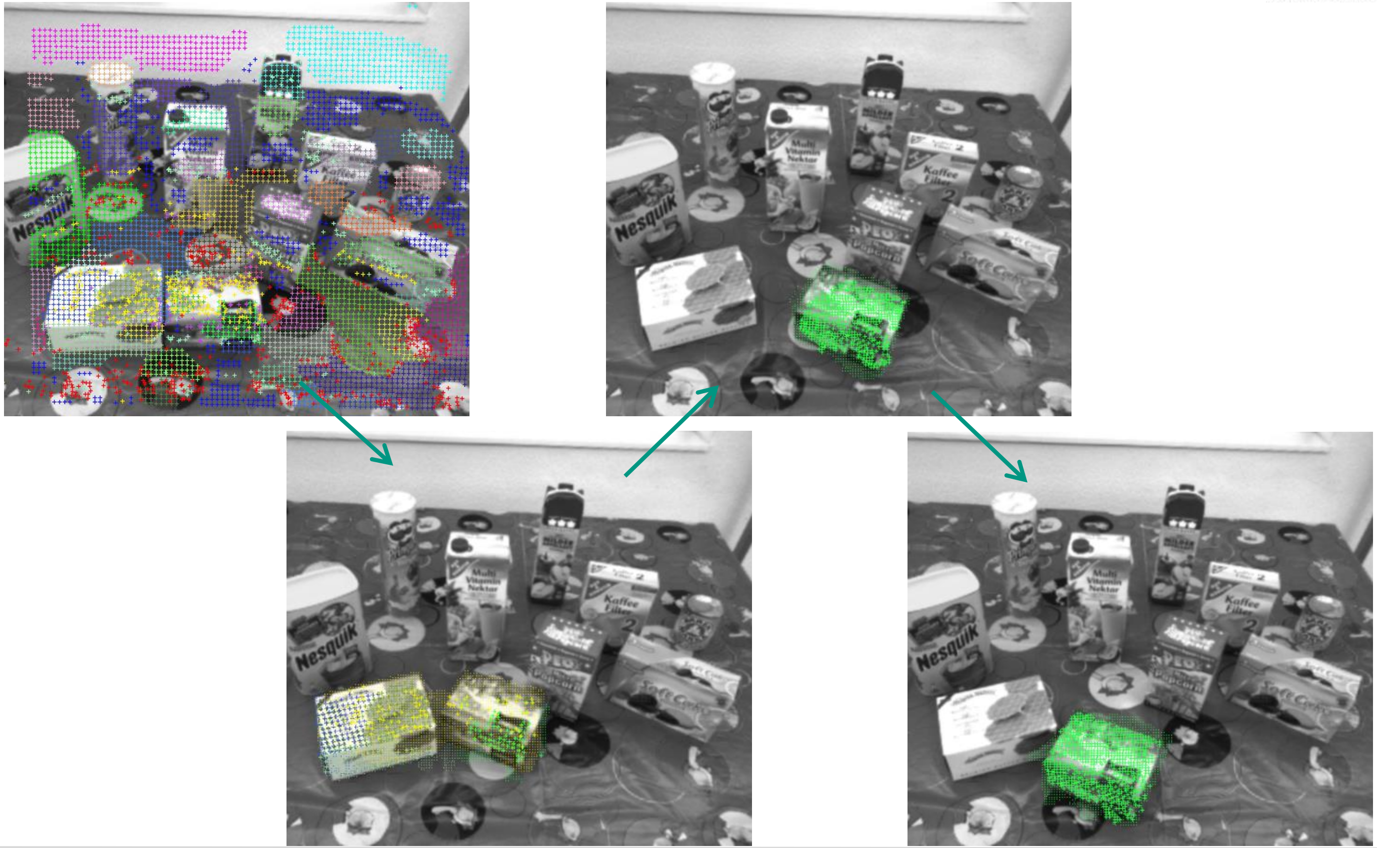


Changed image regions


Crosses are confirmed points, dots newly added candidates

# Object segmentation example

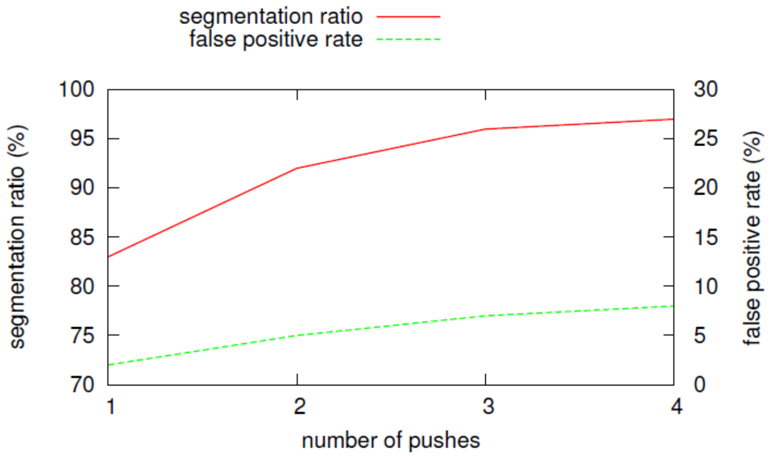




# Object Learning for Recognition: Results

- Segmentations usually correct and complete

- Proof of concept: Simple object descriptor created based on the segmentation
  - $\Rightarrow$ 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

Red: relevant for the exam

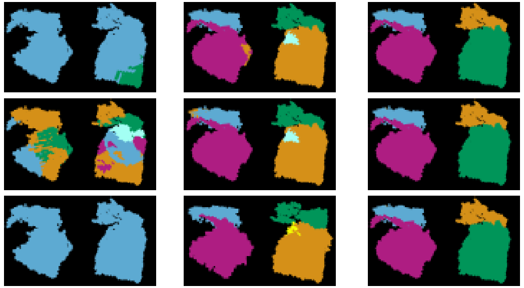
- D. Schiebener, A. Ude and T. Asfour, *Physical Interaction for Segmentation of Unknown Textured and Non-textured Rigid Objects*, IEEE International Conference on Robotics and Automation (ICRA), 2014
- D. Schiebener, J. Morimoto, T. Asfour and A. Ude, *Integrating visual perception and manipulation 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

# 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.



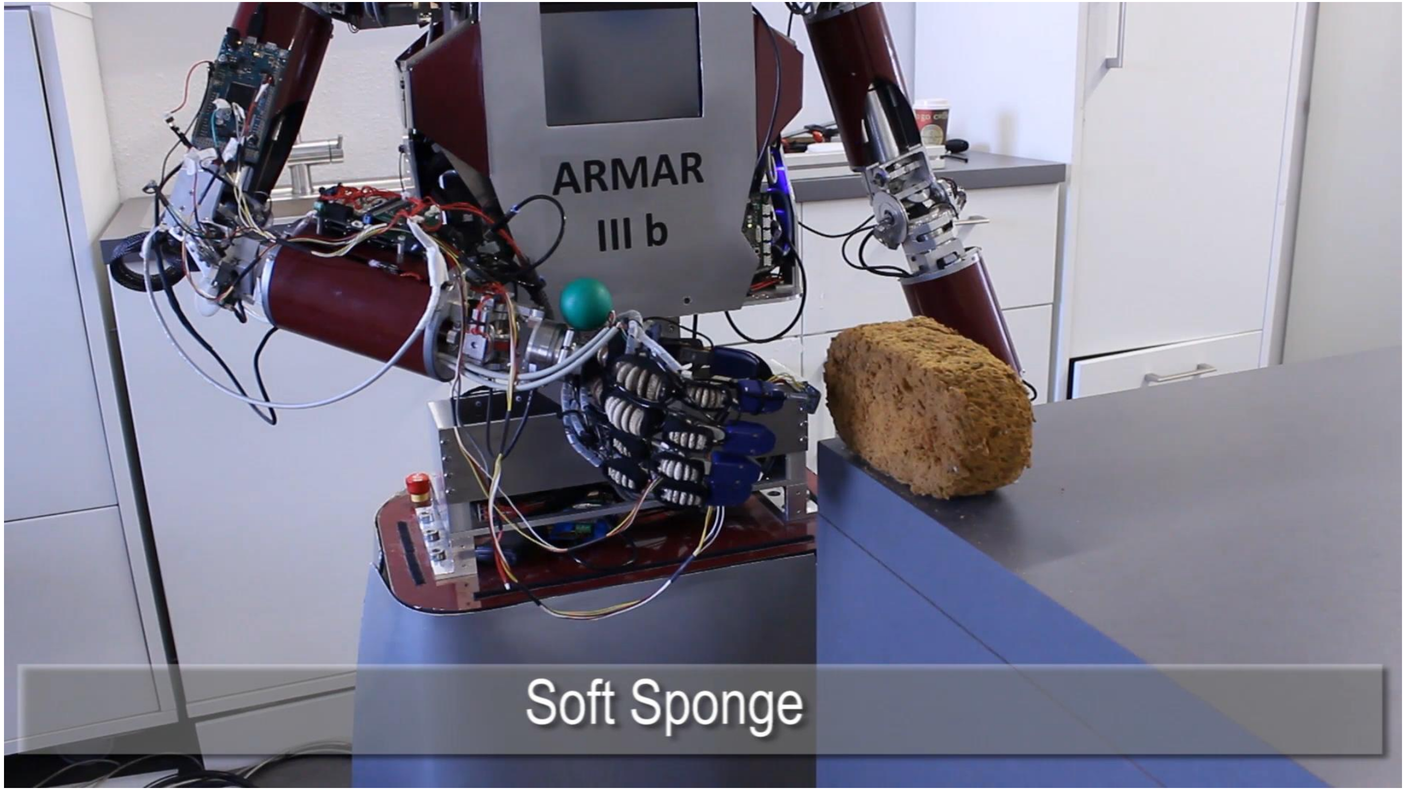
Prior samples (0 actions) | Posterior samples (5 actions) | Posterior samples (15 actions)

# Outline of This Lecture

- Introduction to Active Perception
  - Definition of Active Perception
  - The Human Eye
  - Human Visual Attention
- Active Visual Perception
  - Gaze Control & Stabilization
  - Object Discovery and Segmentation **(by pushing)**
- Active Haptic Perception
  - Tactile Exploration
  - Visuo-Haptic Grasping
- Active Hearing

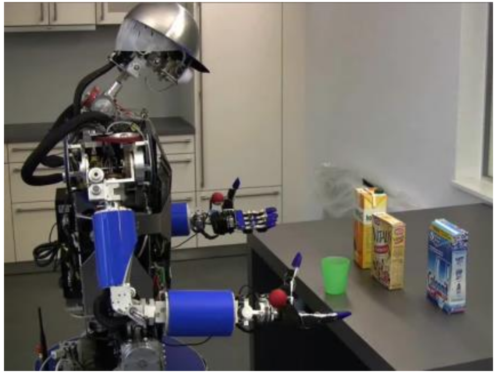
# Active Haptic Perception = Haptic Exploration

# Tactile exploration of a sponge using ARMAR-IIIb



# Motivation

Karlsruher Institut für Technologie

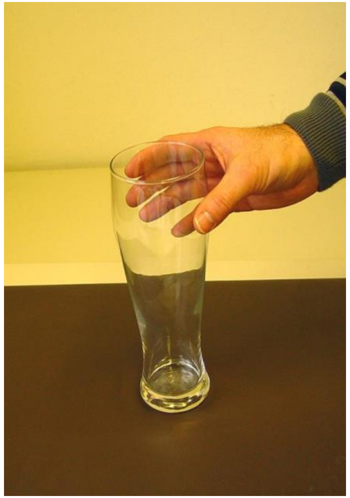


- How to grasp and manipulate unknown objects?
- How to acquire visual object knowledge?
- How to augment visual object information?

→ Haptically explore unknown objects

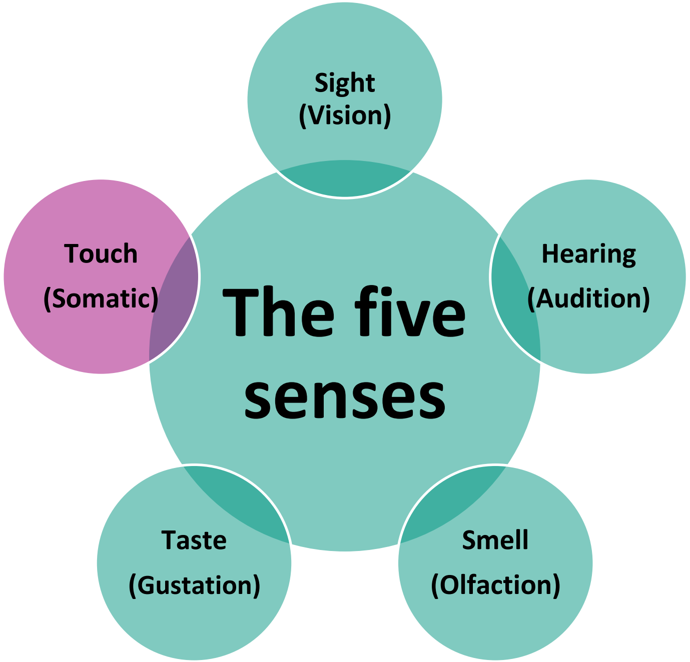
- Active touch information from haptic exploration enables human to “discover” object properties (share, surface, …)

  → Hints for classification, recognition and manipulation



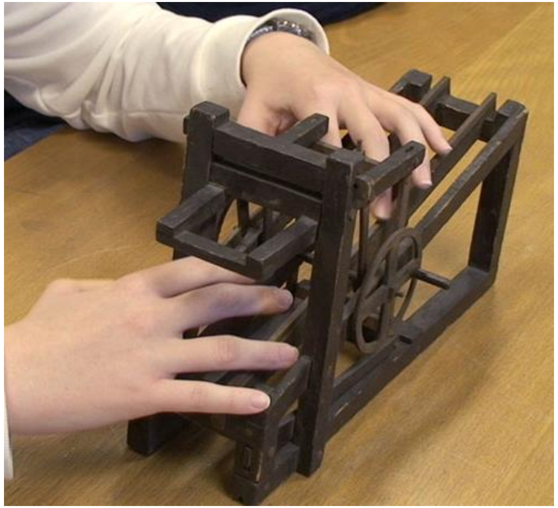
# What is Haptics?

- The sense of touch
- Any form of nonverbal communication involving touch
- Body contact is fundamental!
  - Object exploration
  - Hand shake
  - Communication of feelings



# What 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

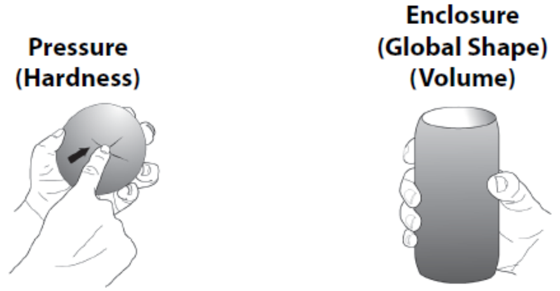
- Tactile / Cutaneous:
  - temperature, pressure, vibration, slip, pain
  - Sensation arising from stimulus to the skin

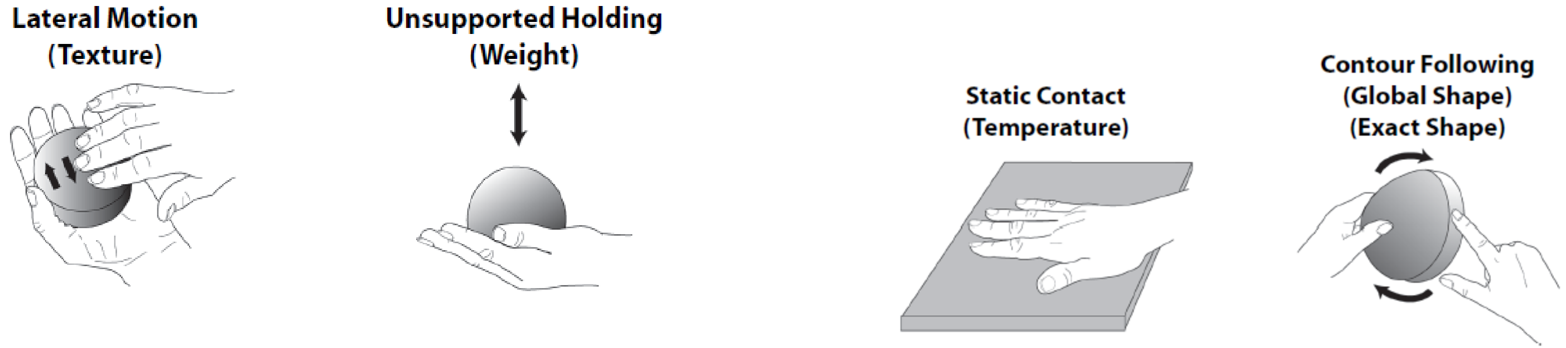
- Proprioception / kinesthesia:
  - Limb position/location, motion, force
  - End organs located in muscles, tendons, and joints
  - Stimulated by body movement

**Haptics = Tactile + Proprioception**

# Haptic 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

# Important questions in haptic perception

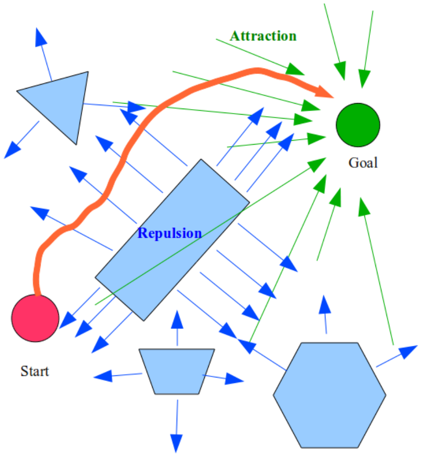
- Haptic sensor technologies
- Object shape estimation based on collected sparse haptic data
- Applications, e.g. grasping unknown objects

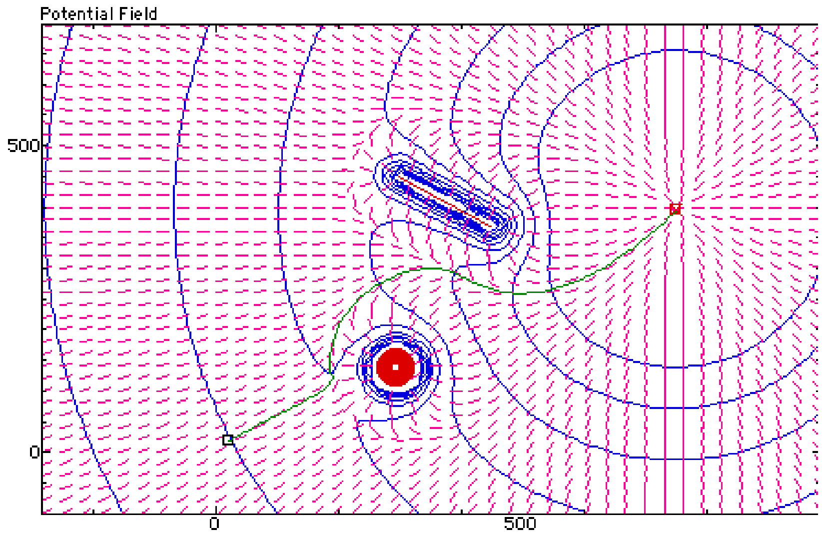
# Potential Field Based Exploration

Method originally developed for

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

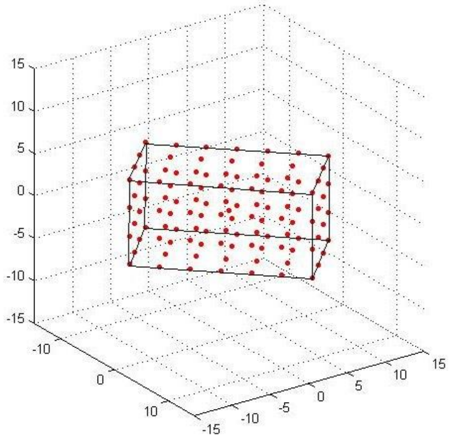
*Bierbaum, A., Rambow, M., Asfour, T., Dillmann, R. Grasp Affordances from Multi-Fingered Tactile Exploration using Dynamic Potential Fields. In IEEE/RAS International Conference on Humanoid Robots, 2009.*



# 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 based on tactile response
- Superposition of individual potential sources

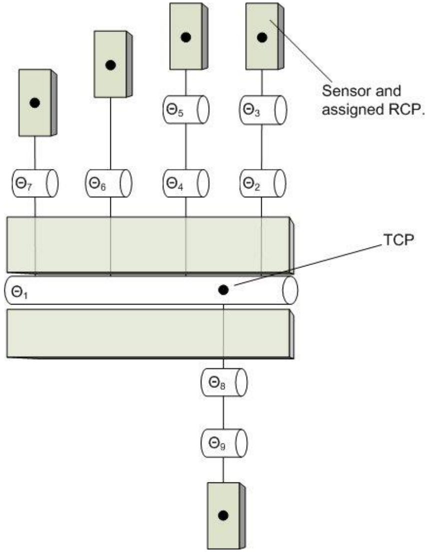



$$\Phi(x) = \sum_i \Phi_{r,i}(x) + \sum_j \Phi_{a,j}(x)$$

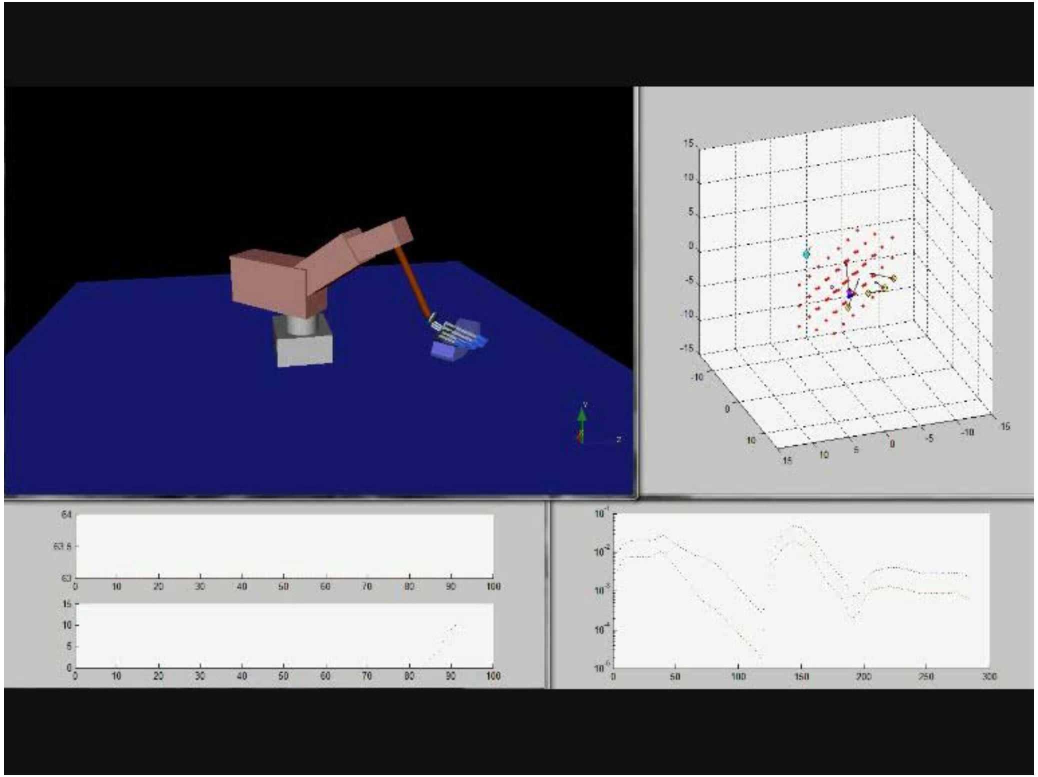
- Field initialization from pose and extension estimation of target object, e.g. by computer vision.

# Exploration using dynamic potential fields

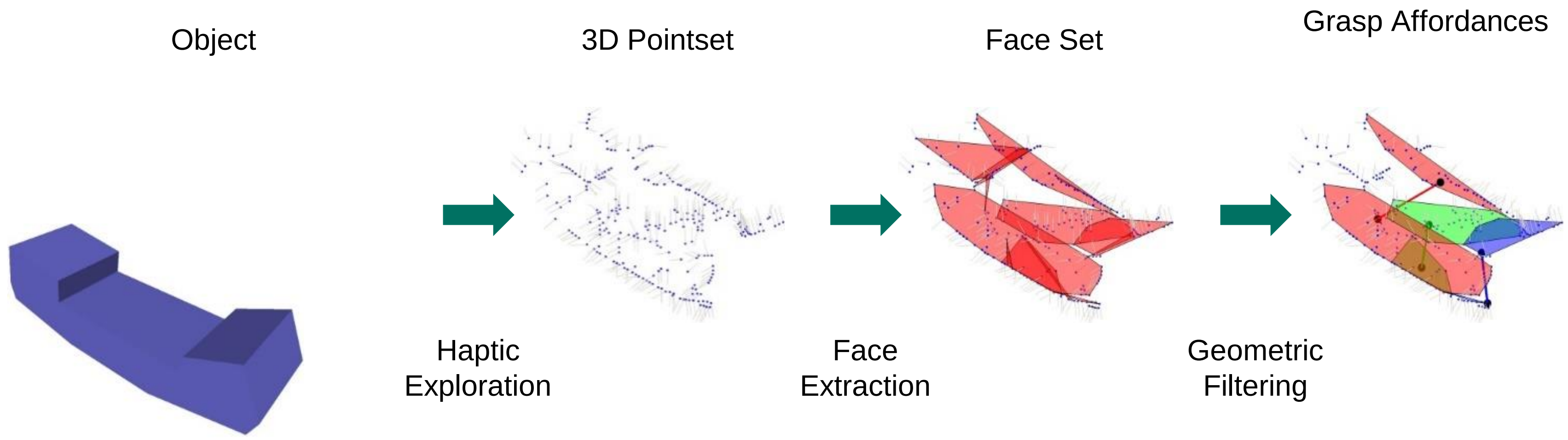
- Generation of trajectories for multi-point end-effectors (Robot Control Points, RCPs) using real-time gradient calculation [Khatib 1986]
- Harmonic potential functions to minimize number of local minima
- Reconfiguration strategy for resolving structural local minima of the hand
- Real-time inverse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



# Haptic Exploration with Movemaster (RM-501)



# Extracting Grasp Hypotheses

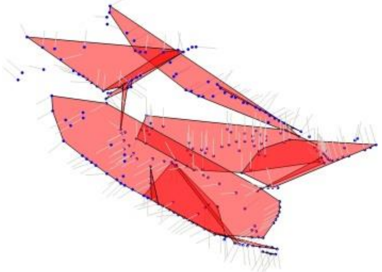


# Geometric Filtering and Grasp Computation

- 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

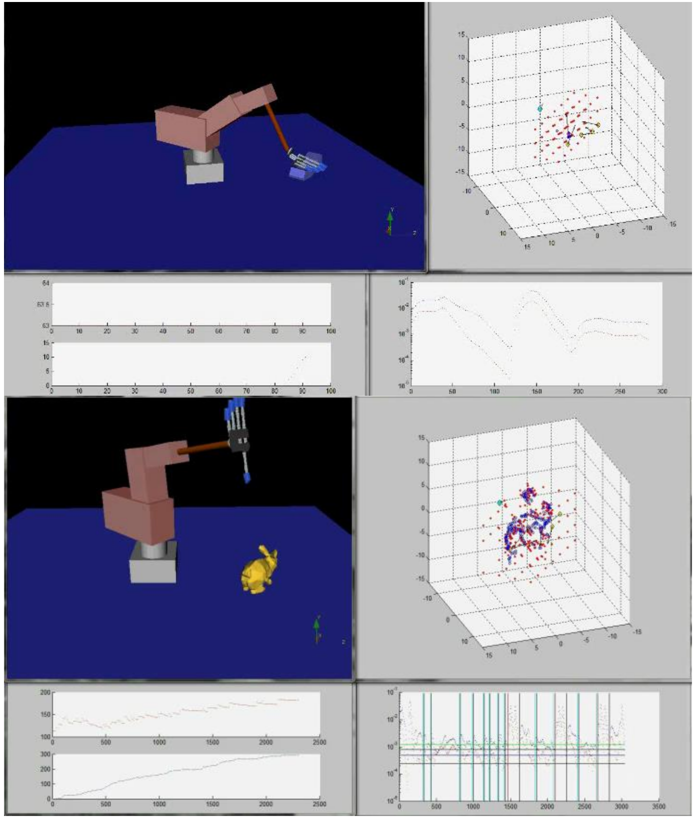
$$s(f_1, f_2) = \prod_{i=1}^{4} o_i(f_1, f_2)$$



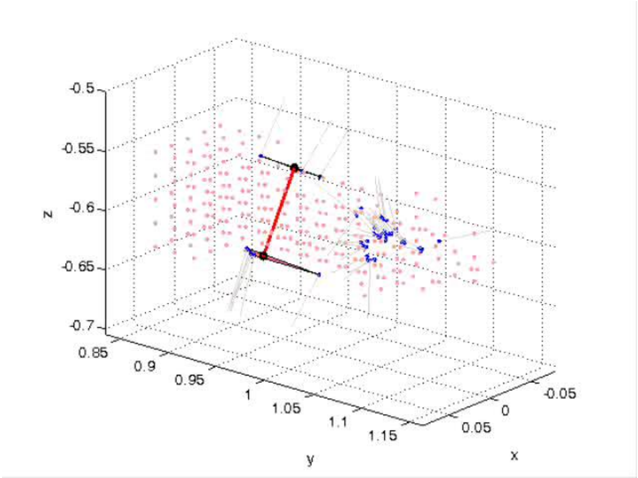
# Tactile Object Exploration

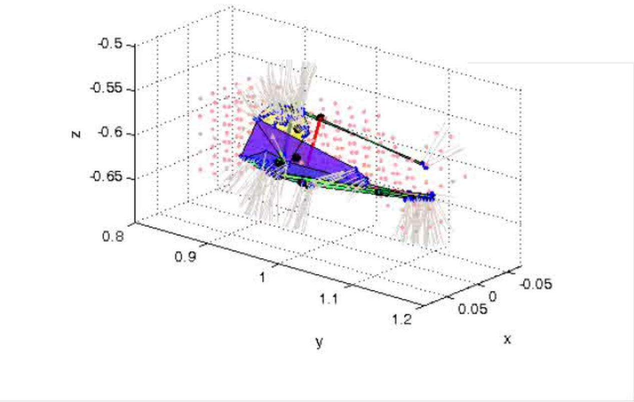
- Potential field approach to guide the robot hand along the object surface
- Oriented 3D point cloud from contact data
- Extract faces from 3D point cloud in a geometric feature filter pipeline
  - Parallelism
  - Minimum face size
  - Face distance
  - Mutual visibility

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

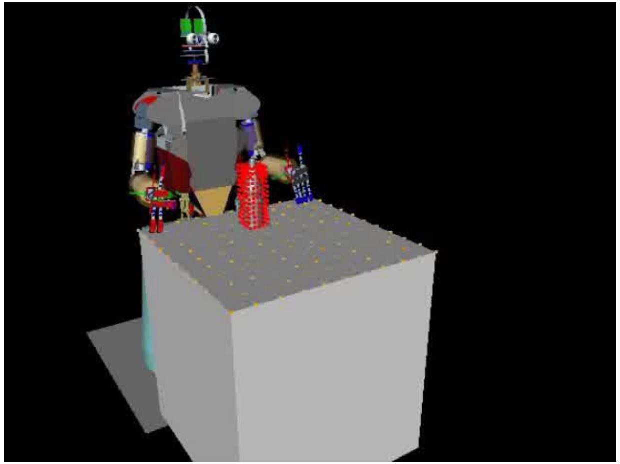


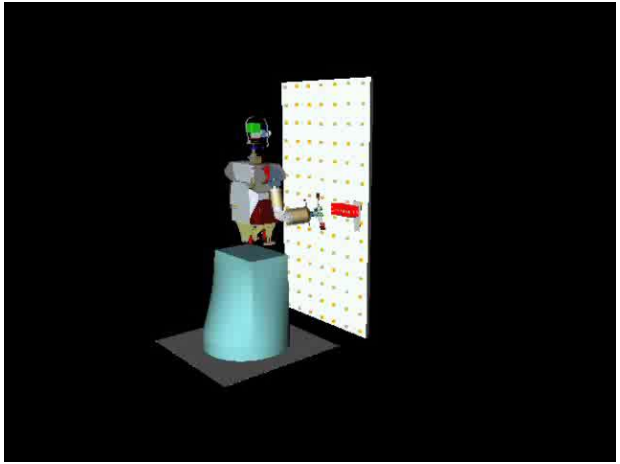
# Examples: Bottle



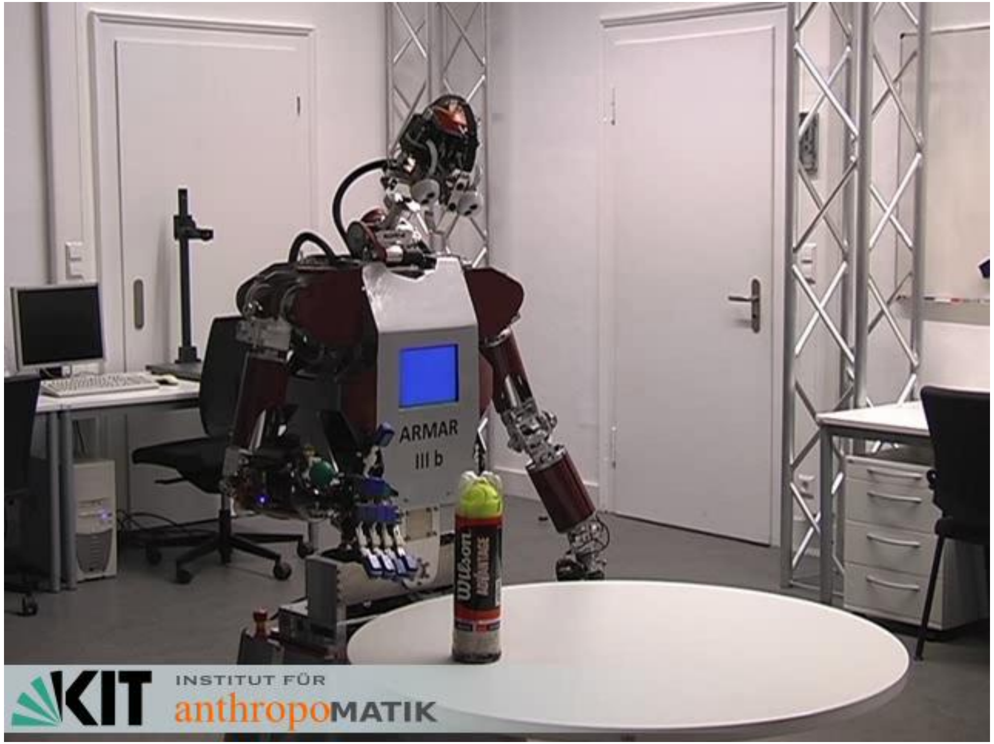
# Visually guided exploration on ARMAR





- 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

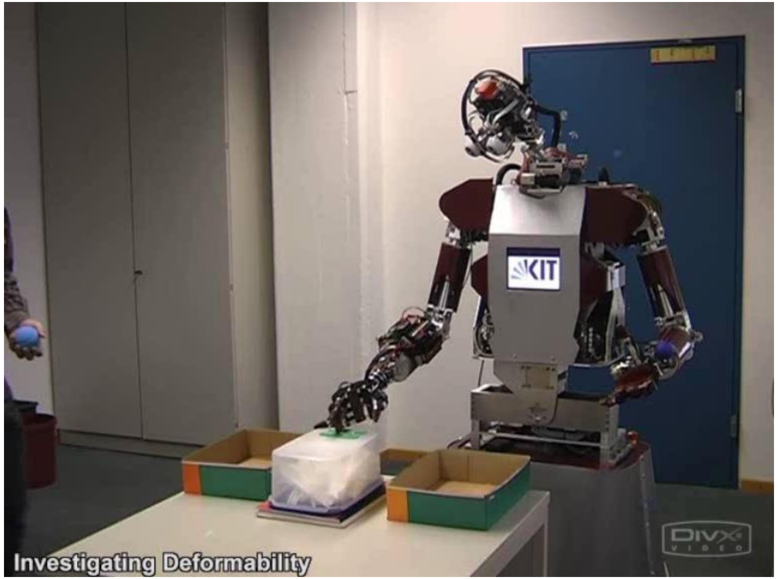


# Combining Vision and Haptics

# Visually-guided haptic exploration

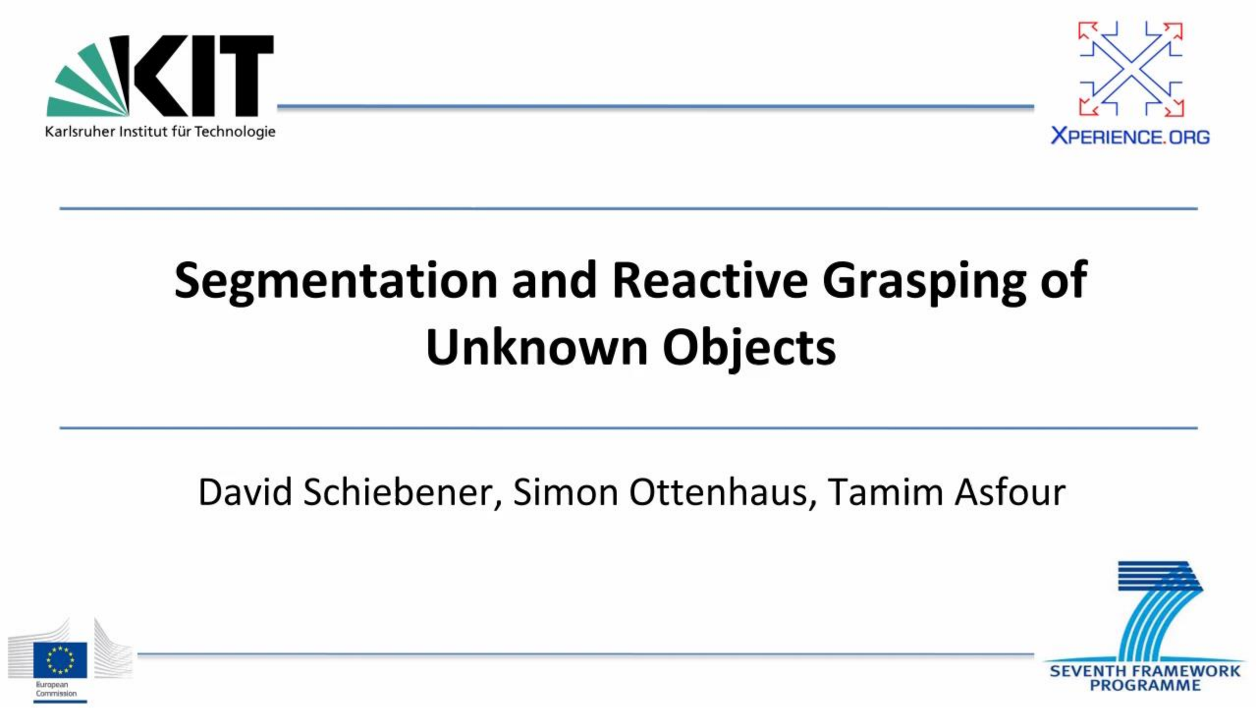
Fusion of tactile, proprioceptive and visual sensor data with a five-fingered hand

Verification of object size

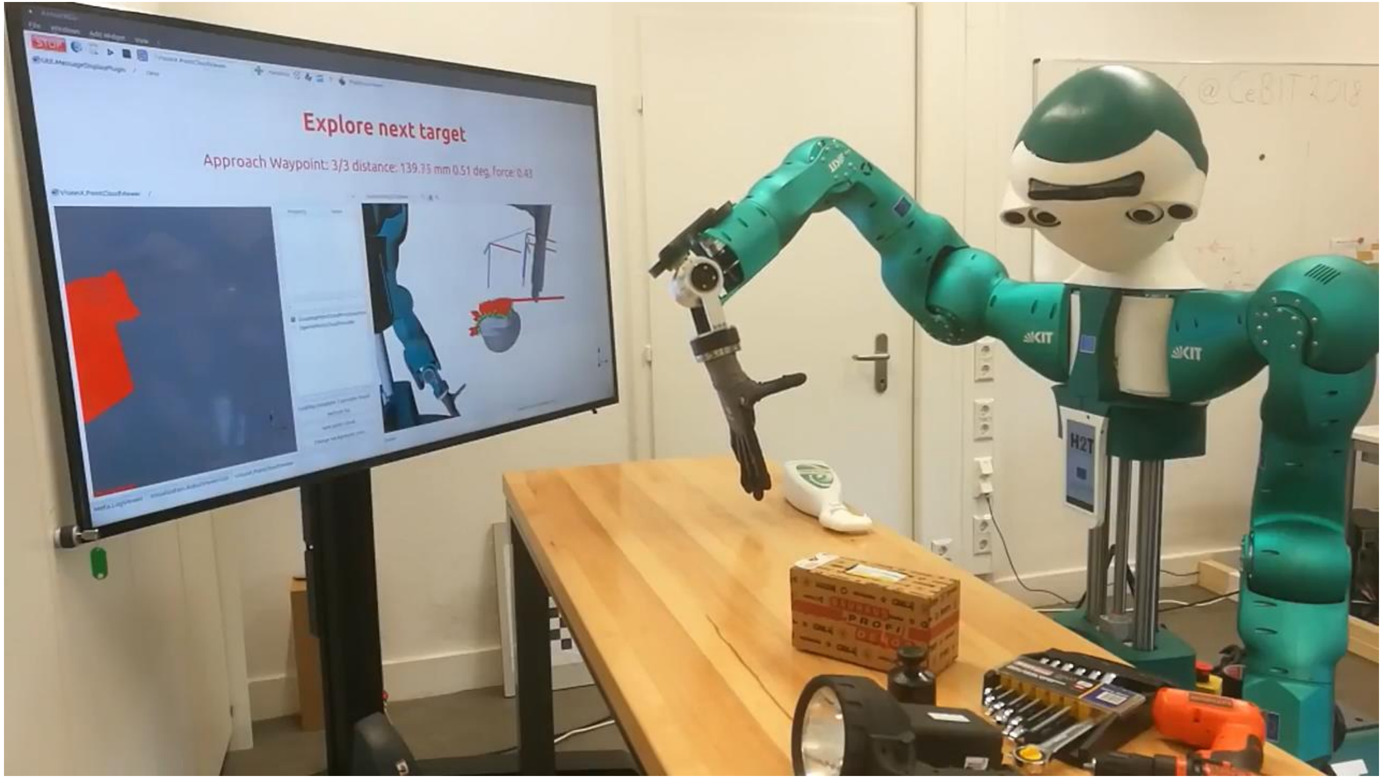



Verification of object deformability

# Combining vision and haptics for grasping



# Visuo-Haptic Grasping of unknown Objects



# Haptic Exploration for Grasping

How to maximize information gain during haptic exploration?

**Exploration procedure: Requirements**

1. Estimate object surface in a data efficient manner
   ➔ Data efficient surface model

2. Plan exploration actions efficiently
   ➔ Maximize information gain per cost

3. Gather as much information per contact as possible
   ➔ Maximize information gain per contact

# Data Efficient Surface Model

- Estimate a surface using **Gaussian Processes (GP)** as the implicit surface potential (ISP)

$$F(x) = k_*^T (K + \sigma^2 I)^{-1} y$$

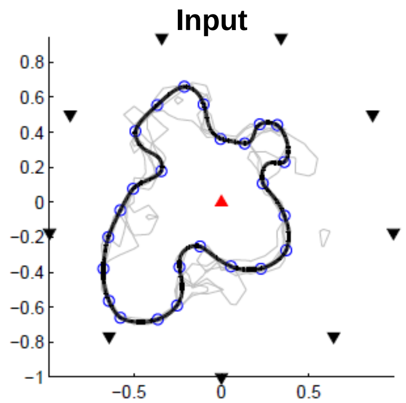
- **Implicit Shape Potential (ISP):** Samples on the surface (=0), inside (=1) and outside (=-1) of the object
- Estimated surface: 0-level set of the GP

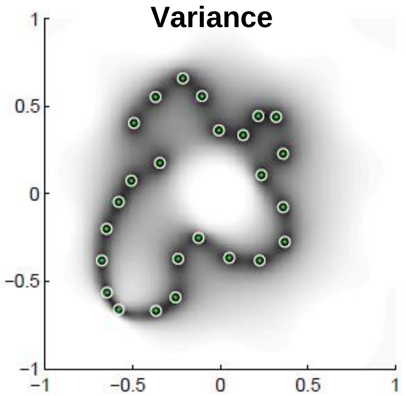
$$S = \{x \mid F(x) = 0\}$$

- The estimated variance of the GP can be used as an uncertainty measure of the surface

→ Gaussian Process Implicit Surfaces (GPIS)

[Williams et al. 2007]



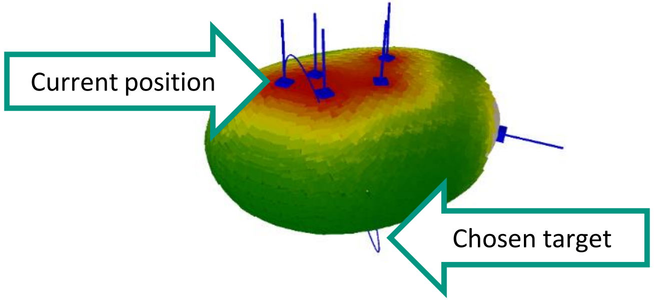



# Next-Best-Touch for Tactile Exploration

- What is the next best target for exploration?

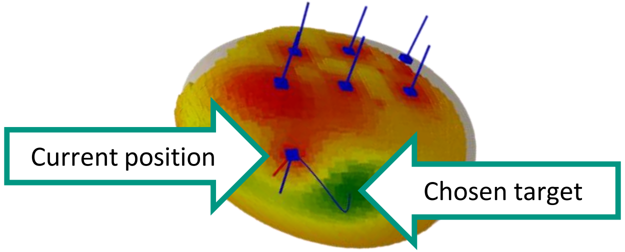
## State-of-the-art: Gaussian process variance (GP-V)

- Maximize information
- Ignore path cost




## Tactile exploration @ H²T

- Information Gain Estimation Function (IGEF)
- Minimize uncertainty
- Minimize path cost
- Maximize locality




Ottenhaus, S., Kaul, L., Vahrenkamp, N. and Asfour, T., *Active Tactile Exploration Based on Cost-Aware Information Gain Maximization*, International Journal of Humanoid Robotics, vol. 15, no. 1, pp. 1-21, 2018

# Plan Exploration Actions Efficiently

1. **Maximize Δ information:** Reduce uncertainty in previously unseen regions by exploring distant candidates

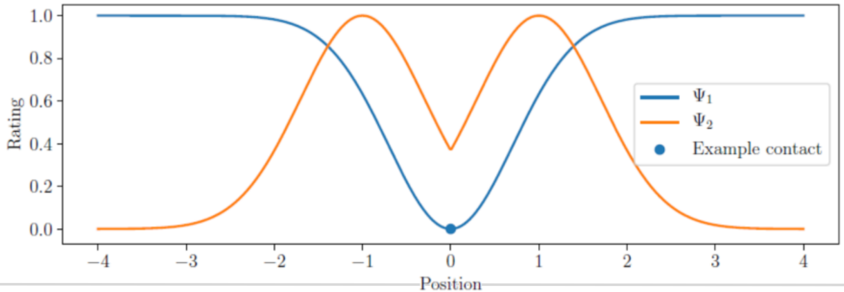
$$\boldsymbol{\Psi_1(x) = \min_{c \in C}\left(1 - \exp\left(-\frac{\|x - c\|^2}{\sigma^2}\right)\right)}$$

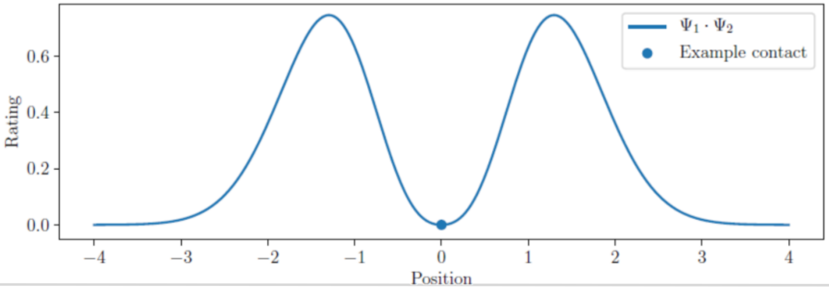
**2. Stay local:** Prefer targets, that are close (Gaussian kernel)

$$\boldsymbol{\Psi_2(x) = \exp\left(-\frac{(\|x - c\| - \mu)^2}{\sigma^2}\right)}$$

| Symbol | Description |
|---|---|
| $x$ ; $x_n$ | Query position; normal |
| $r$ ; $r_n$ | Current position; normal |
| $c \in C$ | Set of explored points |
| $\sigma, \mu$ | Scaling factors |

**Example:** One contact point at 0; $\sigma$ and $\mu$ are set to 1



# Plan Exploration Actions Efficiently

**3. Minimize path cost:** Minimize movement and rotation of the hand

$$\boldsymbol{\Psi_{3,pos}(x) = \frac{1}{\|Path(r,x)\|}} \qquad \boldsymbol{\Psi_{3,rot}(x) = exp\left(-\frac{sin^2\left(\frac{1}{2}arccos(r_n \cdot x_n)\right)}{\sigma^2}\right)}$$

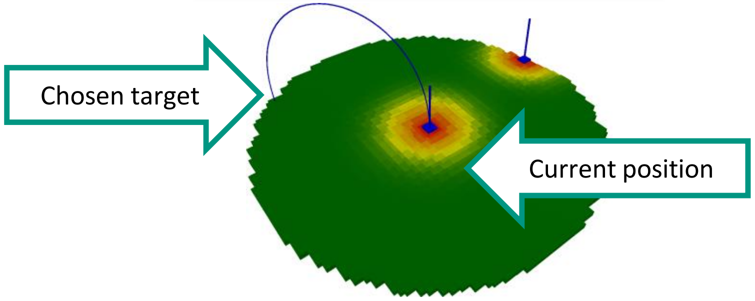
## Resulting Information Gain Estimation Function (IGEF)

IGEF: $\Psi = \Psi_1 \cdot \Psi_2 \cdot \Psi_{3,pos} \cdot \Psi_{3,rot}$

# Next-Best-Touch for Tactile Exploration

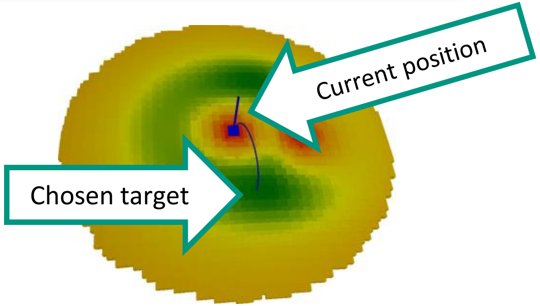
- Consider different goals during exploration
  - **Uncertainty:** Explore unknown regions
  - **Cost:** Minimize path cost in distance and rotation
  - **Locality:** Prefer exploration targets that are in proximity of explored regions

$$Maximize(\Delta Information)$$




State of the art
Gaussian Process Variance

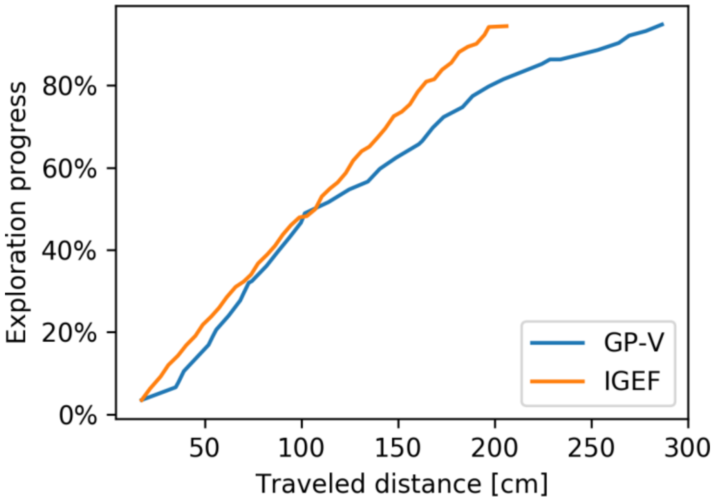
$$Maximize \frac{\Delta Information}{\Delta Cost}$$

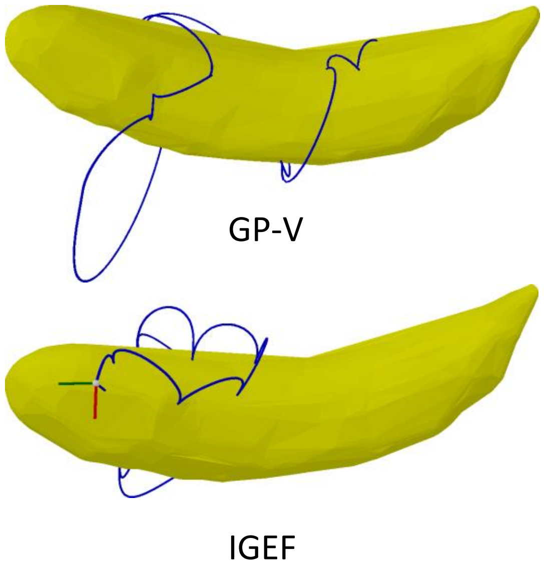



Proposed method (H2T)
Information Gain Estimation Function

# Next-Best-Touch: Evaluation (I)

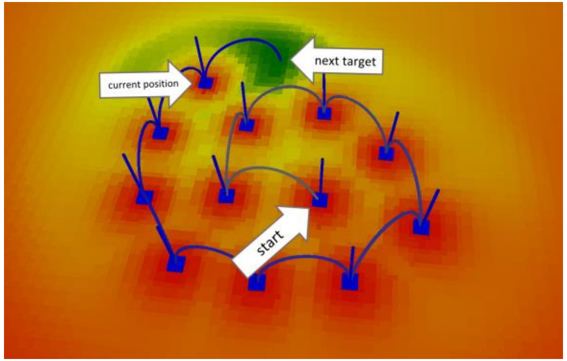
- Exemplary exploration path for the YCB banana
- Path generation using Bezier curves
- GP-V is greedy ➔ large steps
- IGEF stays local ➔ smaller steps

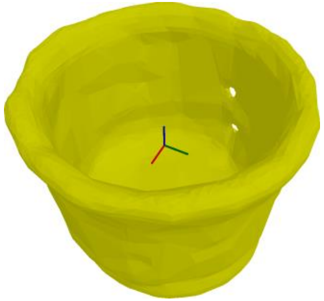


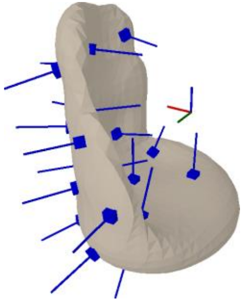



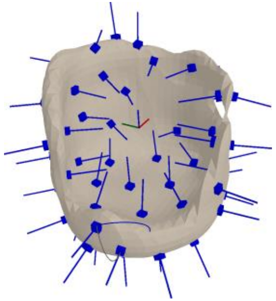
# Next-Best-Touch: Evaluation (II)

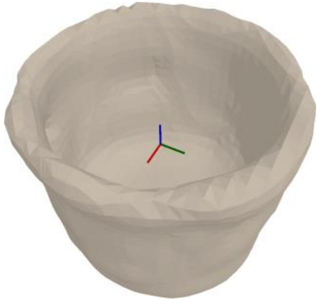
- Combined goals lead to a systematic exploration of the object
- Spiral exploration pattern on a plane
- Also for complicated surfaces





Ground truth mesh


27 oriented contacts

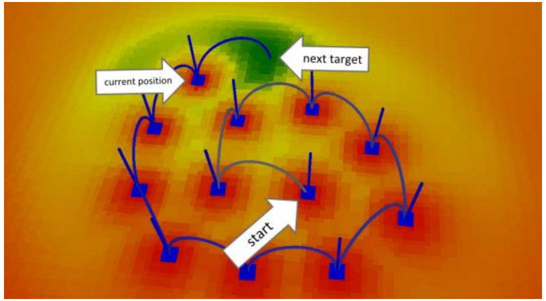

73 oriented contacts

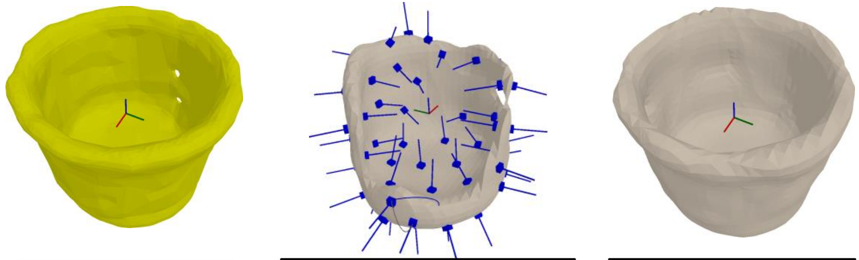

117 oriented contacts

# Next-Best-Touch: Evaluation (III)

- Comparison against state-of-the-art (GP-V)
- 109 objects ( KIT and YCB object set)

Spiral exploration pattern



Ground truth | Partially explored | Fully explored

| Metric (average) | SotA | Proposed Method | Improvement |
|---|---|---|---|
| Traveled distance | 246 cm | 94 cm | **- 62%** |
| Reconstruction RMSE | 0.77 mm | 0.59 mm | **- 17%** |

Ottenhaus, S., Kaul, L., Vahrenkamp, N. and Asfour, T., *Active Tactile Exploration Based on Cost-Aware Information Gain Maximization*, International Journal of Humanoid Robotics (IJHR), vol. 15, no. 1, pp. 1-21, 2018

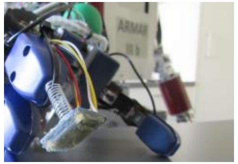
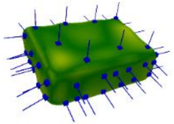
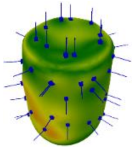
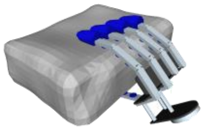
# Haptic Exploration for Grasping

## Exploration procedure: Summary

- Data efficient surface model → GPIS
- Maximize information gain per cost → Proposed procedure
- Maximize information gain per contact → Position & normal

Object → Tactile exploration → GPIS model → Grasp planning




→ Grasp synthesis possible

→ **Are all these exploration actions needed for grasping?**

# How many touches are necessary for grasping?

**Test set: unseen objects**

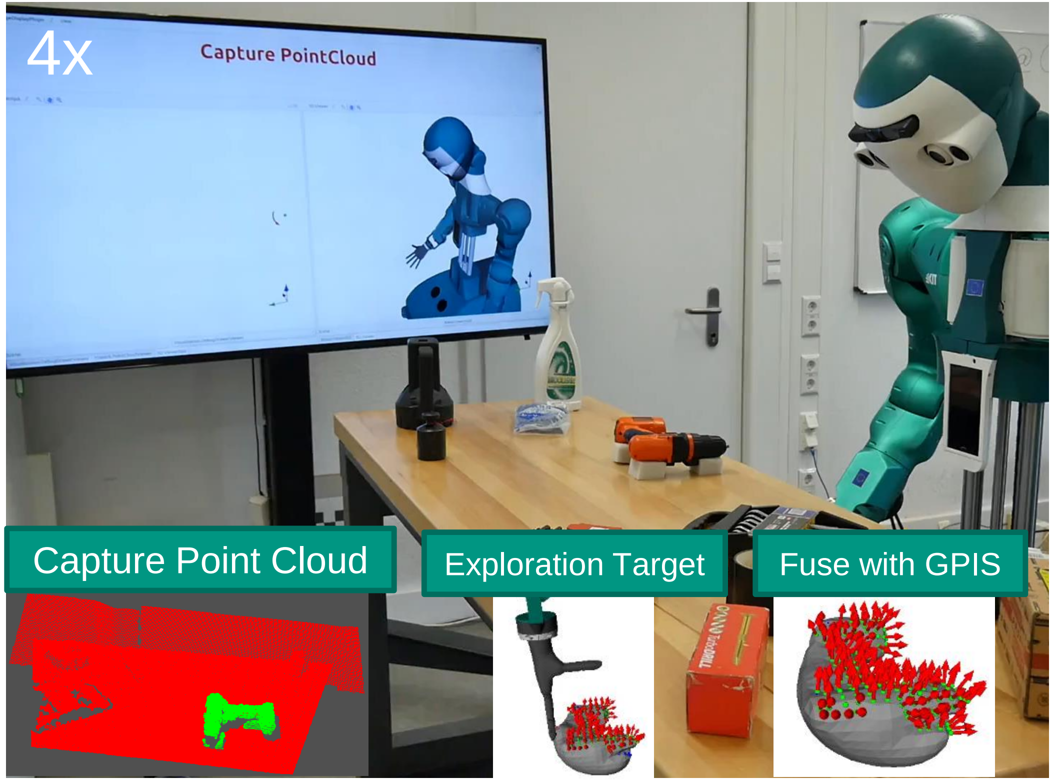

Exploration takes time: How many touches are necessary for **grasping**?

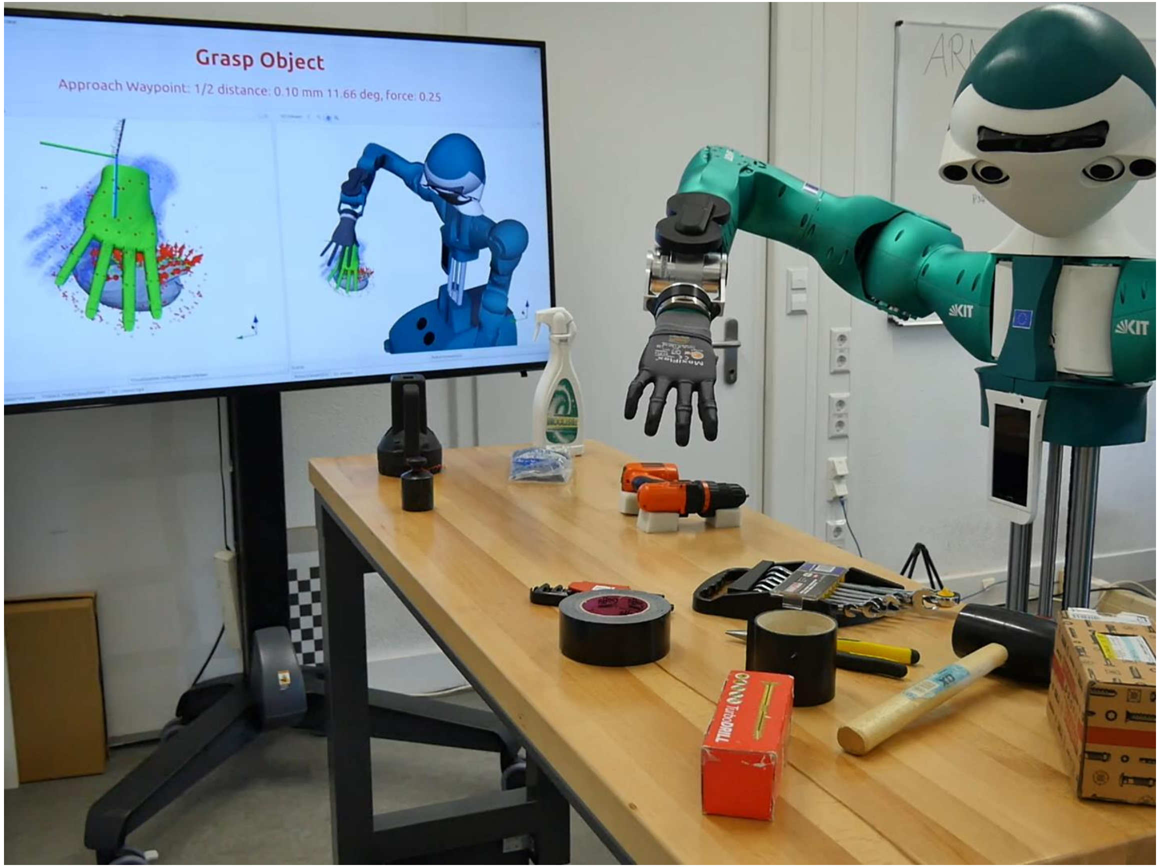
**Exploration evaluation**

| #Touches | 0 | **1** | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Success | 89% | **94%** | 94% | 94% | 95% | 95% |

- → **First touch yields most information**
- → **One touch for robot experiment**
- → **All objects lifted successfully**

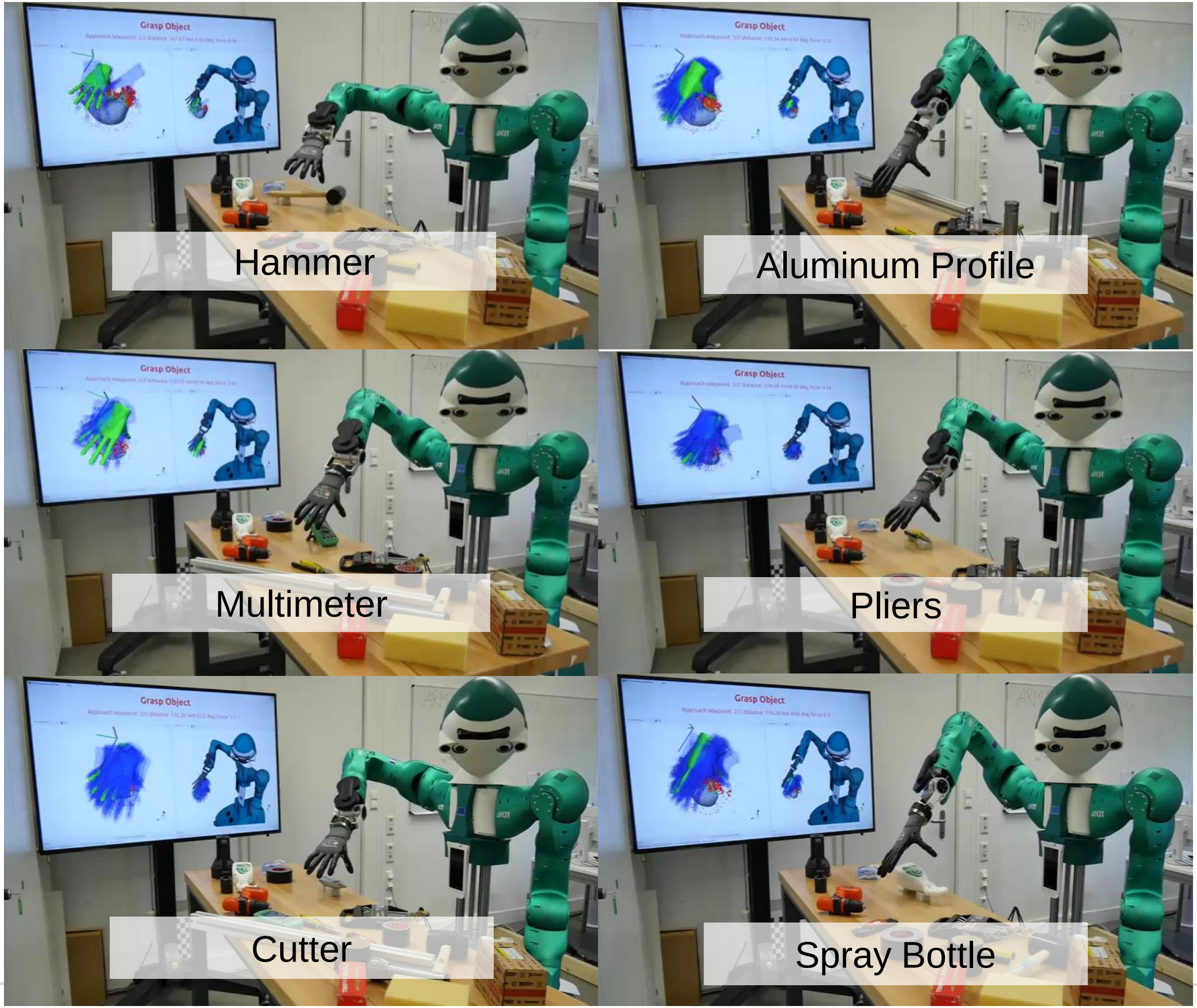
# Tactile Exploration



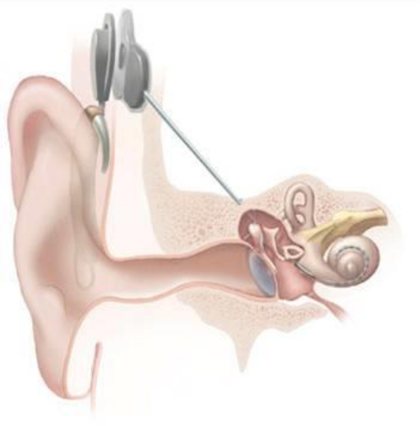



Lift Object

FT Sensor active


Hammer
Aluminum Profile
Multimeter
Pliers
Cutter
Spray Bottle

# Active Hearing



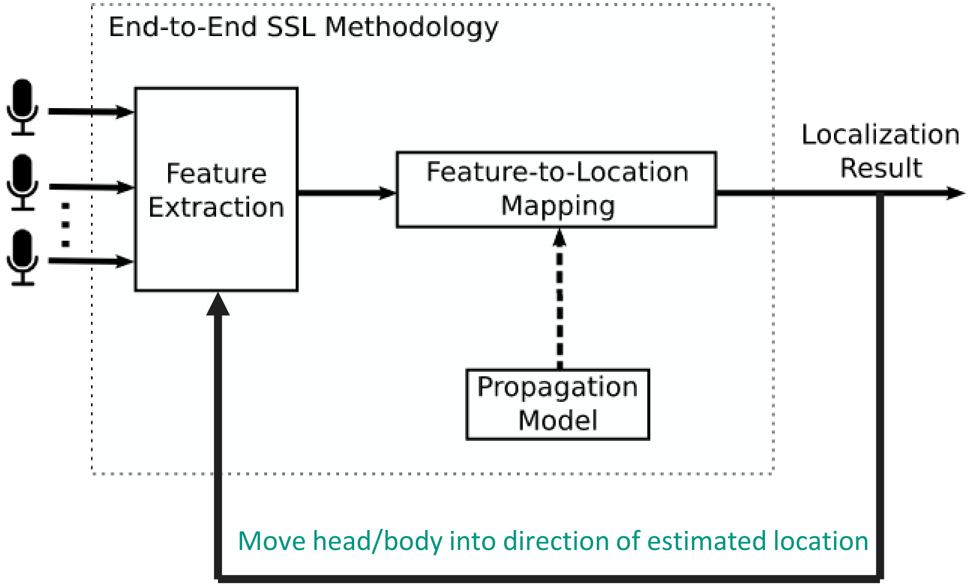
**How to move the body to improve the quality of the perceived sound?**

# Fundamental questions in active hearing

Karlsruher Institut für Technologie

- What are the spatial filtering properties of the recording system
- How does the environment transform the signal (e.g. pulse responses)
- Does the robot produce self-noise?

# Approach using Sound source localization (modified)




C. Rascon, I. Meza, *Localization of sound sources in robotics: A review*, Robotics and Autonomous Systems 96 (2017), pp. 184–210

# Further research topics

- Auditory Scene analysis
  - How to identify / separate different sound sources
  - E.g. Talking to other persons in a noisy room
- Sound Understanding
  - Use knowledge of perceived signal (e.g. human voice, phone bell, door knocking…)
- Sound Reasoning
  - Use the sound signal to reason about the environment or the sound source (e.g. echo)

Huang, J. and Supaongprapa, T. and Terakura, I and Wang, F. and Ohnishi, N. and Sugie, N., "A Model Based Sound Localization System and Its Application to Robot Navigation", *Robotics and Autonomous Systems (Elsevier Science)*, Vol.27, No.4, pp.199-209, 1999.

Huang, J. and Ohnishi, N. and Sugie, N., "Building Ears for Robots: Sound Localization and Separation", *Artificial Life and Robotics (Springer-Verlag)*, Vol.1, No.4, pp.157-163, 1997.