Going beyond vision: multisensory integration for perception and action

Heinrich H. Bülthoff

Cognitive & Computational Psychophysics
Max Planck Institute for Biological Cybernetics

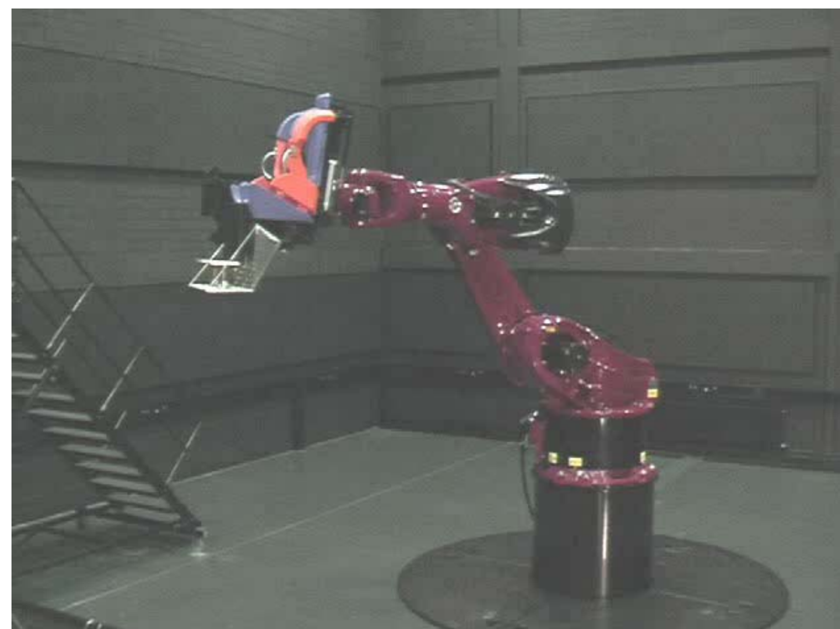
Overview

- The question of how the **human brain "makes sense"** of the sensory input it receives has been at the heart of cognitive and neuroscience research for the last decades.
- One of the most fundamental perceptual processes is **categorization** – the ability to compartmentalize knowledge for efficient retrieval.
- Recent advances in **computer graphics and computer vision** have made it possible to both produce highly realistic stimulus material for **controlled experiments in life-like environments** as well as to enable highly **detailed analyses of the physical properties** of real-world stimuli.

Research Philosophy

- Study perception and action with stimuli as close as possible to the real world, using
  - Computer Graphics
    to generate natural but well controlled stimuli of objects and scenes

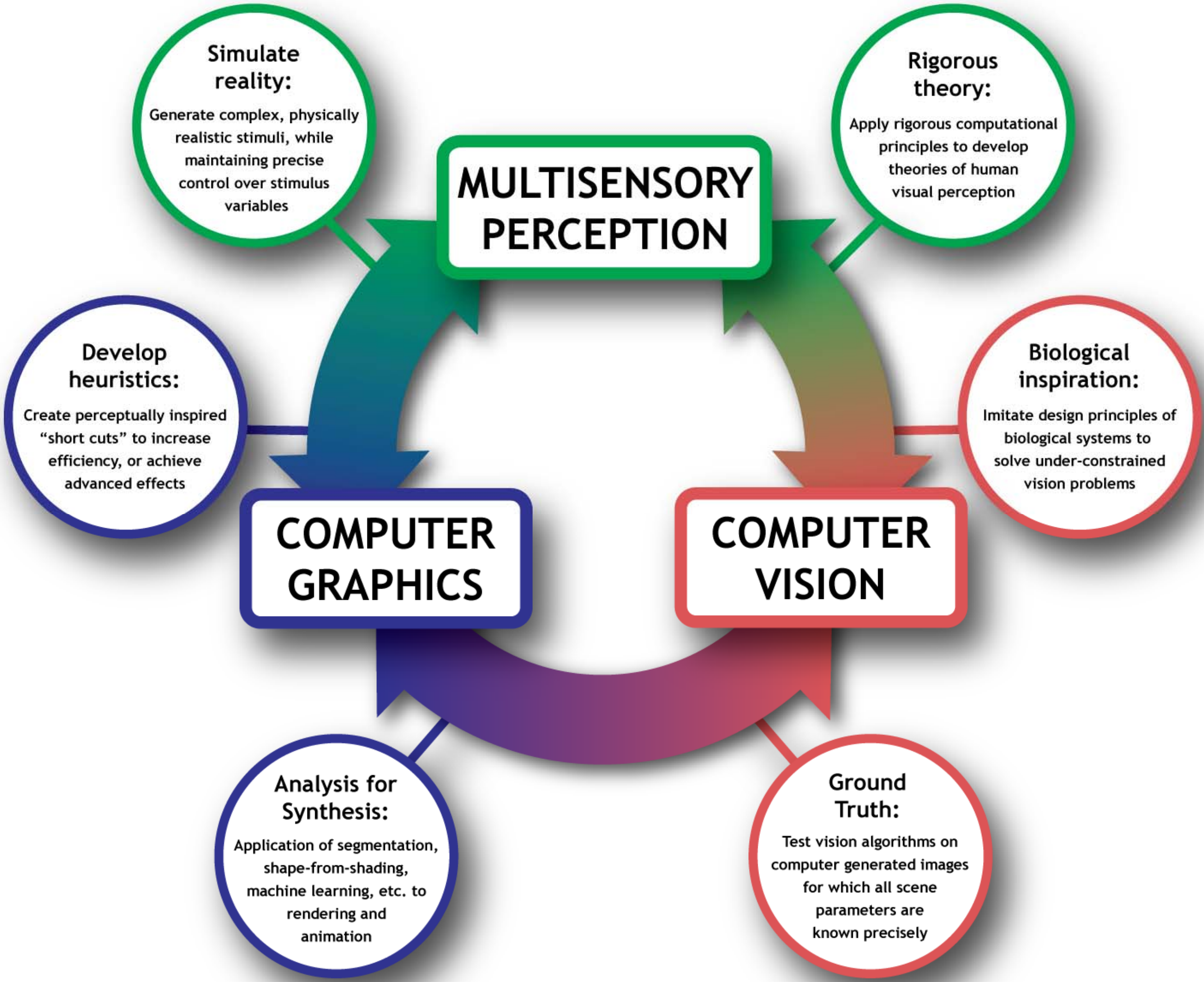
  - Virtual Reality
    - www.cyberneum.de
    - motion simulators
    - haptic simulators
    - walking simulators
    - immersive environments
    - panoramic projections
    - EU-projects: JAST, BACS, CyberWalk, Immersense, Wayfinding



Overview

- In this talk, we will review some of the **key challenges** in understanding categorization from a combined cognitive and computational perspective:
  - the need for spatio-temporal representations
  - perception of material properties
  - multi-modal/multi-sensory aspects of object categorization
  - coupling of perception and action

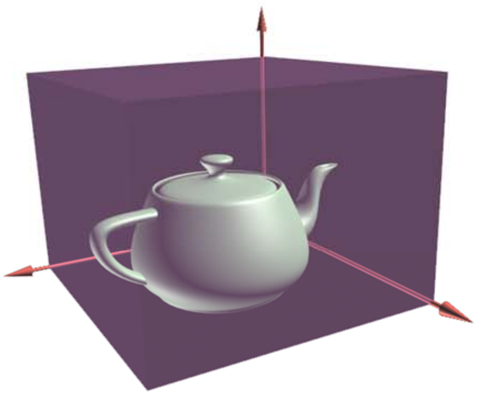
Research Paradigm

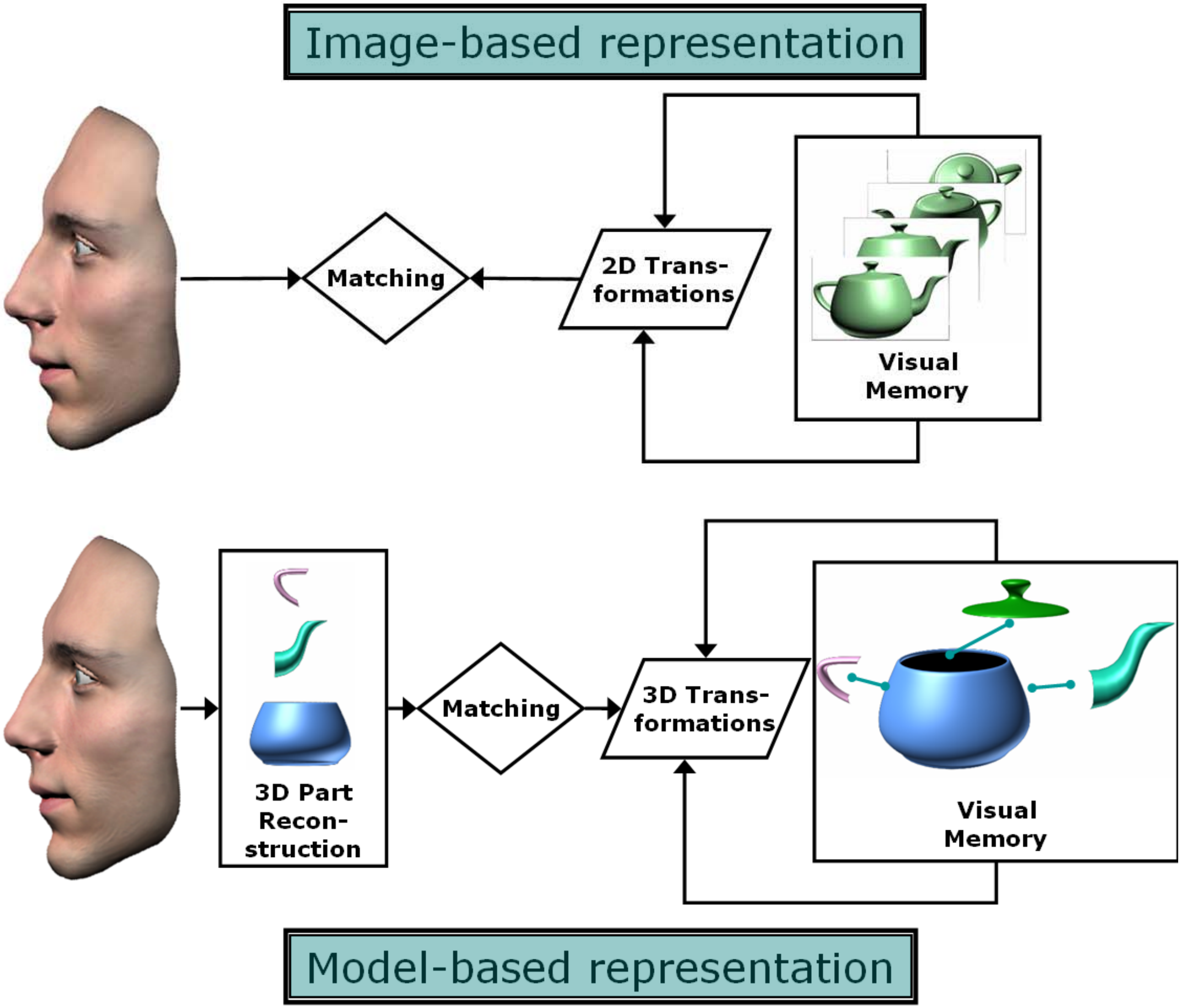


Overview

- The talk will focus on issues that so far have only started to be addressed but that are crucial for a deeper understanding of perceptual processes:
  - **the need for spatio-temporal representations**
  - perception of material properties
  - multi-modal/multi-sensory aspects of object categorization
  - coupling of perception and action

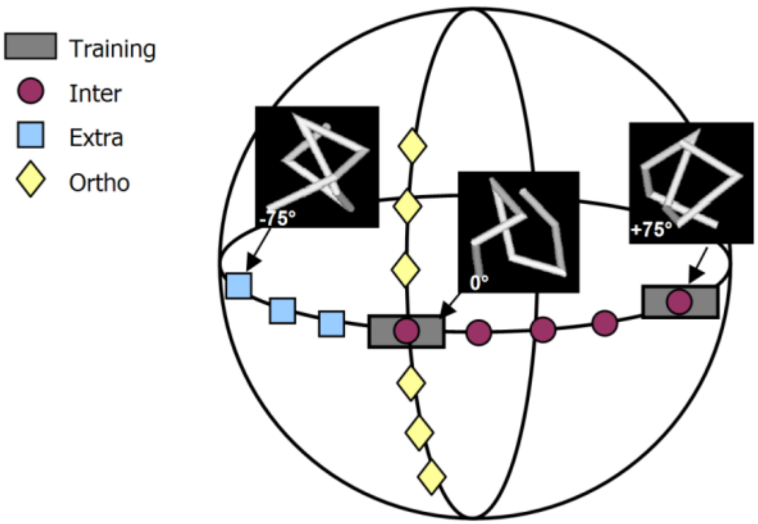
Representing objects: two models



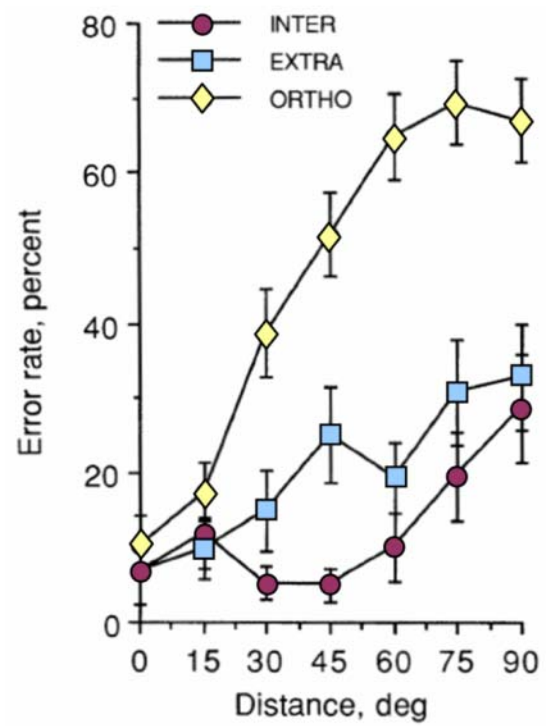


Representing objects: image-based recognition
Bülthoff and Edelman [*PNAS*, 1992]

- Recognition of novel objects depends on the viewing conditions (→ image-based recognition)



Design

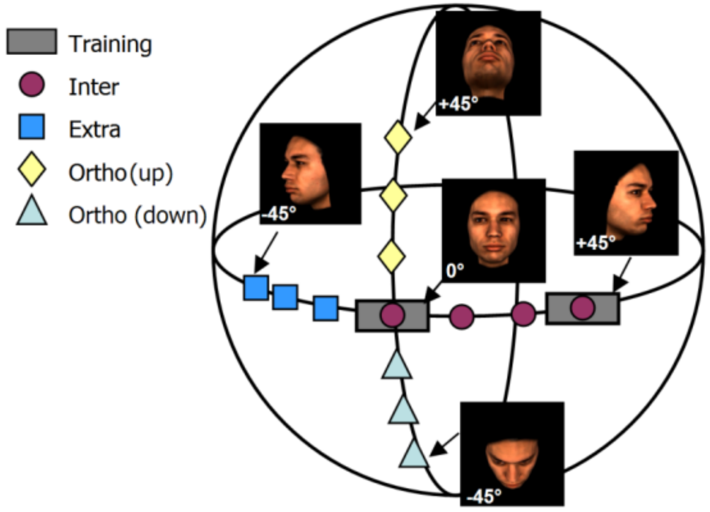


Inter > Extra > Ortho

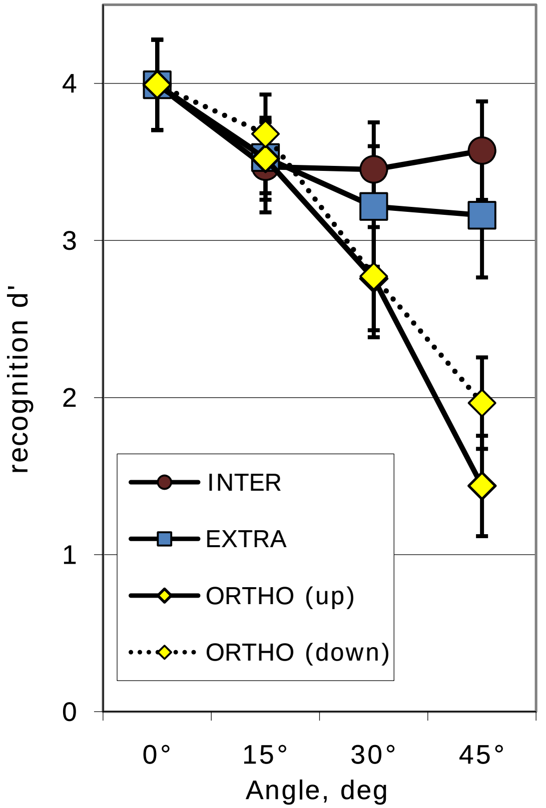
Representing faces: image-based recognition

Wallraven, Schwaninger, Schumacher, Bülthoff [*BMCV*, 2002]

- Recognition of novel **and familiar** objects depends on the viewing conditions (→ image-based recognition)



Design



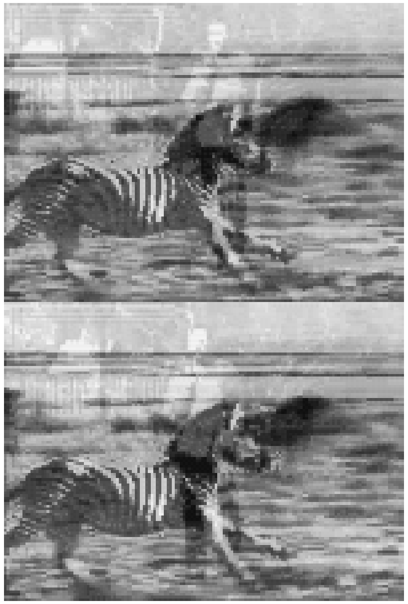
Inter > Extra > Ortho

The role of motion in recognition

1. Familiar motion facilitates person identification

Pilz, Vuong, Bülthoff, Thornton [*JEP: HPP*, subm]

2. Motion facilitates human target detection



Vuong, Hof, Bülthoff, Thornton [*Journal of Vision,* 2006]

3. Non-rigid motion is encoded as identity cue

Chuang, Vuong, Thornton, Bülthoff [*Visual Cognition,* 2006]

Quick summary (Spatio-temporal representations)

- Objects and faces are represented in an image-based fashion
- The temporal properties of objects play an important role during learning and recognition
- Object representations are spatio-temporal

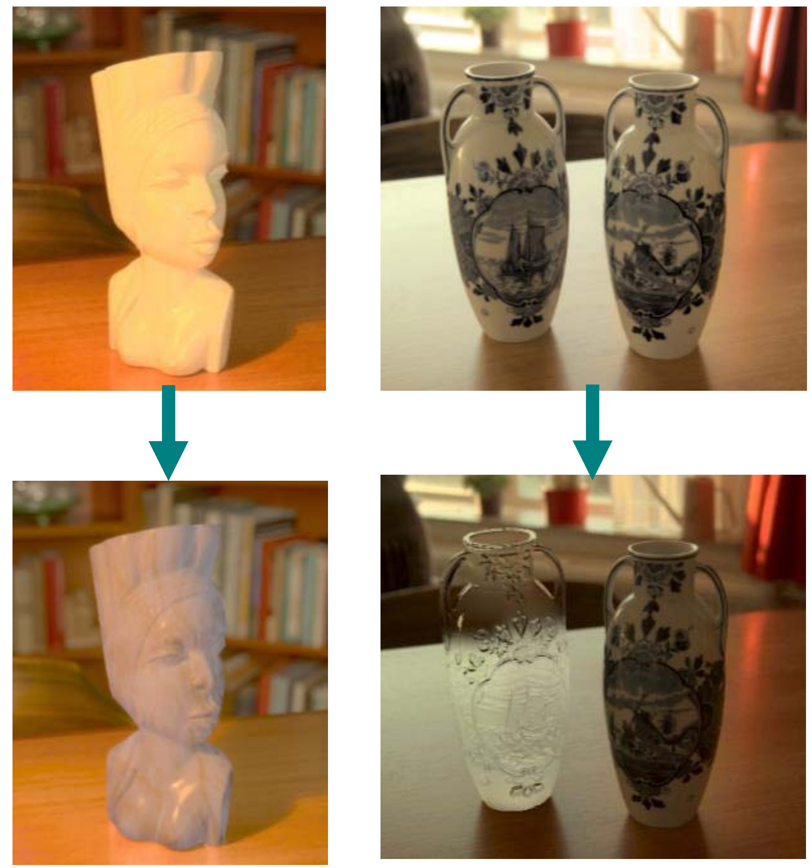
Overview

- In this talk, we will review some of the **key challenges** in understanding categorization from a combined cognitive and computational perspective:
  - the need for spatio-temporal representations
  - **perception of material properties**
  - multi-modal/multi-sensory aspects of object categorization
  - coupling of perception and action

Image-based material editing

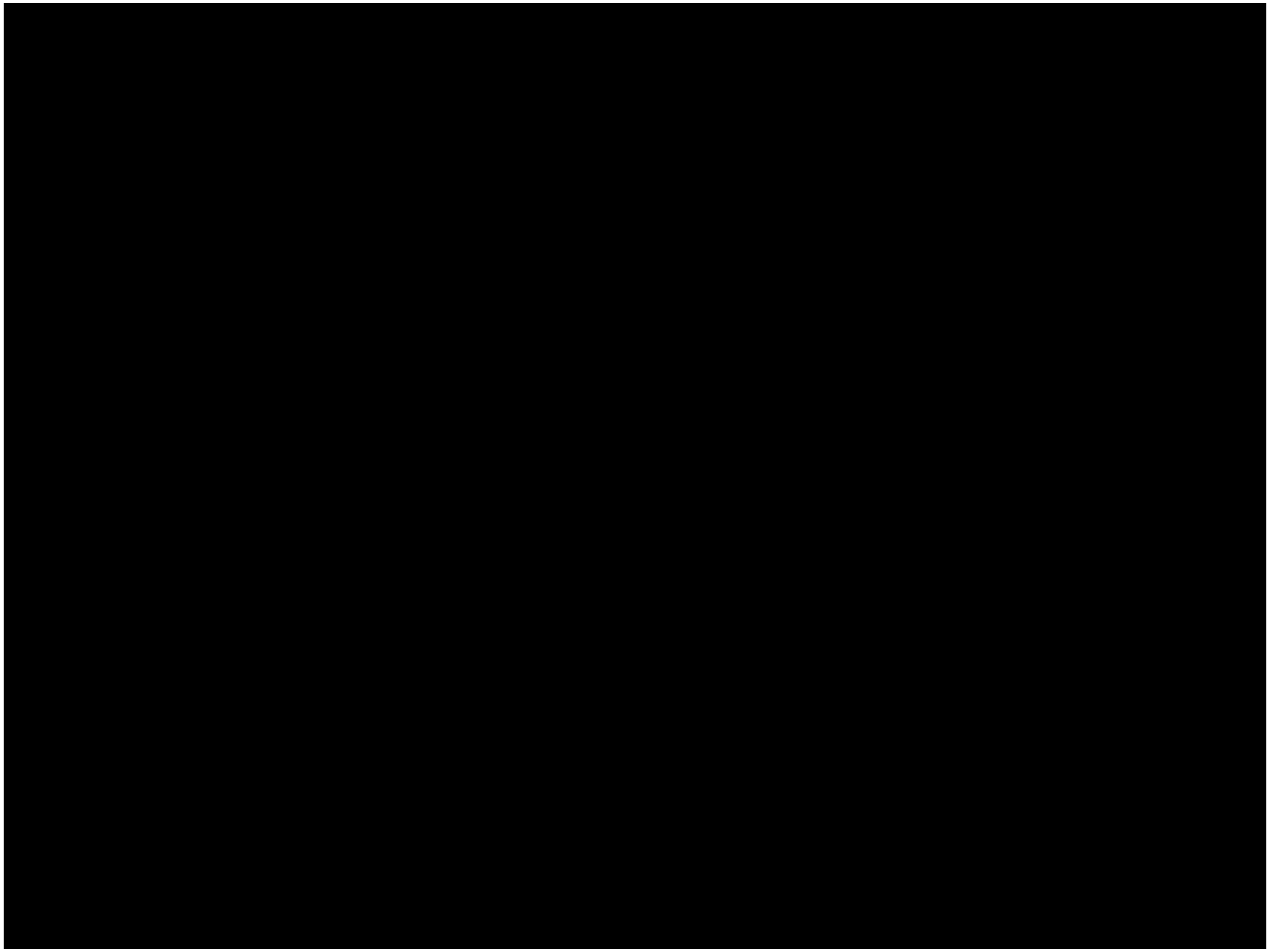
Kahn, Reinhard, Fleming, Bülthoff [*SIGGRAPH*, 2006]

- Goals:
  - How do humans perceive materials?
    - Ill-posed problem
  - Can we exploit perceptual tricks to change materials in a photograph (without a 3D-model)?
- Methods:
  - Crude 3D shape reconstruction using bilateral filter (dark means deep – SFS)
    - Exploits **generic viewpoint assumption** as an image is consistent with many 3D models
  - Simple background-inpainting for transparency
    - Exploits **masking**
    - **weak model of refraction**
- Results:
  - Re-texturing
  - Medium gloss to matte or glossy
  - Opaque to transparent or translucent



re-textured **transparency**

Image-based material editing
Kahn, Reinhard, Fleming, Bülthoff [*SIGGRAPH,* 2006]



Quick summary (Material Perception)

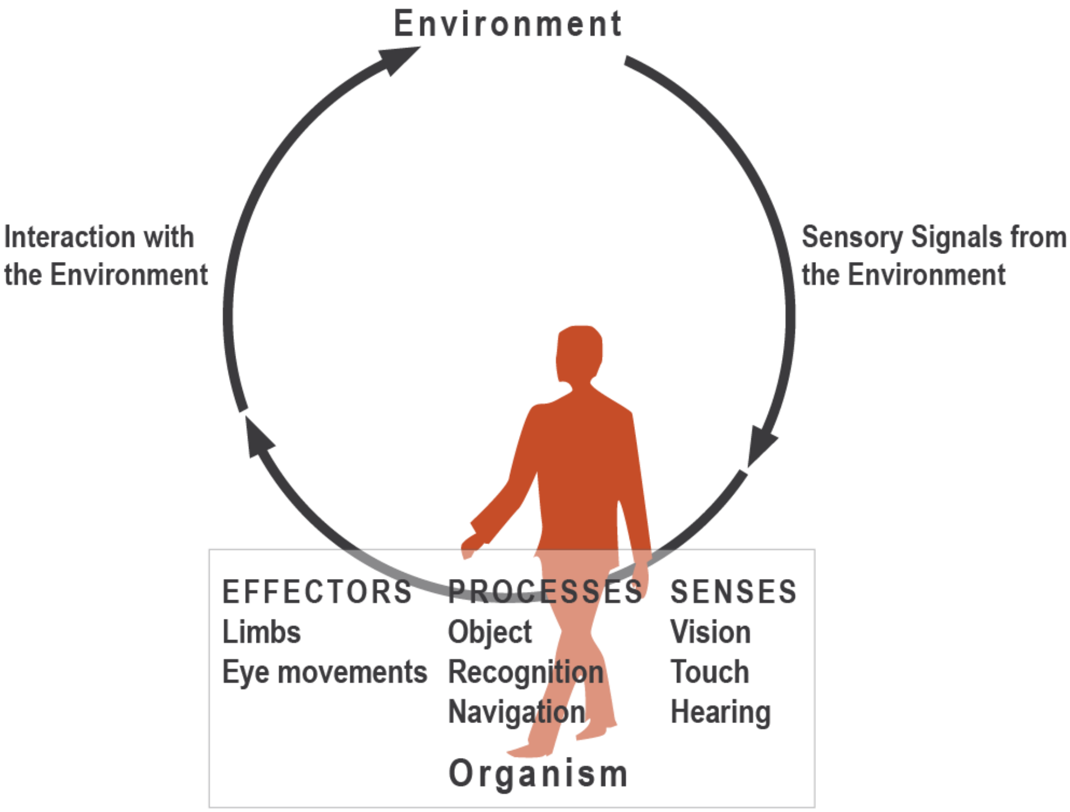
- The brain does **not** use an inverse physics approach to perception
- Rather, the brain uses (complex) heuristics to estimate
  - Material properties
  - Shape
- By exploiting these heuristics one can create simple, but effective work-arounds to control these properties

Overview

- In this talk, we will review some of the **key challenges** in understanding categorization from a combined cognitive and computational perspective:
  - the need for spatio-temporal representations
  - perception of material properties
  - **multi-modal/multi-sensory aspects of object categorization**
  - coupling of perception and action

Sensory integration

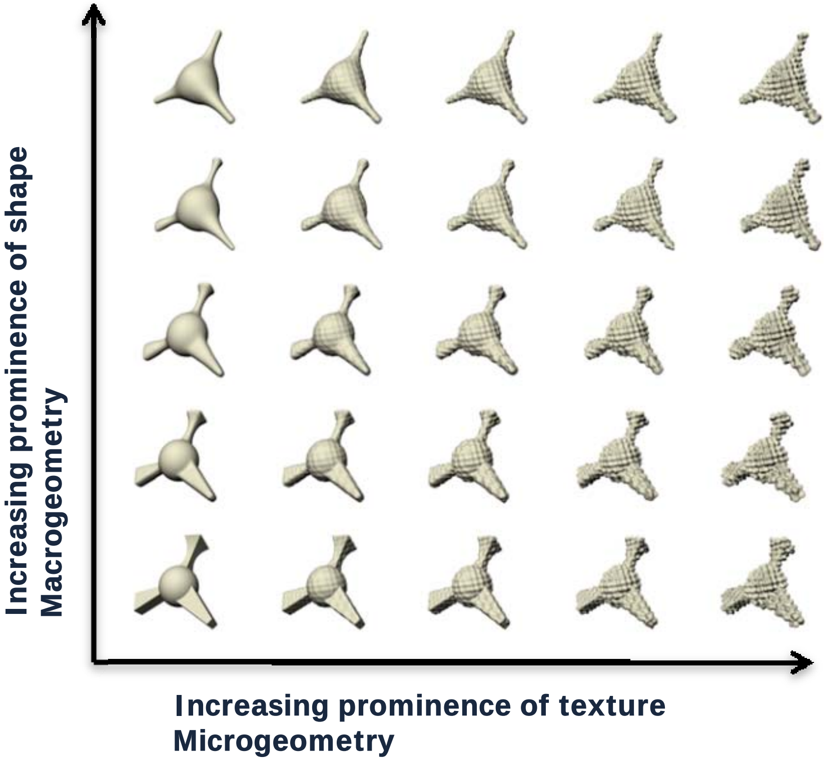
- Humans act upon objects in order to interact with the world.
- The following studies addressed the questions to what degree object representations are multi-modal.

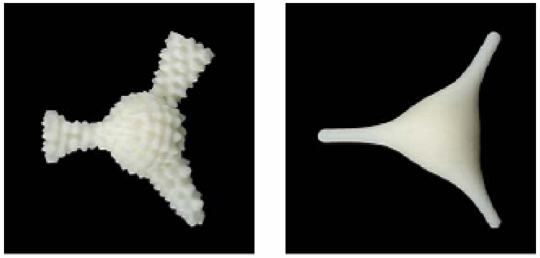


Multi-modal similarity and categorization of novel, 3D objects

Cooke, Jäkel, Wallraven, Bülthoff [*Neuropsychologia*, 2007]

- Goal:
  - Develop framework for understanding multi-sensory (visuo-haptic) object perception
- Methods:
  - Controlled space of visuo-haptic stimuli printed in 3D
  - Multi-Dimensional-Scaling for finding perceptual space for haptic, visual and bimodal exploration

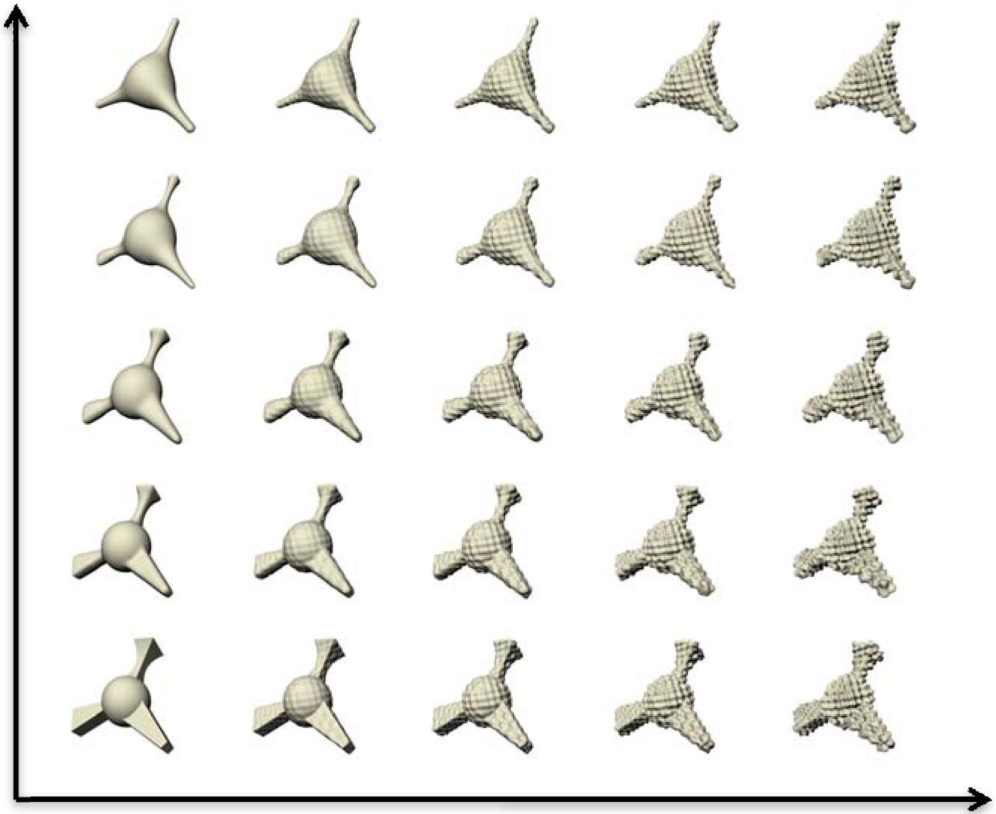




Photographs of printed 3D objects

The tools: Parametrically-defined stimuli & 3D printer

Cooke, Jäkel, Wallraven, Bülthoff [*Neuropsychologia*, 2007]



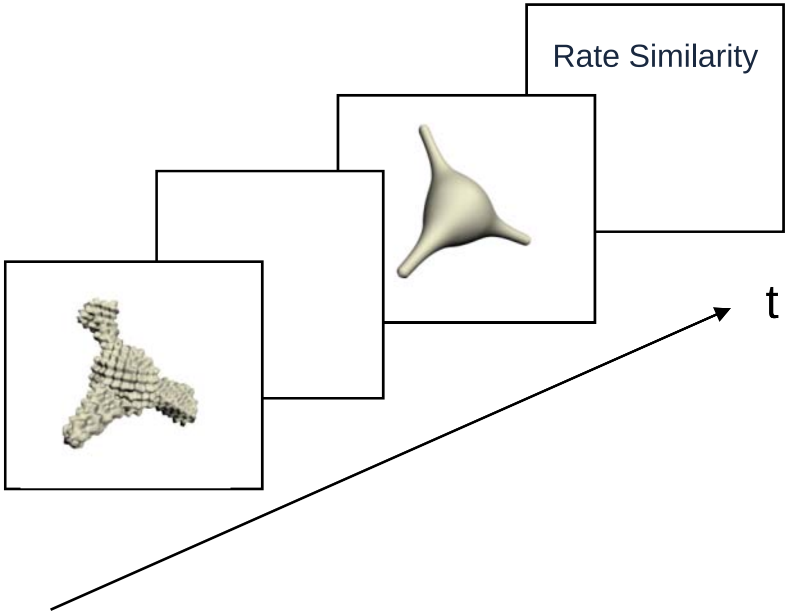
3D printer

Printed object

The experiment: Multi-sensory similarity

Cooke, Jäkel, Wallraven, Bülthoff [*Neuropsychologia*, 2007]

Experimental Setup

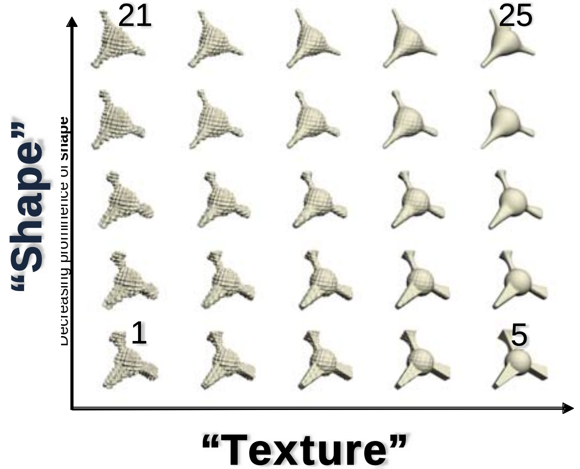


Trial Sequence

- 10 subjects x 3 conditions: Visual (V), Haptic (H), Visuohaptic (VH)
- Task : Similarity ratings

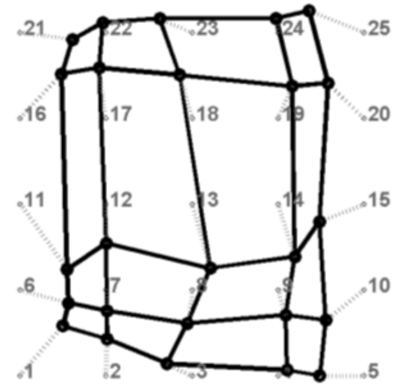
Results: Modality Effects

Cooke, Jäkel, Wallraven, Bülthoff [*Neuropsychologia*, 2007]



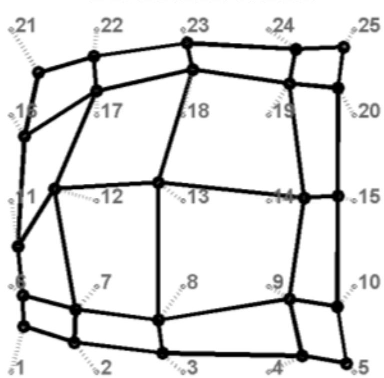
Visual

2D Stress 0.157



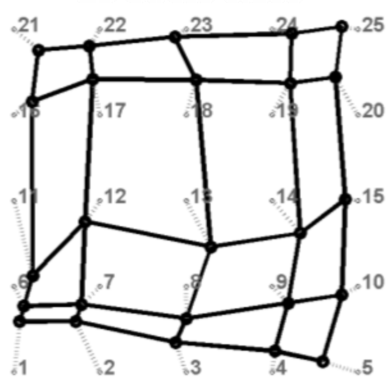
Haptic

2D Stress 0.168



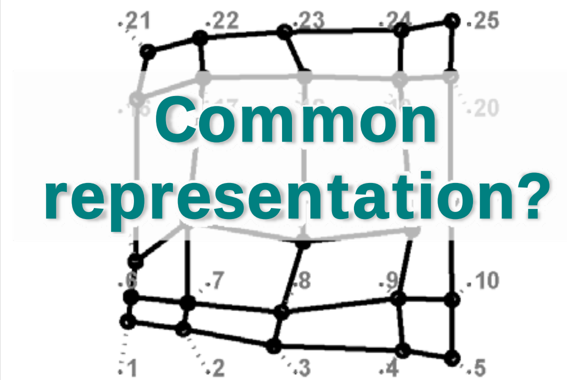
Visuohaptic

2D Stress 0.160

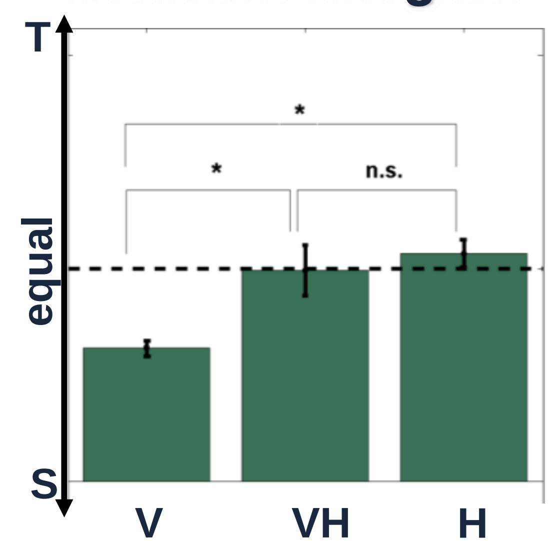


All Data

2D Stress 0.167



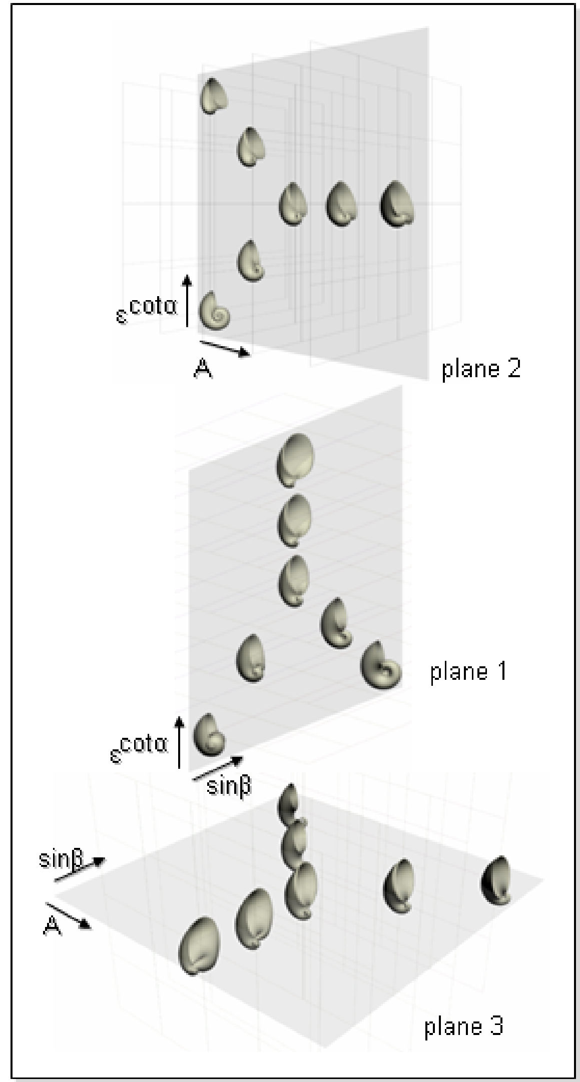
Relative Weights



Multi-modal similarity and categorization of novel, 3D objects

Gaißert, Wallraven, Bülthoff (2007,2008)

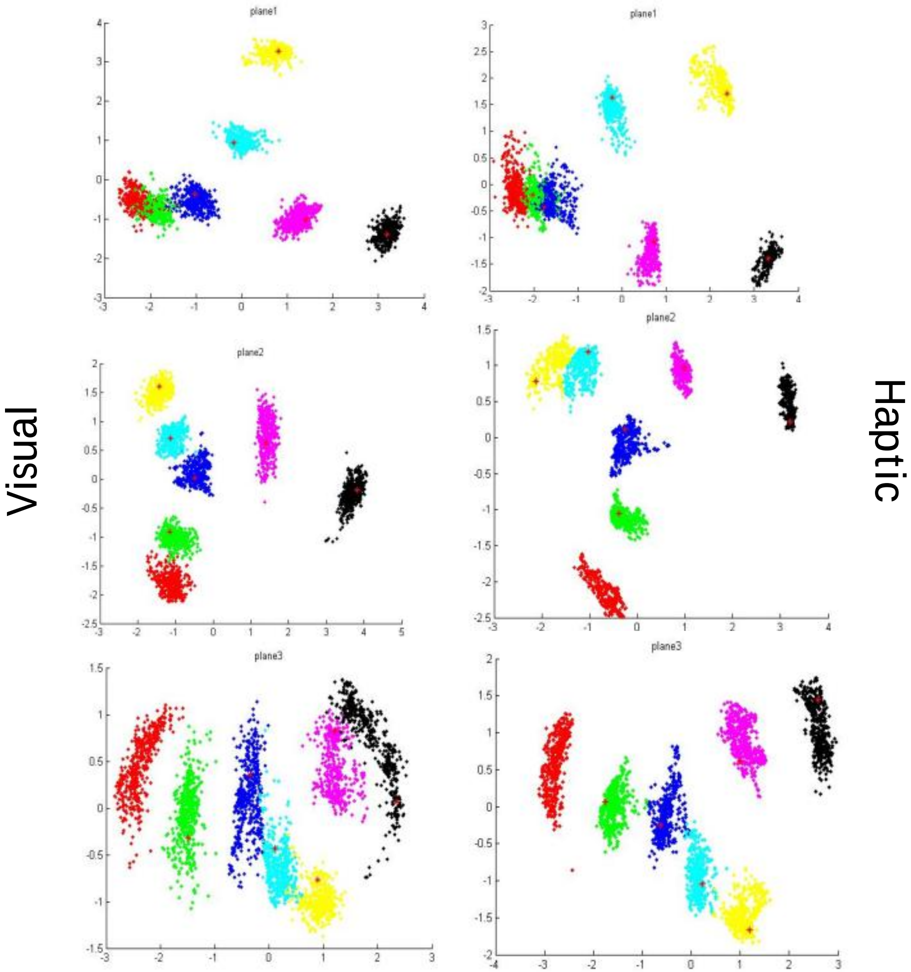
- Goal:
  - Refine framework for understanding multi-sensory (visuo-haptic) object perception
- Methods:
  - 3D printer
  - Controlled space of visuo-haptic stimuli with physical properties that are **less intuitive** than global shape and local texture
    →Parametric model of shells
    Fowler, D.R., Meinhardt, H. and Prusinkiewicz, P. (1992): ACM Transactions on Computer Graphics 26(2), 79-387
  - Similarity Ratings (as before)
  - MDS for finding perceptual space for haptic and visual exploration



Multi-modal similarity and categorization of novel, 3D objects

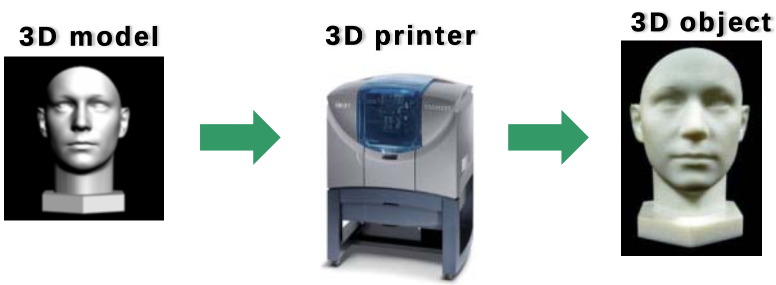
Gaißert, Wallraven, Bülthoff (2007,2008)

- Results:
  - The perceptual maps are again two-dimensional
  - Visual and haptic representation show the Y-shaped pattern of the stimulus space
  - **This is a good indication that, indeed, object representations might be shared across modalities**



Haptic face recognition
Dopjans, Wallraven, Bülthoff [2007]

- Research questions:
  - How well can people haptically distinguish, learn and recognize faces?
  - Can we generalize from haptically learned faces to the visual domain and vice versa?
  - How orientation sensitive is haptic face recognition?
- Methods:
  - MPI face database + 3D printer
  - Psychophysical recognition experiments
- Results:
  - Participants can recognize faces haptically
  - Clear cross-modal transfer: given haptic training, participants can recognize faces visually **and** vice versa surprisingly well
  - We found **no evidence** for a face inversion effect for haptic recognition



Quick summary (Sensory Integration)

- Object representations can incorporate multi-sensory information
- We found evidence for a common representation for vision and haptics
  - Shown for face recognition, object categorization
  - Cross-modal transfer between vision and haptics
    Newell, F., M. O. Ernst, B. S. Tjan and H. H. Bülthoff *Psychological Science* [2001]
- This has important applications in computer vision, where multi-sensory information can be used to improve object learning and recognition.
  - See e.g. the integration of proprioception and vision for object learning (Wallraven, C. and H.H. Bülthoff *Object Recognition, Attention, and Action* [2007])

Overview

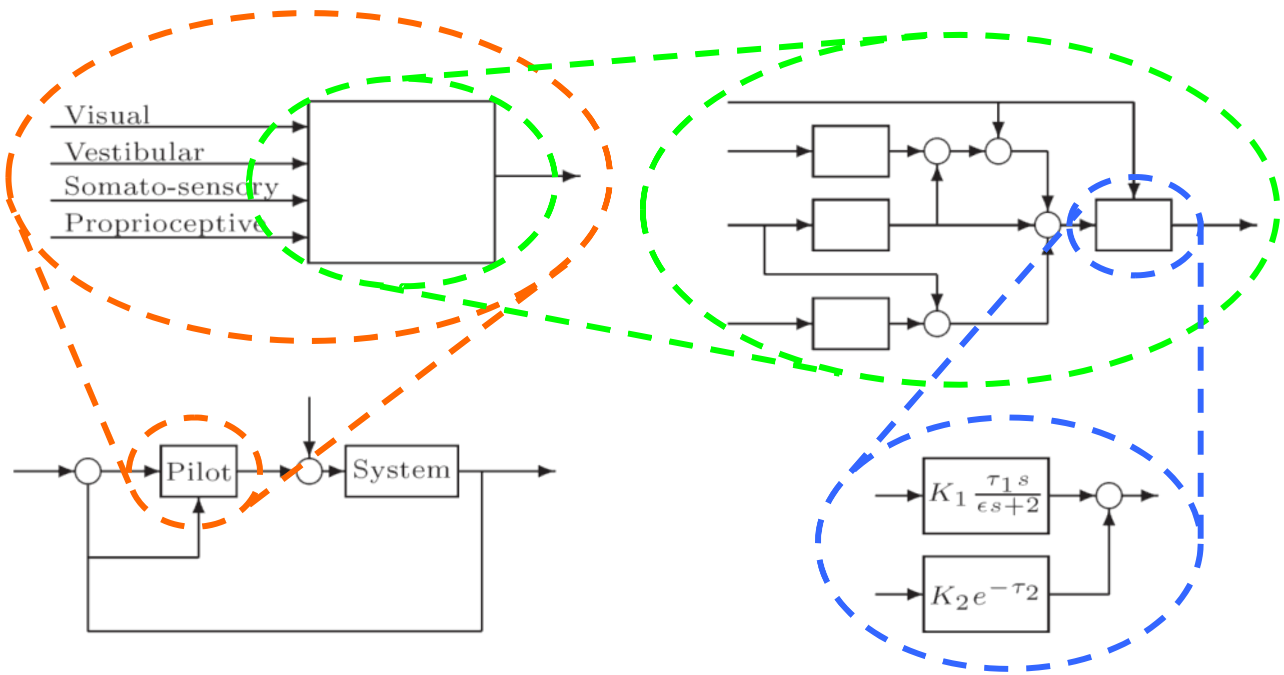
- In this talk, we will review some of the **key challenges** in understanding categorization from a combined cognitive and computational perspective:
  - the need for spatio-temporal representations
  - **perception of material properties**
  - multi-modal/multi-sensory aspects of object categorization
  - coupling of perception and action

Multisensory Integration for Control tasks

- control task pose a whole new set of problems for multisensory integration
- new research direction of our lab
- how are cues integrated during active control of orientation in space
  - 3D maze navigation *(Vidal & Berthoz, 2005)*
  - body sway (*Cunningham et al, 2006*)
  - helicopter hover control *(Berger et al, 2007)*
  - **helicopter side-step maneuver***(Beykirch et al, 2007, 2008)*

Cybernetic Approach to Perception and Action

- Develop a deeper understanding of the processing of self-motion information by considering the brain as a complex control system, which has sub-components, but which is also part of a larger system



Helicopter Control

- Why helicopter control?
- helicopter control is an interesting problem for multisensory integration and self motion perception

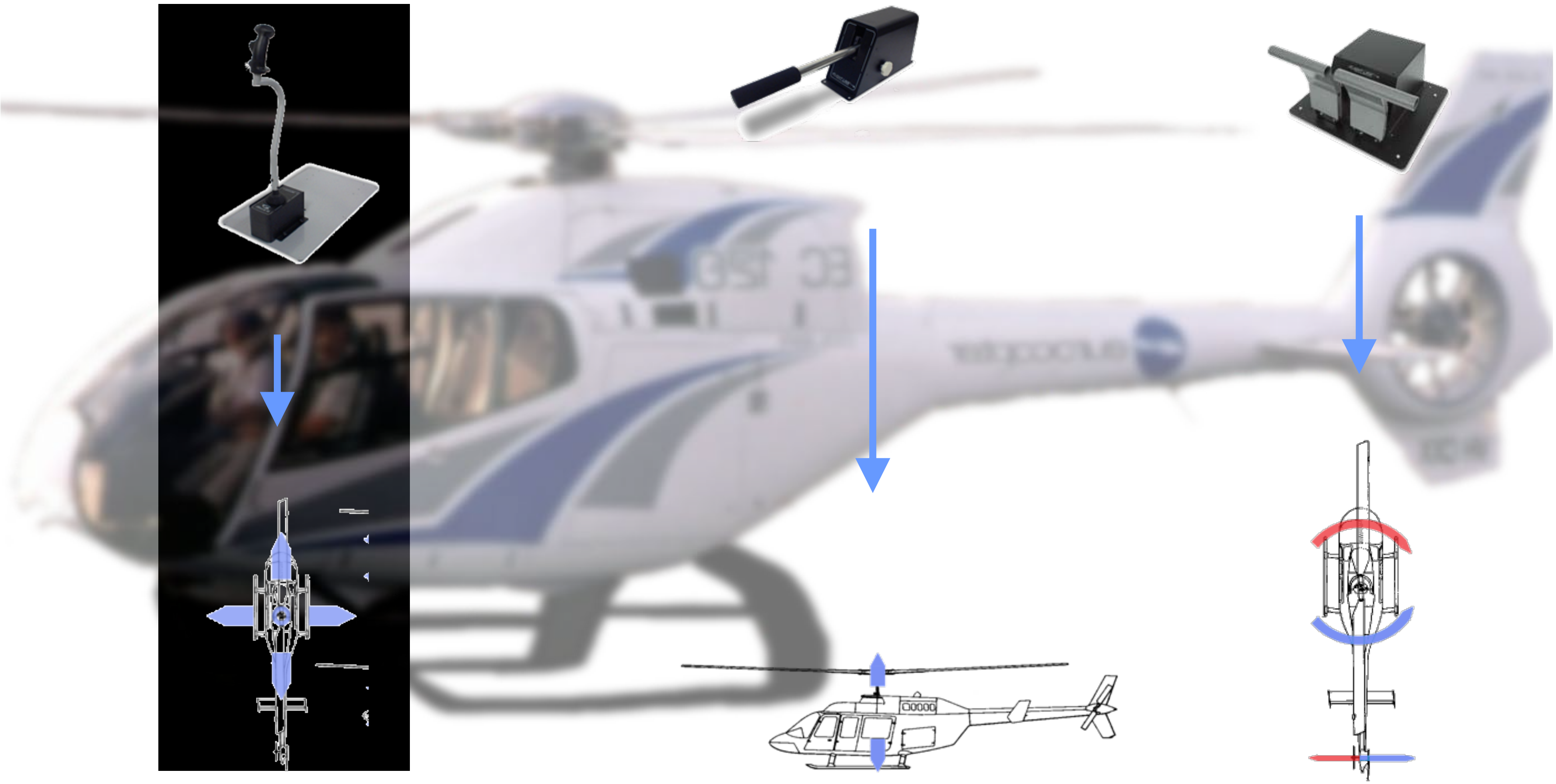
- a helicopter behaves like an inverse pendulum
- accelerates roughly in the direction it is tilted to
- different axes are dynamically coupled, so compensation for one axis effects other axes

Helicopter Control Devices

Cyclic stick

Collective stick

Pedals

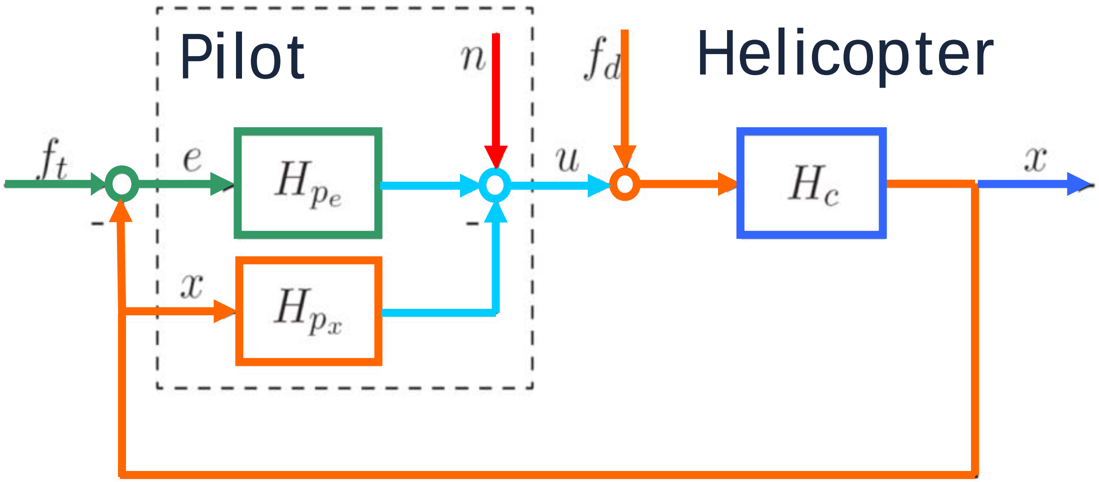


Horizontal movement

Vertical movements

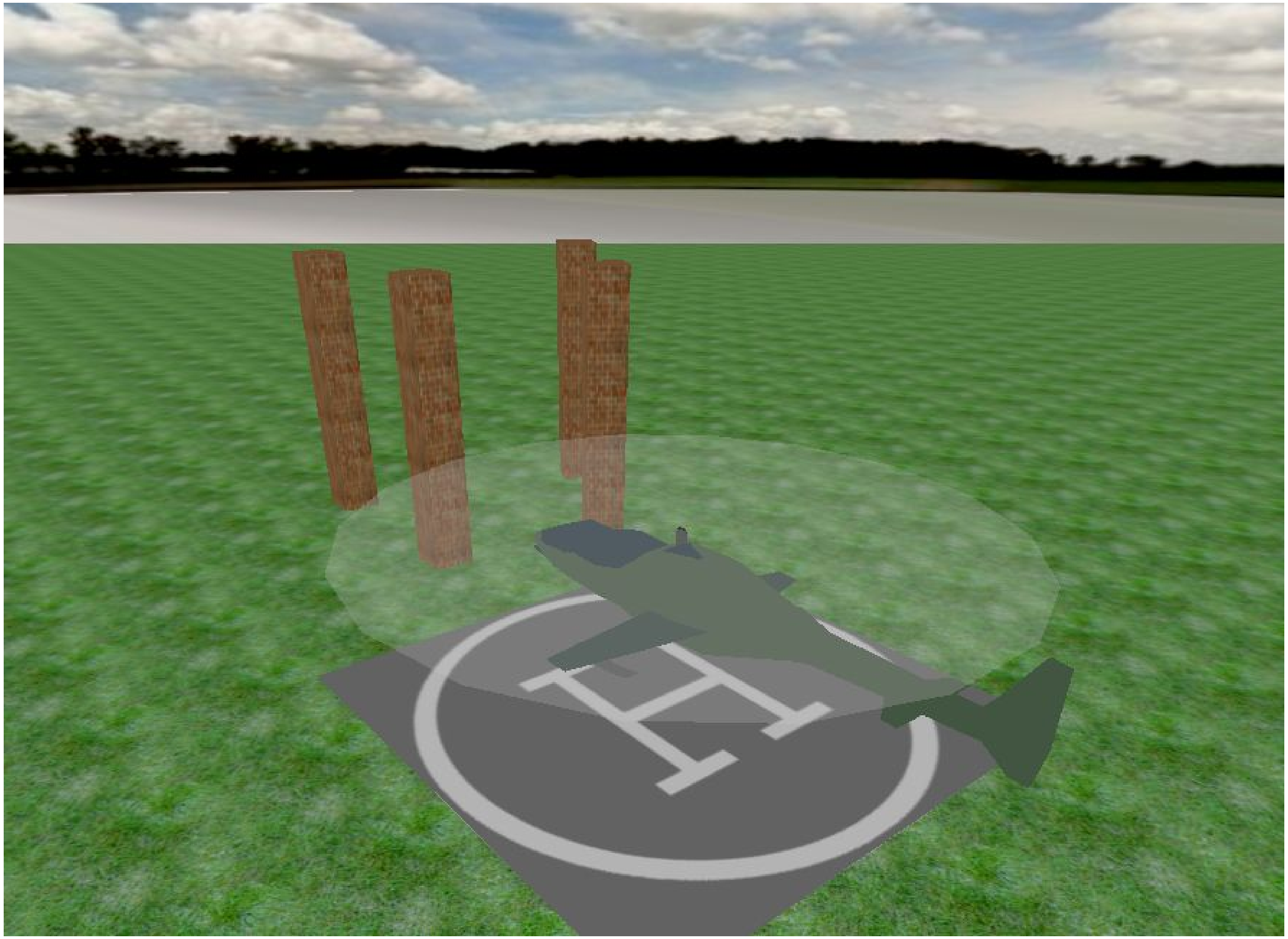
Yaw rotation

Experimental Question



- How are cues from multiple modalities integrated for *action* in a control task with the human 'in-the-loop'?
- How do we build an internal model of a physical system?

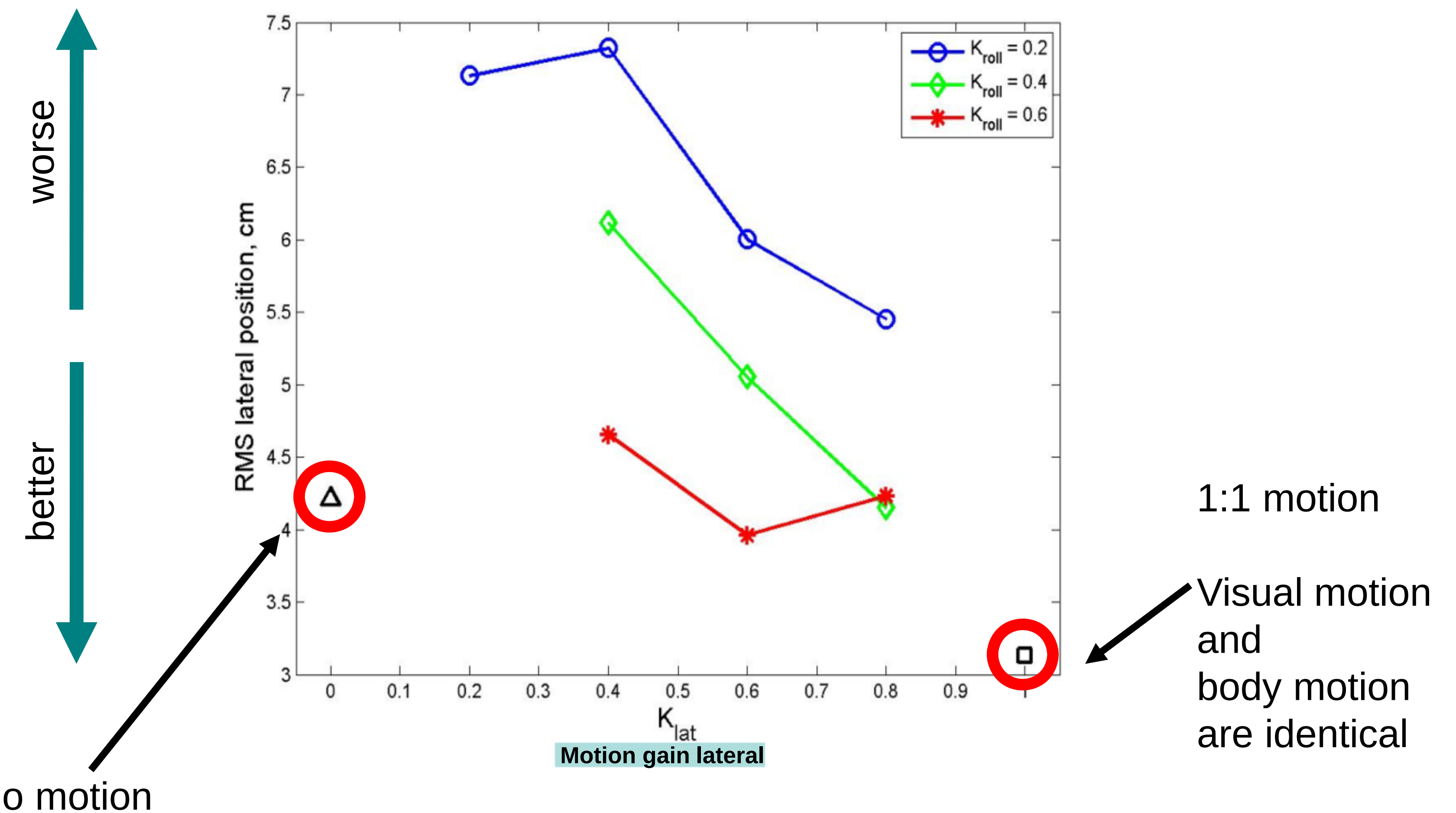
Helicopter side-step maneuver



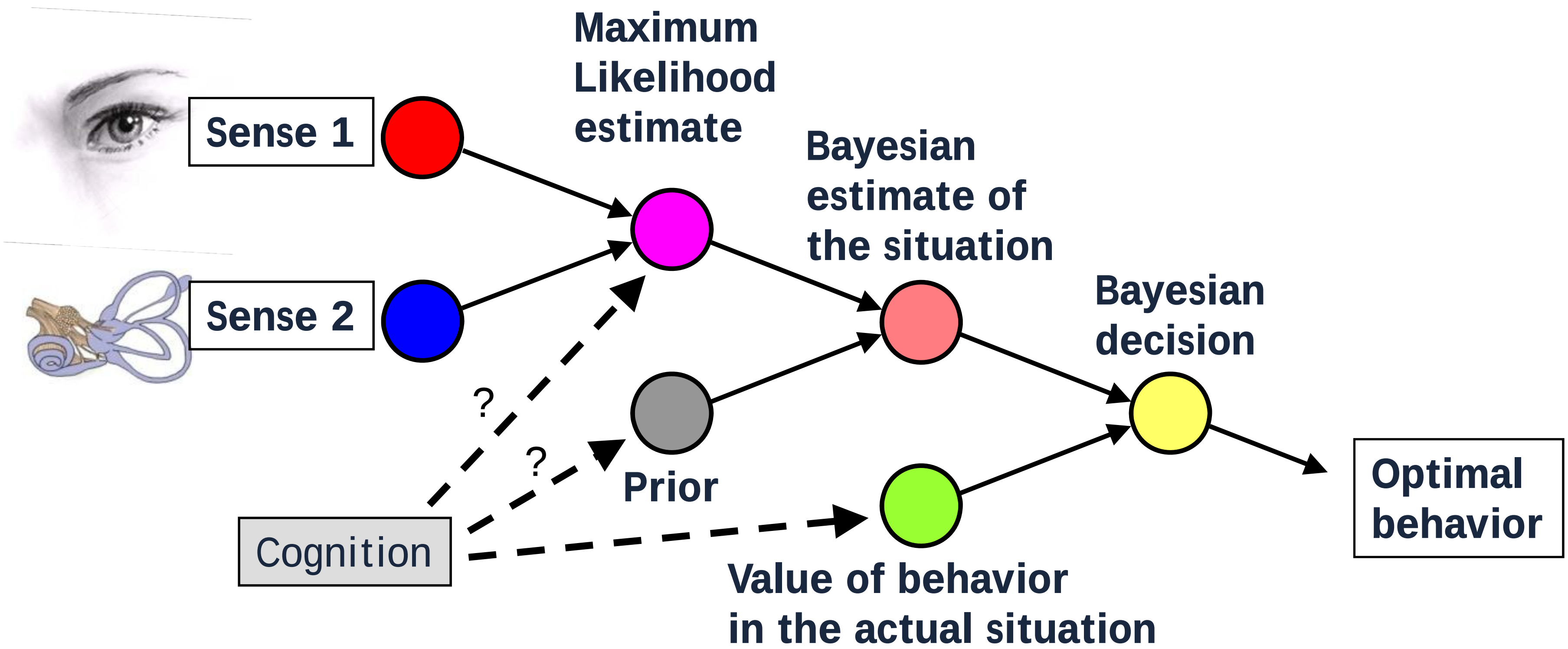
Helicopter side-step maneuver

Results

- **Pilot performance**



Better perceptual models:
Bayes as the basis for perception and action



Bayesian Decision Theory
Bülthoff & Yuille (1989-1993)
Ernst, Banks & Bülthoff (2000, ...)

Conclusion

- These recent results highlight the importance of investigating multisensory integration from the perspective of self-motion in **large-scale controlled** (**VR**) *Natural Environments*
- Studying **closed-loop** behavior offers new insights into how humans interact with the environment and solve difficult control problem
- Psychophysical experiments evaluating the impact of the different sensory cues on the perception of self-motion are valuable both to **understanding the human observer** and for **improving the technology** (eg., motion simulators)

Some open questions

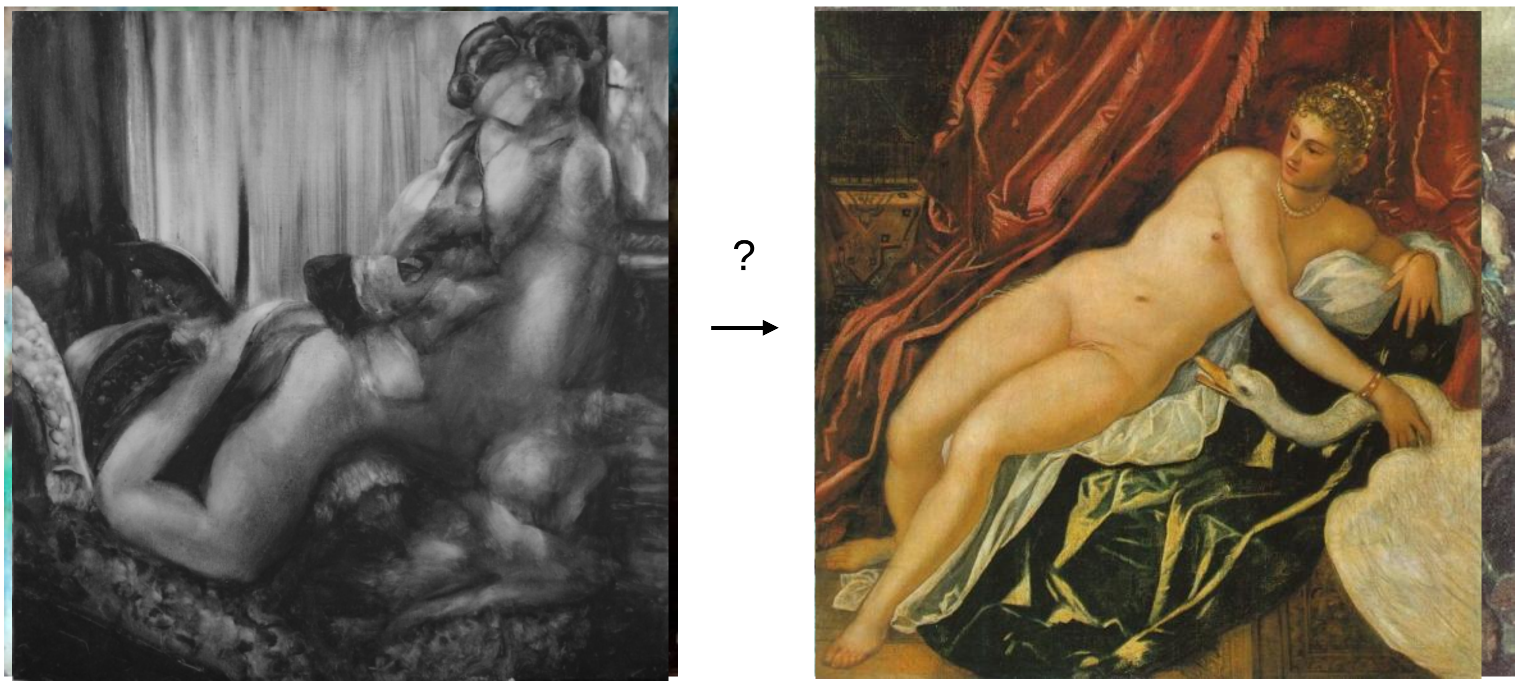
- Computer vision
  - Can we go beyond image fragments (“bags of words”)?
  - Do the current approaches scale to 1000s of categories?
  - How do we incorporate other modalities?
- Computer graphics
  - What is perceptual realism?
  - How can we make better animations?
  - Can we learn graphics?
- Perception research
  - Can we come up with a quantitative model for object recognition?
  - Does optimal integration hold everywhere – where does it break?
  - What is the psychophysics of higher-level cognitive functions?

Challenges

- The "Chair" challenge

Challenges

- The "Art" challenge: build a computer vision system that learns to interpret art images
  - Such a system would need to deal with abstraction



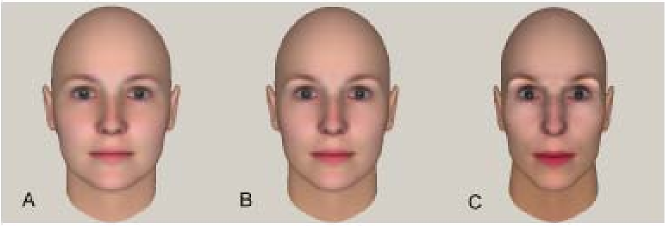
Images (c) by Robert Pepperell, see Wallraven et al. [*APGV,* 2007]

Challenges

- The "Pawan Sinha" challenge
  - build a computer vision system that integrates the 20 results every CV researcher should know about face recognition
    http://web.mit.edu/bcs/sinha/papers/20Results_2005.pdf

Eyebrows as important features

Recognition under distortions



Caricature effect for recognition

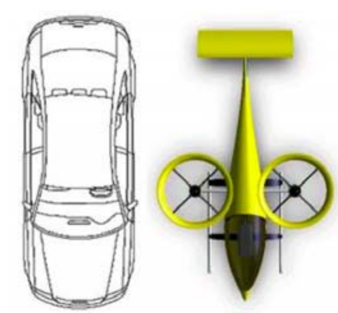
Challenges

- The “Personal Air Transport” challenge
  - Build a Personal Aerial Vehicle which makes flying as easy as driving
  - A pioneering research project incorporating novel ideas from
    - Automation, computer vision, human-machine interfaces, flight control







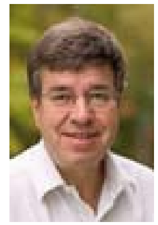


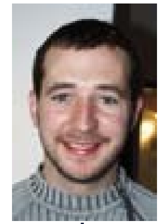
Thanks to members of the perception-action lab


Daniel Berger


Karl Beykirch


Jean Pierre Bresciani


Heinrich Bülthoff


John Butler


Jenny Campos


Franck Caniard


Astros Chatziastros


Marc Ernst


Reinhard Feiler

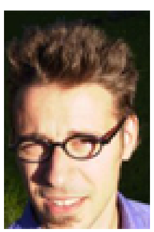

Cora Kürner


Michael Kerger

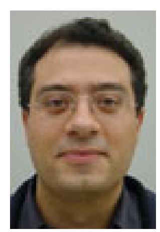

Betty Mohler


Hans-Günther Nusseck


Cengiz Terzibas


Tobias Meilinger


Frank Nieuwenhuizen


Paolo Pretto


Jörg Schulte-Pelkum


Harald Teufel


Michael Weyel