A Cognitive Ego-Vision System for Interactive Assistance

Marc Hanheide

Bielefeld University

January, 2007

euCognition
www.euCognition.org

not PDAs or cell-phones

- pure databases
- "blind"

... but cognitive abilities

- situated
- attentive
- (inter-)active

VAMPIRE
Visual Active Memory Processes and Interactive REtrieval

- **"human in the loop" systems**
  - shared attention and perception
  - mediated embodiment
  - **user controls system's perception**
- **augmented reality**
  - **system controls user's perception**
- **on-board sensors only: mobile**
- **challenges for perception**
  - no ambient sensors
  - arbitrary motions
  - limited field of view

- **assistance in object manipulation**
  - step-wise instructions
  - guidance to object positions
  - supervision and correction

- **evaluation scenario**
  - "cocktail assistant"

- **cognitive assistance system**

- ***ego-vision*** paradigm

- ***perception*** & interaction

- ***memory*** for cognitive systems

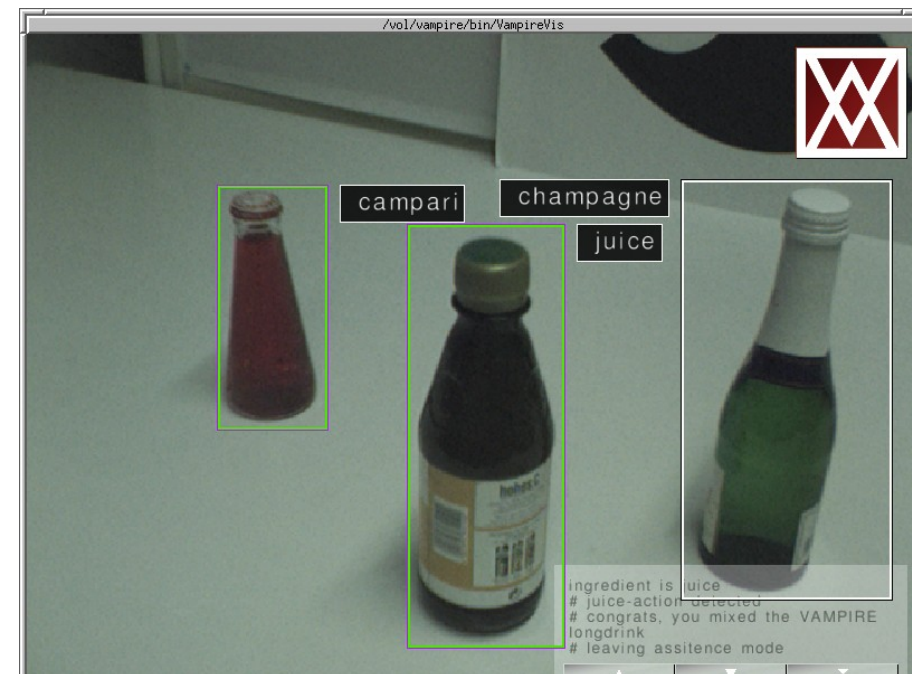[Search Recipes] [Submit a Recipe] [Top 100 Recipes]

Home >> Cocktails >> Vampire Gimlet

Cocktail Recipe: Vampire Gimlet

1 1/2 oz Vodka
　　1 oz Rose's Concentrated Lime Juice
　　1 oz Peppermint Schnapps
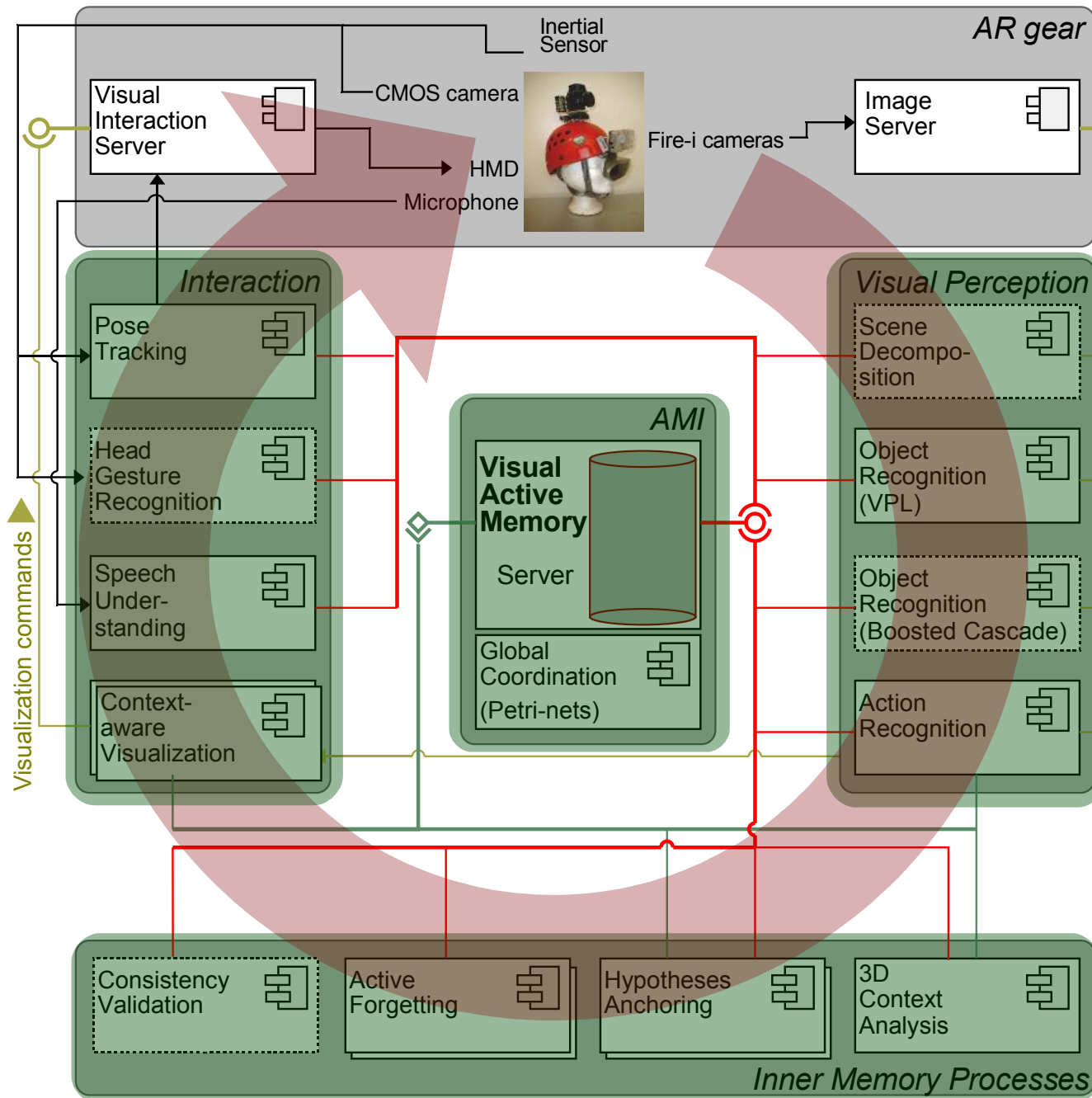
Combine ingredients with ice, shake, strain into cocktail glass.

© barmeister.com

AR gear

- Inertial Sensor
- CMOS camera
- Fire-i cameras
- HMD
- Microphone

Visual Interaction Server

Image Server

(stereo) images

Interaction

- Pose Tracking
- Head Gesture Recognition
- Speech Understanding
- Context-aware Visualization

Visualization commands

AMI

- Visual Active Memory Server
- Global Coordination (Petri-nets)

Visual Perception

- Scene Decomposition
- Object Recognition (VPL)
- Object Recognition (Boosted Cascade)
- Action Recognition

Inner Memory Processes

- Consistency Validation
- Active Forgetting
- Hypotheses Anchoring
- 3D Context Analysis

- follows EVS loop
- visual active memory
  - integration
  - memory processes
- perception
- inner processes
  - cognitive functions
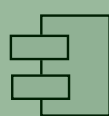 - context
- interaction
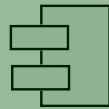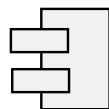 - visualization

Visual Perception

Scene Decompo-sition

Object Recognition (VPL)

Object Recognition (Boosted Cascade)

Action Recognition

- account for ego-vision challenges
- scene decomposition
  - mosaicing approach[1]

- *what is where?*
- two opposed appearance-based solutions
  - VPL classifier[2]
    - online training
  - boosted cascade[3]
    - cluttered background

[1] Gorges et al. Mosaics from Arbitrary Stereo Video..., DAGM 2004
[2] Bekel et al. Adaptive computer vision: online learning..., DAGM 2004
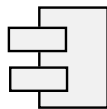[3] Viola and Jones. Rapid object detection..., CVPR 2001
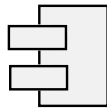
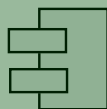## Visual Perception

- Scene Decompo-sition
- Object Recognition (VPL)
- Object Recognition (Boosted Cascade)
- Action Recognition

- *what is going on?*
- supervision in assistance
  - step-wise instructions
  - correction
- has the action been performed correctly?
- ego-vision
  - simple model of attention
  - user looks where she acts
- analyze trajectory of manipulated objects
  - track object
  - classify trajectory: pouring, shaking,...

$$p_{end}(\mu_1)$$

$$p_{end}(\mu_2)$$

$$p_{thres}$$

probability

time

- initialized on centered objects

- capture object trajectory

- kernel-based tracker[1]
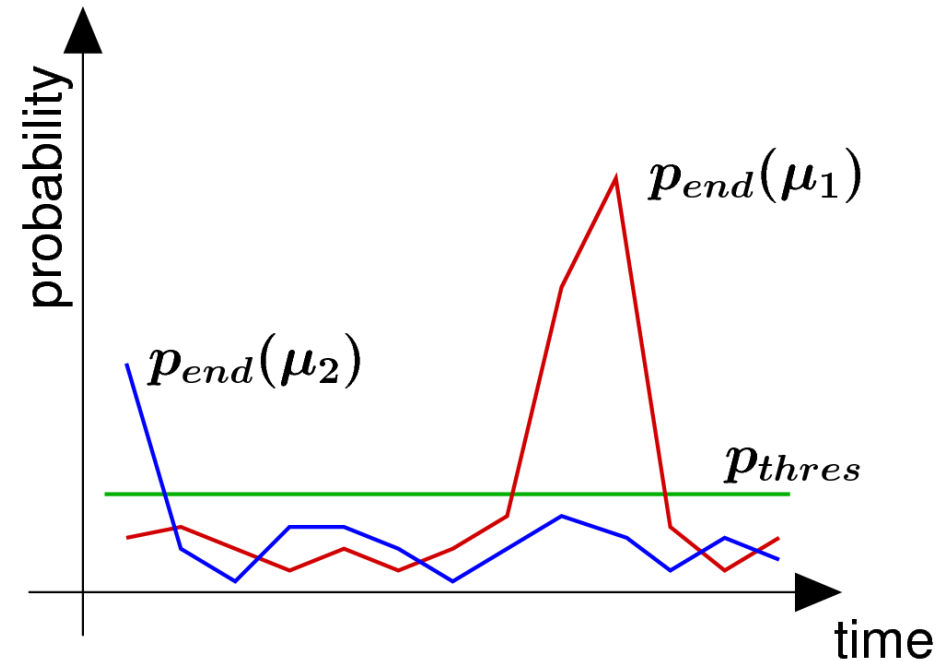
- *condensation algorithm*
  - segmentation-free classification

- originally for hand trajectories[2]

- detection on threshold
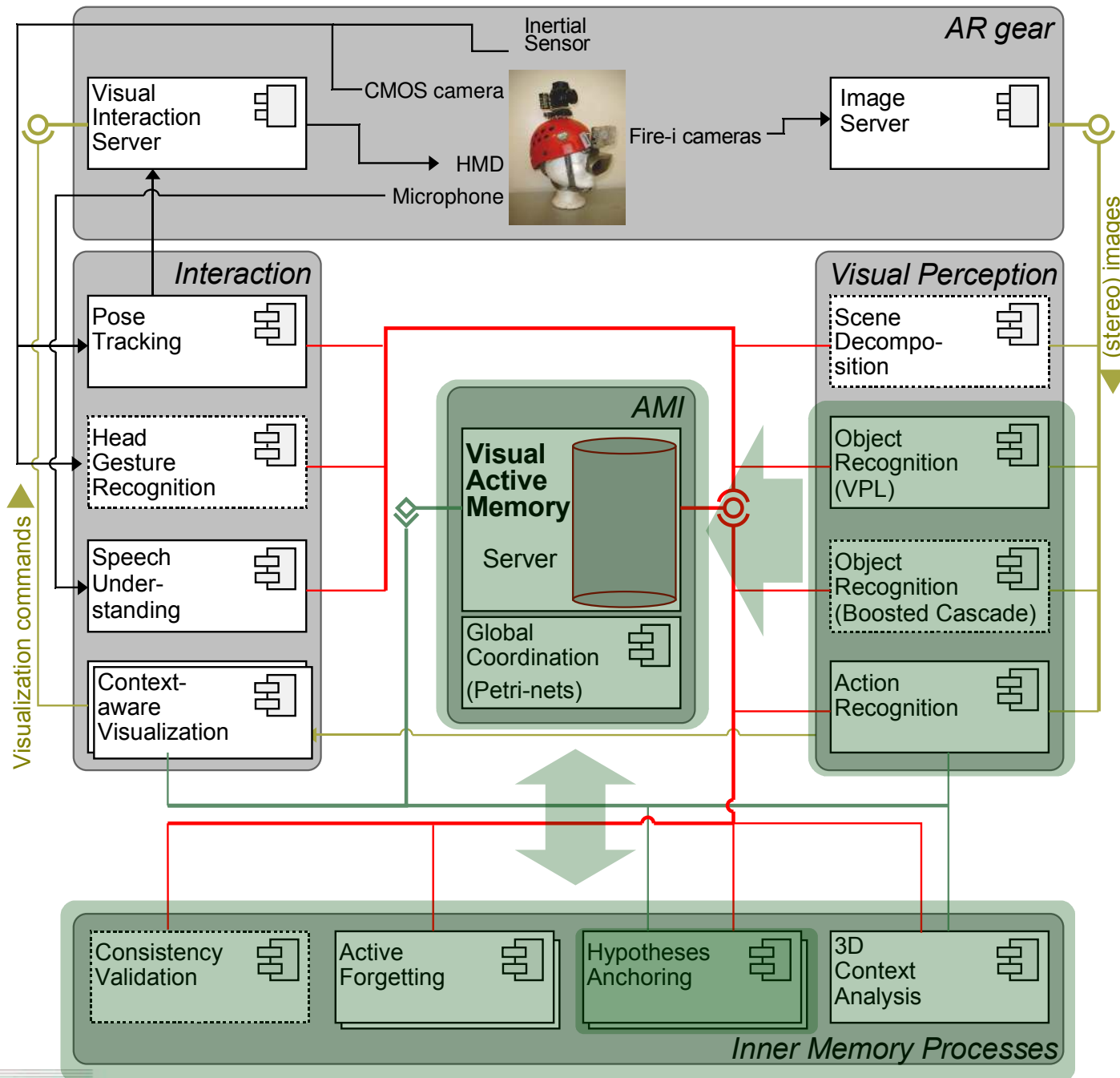
[1] Comaniciu et al. Kernel-Based Object Tracking, PAMI, 2003
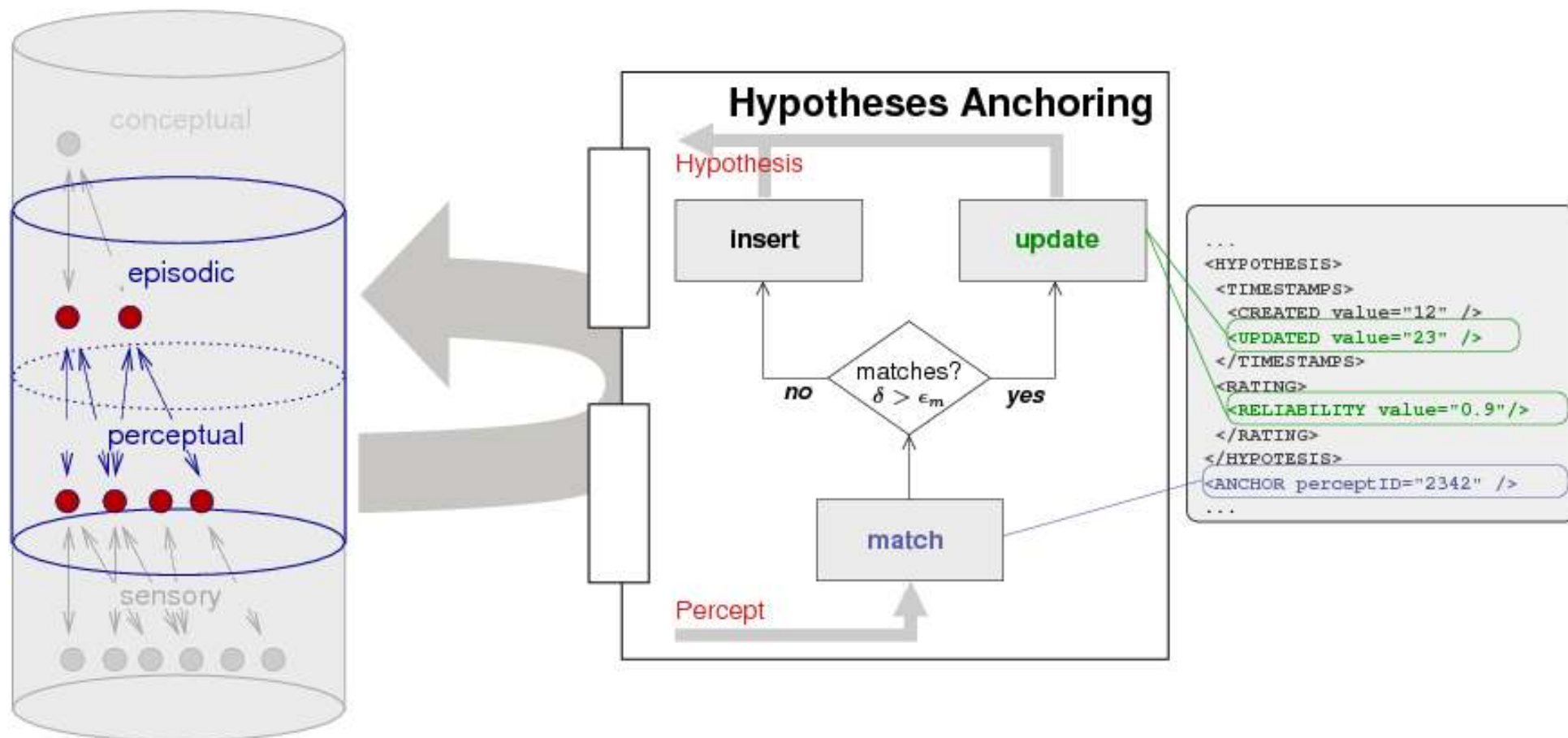[2] Fritsch, Hofemann et al. Combining [...] for Gesture Recognition, ICPR, 2004

- **perceptions into memory**
- **memorize & analyze**
  - what?
  - where?
- **inner processes**
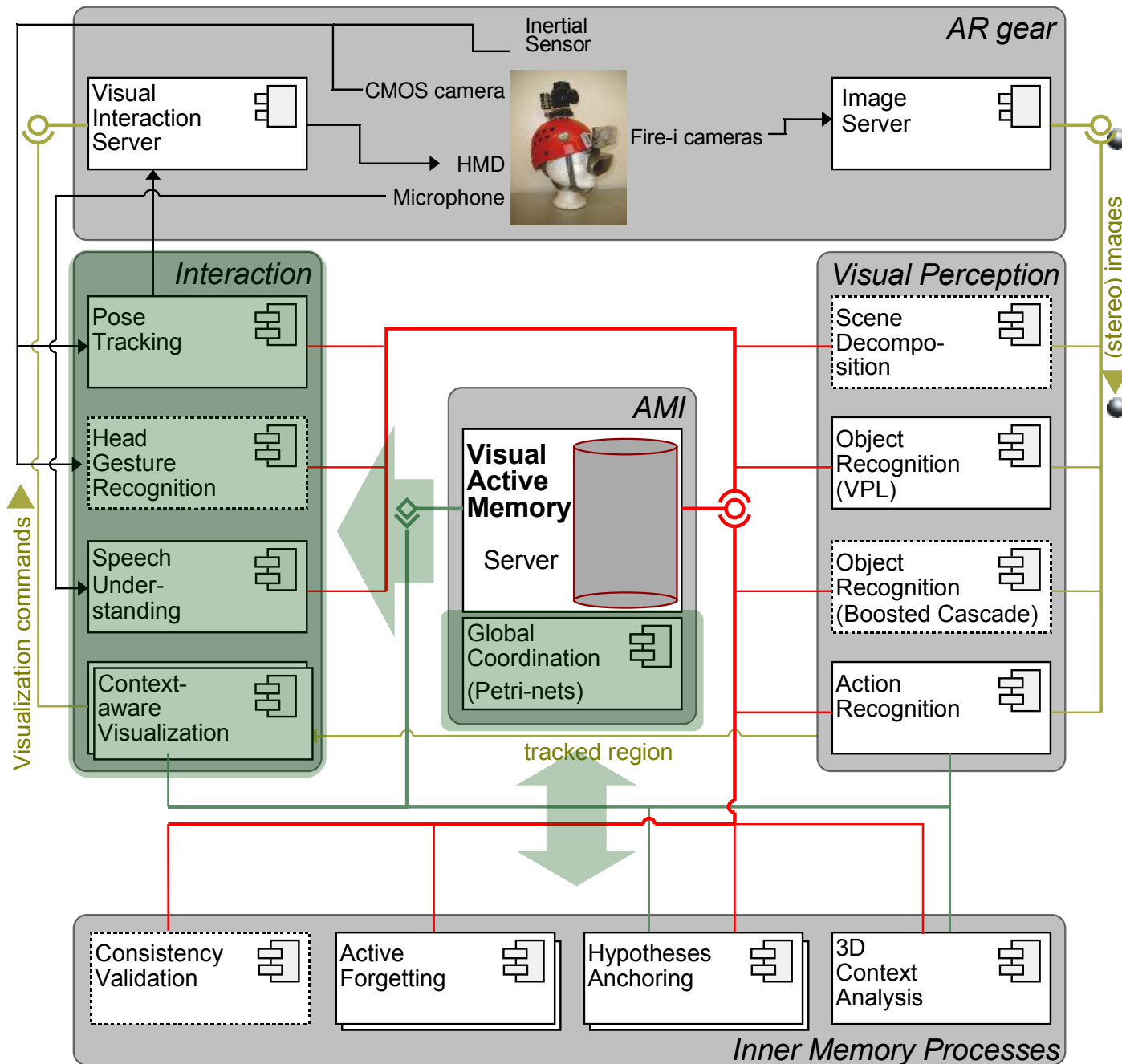 - context analysis
  - forgetting
  - consistency validation
  - anchoring

- memory is layered
- map percepts to episodes
- related to anchoring
  [Coradeschi & Saffiotti]

- episodic <-> real world
- generic process
  - actions & objects
- hypotheses: support & doubt

global coordination
- recipe handling
- state-based control
- error handling

interaction
- speech
- head gestures
- visualization
- answer queries

- **individual components**
  - recognition rate
    (~80% in action recognition)
  - others

- **user studies**
  - online, real world
  - task oriented
    - "teach some objects!"
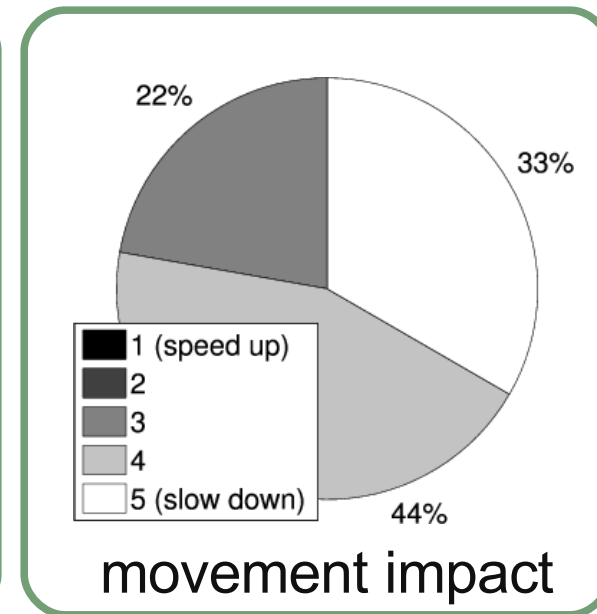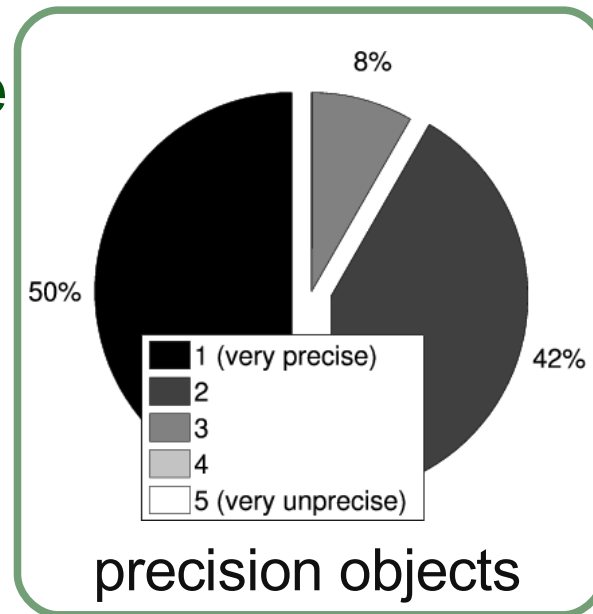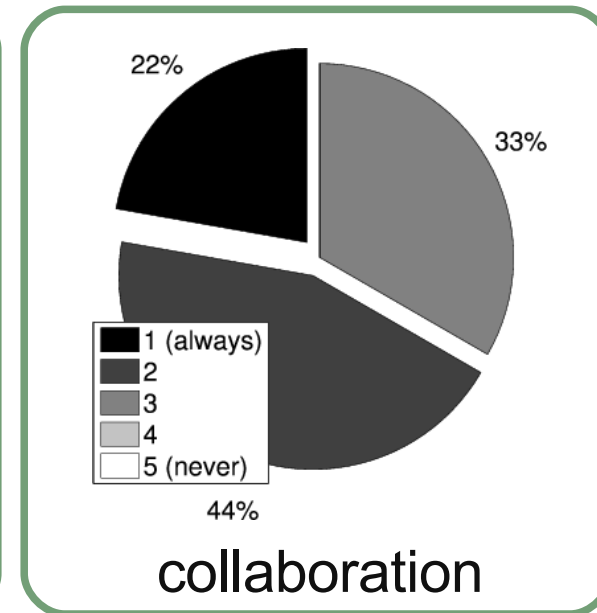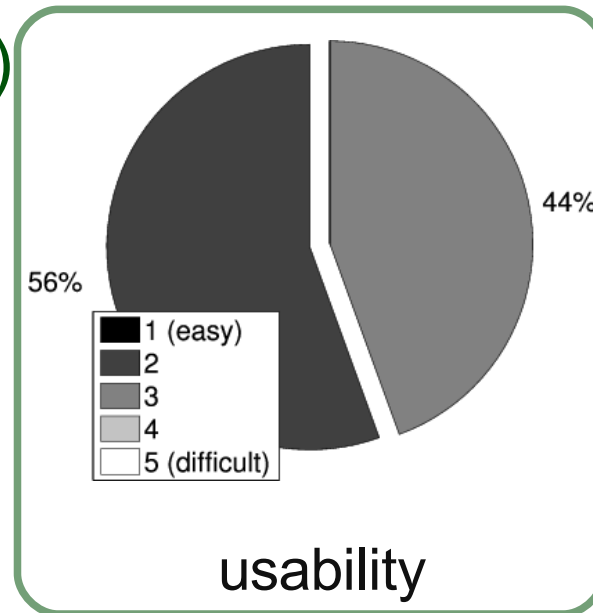   - "prepare a cocktail!"

- **11 first time users**

- **all solved the task**
  - some explanation required

- questionnaire (24 answers)
  - usable
  - collaboration works
  - object highlight precise
    - reactive
    - reliable
  - impact on movement
- **"fun to use!"**
- **ego-vision for assistance**

usability

Legend:
- 1 (easy)
- 2
- 3
- 4
- 5 (difficult)

56% / 44%

collaboration

Legend:
- 1 (always)
- 2
- 3
- 4
- 5 (never)

22% / 33% / 44%

precision objects

Legend:
- 1 (very precise)
- 2
- 3
- 4
- 5 (very unprecise)

50% / 8% / 42%

movement impact

Legend:
- 1 (speed up)
- 2
- 3
- 4
- 5 (slow down)

22% / 33% / 44%

Ego Vision

- mediated embodiment

- shared
  - perception
  - attention

- mutual control
  - challenges for perception

Active Memory

- architectural concept

- memory processes
  - perception
  - anchoring

Outlook

- Active Memory: towards a cognitive architecture (robotics)

- Ego-Vision: focus attention and collaboration

- Assistance System:
  - lighter hardware
  - challenging scenario

Thank you!

Questions?

credits:
Ingo Bax, Holger Bekel, Nicolas Gorges, Christoph Gräßl, Nils Hofemann, Marcel Richter, Jan Schäfer, Hannes Siegl, Christoph Stock, Sebastian Wrede, Timo Zinßer