Vision in Man and Machine

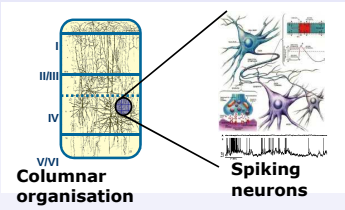
Part 9
Cognitive Vision and Online Learning

Heiko Wersing

Honda Research Institute Europe GmbH

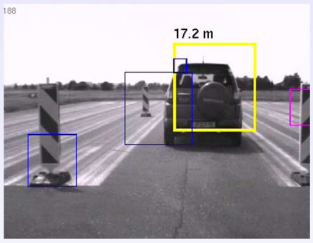
Research at HRI Europe, Offenbach, Germany

Centre of excellence for Honda in the area of **Intelligent Systems Research**
~70 Researchers doing only fundamental research

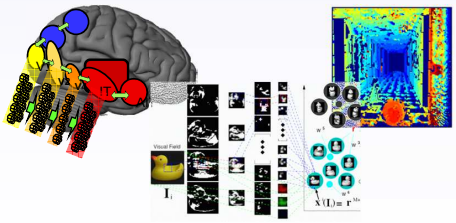


Computational Neuroscience

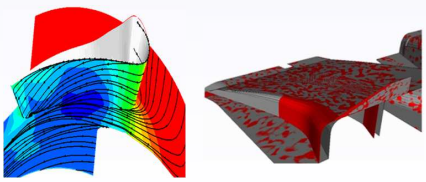
Cognitive Robotics



Driver Assistance



Biologically Motivated Computer Vision



Evolutionary Optimization

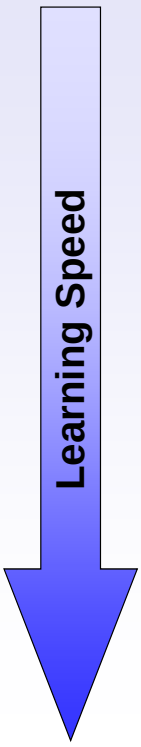
Object Learning in Humanoids

- Key ability for a cognitive robot in a changing environment
- Interactive training dialogue with a human partner
- Life-long buildup of knowledge representation

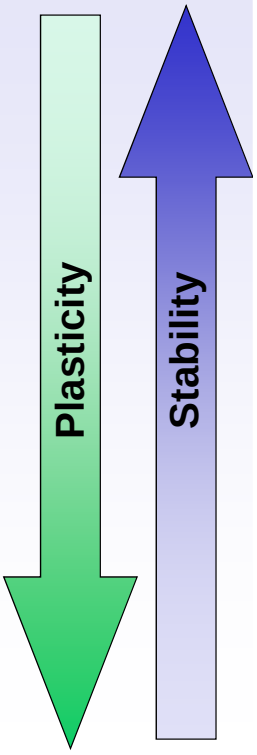
Challenging problems:

- When and what should be learned
- Efficient object representations (dimensionality problem)
- Stability-plasticity dilemma of incremental learning

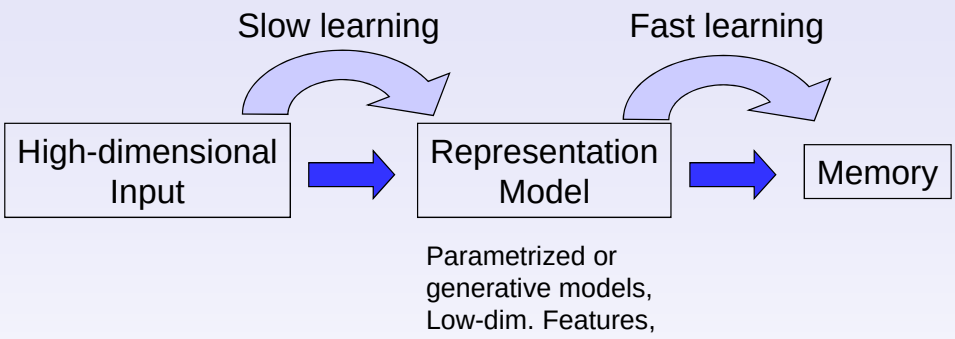
Time scales of learning



- Offline or batch learning
  - Replace programmed functions
  - Use large data collections for statistical learning
- Online calibration and parameter adaptation
  - Adapt simple behaviour parameters
  - Local learning in single functional modules
- Online and one-shot learning
  - Interactive feedback
  - Situation- and context dependent learning

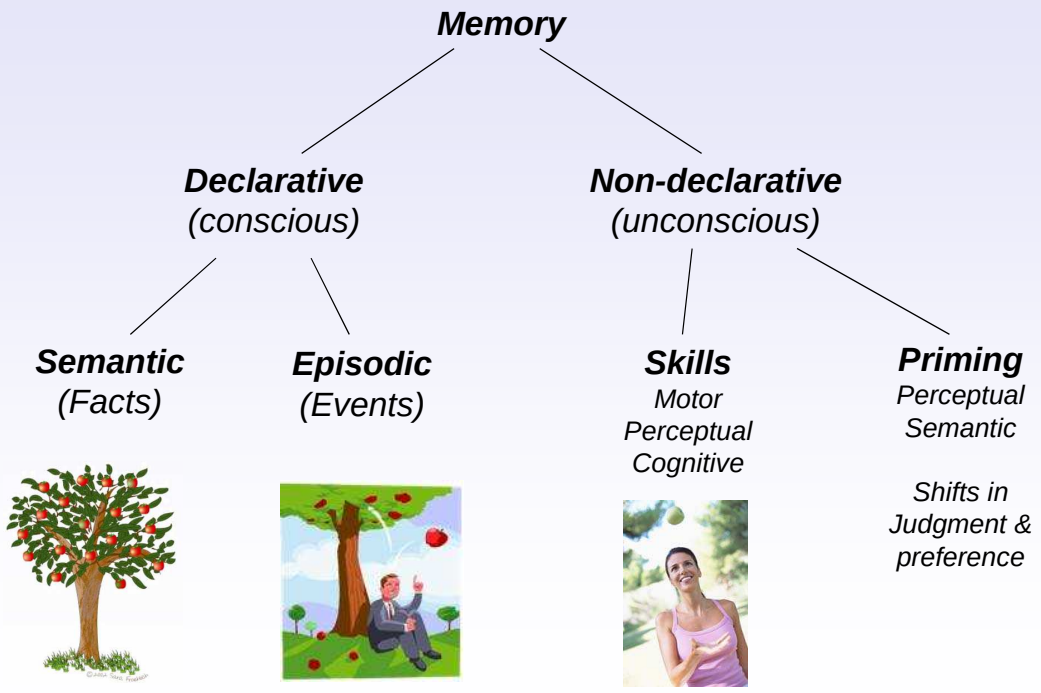


Concepts



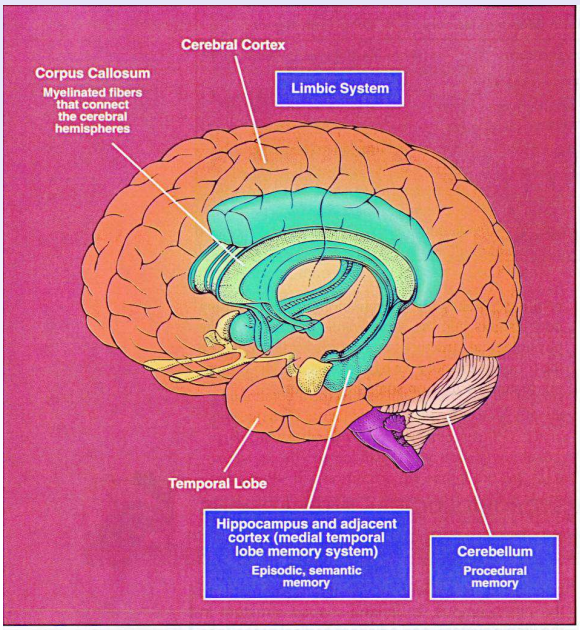
- Obtain representation models by slow/offline learning or heuristics
- Use low-dimensional representations for fast learning
- Lifelong learning needs integration of both → Stabiliy-Plasticity-Dilemma
  - Stability: Keep all old knowledge
  - Plasticity: Add and modify knowledge according to learning

Human memory taxonomy (Schacter & Tulving 1994)



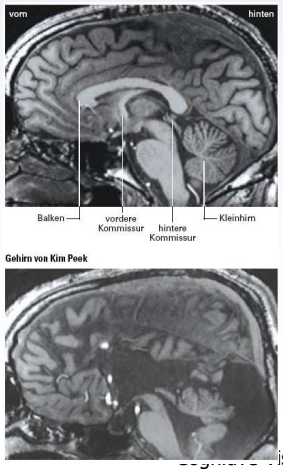
Human memory systems

- Sensory memory
  - 1-3 seconds
  - Reverberating activity in sensory areas
- Working memory
  - Only for current task
  - Storage of few entities for current behaviour
  - Prefrontal cortex
- Short-term memory
  - Up to hours
  - Preservation of information for LTM transfer
  - Mediotemporal lobe (Hippocampus) in interaction with higher neocortical areas
- Long-term memory
  - Lasts for a life-time
  - Higher neocortical areas, Transfer requires MTL structures



Kim Peek, an extraordinary mind

- Born 1951 in Salt Lake City
- Knows ~9000 books in full text (only facts, no fiction)
- Inspired the „Rain man", but is not autistic
- Has complete access to all his knowledge
- Missing corpus callosum → „Split brain"
- Hypothesis: No dominance of left hemisphere over the right



Other Savants

- Orlando Serrel
  - Has a *complete* episodic memory of his life until the age of 10
  - No particular brain anomaly could be identified so far

- Daniel Tammet
  - Recited 22514 digits of π
  - Can learn a new language in one week
  - Performs large number calculations directly

Trained brains

- Chao Lu: World record holder with 67890 digits of π in 24 hours, 4 minutes
- Technique of „mental walk" (Simones of Ceos, 477 BC)
- Strong hippocampal activity can be measured during training, but no persistent structural change
  (Maguire et al. Nature Neuroscience 2002)

Learning of Sensory Representations (LSR)

- Visual object memory is a special case of memory
  - Mixture of declarative and non-declarative memory
  - High dimensionality, abstraction of representation unclear
  - Generalization and invariance are very important
  - Rather difficult properties for memory experiments

Standard object learning: Offline learning

Image Data

Long training procedure

Can Classifier

Isolated „Black Box“
Closed solution

Recognition Module

Face Classifier | Can Classifier | Cup Classifier

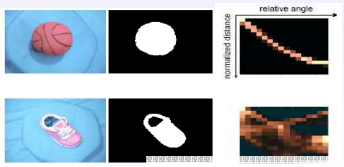
• No flexible online-learning

⇨ Unsuitable for autonomous learning in robotics

Cognitive Vision: Learning Object Associations

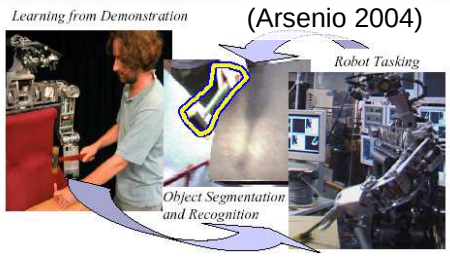
- Steels & Kaplan (2001)
  - Color-histogram-based object representations
  - Focus on social learning in interaction with AIBO
- Roy & Pentland (2002)
  - Multi-dimensional shape and color histograms
  - Short-term and long-term memory model
- Arsenio (2004)
  - Task-specific learning
  - Object segmentation from periodic motion → Hash tables

(Steels & Kaplan 2001)



(Roy & Pentland 2002)

(Arsenio 2004)

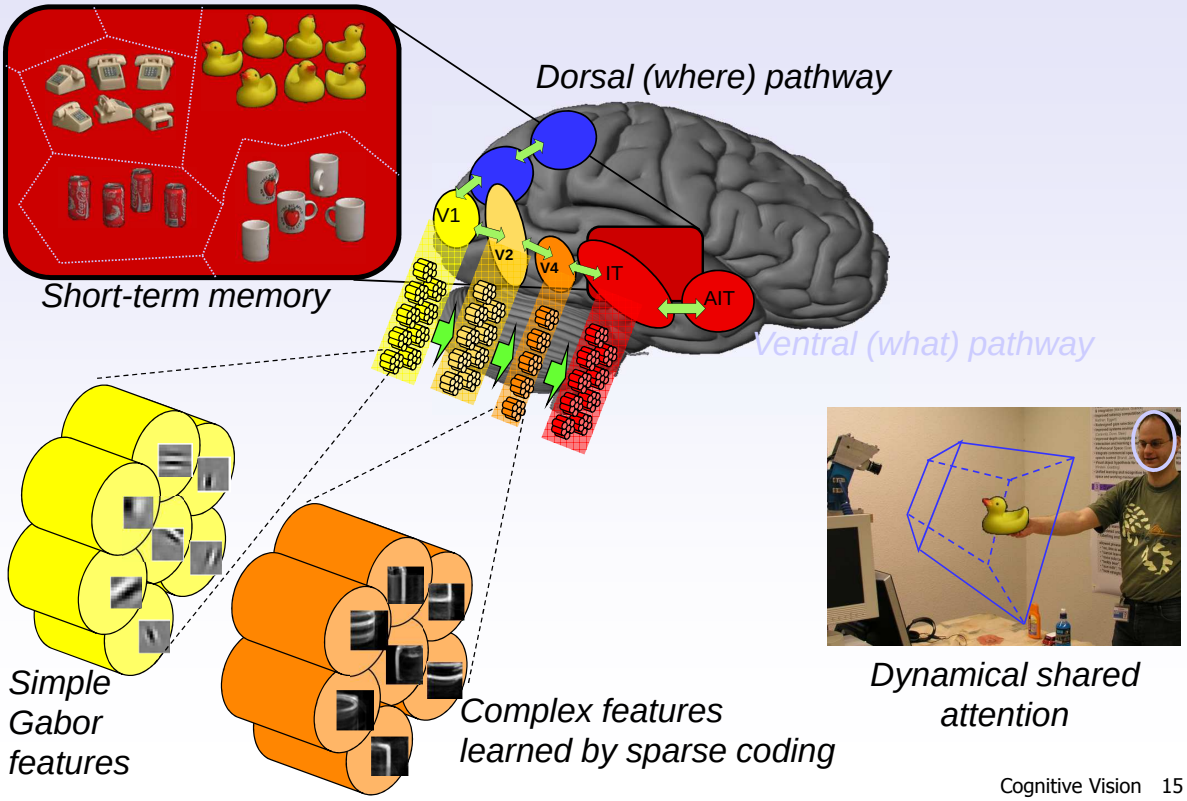


Online Learning in Object Recognition

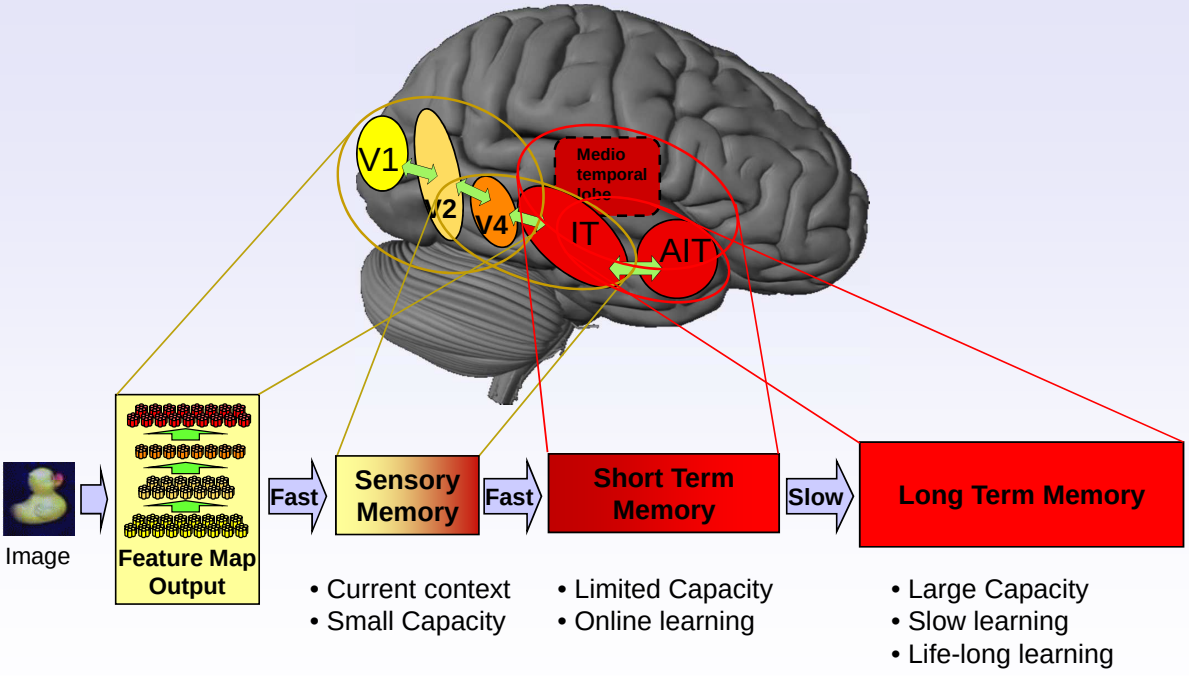
- Bekel, Bax, Heidemann, & Ritter (2004)
  - Dimension reduction, local PCA
  - Local linear map classifier
- Wrede, Hanheide, Wachsmuth, & Sagerer (2006)
  - Integrated system for interactive learning

- Roth, Donoser, & Bischof (2006)
- Roth, Grabner, Skocaj, Bischof, & Leonardis (2005)
  - Incremental PCA for dimension reduction and model learning
  - Online AdaBoost classifier
  - Constant background model

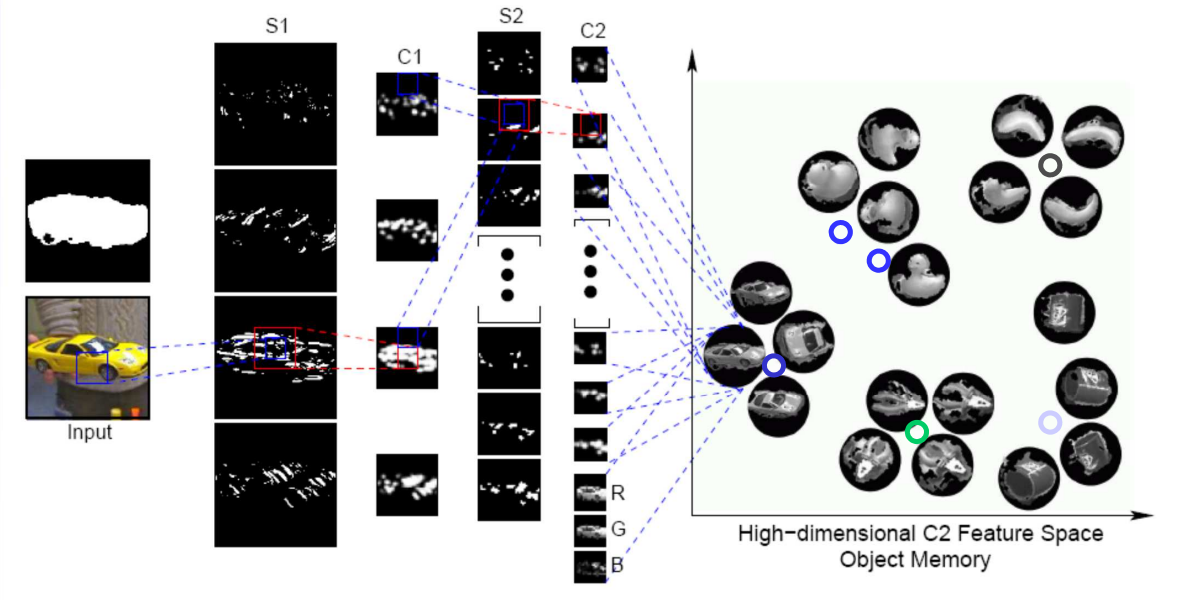
Our biologically motivated approach



Visual Memory Architecture

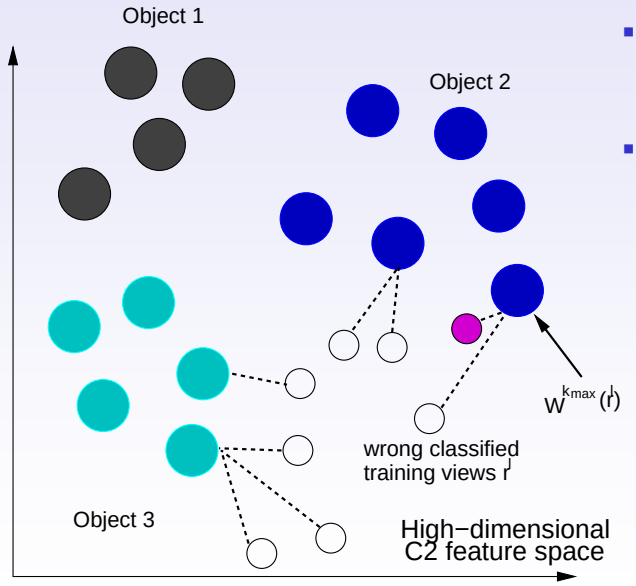


Recognition architecture



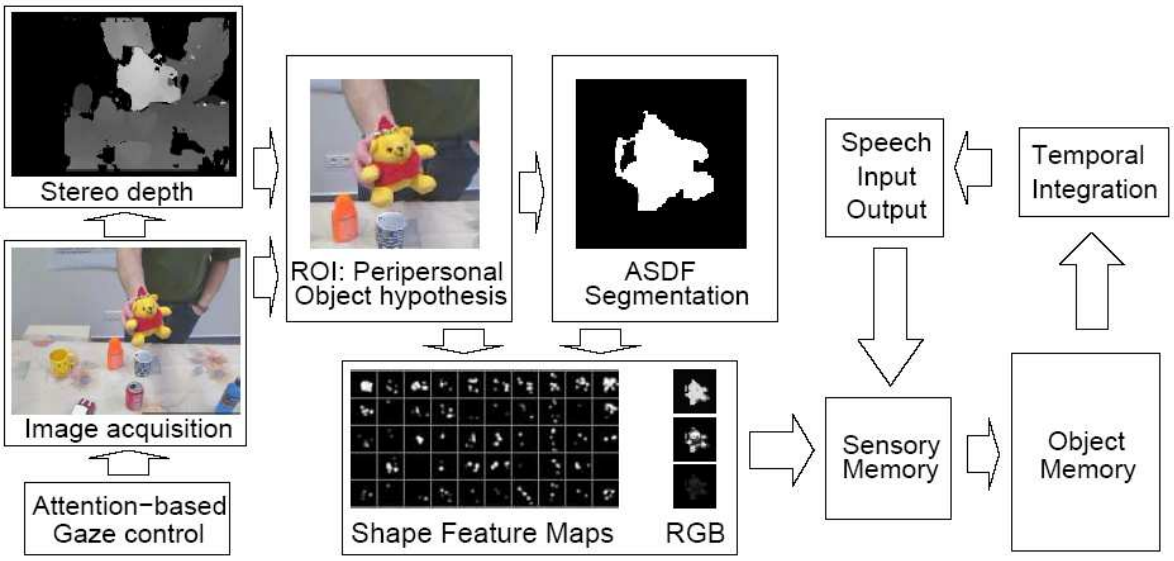
Short-term
→ Long-term memory

LTM Model – Insertion Rule



- False classified training vectors and their wrong winning nodes are collected
- After a sufficient number of errors for each wrong classified class one new node is inserted
- New nodes are mainly inserted near class borders

Online Learning System Architecture



Cognitive Vision: Online Learning of Objects

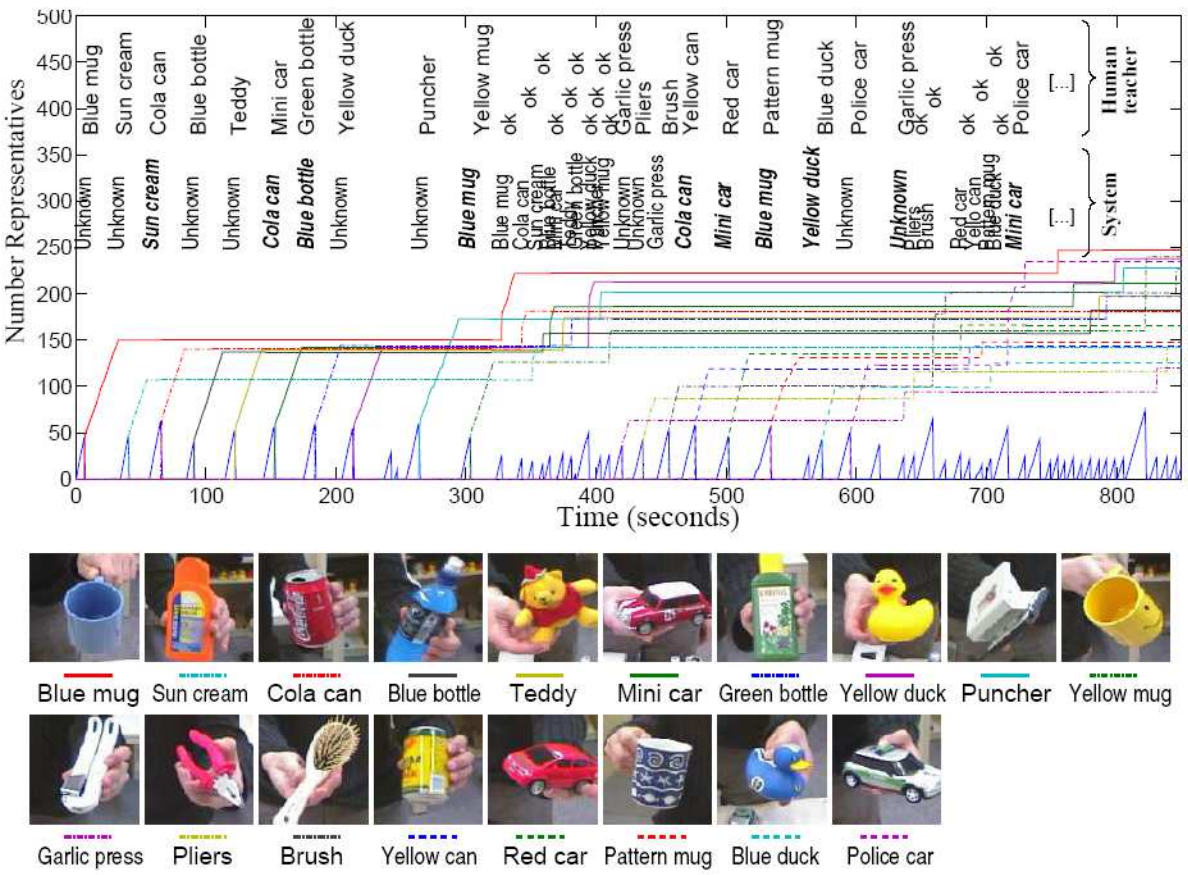
Biologically motivated architecture

Attention
Gaze control
Online object learning

Contributors:

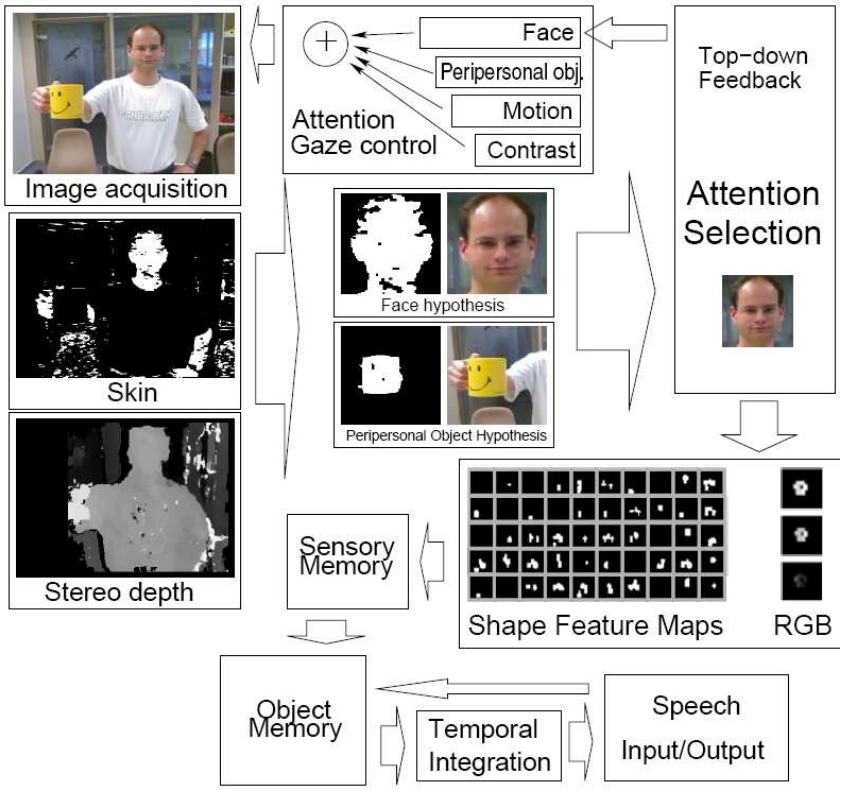
C. Goerick
H. Wersing
I. Mikhailova
S. Kirstein
J. Eggert
A. Ceravola
M. Dunn
S. Rebhan
M. Weissbach
J.J. Steil
M. Götting
H. Janssen
H. Brandl

Demonstration System at Honda Research Institute (2006)



ALIS Autonomous Learning and Interacting System

Online Learning System Architecture



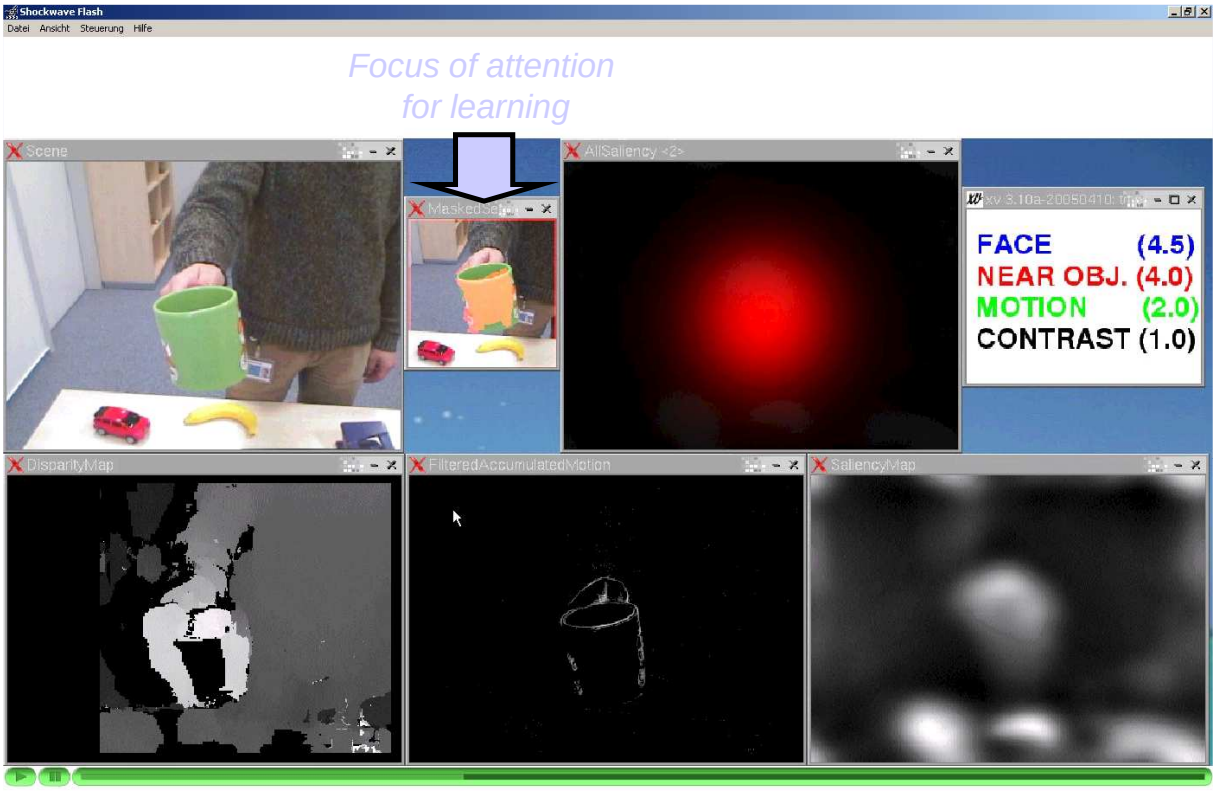
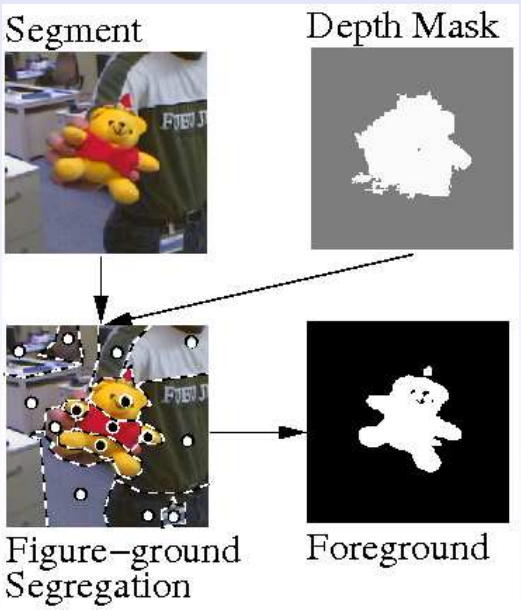


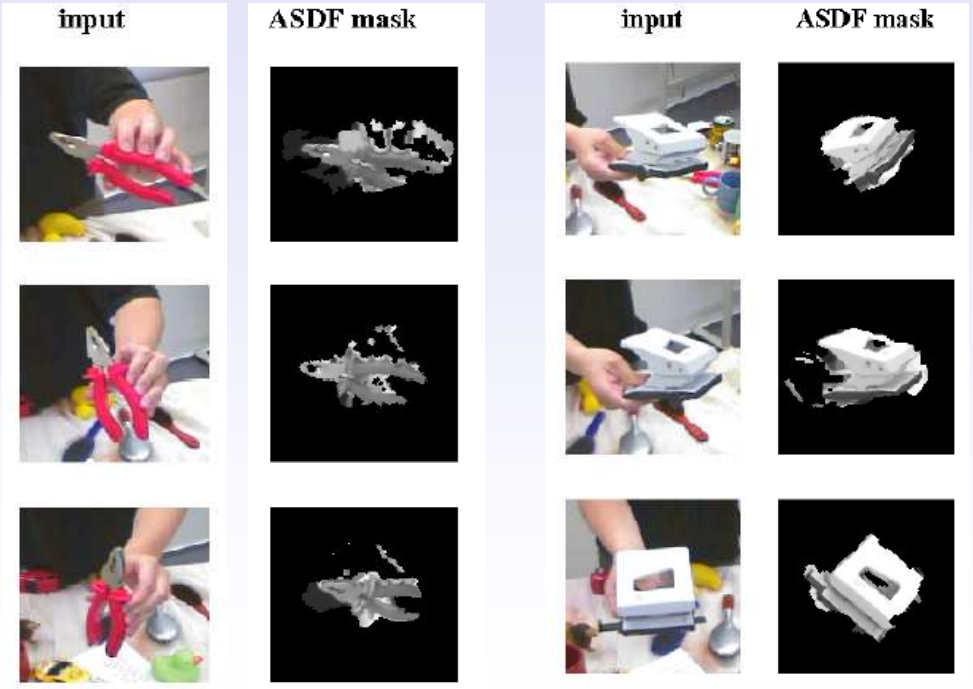
Figure Ground Segregation



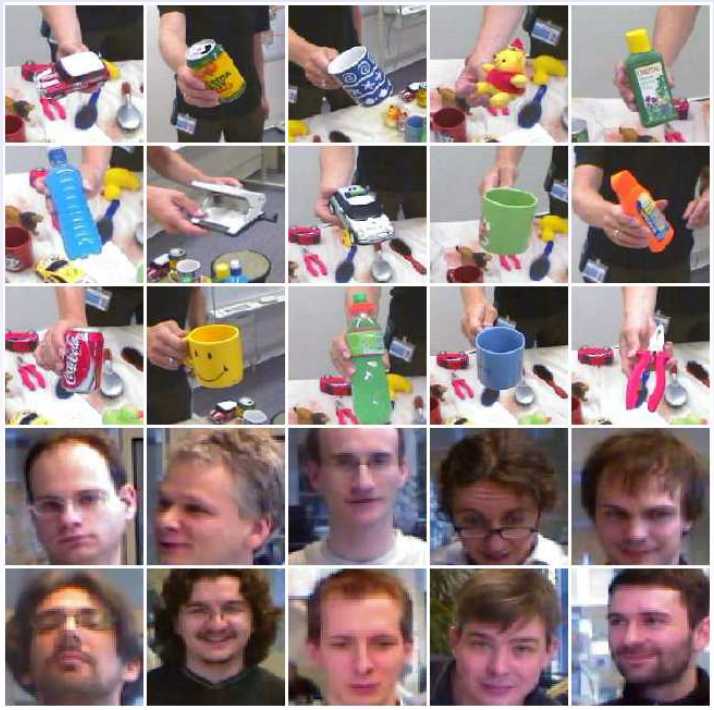
- Supervised learning vector quantization approach (LVQ)
- Initialization with noisy depth blob
- Learn predefined number of foreground and background prototypes
- Additional relevance factors are calculated based on generalized matrix LVQ
- Output is binary foreground mask → only foreground pixels are used for feature extraction

Denecke et al. ESANN (2008)

Segmentation + Rotation Normalization

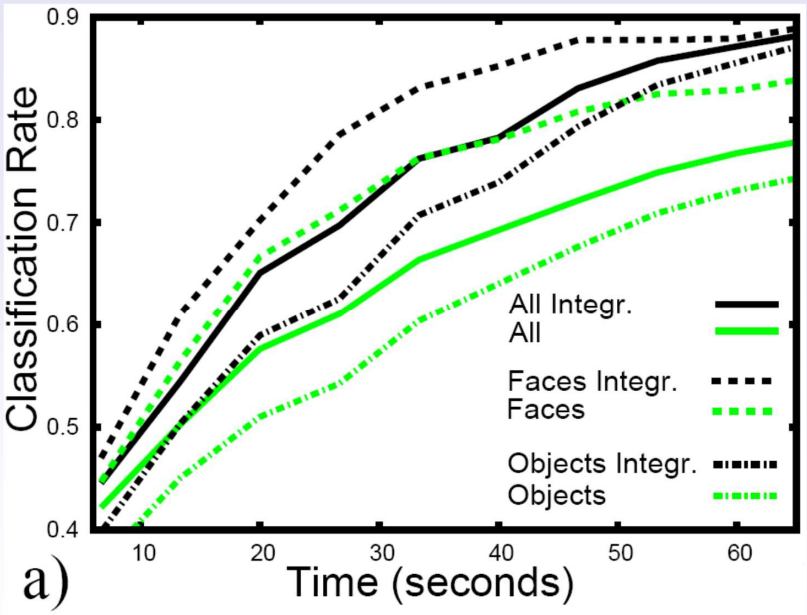


Set of 15 objects + 10 Faces



ROI Size: 144x144 pixels, Object size: 30-100 pixels

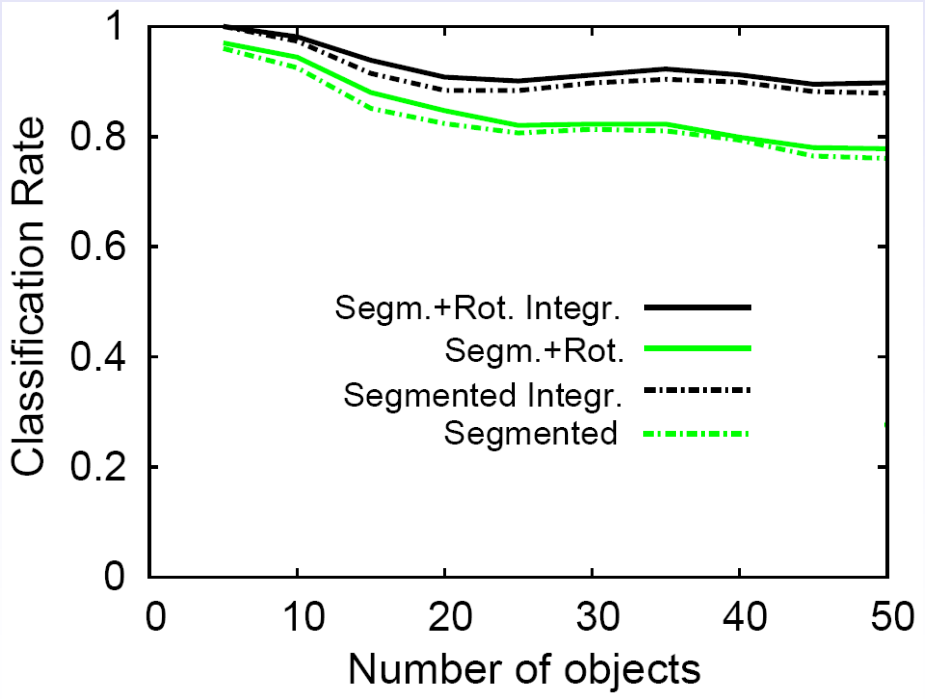
Online performance



Average classification rate for training the 15th object, after 14 were already trained

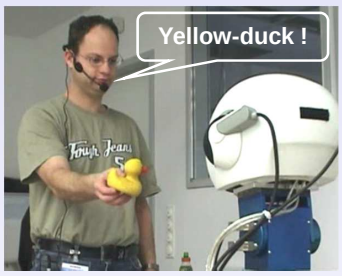
Scaling to 50 objects

Scaling to 50 objects



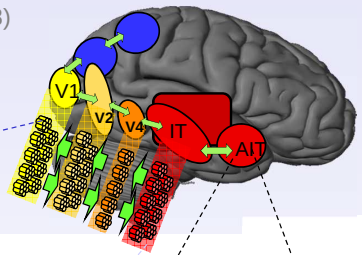
- → BBC2 Video at 38:00

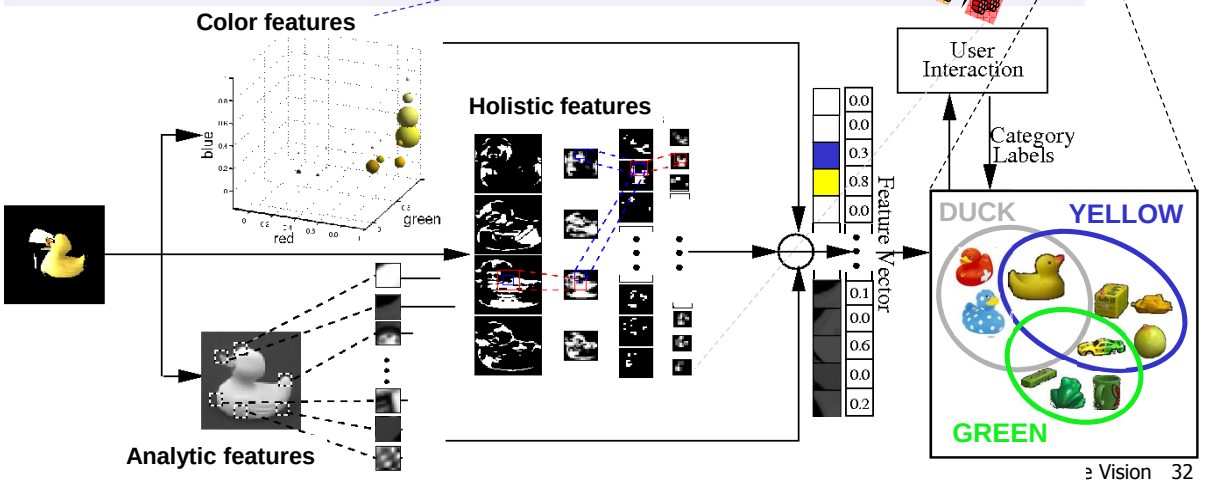
Online Category Learning



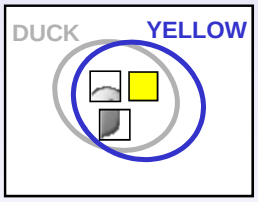
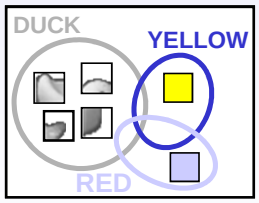
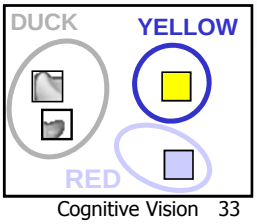
Demo: First realization of online learning of multiple category properties for several complex objects

Kirstein, Wersing, Gross, & Körner (2008)
Best paper award at Int. Conf. on Neural Information Processing 2008, (of 260 accepted papers)





Feature Selection For Category Learning

	Image Input	Feature Input	BRAVO System Output	User Training	BRAVO Category Representation
Step 1			"unknown category"	"yellow duck"	
Step 2			"yellow duck"	"red duck"	
Step 3			"yellow"	"OK !"	

Category training

BRAVO Online Category Learning Result

BRAVO Active object localization

Recognition for driver assistant systems