An Integrative view of Cognitive Architecture: Process/Structure/Behaviours

Gordon Cheng
Department of Humanoid Robotics and Computational Neuroscience,
ATR Computational Neuroscience Laboratories
and
JST-ICORP Computational Brain Project
Japan Science and Technology Agency,
Kyoto, Japan

Department of Humanoid Robotics and Computational Neuroscience (HRCN)

Department Head/Group Leader:
Gordon Cheng (ATR/ICORP)
Assistants:
Toyoko Morihisa
Keiko Katsushima

Human Movements Transfer
Stefan Schaal (USC)
Auke Ijspeert (EPFL)
Jessica Hodgins (CMU)
Chitomi Kadowaki (NAIST)
Yuka Ariki (NAIST)
Marcia Riley (GeorgiaTech)

Tasks and Skills Transfer
Christopher Atkeson (CMU)
Darrin Bentivegna (ICORP)
Nancy Pollard (CMU)

Understanding Human Activities
Ales Ude (JSI)
Erhan Oztop (ICORP)
Joshua Hale (ICORP)
Julie Digne (ENST)
Aude Billard (EPFL)
Kai Welke (Karlsruhe)
Thomas Gumpp (Karlsruhe)
Pedram Azad (Karlsruhe)

Biologically-inspired Biped Locomotion
Jun Morimoto (ICORP)
Jun Nakanishi (ICORP)
Gen Endo (SONY)
Sang-Ho Hyon (ICORP)
Takamitsu Matsubara (NAIST)

Motivation

- How do humans handle all different types of interaction with ease - in such a competent manner?
- How such a rich system can be built?
- What are the underlying mechanisms?
  - Understanding the Processes and controls
- Explore and exploit what has been learnt from other fields of discipline: philosophy, neuroscience, psychology, physiology etc.

Our View: Understanding Through Creating

- we share the view that human-like response can only be yielded through a richly integrated humanoid system
- an appropriate tool will automatically enhance the way we think [Dennett, Gregory] - thus, a humanoid will help us think deeper about humans and their behaviours
- gain a better understanding through reproducing Human-like behaviours on Humanoid Robots

Human Activities Motivation

- involve many sensory receptors:
  - seeing,
  - hearing,
  - touching etc.
- bombard us continuously on multiple basis [Berthoz, Damasio, Dennett, Kelso]
- producing varieties of similar responses

An Integrated Human-oid

- Why an integrated system?
  - Humans are complete systems, therefore it only makes sense to conduct our studies with an Integrated System
- What do we get from an integrated system like humans?
  - make available a rich sensory system for research – with concurrent and continuous processing
  - enable to respond to the large and numerous types of inputs with a wide range of responses - in a richer manner
  - ensure mutual human-like interaction is embraced (due to expectation)

An Integrated Human-oid

- What do we get from an integrated system like humans?
  - provide a vast number of alternatives in which problems can be solved, e.g. exploiting the rich sensory system available
  - a diverse system will yield a highly robust and redundant system - allowing it to adjust to failure
  - extend the boundary and limitation of current methodology – increasing the variation of possible solutions

Consideration for Humanoid Interaction

- Integrated point of view
  - seamlessness/continuity
    - which allows the system to interact with the environment in a continuous manner. The system should be able to interact continuously in a responsive and timely manner, thus without stopping or re-starting.
  - adaptivity/redundancy
    - the mechanism used should provide the system with redundancy by being adaptive. The system should be able to adapt to failure of sub-systems or components, for example if one component fails another is able to function without any dependency. Thus, a system will be able to support Self-Preservation.

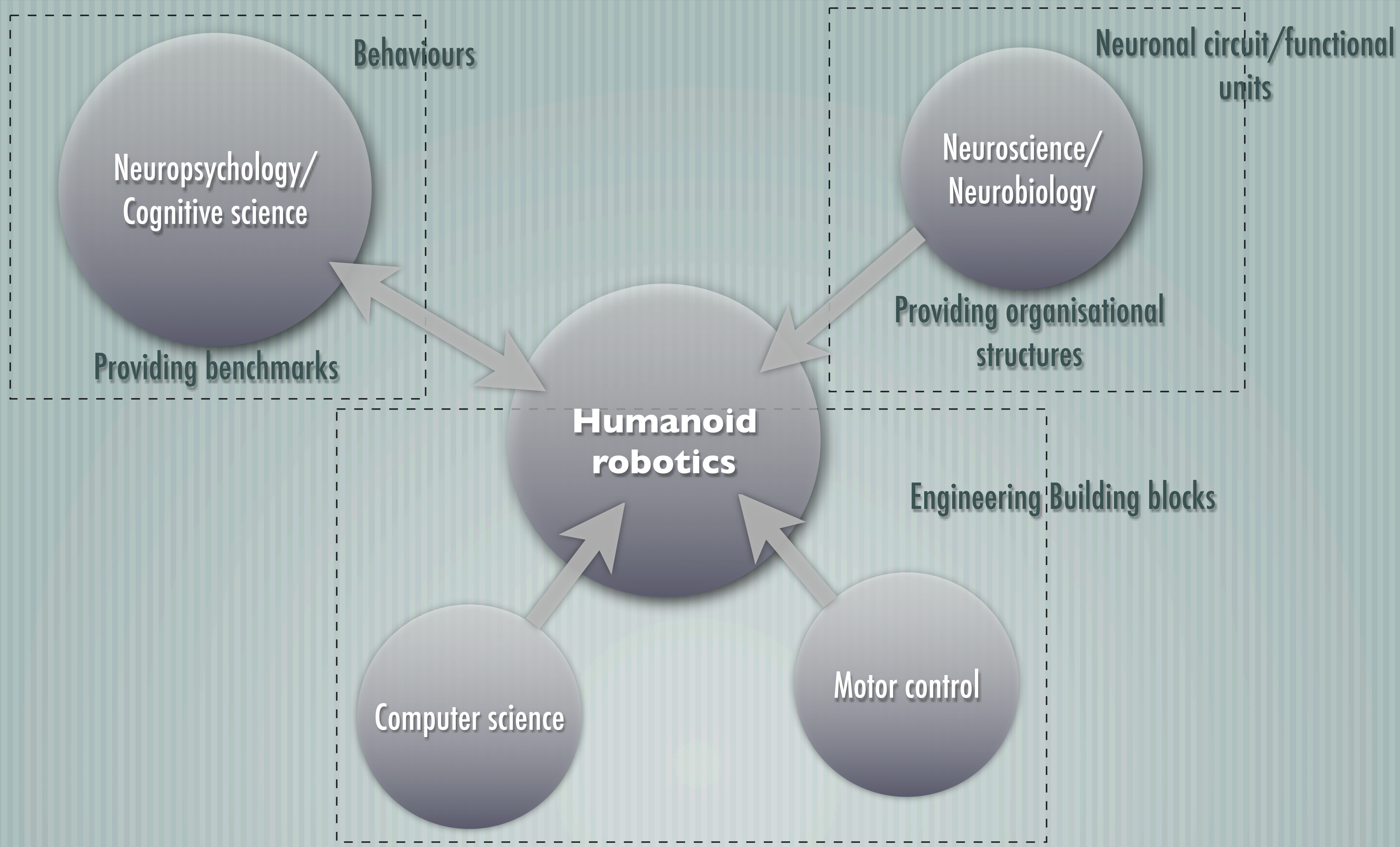
Consideration for Humanoid Interaction

- dominance/competition/cooperation
  - channelling all inputs into an integrated system, in a competitive manner. [Damasio, Dennett, Kelso]
- flexibility
  - additional components should be able to be introduced in a simple manner. Flexibility is an important attribute for the process of integration. E.g. Neural-based architecture

Emulating: ecologically driven

- natural environment
  - the environment in which the humanoid occupies should remain unmodified
    - unmodified in anyway to accommodate for any special perceptual need.
- multiple input/multiple output
  - consideration should be taken into account for a large number of sensors, and a wide range of concurrent responses should be exhibitable.

Multi-disciplinary integration



Biological-based framework for the studies of Human

- “Beyond simple Neural Network modelling”: Neuroscience/Neurobiology providing guideline for the low-level construction of “modulars/pathways” - allows spatial-temporal studies of organisational structures
- “Beyond simple behaviours” as Benchmarking: behavioural outcomes can be taken as benchmarks, evaluated within nature environment - ecological
- “Systematic studies of complex models”: even for impairments and defects - “lesion studies”

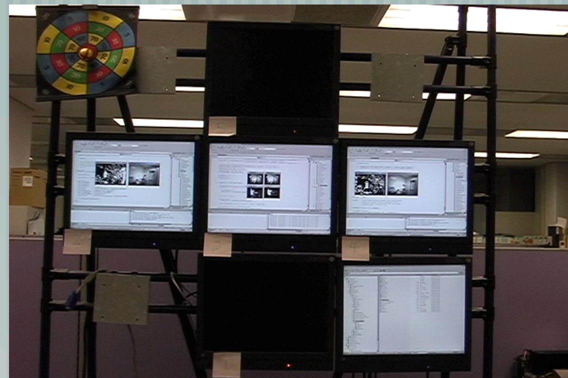
Integration is about process structure

- Naturally the process elements/modules plays a key role
- The underlying structure being important for:
  - Behavioural outcome of the system
  - Facilitate the understanding of the overall system/sub-modules
- More recent ideas tends to converge on the idea that behaviours being the end goal, without accounting for internal organisations

Key aspects

- Temporal adjustment: controlling of delays
- Allow connection between functional units: mechanisms for bottom-up inputs and top-down influence to bottom-up processing
- Providing: rich inputs (low-level inputs, visual/auditory/etc); and diverse outputs (control of motor);

Exploring Cognition and beyond

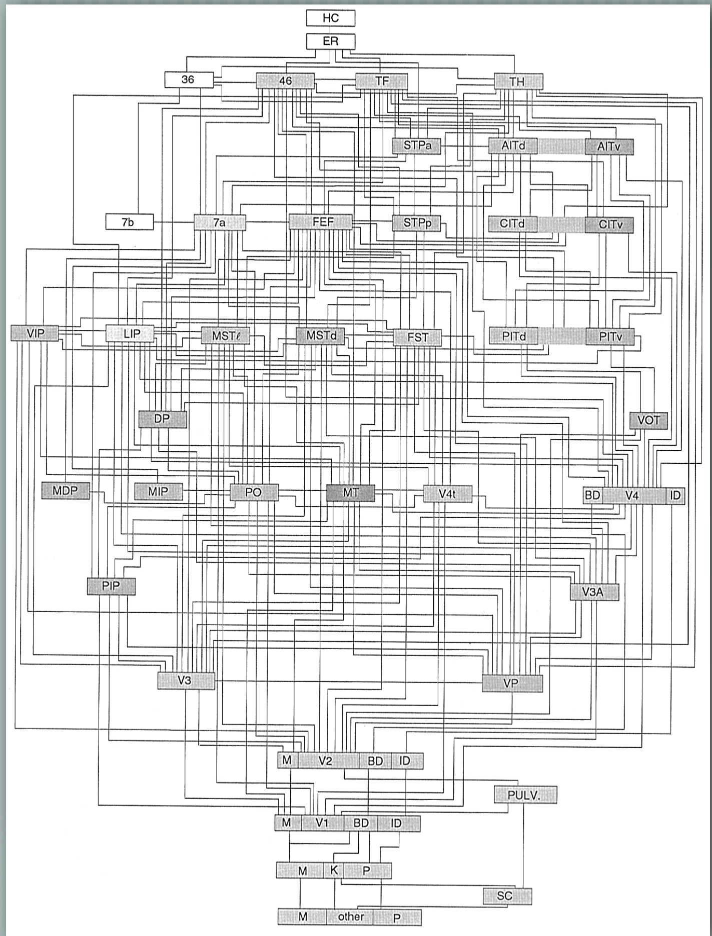


- A framework for combining Brain/Body for:
  - Perception
  - Action
  - Learning

Human-oid Response

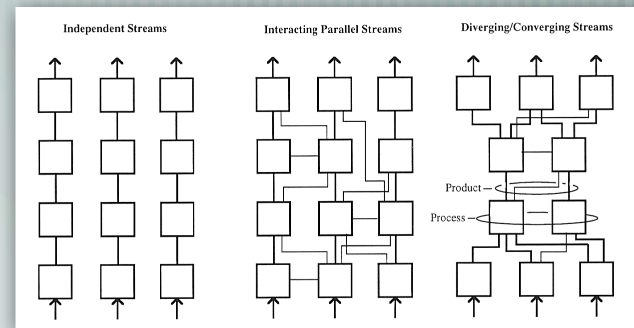
- Having the body of a human automatically gains:
  - Induce and deduce expectation of human-like response(s)

Scaleable and Distributed Computation Architecture



- The supporting architecture need to be flexible/scaleable to allow the construction of complexly systems

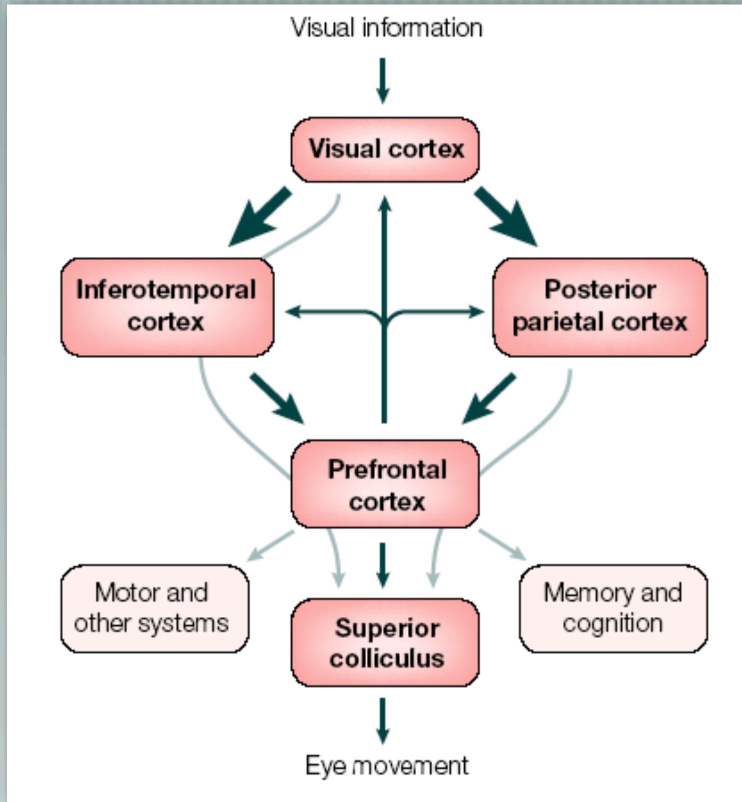
- Example: Hierarchy of the visual areas in the macaque - from "Concurrent Processing in the Primate Visual Cortex" by Van Essen and Deyoe, 1992



Looking at Synthesising Neuronal mechanisms

- Cortical areas along the ‘dorsal stream’ (including the posterior parietal cortex; PPC) are primarily concerned with spatial localisation and directing attention and gaze towards objects of interest in the scene

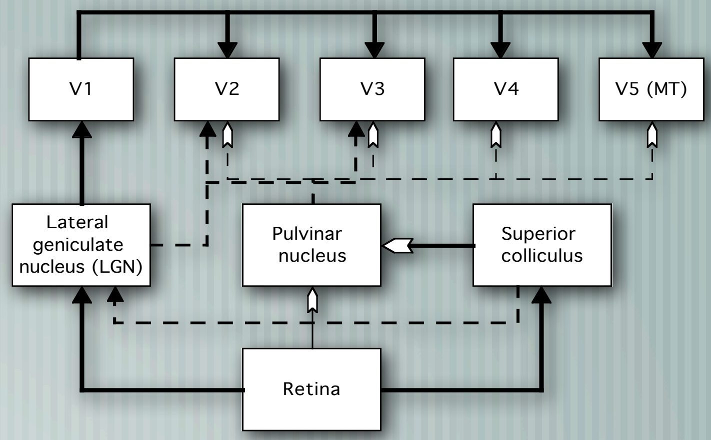
- Cortical areas along the ‘ventral stream’ (including the inferotemporal cortex; IT) are mainly concerned with the recognition and identification of visual stimuli



Itti et. al.

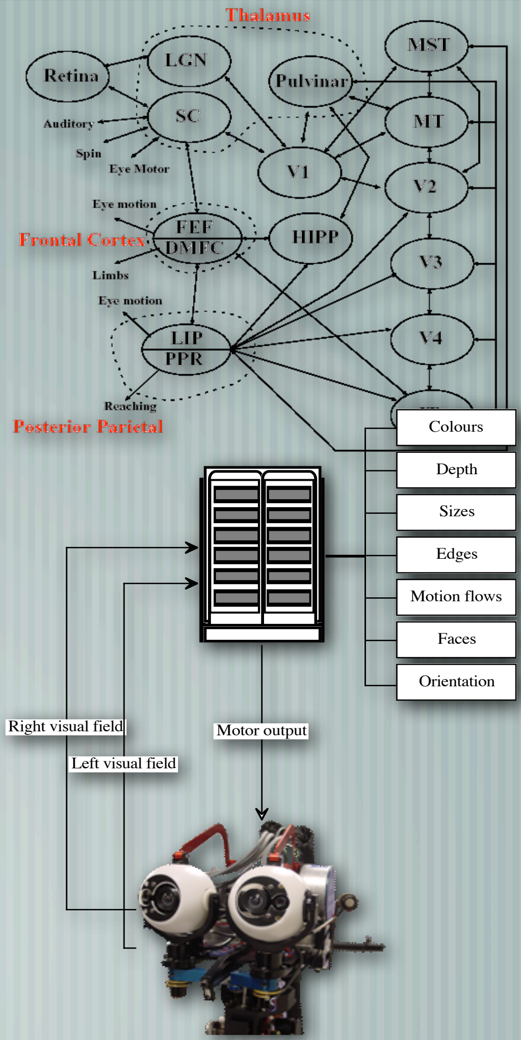
Network-based vision processing

- Attempt to emulate parts of the visual pathways
  - V1 (Primary visual cortex):
    - Retinal disparity (difference in the relative position of the stimuli in the visual field of the two eyes)
    - Colour
  - V5:
    - Motion and depth
- Initially simplified into two pathways:
  - What: e.g. V1 projection to the temporal lobes



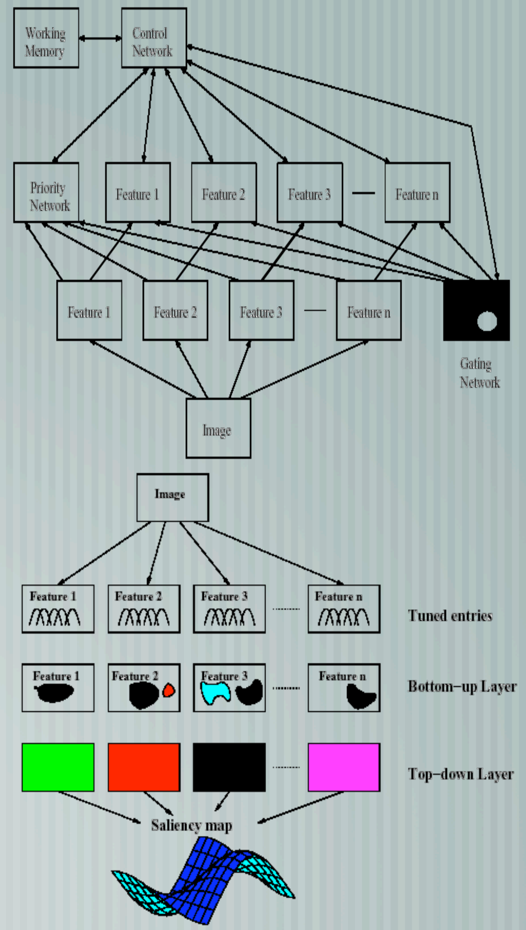
Distributed architecture

- Synthesising the human visual system
- To make better sense of human sensory processing
- More closely match the ways people process visual stimuli
- Network-based: nodes are connected together to form the overall structure
- Specialised vision processing at each node: colour, motion, depth, etc
- Basic system, a feature-based visual mechanism

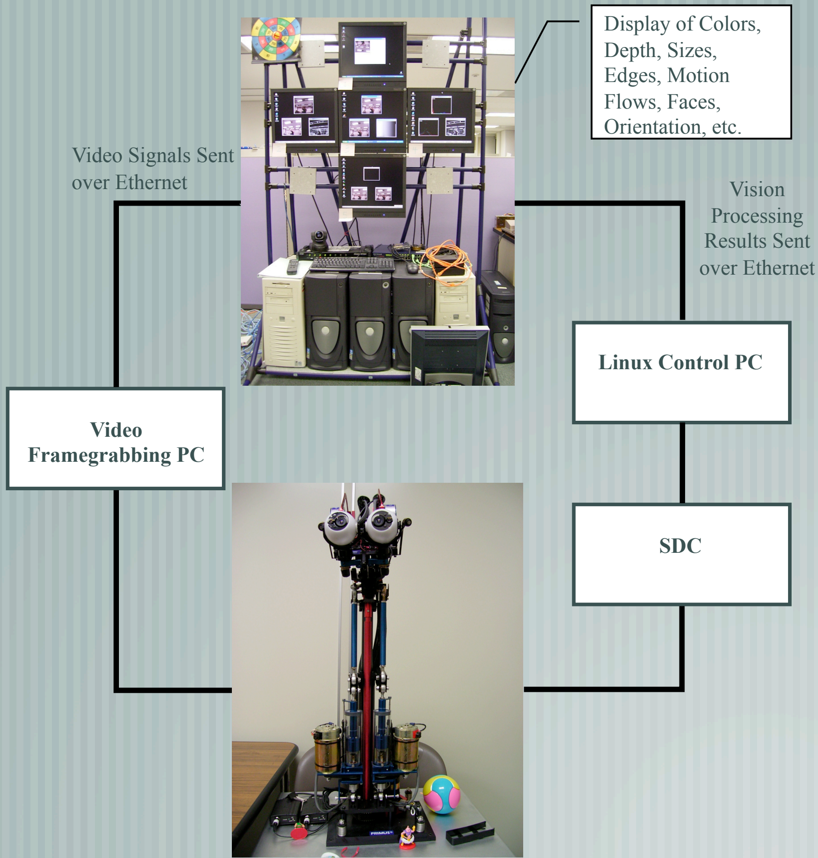


Feature-based network

- Channelling features together
    - For object encoding
- Bi-directional: encoding/ decoding
- Start with performing spatial orientation towards stimulus

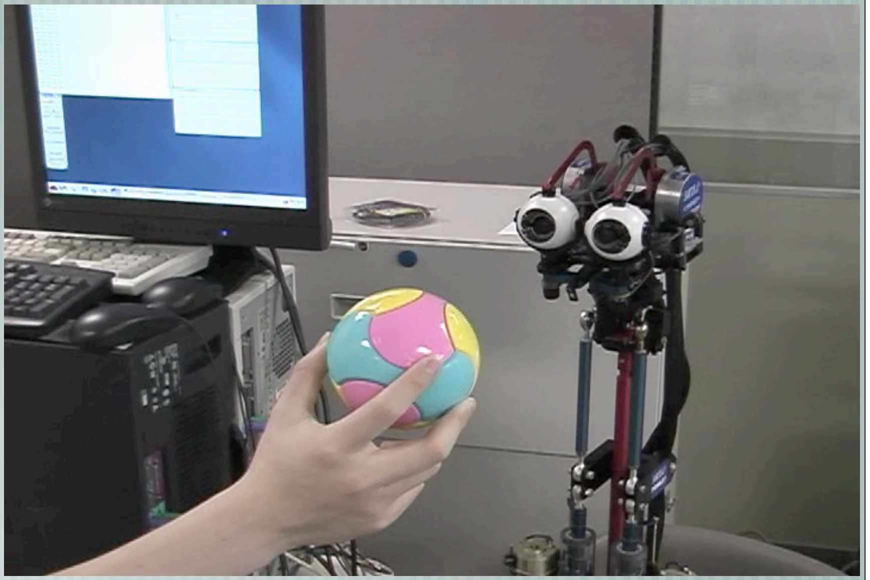


Overall structure



Distributed Architecture

- Distributed processing
  - Connection oriented i.e. network of connection
- Low latency
- Scalable across platform
- Distributed
  - Sensory processing
  - Control (work across robotics platforms)

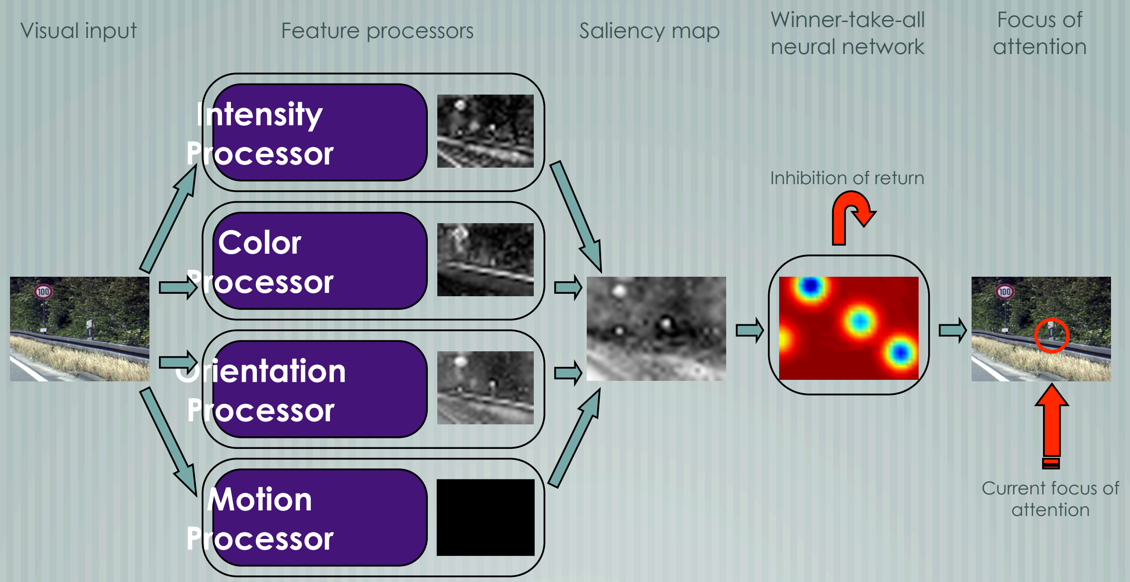


Gordon Cheng et. al.

Saliency-based Visual Attention System

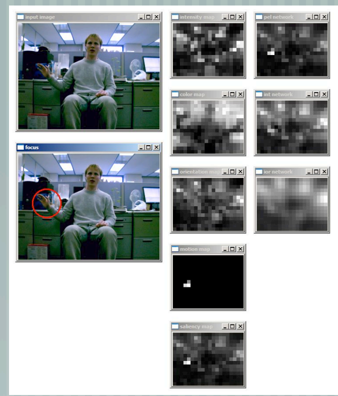
- We want a robot to be able to interpret a complex scene within a small amount of time, by focusing its processing power on a limited region, selected as interesting, at a time.
- We build a visual attention system inspired by the architecture and the behaviour of the primate primary visual cortex (V1).
  - General architecture originally proposed by L. Itti and C. Koch (1998).
  - By interesting, we mean considered as salient by the receptive cells of V1.

Visual Attention processing Model



Biologically motivated perception and action

- Gabor Filters have similar shape as the receptive fields of simple cells in the visual cortex
- Utilised for visual attention (dorsal stream/where pathway)
- Utilised for early vision to support recognition (ventral stream/what pathway)



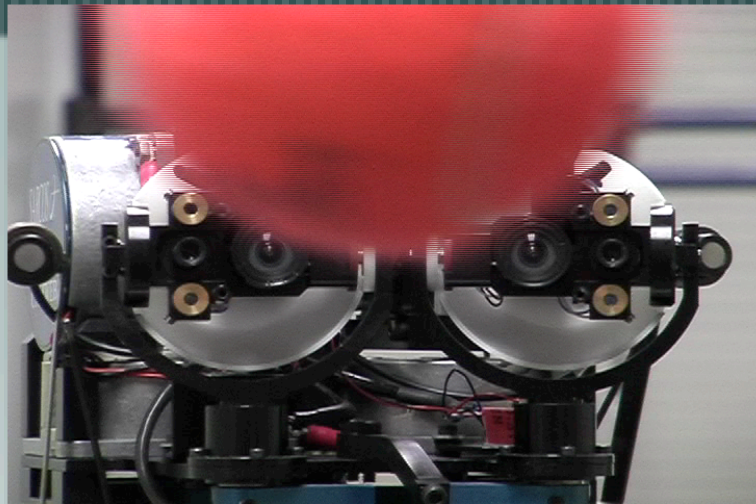
Ales Ude et. al.

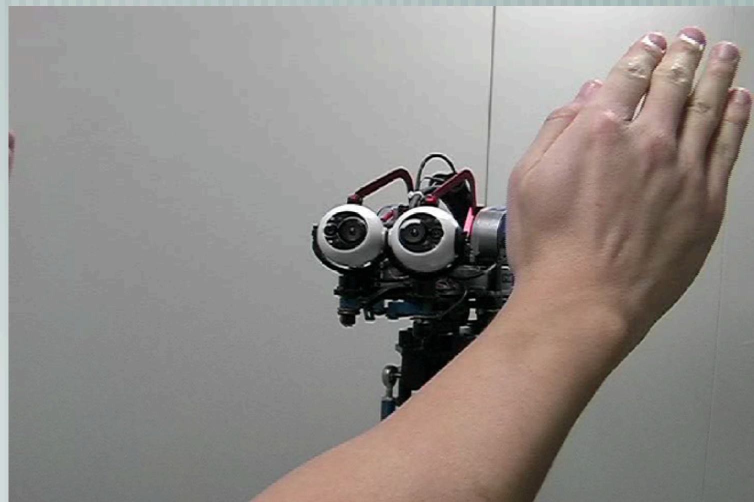
Active Vision Outline

- Version (keeping the target centered in the visual field of each eye)
- Vergence (minimizing target disparity by symmetric eye movement)
- Saccadic eye motion (quick knee-jerk type eye movements to detected motion)
- VOR (gaze stabilisation by compensating for externally induced head movements)
- Coupling of eye movements with head movement
- Saccade followed by smooth pursuit of target in an integrated control environment

Eyes movements

- Eye movements
  - Smooth pursuit (visual tracking)
  - Version
  - Vergence
  - Saccadic eye movement
- Visual processing:
  - Stereo vision
  - Motion detection
  - Colour processing





Gordon Cheng et. al.

Seeing & Reaching Development

- Human-like reaching
  - A head with two eyes
    - Look at the target for accuracy
    - See the hand when it nears the target
  - Reach using the whole body
  - Cope with change and disturbances
- Our implementation
  - Combine endpoint closed-loop and endpoint open-loop visual servoing
  - Learned motor-motor mapping

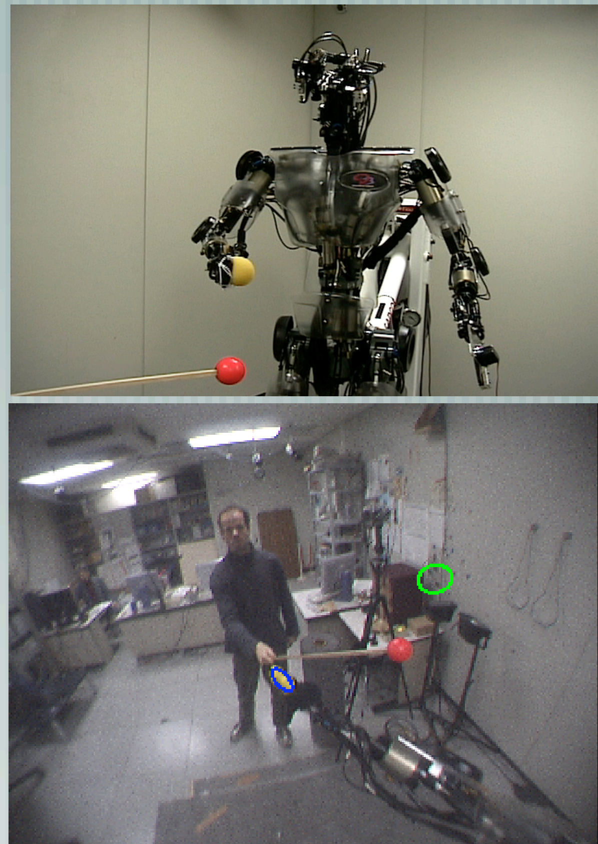
Development of Human Reaching

- Babies spend a lot of time reaching out and looking at their hands
- Perhaps a sensor-motor map between view of the hand and the position of the hand is learnt first... – retinal coordinate system
- Then, based on the sensor-motor map, a motor-motor map could be learnt that allows more awareness of the body e.g. proprioception, and reaching when the hand is not visible

  – joint coordinate system

Exploring Human-like Behaviour

- Reproducing Human-like Behaviour:
  - Seeing & Reaching
  - Real-time vision/Active Stereo Tracking
  - Full-body movements
    - To maximise coverage
- Developmental/Interactive approach
  - Learning through interaction
  - Motor-motor mapping
    - From Close-loop to Open-loop representations are learnt



Chris Gaskett et. al.

Higher-level vision

- Visual search for detection / visual attention, saccadic movements (peripheral vision)
- Tracking and smooth pursuit (peripheral / foveal vision)
- Recognition (foveal vision) and action

Visual goal selection

- Combining
  - Peripheral and Foveal vision
  - Colour Tracking
  - Object Recognition
- Better Recognition is Gained by using Foveal view of the object - through visual acuity

Ales Ude et. al.

Robotic Hand as a Trestbed

- Allows testing of various computational models on a physical platform
  - e.g. Imitation, skill acquisition via reinforcement learning, and affordance learning
- Our first project using the Gifu Hand infrastructure as a model testing tool:
  - LEARNING TO IMITATE

Ways to Learn to Imitate

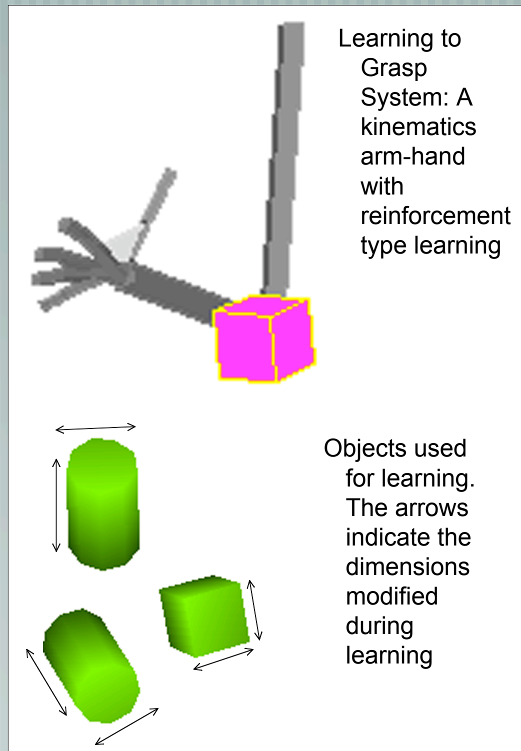
- Self observation
  - Agent produces action (A)
  - Agent sees consequence of the action (V)
  - Agent associates A and V
- Social learning
  - Through interacting with other agents in the environment.

SOCIAL LEARNING FOR IMITATION

- Social (reinforcement) learning
  - Agent observes action (V)
  - Agent generates an action (A)
  - If social reward is collected,
    - agent associates A and V
- Social (supervised) learning
  - Agent shows action (A)
  - Agent sees teacher’s imitation (V)
  - Agent associates A and V

Simulation of a Learning to Grasp system

- The monkey parietal area AIP has been shown to contain neurones that extract object features relevant for grasping such as the width and the height.
- AIP is strongly interconnected with the ventral pre-motor area F5, that is involved in grasp planning and execution.
- We present evidence supporting our simulation results from infant developmental studies.
- The focus of this study is to investigate the formation (emergence) of AIP neurons during grasp development



Erhan Oztop et. al.

Exploring brain mechanisms for dexterity

- Motion capture studies of human dexterity
- Emergent Grasp Affordance Encoding via Grasp Learning
  - "Affordances" are properties of the object relevant for action (for example grasping)
  - AIP cells encode "affordances" for grasping from the visual stream and sends these

    on to area F5
  - These ideas from biology will be tested on a real robotic system (Gifu Hand III)

Erhan Oztop et. al.

Associative memory hypothesis of imitation bootstrapping

- study a range of topics ranging from manual skill learning, imitation and affordance learning
- we have explored the "associative memory" hypothesis of imitation bootstrapping

Erhan Oztop et. al.

Conclusions

- A framework to support the investigation/development of Embodied Cognition has been proposed
  - Merging Brain and Body together
- An architecture for the construction of components and more importantly the connective organisation of these components
- Some initial results (at various levels) have been proposed/developed