RobotCub

Building a humanoid robotic platform

The 1st RobotCub Open Day
Genoa, Italy, July 14th 2005
Outline

• Our motivations
  – Why do we do what we do?
• Building what
  – A humanoid robot
• Our goals
  – Understanding cognition, building cognition
Two keywords

“Perception, cognition and motivation develop at the interface between neural processes and actions. They are a function of both these things and arise from the dynamic interaction between the brain, the body and the outside world”

Von Hofsten, TICS 2004
• Development: to replicate something requires to know how to build it
  - Corollary: “building” is not entirely like “understanding”

• Action: interaction in the real world requires a body
  - Corollary: the shape of the body determines the affordances that can be exploited
What is changing?
• The controller is changing, coordination is changing

• Konczak et al. for instance showed that it is not a problem of peak “torque” generation but one of control
Action is important
The perception of actions happens through the mediation of the action system

i.e. perception is not the private affair of the sensory systems
Active perception

LIRA-Lab, 1991 or so
Also, objects come to existence because they are manipulated.

Fixate target
Track visual motion…
(...including cast shadows)
Detect moment of impact
Separate arm, object motion
Segment object

Which edge should be considered?
Maybe some cruel grad-student glued the cube to the table

Color of cube and table are poorly separated
Cube has misleading surface pattern
Exploring an affordance: rolling

A toy car: it rolls in the direction of its principal axis

A bottle: it rolls orthogonal to the direction of its principal axis

A toy cube: it doesn’t roll, it doesn’t have a principal axis

A ball: it rolls, it doesn’t have a principal axis
An old video...
The MIRROR project

2 cameras

Frame grabbers

Images

To disk

Cyber-glove

RS232

Tracker

RS232

40 msec

Other sensors

To disk

Tactile sensors
Bayesian classifier

\{G_i\}: set of gestures  
\(F\): observed features  
\{O_k\}: set of objects

\[ p(G_i | O_k) \]: priors (affordances)  
\[ p(F | G_i, O_k) \]: likelihood to observe \(F\)

\[
p(G_i | F, O_k) = p(F | G_i, O_k) p(G_i | O_k) / p(F | O_k)
\]

\[ \hat{G}_{MAP} = \arg \max_{G_i} (G_i | F, O_k) \]
Two types of experiments

Learned by backpropagation ANN
Has motor information anything to do with recognition?

Object affordances (priors)

Visual space

Motor space

Classification (recognition)

Grasping actions
### Some results...

<table>
<thead>
<tr>
<th></th>
<th>Exp. I</th>
<th>Exp. II</th>
<th>Exp. III</th>
<th>Exp. IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(visual)</td>
<td>(visual)</td>
<td>(visual)</td>
<td>(motor)</td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Sequences</td>
<td>16</td>
<td>24</td>
<td>64</td>
<td>24</td>
</tr>
<tr>
<td># of view points</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Classification rate</td>
<td>100%</td>
<td>100%</td>
<td>97%</td>
<td>98%</td>
</tr>
<tr>
<td># Features</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td># Modes</td>
<td>5-7</td>
<td>5-7</td>
<td>5-7</td>
<td>1-2</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Sequences</td>
<td>8</td>
<td>96</td>
<td>32</td>
<td>96</td>
</tr>
<tr>
<td># of view points</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Classification rate</td>
<td>100%</td>
<td>30%</td>
<td>80%</td>
<td>97%</td>
</tr>
</tbody>
</table>
In all communication, sender and receiver must be bound by a common understanding about what counts; what counts for the sender must count for the receiver, else communication does not occur. Moreover the processes of production and perception must somehow be linked; their representation must, at some point, be the same.”

[Alvin Liberman, 1993]
The ultimate constituents of speech are articulatory gestures (one and the same thing, one concept to rule them all)
Mirror neurons?

<table>
<thead>
<tr>
<th>Vision</th>
<th>Acoustic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulation</td>
<td>Speech</td>
</tr>
<tr>
<td>Motor</td>
<td>Motor</td>
</tr>
<tr>
<td>Watching others</td>
<td>Listening to others</td>
</tr>
</tbody>
</table>
Manipulation, i.e. taking actions → speech
The iCub

• Requirements
  - Hands to manipulate
  - Arms with a large workspace
  - Head with fast camera movements
  - Waist and legs for crawling
• Able to crawl & reach to fetch objects and sit to manipulate them
• Child-like size
Child-like, how much?

Approx 934mm

439mm

369mm

243mm

Avg. 14Kg - 30.8 lb
Well...

- It is going to be heavier: ~23Kg
- 53 degrees of freedom
  - 9 x2 hands
  - 7 x2 arms
  - 6 head
  - 6 x2 legs
  - 3 torso
- Embedded electronics
Sensors

- Cameras
- Microphones
- Gyroscopes, linear accelerometers
- Tactile sensors
- Proprioception
- Torque sensors
- Temperature sensors
Implementation of the cognitive architecture
...and, yes, it is open!

- GPL for all the software, controller, tools, everything that runs on the robot
- FDL for the drawings, electronics, documentations, etc.
- Open to new partners and collaborations worldwide
Meet the iCub
See you in March 2007!