Perception & Cognition: Two Foremost Ingredients toward Intelligent Robots

Prof. Kurosh Madani

Images, Signals & Intelligent Systems Lab. (LISSI / EA 3956)
University PARIS-EST Creteil (UPEC) / Sénart-Fontainebleau Institute of Tech.
Avenue Pierre Point - 77127 Lieusaint – France
madani@u-pec.fr
Univ. PARIS-EST & UPEC

Paris Region

50000 Students
(33000 + 17000)

UPEC

UPEM

Senart-FB Institute of Technology
UPEC / LISSI Lab.

Laboratory of Images Signals & Intelligent Systems
IMS Team of LISSI Lab.

Dr. V. Amarger
Dr. A. Chebira
Dr. A. Chohra
Dr. A. Hazan
Prof. K. Madani
Dr. C. Sabourin

T. Wang (P. R. China)
D. Ramik (Czech Republic)
L. Voysekhovich (Belarus)
A. Kabysh (Belarus)
J. Wang (P. R. China)
D. Kanzari (Tunisia)

Dr. L. Thiaw (Senegal)
Dr. M. Sene (Senegal)
Dr. W. Yu (China)
Dr. A. Bahrammizae (Iran)
IMS team of LISSI (Intelligent Machines & Systems) works on exploitation of bio-inspired mechanisms in order to realize and implement « intelligent artificial systems »: application to Real-World problems.

Privileged areas & applications are:

1. complex information processing
2. “industrial” & “Real-World” problems
3. modeling & implantation of complex systems (Humanoid robots, Collective & Socials Systems, Self-Organizing Systems, etc...).
The robots’ (and machines’) autonomy is one of primary challenges of the present decade.

Such robots have to be self-sufficient enough in order to be able to operate in cooperation with users who have no a-prior technical skills.

Among vital capabilities for such robot:

✓ understanding its environment
✓ deciding (& modifying) actions regarding its environment
“Cognition”: refers to the ability for the processing of information applying knowledge.

In the field of computer science: often intends the artificial intellectual activities and processes relating the “machine awareness” realized knowledge-based “intelligent” artificial functions.

However, “awareness” & “knowledge construction” require ability to perceive information from surrounding environment.
Thus, “Cognition” and “Perception” remain inseparable ingredients toward autonomous machines (robots…)

Inspired from human’s “early-ages cognitive skills” development:

Multi-layer cognitive “Perception-Motion” architecture for autonomous robots
The “human’s walking” ability is not an automatic process but a cognitive development since the early ages.

(Dr. Marianne Barbu-Roth : CNRS – Univ. Paris Descartes)
Cognitive Perception-Motion Scheme

Conscious Cognitive Level

Unconscious Cognitive Level

Conscious Cognitive Visual Functions (Conscious Cognitive Level)

Unconscious Cognitive Visual Functions (Unconscious Cognitive Level)

Conscious Cognitive Motion Functions (Conscious Cognitive Level)

Unconscious Cognitive Motion Functions (Unconscious Cognitive Level)

Robot-CCF Channels

Robot-UCF Channel

Inter-levels Channel

Perception

Motion
Cognitive level & Elementary Functions

Cognitive Level

Elementary Function k

Action Elementary

Action Elementary

Action Elementary Component (AEC)

AEC: Action Elementary Component

DEC: Decision Elementary Component

Input: \( \Psi_k \)

Output: \( O_k \)
**Cognitive level & Elementary Functions**

**Elementary Function k**

**Input:** $\Psi_k$

**Action Elementary**

**Action Elementary Component (AEC)**

**Output:** $O_k$

**Learning process**

**Training model’s parameters**

$\varepsilon (O_k, O^d)$

$O^d$ Desired Output
Unconscious Cognitive Function Example

"Average-Velocity’s control UCF"

Trunk’s angular position’s control

Update unit DEC

"flat-ground walking" UCF

AEC

CMAC

CMAC

CMAC

CMAC

CMAC

Fuzzy DEC

"Unconscious Cognitive Function Example"
Unconscious Cognitive Function Example

- IF $V_M$ is Very-Small THEN $Y = o_1$
- IF $V_M$ is Small THEN $Y = o_2$
- IF $V_M$ is Medium THEN $Y = o_3$
- IF $V_M$ is Big THEN $Y = o_4$
- IF $V_M$ is Very-Big THEN $Y = o_5$

<table>
<thead>
<tr>
<th></th>
<th>$V_M$ (m/s)</th>
<th>$q_0$ (°)</th>
<th>$q_{1j}$ (°)</th>
<th>$q_{2j}$ (°)</th>
<th>$L_{step}$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMAC$^1$</td>
<td>0.4</td>
<td>20</td>
<td>-7</td>
<td>3.5</td>
<td>0.23</td>
</tr>
<tr>
<td>CMAC$^2$</td>
<td>0.5</td>
<td>25</td>
<td>-10</td>
<td>3</td>
<td>0.28</td>
</tr>
<tr>
<td>CMAC$^3$</td>
<td>0.6</td>
<td>30</td>
<td>-15</td>
<td>2.5</td>
<td>0.31</td>
</tr>
<tr>
<td>CMAC$^4$</td>
<td>0.7</td>
<td>35</td>
<td>-20</td>
<td>8</td>
<td>0.36</td>
</tr>
<tr>
<td>CMAC$^5$</td>
<td>0.8</td>
<td>40</td>
<td>-25</td>
<td>8</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Unconscious Cognitive Function Example
Emergent Behavior

Walking before force perturbation

Walking after force perturbation

Time (s)

Average velocity

Desired velocity

Velocity (m/s)
Generalization: Same structure

- IF Slope is Big-Negative THEN $Y = o_1$
- IF Slope is Small-Negative THEN $Y = o_2$
- IF Slope is Zero THEN $Y = o_3$
- IF Slope is Small-Positive THEN $Y = o_4$
- IF Slope is Big-Positive THEN $Y = o_5$
Perception (Vision) Scheme

Conscious Cognitive Visual Functions (Conscious Cognitive level)

Unconscious Cognitive Visual Functions (Unconscious Cognitive level)

Robot-UCF Channel

Robot-CCF Channels

Inter-levels Channel

Unconscious Cognitive Level

Conscious Cognitive Level

new image acquired in different conditions

object detector

incremental learning

unknown object

perceived image

saliency map

salient objects

extraction

weak classifier set

extracted object

noise

acquired knowledge use

object detected on a new image

Unconscious

Conscious
We define “Luminance Saliency” as:

$$S_Y(x, y) = \| I_{\mu Y} - I_{\omega hc \ Y}(x, y) \|$$

We define “Chromatic Saliency” as:

$$S_{CC}(x, y) = \| I_{\mu CC} - I_{\omega hc \ CC}(x, y) \|$$

in YCrCb color space: “Y” for Y channel and “CC” for Cr & Cb channels.
Salient objects detection

Final Saliency: Neighborhood

\[ S_{final}(x, y) = f \left( S_Y, S_{CC}, \text{Neighbours- pixels} \right) \]

Original image (left), Salient map example (middle), salient objects (right)
Salient Objects’ Finding

Original image (left), Salient region detection (middle), segmentation (right)

Examples of salient objects’ detection
Learning & Knowledge Construction

**Acquire** image

**Extract** fragments by salient object detector

**For** each fragment $F$

- **If** ($F$ is classified into one group) **Then**
  - populate the group by $F$
- **If** ($F$ is classified into multiple groups) **Then**
  - populate by $F$ the closest group by Euclidian distance of features
- **If** ($F$ is not classified to any group) **Then**
  - create a new group and place $F$ inside

**Select** the most populated group $G$

**Use** fragments from $G$ as learning samples for object detection algorithm
Classifier

A combination of 4 classifiers: \( \{w_1, w_2, w_3, w_4\} \)

Each classifying a fragment “F” as “belonging” or “not belonging” to a certain class

**Area** \( w_1 \): separates “too different” areas.

**Aspect** \( w_2 \): separates fragments with “too different” aspects.

**Chromaticity** \( w_3 \): separates fragments with “too different” chromaticity.

**Texture** \( w_4 \): separates fragments with “too different” textures.
Some Results

Table 1. Percentage of correct detections of object over testing image set using Viola-Jones detection framework.

<table>
<thead>
<tr>
<th>Viola-Jones</th>
<th>apple</th>
<th>beer</th>
<th>coke</th>
<th>khepera</th>
<th>mouse</th>
<th>mug</th>
<th>pda</th>
<th>shoe</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of correctly</td>
<td>98.0</td>
<td>88.2</td>
<td>77.3</td>
<td>60.0</td>
<td>76.0</td>
<td>89.1</td>
<td>80.0</td>
<td>81.8</td>
</tr>
<tr>
<td>detected instances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examples of objects’ detection in current environment.
Examples of objects’ detection in current environment.
Conscious Cognitive Function
Example
Implementation on NAO Robot

The process

1) get an image from Nao's camera
2) find the desired object on it using image processing algorithms
3) approximate its distance and position with respect to Nao with some human inspired algorithms
4) fetch the object

NB: Nao uses its voice synthetiser to inform us about its actions ...
Implementation on NAO Robot

Learning a book (passive scenario):
The object is shown to the robot by the user

Recognition results: searching the book

the book is extracted and learned

Recognition results: searching the book

robot >> I am near the book
Final Words

Machine intelligence Architecture based on :
“Cognition-Perception” duality

- obtained results show viability of the proposed architecture
- obtained results show also the pertinence of the used methods

Probably a promising aspect is related to “Emergent behaviour”: part of “intelligent skill”
However, today, the most important is of course:

Your Opinion
Thank you for your attention