Deep Hierarchies in Human and Computer Vision

Norbert Kruger
University of Southern Denmark
Cognitive and Applied Robotics Group
Overview

• Some annoying prior remarks

• The primate’s vision system: A deep Hierarchy

• SotA and Problems of research on deep hierarchical systems

• Reflections
Deep Hierarchies in the Primate Visual Cortex: What Can We Learn For Computer Vision?

Norbert Krüger, Peter Janssen, Sinan Kalkan, Markus Lappe, Aleš Leonardis, Justus Piater, Antonio J. Rodríguez-Sánchez, Laurenz Wiskott

Abstract—Computational modeling of the primate visual system yields insights of potential relevance to some of the challenges that computer vision is facing, such as object recognition and categorization, motion detection and activity recognition or vision-based navigation and manipulation. This article reviews some functional principles and structures that are generally thought to underlie the primate visual cortex, and attempts to extract biological principles that could further advance computer vision research. Organized for a computer vision audience, we present functional principles of the processing hierarchies present in the primate visual system considering recent discoveries in neurophysiology. The hierarchal processing in the primate visual system is characterized by a sequence of different levels of processing (in the order of ten) that constitute a deep hierarchy in contrast to the flat vision architectures predominantly used in today’s mainstream computer vision. We hope that the functional description of the deep hierarchies realized in the primate visual system provides valuable insights for the design of computer vision algorithms, fostering increasingly productive interaction between biological and computer vision research.

Index Terms—Computer Vision, Deep Hierarchies, Biological Modeling

1 INTRODUCTION

The history of computer vision now spans more than half a century. However, general, robust, complete satisfactory solutions to the major problems such as large-scale object, scene and activity recognition and categorization, as well as vision-based manipulation are still beyond reach of current machine vision systems. Biological visual systems, in particular those of primates, seem to accomplish these tasks almost effortlessly and have been, therefore, often used as an inspiration for computer vision researchers.

Interactions between the disciplines of “biological vision” and “computer vision” have varied in intensity throughout
Flat versus deep Hierarchies

**Deep Hierarchy**

<table>
<thead>
<tr>
<th>Task V1</th>
<th>Task V2</th>
<th>Task V3</th>
<th>Task Vn</th>
<th>Task D1</th>
<th>Task D2</th>
<th>Task D3</th>
<th>Task Dn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 5 (ventral)</td>
<td>Level 5 (dorsal)</td>
<td>Level 4</td>
<td>Level 3</td>
<td>Level 2</td>
<td>Level 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Flat Hierarchy**

<table>
<thead>
<tr>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
<th>Task 5</th>
<th>Task 6</th>
<th>Task 7</th>
<th>Task 8</th>
<th>Task n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some kind of Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Overview

• Some annoying prior remarks

• The primate’s vision system: A deep Hierarchy

• SotA and Problems of research on deep hierarchical systems

• Reflections

The Nobel Prize in Medicine 1981

David Hubel and Torsten Wiesel
Some remarks on the interaction of human vision research and computer vision

- David Marr 1982: Vision: A computational investigation into the human representation and processing of visual information
- 3 Stages
  - Primal Sketch: Multi-scale Edge Detection
  - 2.5D Sketch: Viewer centered Scene Representation
  - 3D Sketch: Object Centered Representation
<table>
<thead>
<tr>
<th>Viewer centred</th>
<th>Object centred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Image</td>
<td>3-D Model Representation</td>
</tr>
<tr>
<td>Perceived intensities</td>
<td>3-D models hierarchically organised in</td>
</tr>
<tr>
<td></td>
<td>terms of surface and volumetric</td>
</tr>
<tr>
<td></td>
<td>primitives</td>
</tr>
<tr>
<td>Primal Sketch</td>
<td>Local surface orientation and</td>
</tr>
<tr>
<td></td>
<td>discontinuities in depth and in</td>
</tr>
<tr>
<td></td>
<td>surface orientation</td>
</tr>
<tr>
<td></td>
<td>Zero crossings, blobs, edges, bars,</td>
</tr>
<tr>
<td></td>
<td>ends, virtual lines, groups, curves</td>
</tr>
<tr>
<td></td>
<td>boundaries.</td>
</tr>
</tbody>
</table>

**Input Image**
- Edge Image
- 2\(\frac{1}{2}\)-D Sketch
- 3-D Model

**Diagrams**
- Input image
- Edge image
- 2\(\frac{1}{2}\)-D sketch
- 3-D model
Why did that ‘fail’? Two reasons

• The project was too ambitious at Marr’s time
  • Lack of knowledge on low-level modalities
    • Optic flow
    • Edge detection
    • Stereo
    • Structure-from-Motion

• Lack of computational resources
  • Slow clock frequency
  • No GPUs
‘Computer Vision’ and ‘Biological Vision’

- In the 80th and 90th there was a strong link.
- This link has been kind of diluted from ‘both sides’
  - Computer Vision became a sub-discipline of Machine Learning
  - Many neurophysiologists have given up on understanding the brain on a functional level
- ‘Biologically inspired’ got a somehow bad reputation
  - Not efficient
  - Everything could somehow be biologically inspired
Maybe a restart is worthwhile

- Much better understanding of early vision
- Significantly larger computational resources
- Still many unsolved problems in CV
- Aim of the paper
  - Distill essential knowledge on the human visual system for Engineers
Overview

• Some annoying prior remarks
• The primate’s vision system: A deep Hierarchy
• SotA and Problems of research on deep hierarchical systems
• Reflections
Basic facts

- 55% of the neo-cortex of the primate brain is concerned with vision

- Devision in
  - Occipital Cortex
  - Dorsal Pathway
  - Ventral Pathway
Dr. Alesha Sivartha in the late 1800s (published in his metaphysical book *The Book of Life: The Spiritual and Physical Constitution of Man*)

From: van Essen 1992
Dr. Alesha Sivartha in the late 1800s (published in his metaphysical book *The Book of Life: The Spiritual and Physical Constitution of Man*)

From: van Essen 1992
Gall (1758–1828): Phrenology
# Basic Facts

<table>
<thead>
<tr>
<th>Area</th>
<th>Size (mm²)</th>
<th>RFS</th>
<th>Latency (ms)</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>CI/SI/PI/OI</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retina</td>
<td>1018</td>
<td>0.01</td>
<td>20-40</td>
<td>bl</td>
<td>+/-/-</td>
<td>+/-/-</td>
<td>sensory input, contrast computation</td>
</tr>
<tr>
<td>LGN</td>
<td>0.1</td>
<td></td>
<td>30-40</td>
<td>co</td>
<td>+/-/-</td>
<td>+/-/-</td>
<td>relay, gating</td>
</tr>
</tbody>
</table>

Sub-cortical processing

<table>
<thead>
<tr>
<th>Area</th>
<th>Size (mm²)</th>
<th>RFS</th>
<th>Latency (ms)</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>CI/SI/PI/OI</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>1120</td>
<td>3</td>
<td>30-40</td>
<td>co</td>
<td>+/+/-</td>
<td>+/-/-</td>
<td>generic feature processing</td>
</tr>
<tr>
<td>V2</td>
<td>1190</td>
<td>4</td>
<td>40</td>
<td>co</td>
<td>+/+/-</td>
<td>+/-/-</td>
<td>generic feature processing</td>
</tr>
<tr>
<td>V3/V3A/VP</td>
<td>325</td>
<td>6</td>
<td>50</td>
<td>co</td>
<td>+/+/-</td>
<td>+/-/-</td>
<td>generic feature processing</td>
</tr>
<tr>
<td>V4/VOT/V4t</td>
<td>650</td>
<td>8</td>
<td>70</td>
<td>co</td>
<td>+/+/-</td>
<td>+/-/-</td>
<td>generic feature processing / color motion</td>
</tr>
<tr>
<td>MT</td>
<td>55</td>
<td>7</td>
<td>50</td>
<td>co</td>
<td>+/+/-</td>
<td>+/-/-</td>
<td>motion</td>
</tr>
</tbody>
</table>

Occipital / Early Vision

<table>
<thead>
<tr>
<th>Area</th>
<th>Size (mm²)</th>
<th>RFS</th>
<th>Latency (ms)</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>CI/SI/PI/OI</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEO</td>
<td>590</td>
<td>3.5</td>
<td>70</td>
<td>co</td>
<td>(+/-/-/+)</td>
<td>+/-/-?</td>
<td>object recognition and categorization</td>
</tr>
<tr>
<td>TE</td>
<td>180</td>
<td>10-20</td>
<td>80-90</td>
<td>bl</td>
<td>+/-/-</td>
<td>+/-/-?</td>
<td>object recognition and categorization</td>
</tr>
<tr>
<td>Sum</td>
<td>770</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+/-/-?</td>
<td></td>
</tr>
</tbody>
</table>

Ventral Pathway / What (Object Recognition and Categorization)

<table>
<thead>
<tr>
<th>Area</th>
<th>Size (mm²)</th>
<th>RFS</th>
<th>Latency (ms)</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>CI/SI/PI/OI</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>MST</td>
<td>60</td>
<td>&gt;30</td>
<td>60-70</td>
<td>bl</td>
<td>+/-/-</td>
<td>I</td>
<td>optic flow, self-motion, pursuit</td>
</tr>
<tr>
<td>VIP</td>
<td>40</td>
<td>10-30</td>
<td>50-60</td>
<td>bl</td>
<td>+/-/-</td>
<td>+/-/-?</td>
<td>optic flow, touch, near extra personal space</td>
</tr>
<tr>
<td>7a</td>
<td>115</td>
<td>&gt;30</td>
<td>90</td>
<td>bl</td>
<td>+/-/-</td>
<td>+/-/-?</td>
<td>Optic flow, heading</td>
</tr>
<tr>
<td>LIP</td>
<td>55</td>
<td>12-20</td>
<td>50</td>
<td>cl</td>
<td>+/-/-</td>
<td>+/-/-?</td>
<td>salience, saccadic eye movements</td>
</tr>
<tr>
<td>AIP</td>
<td>35</td>
<td>5-7</td>
<td>60</td>
<td>bl</td>
<td>?+/+/?</td>
<td>?+/+/?</td>
<td>grasping</td>
</tr>
<tr>
<td>MIP</td>
<td>55</td>
<td>10-20</td>
<td>100</td>
<td>co</td>
<td>+/-/-?</td>
<td>I</td>
<td>reaching</td>
</tr>
</tbody>
</table>

Dorsal Pathway / Where and How (Coding of Action Relevant Information)

<table>
<thead>
<tr>
<th>Area</th>
<th>Size (mm²)</th>
<th>RFS</th>
<th>Latency (ms)</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>CI/SI/PI/OI</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum</td>
<td>585</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+/-/-?</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1**

Basic facts on the different areas of the macaque visual cortex based on different sources [44], [28], [95], [141], [161].

First column: Name of Area. Second column: Size of area in mm². ‘?’ indicates that this information is not available. Third column: Average receptive field size in degrees at 5 degree of eccentricity. Fourth column: Latency in milliseconds. Fifth Column: Contra versus bilateral receptive fields. Sixth Column: Principles of organization: Retinotopic (rt), spatiotopic (st), clustered (cl) columnar (co). Seventh Column: Invariances in representation of shape: Cue-Invariance (CI), Size Invariance (SI), Position Invariance (PI), Occlusion Invariance (OI). ‘I’ indicates that this entry is irrelevant for the information coded in these areas. Eighth Column: Function associated to a particular area.
Basic Terms

- **Retinotopic/Spatiotopic**
- **Different kinds Of Invariances**
  - Cue Invariance
  - Size Invariance
  - Position Invariance
  - Occlusion Invariance

### Table: Sub-cortical processing and Occipital / Early Vision

<table>
<thead>
<tr>
<th>Area</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>Cl/SI/PI/OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retina</td>
<td>bl</td>
<td>+/-/-/-</td>
<td>-/-/-</td>
</tr>
<tr>
<td>LGN</td>
<td>co</td>
<td>+/-/-/-</td>
<td>-/-/-</td>
</tr>
</tbody>
</table>

#### Occipital / Early Vision

<table>
<thead>
<tr>
<th>Area</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>Cl/SI/PI/OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>co</td>
<td>+/-/-/-</td>
<td>-/-/-</td>
</tr>
<tr>
<td>V2</td>
<td>co</td>
<td>+/-/-/-</td>
<td>-/-/-</td>
</tr>
<tr>
<td>V3/V3A/VP</td>
<td>co</td>
<td>+/-/-/-</td>
<td>-/-/-</td>
</tr>
<tr>
<td>V4/VOT/V4t</td>
<td>co</td>
<td>+/-/-/-</td>
<td>+/-/-</td>
</tr>
<tr>
<td>MT</td>
<td>co</td>
<td>+/-/-/-</td>
<td>+/-/-</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Pathway / What (Object Recognition and)

<table>
<thead>
<tr>
<th>Area</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>Cl/SI/PI/OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEO</td>
<td>co</td>
<td>(+)/-/-/+</td>
<td>?/-/-</td>
</tr>
<tr>
<td>TE</td>
<td>bl</td>
<td>-/+/-/+</td>
<td>+/-/+(-)</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### / Where and How (Coding of Action R

<table>
<thead>
<tr>
<th>Area</th>
<th>co/bi lat.</th>
<th>rt/st/cl/co</th>
<th>Cl/SI/PI/OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MST</td>
<td>bl</td>
<td>+/-/-/+</td>
<td>I</td>
</tr>
<tr>
<td>CIP</td>
<td>bl</td>
<td>+/-/?/?</td>
<td>+/-/?/?</td>
</tr>
<tr>
<td>VIP</td>
<td>bl</td>
<td>-/+/-/-</td>
<td>I</td>
</tr>
<tr>
<td>7a</td>
<td>bl</td>
<td>(+)/-/-/-</td>
<td>?/+/?</td>
</tr>
<tr>
<td>LIP</td>
<td>cl</td>
<td>+/-/-/-</td>
<td>?/-/-</td>
</tr>
<tr>
<td>AIP</td>
<td>bl</td>
<td>?/+/?/-</td>
<td>?/+/-</td>
</tr>
<tr>
<td>MIP</td>
<td>bl</td>
<td>+/-/-/-</td>
<td>I</td>
</tr>
<tr>
<td>Sum</td>
<td>co</td>
<td>+/-/-/-</td>
<td>I</td>
</tr>
</tbody>
</table>
Pre-cortical Areas

- hippocampus (memory
- prefrontal cortex (non-motor)
- FEF, frontal cortex, occulomotor
- SC, brain stem
- PMD, PMV, prefrontal cortex (hand control, rule-based behaviors)

Ventral pathway:
- TE (A1)
- TEO (PIT)
- V3/V3A
- V4

Dorsal pathway:
- AIP
- MST
- VIP
- MI

Object recognition pathway:
- space/action pathway
- Vestibular information about arm, eye, and head position

Occipital cortex:
- V1
- V2

 VISUAL CORTEX

Retina

Dorsal lateral geniculate nucleus

Inferior temporal cortex: Second level of visual association cortex

Dorsal stream

Striate cortex (primary visual cortex)

Ventral stream

The Mærsk McKinney Møller Institute
University of Southern Denmark
Precortical Areas

Retina

LGN

• No Feature Transformation
• Preparing for Stereo
Occipital Cortex

Hippocampus
Prefrontal cortex (non motor)
FEF, Frontal Cortex, Occulomotor
SC, Brain Stem
PMD, PMV, Prefrontal cortex (Hand control, rule based behaviours)

Ventral pathway

TE (A1T)
TEO (PIT)
Arm control, PMD

Ventral stream

Dorsal pathway

V4
V3/V3A
V2
V1

Occipital cortex

LGN
Retina

Second level of visual association cortex in parietal lobe
Dorsal lateral geniculate nucleus
Thalamus
Eye
Optic nerve
Inferior temporal cortex: Second level of visual association cortex

Dorsal Stream

Striate cortex (primary visual cortex)
Extrastriate cortex
Occipital Cortex

- More than 70% of the visual cortex
  - Occipital Cortex 3340mm²
  - Ventral Pathway 770mm²
  - Dorsal Pathway 585mm²

- Processing
  - Task unspecific generic scene representation

Retinotopic Organization
Occipital Cortex: V1 and V2
Concept of Hue as Object Property
Linguistic Concept of ‘red’ or ‘blue’
Ventral Pathway

- **V1 (VISUAL CORTEX)**
  - Retina
  - LGN
- **V2**
  - Occipital cortex
- **V3/V3A**
- **V4**
  - Hippocampus
  - Prefrontal cortex (non motor)
- **Ventral pathway**
  - TE (AIT)
  - TEO (PIT)
  - Arm control, PMD
- **Dorsal pathway**
  - PMD, PMV, Prefrontal cortex (Hand control, rule based behaviours)
  - 7a
  - AIP
  - Dorsal lateral geniculate nucleus
  - Thalamus
  - Dorsal Stream
  - Striate cortex (primary visual cortex)
- **Ventral Stream**
  - Inferior temporal cortex: Second level of visual association cortex
  - Extrastriate cortex
  - Ventral Stream
- **Ventral Pathway**
  - Vestibular information about arm, eye and head position
  - Dorsal Stream
  - Second level of visual association cortex in parietal lobe

The Mærsk McKinney Møller Institute
University of Southern Denmark
Ventral Pathway

- More than 70% of the visual cortex
  - Occipital Cortex 3340mm²
  - Ventral Pathway 770mm²
  - Dorsal Pathway 585mm²

- Processing
  - Object Recognition and Categorization
  - Many suggestions for how to divide into areas
Ventral Pathway: TEO and TE

- TEO
- TE

Hyppocampus Memory
Prefrontal cortex (non motor)
FEF, Fror

Arm control,

Ventral pathway

Tanaka
Dorsal Pathway

- Hypocampus: Memory
- Prefrontal cortex: (non-motor)
- FEF, Frontal eye fields
- Cortex, Occulomotor SC, Brain Stem
- PMD, PMV, Prefrontal cortex: (Hand control, rule based behaviours)
- AIP
- Arm control, PMD

Ventral pathway

- TE (AIT)
- TEO (PIT)
- V4
- V3/V3A
- MT
- Occipital cortex
- V2
- V1
- VISUAL CORTEX
- LGN
- Retina

Dorsal Stream

- Second level of visual association cortex in parietal lobe
- Striate cortex: (primary visual cortex)
- Dorsal lateral geniculate nucleus
- Thalamus
- Extrastriate cortex
- Inferior temporal cortex: Second level of visual association cortex
- Ventral Stream
- Eye
- Optic nerve
Dorsal Pathway

- **More than 70% of the visual cortex**
  - Occipital Cortex 3340mm²
  - Ventral Pathway 770mm²
  - Dorsal Pathway 585mm²

- **Processing**
  - Much less known than Ventral Pathway
  - Many more distinguished areas
  - Coding visual information related to action and position in space
Dorsal Pathway

CIP

Cue invariant 3D shape

MST

Ego-motion

AIP

Hand shape and affordances

MIP

Reaching

VIP

Ego-space

LIP

Saccadic related retinotopic repr.
Vertical View
What do we know about primate’s vision which is relevant for engineers?

- Richness of representation
- Deep Hierarchy versus flat Architectures
- Separation of information
Richness of representation

- The occipital cortex provides a huge variety of visual aspects at different levels of granularity and different levels of abstractions
  - Zoo of features
  - Challenge: Designing/learning this hierarchy is difficult but maybe required
- What is important for learning a certain task or category is unclear
  - Challenge: Learning algorithms that are able to deal with such a huge and at the same time highly structured input space
What do we know about primate’s vision which is relevant for engineers and linguists?

- Richness of representation
- Deep Hierarchy versus flat Architectures
- Separation of information
Deep Hierarchy

- Richness of representation
- Deep Hierarchy versus flat Architectures
- Separation of information
- Feedback
- Learning versus hard-wiring

Ventral pathway

Object recognition pathway

Space/Action pathway

Ventricular information about arm, eye and head position

Hyppocampus Memory
Prefrontal cortex (non motor)
FEF, Frontal Cortex, Occulomotor SC, Brain Stem
PMD, PMV, Prefrontal cortex (Hand control, rule based behaviours)

Arm control, PMD

V4

V3/V3A

V1

V2

Occipital cortex

VISUAL CORTEX

LGN

Retina
Flat versus deep Hierarchies

### Deep Hierarchy

<table>
<thead>
<tr>
<th>Level 5 (ventral)</th>
<th>Task V1</th>
<th>Task V2</th>
<th>Task V3</th>
<th>Task Vn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task D2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task D3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Dn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Flat Hierarchy

<table>
<thead>
<tr>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
<th>Task 5</th>
<th>Task 6</th>
<th>Task 7</th>
<th>Task 8</th>
<th>Task n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Some kind of Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example of a flat hierarchy

Increasing Level of Abstraction

- Hypocampus: Memory
- Prefrontal cortex (non-motor)
- FEF, Frontal Cortex, Occulomotor
- SC, Brain Stem
- Arm control, PMD
- PMD, PVV, Prefrontal cortex (Hand control, rule-based behaviours)
- Vestibular information about arm, eye, and head position
- Dorsal pathway
- Ventral pathway
- Object recognition pathway

VISUAL CORTEX

V1

V2

V4

V3/V3A

TEO (PIT)

TE (AIT)

Arm control, PMD

PMD, PVV, Prefrontal cortex (Hand control, rule-based behaviours)

Vestibular information about arm, eye, and head position

Space/Action pathway

50° 0° 50° 0° 50° 0°
Increasing Level of Abstraction

- Hippocampus
- Prefrontal cortex
- FEF, Frontal Cortex, Occulomotor
- SC, Brain Stem
- PMD, PMV, Prefrontal cortex
- Hand control, rule based behaviours
- Arm control, PMD

Pathways:
- TE (AIT): Ventral pathway
- TEO (PIT): Object recognition pathway
- V1: Occipital cortex
- V2: VISUAL CORTEX
- V3/V3A
- LGN: Retina

Space/Action pathway

Vestibular information about arm, eye and head position
Flat versus deep hierarchies

- Flat Hierarchies are inefficient
- No sharing of computational resources
- Transfer of experience across tasks is facilitated within the same representations
What do we know about primate’s vision which is relevant for engineers and linguists?

- Richness of representation
- Deep Hierarchy versus flat Architectures
- Separation of information
<table>
<thead>
<tr>
<th>Area</th>
<th>RF size</th>
<th>Color</th>
<th>2D Shape</th>
<th>3D Shape</th>
<th>Motion</th>
<th>RF size</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AIT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIP</td>
<td>7a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PIT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V3/V3A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LGN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retina (ganglion cells)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retina (receptors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ventral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dorsal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V3/V3A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Separation of Information

- Colour, 2D shape, 3D shape and motion become separated and are then up to a certain level of the hierarchy processed largely independently (while in the pixel domain these aspects are deeply intertwined).
- For learning problems this allows for cutting off non-relevant dimensions.
- It allows also to discover relations between different aspects of visual information on a higher level (e.g., motion and 3D shape).
Overview

- Background Information
- The primate’s vision system: A deep Hierarchy
- SotA and Problems of research on deep hierarchical systems
- Reflections
Research on Deep Hierarchies (non-exhaustive)

- **Meta reasoning**
  - Tsotsos, Geman et al., Mel and Fiser,

- **Learning of Hierarchical Vision Systems**
  - Amit, Hawkins, Leonardis, Piater, Ullman, DiCarlo and Cox, Ommer and Buhmann, Serre and Poggio, Bengio, Wiskott, Hinton

- **Design of Hierarchical Vision Systems**
  - Biederman and Hummel, Fukushima, Pugeault and Kruger
Biederman and Fukushima

John E. Hummel and Irving Biederman (1992). Dynamic Binding in a Neural Network for Shape Recognition
Early Cognitive Vision System

Cognitive Vision Lab Robotics Group
Cognitive & Applied Robotics (CARO)
Robotics Lab - RoboL
Vision Lab - CoViL

10-06-2014
Edge and Surface based Grasp Affordances

M. Popović, G. Kootstra, J. A. Jørgensen, D. Kragic and N. Krüger. Grasping Unknown Objects using an Early Cognitive Vision System for General Scene Understanding. IROS 2011 (nominated as one of the finalists for an IROS Award)

Bootstrapping Robots: Grounding objects and grasping affordances


Learning Hierarchies: Work from Ales Leonardis
Layered Graphical Model

- Each vertex represents a (composite or primitive) feature.
- Each edge is annotated with a spatial relation (scale-normalized distance and relative orientation).
Revival of deep neural net working

• Deep Nets seem to recently beat other algorithms on important benchmarks

  • A single neuron's feature is no more interpretable as a meaningful feature than a random set of neurons.
  • Every deep neural network has "blind spots" in the sense that there are inputs that are very close to correctly classified examples that are misclassified.
Some Reflections

- **Vision is probably a quite hard problem**
  - It uses resources occupying more than 50% of our brain
  - It is far from ‘being solved’

- **Of that 70% is generic scene processing**
  - Deep hierarchy with increasing invariant representations
  - It spans a huge feature space as a basis for grounding processes
  - This space has a high degree of structure
    - Motion
    - Spatial Relations

- **We can learn from the human visual system?**
  - It is worthwhile to build/learn deep hierarchical systems
  - Number of levels
  - Receptive field size
  - What features to extract at what stage in the hierarchy