Motor System’s Role in Grounding, Development, and Recognition in Vision

Computational Neuroscience Lab @ Tsinghua University
July 9, 2010

Yoonsuck Choe, Ph.D.
Department of Computer Science & Engineering
Texas A&M University

With N. Smith, H.-F. Yang, and N. Misra.

Why Do We Have the Brain?
• Survival and reproduction? Think again!

Motivation and Overview
Important aspects of vision may be hidden in its intricate coupling with motor function.

1. Grounding of internal representations in the visual system.
2. Development/co-development of visual receptive fields with their grounding.

Part I: Grounding

Choe et al. (2007); Choe and Smith (2006); Choe and Bhamidipati (2004)
What Do Those Green Lights Represent?

- It is hard to get any idea at all.
- Actually, this is how it might be like, looking at the brain’s activity from the inside of the brain.

They Are Visual Cortical Responses to Oriented Lines
They Are Visual Cortical Responses to Oriented Lines

What Is Grounding?

... How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols? ...
– Harnad (1990)

- Given a representation, figure out what it represents/means.
- Given an activity pattern in the brain, figure out what information it carries (decoding, decompression, etc., cf. Zhaoping 2006).

Miikkulainen et al. (2005); Weliky et al. (1995)

Grounding in the Brain

(a) External observer
(b) Internal observer

The problem of grounding, within the brain:

- **External observer** (e.g., a neuroscientist) can figure out how spike $S$ relates to input $I$.
- **Internal observer cannot** seem to, which does not make sense at all.
Possible Solution: Allow Action

- A major problem in the picture is the **passiveness** of the whole situation.
- Adding action can help solve the problem.
- But why and how?

**Experimental Evidence**

- Vibrotactile array linked to a video camera.
- Passive viewing results in **tactile** sensation.
- Moving the camera results in a **vision-like** sensation.
- Sensation as related to voluntary/intentional action may be the key!

Bach y Rita (1972; 1983)

**Approach: Grounding Through Action**

- Direct access to **encoded internal state** (sensory array) only.
- Action is enabled, which can **move the gaze**.
- How does this solve the grounding problem?

**Action for Unchanging Internal State**

- Diagonal motion causes the **internal state** to remain unchanging over time.
- Property of such a movement exactly reflects the property of the input $I$: Semantics figured out through action.
Action for Unchanging Internal State

• Diagonal motion causes the *internal state* to *remain unchanging* over time.

• Property of such a movement *exactly reflects* the property of the input $I$: Semantics figured out through action.

Task

• Given an encoded sensory signal $s$, we want to learn action $a$ that *maximizes the invariance* in the internal state over time.

• The learned action $a$ will give *meaning* to $s$.

• This is basically a *reinforcement learning* task.

Methods: Orientation Response

Sensory state:

$$ s = \arg \max_{1 \leq \theta \leq n} r_\theta. $$
Methods: Reinforcement Learning

- Policy $\pi$: Given reward probability $R(s,a) = P(a|s)$ and state $s$, stochastically generate action $a$ with probability $P(a|s)$.
- Reward: measure similarity between previous and current response vector $r$

$$\rho_{t+1} = r_t \cdot r_{t+1}$$

- Learning $R(s,a)$:

$$R_{t+1}(s_t, a_t) = R_t(s_t, a_t) + \alpha \rho_{t+1},$$

and then normalize over all actions for a given state.

Results: Learned $R(s,a)$

- Synthetic image
  - Initial (a)
  - Ideal (b)
  - Final (c)

- Natural images
  - Initial (a)
  - Ideal (b)
  - Plant (c)
  - Oleander (d)

- Learned $R(s,a)$ close to ideal.

Results: Reward Probability Table

![Reward Probability Table](http://faculty.cs.tamu.edu/choe)

- Reward probability $R(s,a)$ can be tabulated.
- In an ideal case (world consists of straight lines only), we expect to see two diagonal matrices (shaded gray, above).

Results: Gaze Trajectory

- Input (a)
- Initial (b)
- Final (c)
Part I: Summary

• (1) Using invariance as the only criterion, (2) particular action pattern was learned, (3) that has the same property as the input that triggered the sensors.

Theories of RF Formation

Hoyer and Hyvärinen (2000)

Well-developed understanding on how RFs form:


However, how is the resulting code to be used remains a question.
Questions

- The motor-based grounding experiment assumed that receptive fields are given and fixed.
- Can these be learned (developed) along with the grounding process?

Experiments

- Effects of different action policy on RF learning.
  - Random \( R(s, a) \)
  - Ideal \( R(s, a) \)
- Simultaneous learning of RF and action policy.
  - RF learning through normalized Hebbian learning
  - Reinforcement learning of \( R(s, a) \) based on internal-state invariance

Learning RFs along with Their Grounding (Decoding)

- Grounding (decoding): Same as Part I.
- RFs develop through local learning:
  \[
  g_{ij} = \frac{g_{ij} + \alpha (I_{ij} - g_{ij})}{\sum_{mn} g_{mn} + \alpha (I_{mn} - g_{mn})},
  \]
  where \( g_{ij} \) is the afferent connection weight and \( I_{ij} \) the input pixel value.

Effects of \( R(s, a) \) on RF Learning
Simul. Learning of RFs & $R(s, a)$

- Seemingly unordered RFs and $R(s, a)$ results.

Reordering RFs

- The $R(s, a)$ result looks bad because each row's corresponding RF orientation is not ordered.
- Reordering RF orientation reorders the rows in $R(s, a)$.

Reordered RFs and $R(s, a)$

- However, reordering the RFs and their corresponding $R(s, a)$ rows shows the true underlying structure! (Not perfect, but a good start!)

Part II: Summary

- Action policy strongly influences RF properties, by altering the input statistics.
- Certain action policies may give better RFs, faster.
- Receptive fields and action policy can learn simultaneously, from scratch, thus allowing encoding/decoding to evolve together.
Part III: Shape Recognition

Misra and Choe (2007)

Motor System and Object Recognition

Yarbus (1967)

- When we look at objects, our gaze wanders around.
- Could such an interaction be necessary for object recognition?

Learning About Shapes

• For complex objects, a history of sensory activity may be needed (i.e., some form of memory).
• Invariance can be detected in the spatiotemporal pattern of sensor activity.

Advantage of Motor-Based Memory (Habit, or Skill)

• Sensor-based representations may be hard to learn and inefficient.
• Motor-based approaches may generalize better.
• Comparison: Make both into a 900-D vector and compare backpropagation learning performance.
Class Separability

- Comparison of PCA projection of 1,000 data points in the visual and motor memory representations.
- Motor memory is clearly separable.

Part III: Summary

Motor-based representations of shape are
- More separable in the representational space,
- Faster to learn,
- Better at novel tasks (generalization), compared to sensory representations.

Related Works (Selected)

- Pierce and Kuipers (1997): Learning from raw sensor/actuators (See related work on bootstrap learning).
- Miikkulainen et al. (2005): Visual cortical development and function
- Ballard (1991): Animate vision
- Rizzolatti et al. (2001): Mirror neurons
- Salinas (2006): Sensory RF coding dictated by downstream requirements.
- Sejnowski (2006): Importance of “projective fields”.
Discussion

• Main contribution: Discovery of the invariance criterion for sensorimotor grounding, development, and recognition.

• Importance of self-generated action in autonomous understanding.

• Richer motor primitive repertoire can lead to richer understanding.

• Tool use can dramatically augment motor primitive repertoire.

Discussion (cont’d)

• How to extend to more complex properties?: Attention may be needed (cf. Zhaoping 2006, esp. the “selection” part).

• Are the motor primitives innate? Can they also develop?

• How to extend to non-spatial modalities like olfaction?

Conclusions

We must ask how the brain understands itself.

• Action is important for understanding/grounding.

• Simple criterion (state invariance) can help link sensory coding with meaningful action.

• RFs can be developed along with grounding.

• Motor-based representations are more effective for shape recognition.

Credits

• Contributors: Kuncara A. Suksadadi, S. Kumar Bhamidipati, Noah Smith, Stu Heinrich, Navendu Misra, Huei-Fang Yang, Daniel C.-Y. Eng

• Choe et al. (2008, 2007); Choe and Smith (2006); Choe and Bhamidipati (2004)
Other Projects at Texas A&M

- Knife-Edge Scanning Microscope (KESM) Project
- How to utilize V1 response for saliency thresholding
- Flash-lag effect, delay compensation, and facilitating synapses
- Evolutionary precursor of agency: internal state predictability
- And more ...

Knife-Edge Scanning Microscope Project

- Cut and image whole mouse brain at sub-micrometer resolution.
- Fully automated: one mouse brain imaged in less than 2 weeks.
- Resulting data: 2 to 20 TB per mouse brain.
- Analysis of the data is a major issue.

Saliency Thresholding based on V1 Response

- V1 response shows power law (nothing new).
- Finding: Comparing to Gaussian with same variance gives reliable saliency threshold (Sarma and Choe 2006).

FLE, Delay Compensation, & Facilitating Synapses

- Delay in the nervous system on the order of 100 ms.
- Flash-lag effects suggest a compensatory mechanism.
- Facilitating synapses may be the neural substrate.
Evolutionary Precursor of Agency/Self-Awareness

- Agency > authorship > 100% predictability of own action.
- For this, internal state trajectory must be predictable.
- Same task performance but more predictable internal state trajectory have an advantage when the task becomes more difficult.


References


http://faculty.cs.tamu.edu/choe