A Model of the Ventral Visual System Based on Temporal Stability and Local Memory

by Wyss et al. (2006)

CPSC 644

Presented by Yoonsuck Choe

Overview

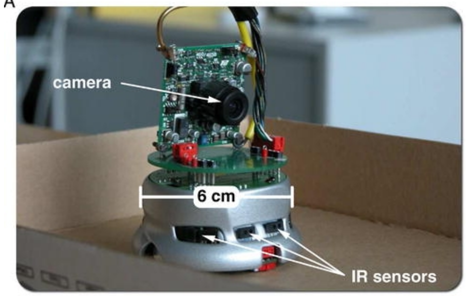
- Hierarchy of visual areas.
- Learning based on stability over time and leaky-integrator memory.
- Results
  - Stability over time.
  - Response property over different regions in the environment.
  - Preference for complex inputs.
  - Network dynamics.
  - Position reconstruction error.
  - Environment manipulation.
  - Input scrambling.

Visual Hierarchy

- The neocortex is remarkably homogeneous.
- Are areas in the visual hierarchy intrinsically different or do they develop into such from initially uniform structure?
- Model based on optimal stability and local memory shows:
  - V1-like RF in lowest layer,
  - More complex feature selectivity in intermediate layers, and
  - Global, “place” selectivity in highest layer develops based on small number of principles.

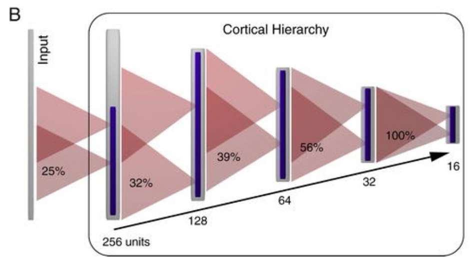
Robotic System

A



- Khepera robot with video camera (16×16) and IR sensors.
- Random translation and rotation of robot within a 31×21 cm^2 arena.
- Avoids obstacle when detected by IR sensor.

Network and Its Activation



- Hierarchical structure, with afferent, lateral (used only for exchanging decorrelation signal), and inter-area feedforward connections (extent of these connections differ).
- Activation is calculated as:

$$A_l^i(t) = f\left(\sqrt{(\vec{I}(t)\cdot\vec{W}_l^{l,i})^2 + (\vec{I}(t)\cdot\vec{W}_2^{l,i})^2}\right)$$

where $\vec{I}$ is the main input, $\vec{W}$ the connection weight, and $f(x) = 1 - \exp(-x^2)$.

Output of Afferent Level l

$$\vec{O}_{l(t)} = \frac{1}{\tau_l}\vec{A}_l{}'(t) + \left(1 - \frac{1}{\tau_l}\right)\vec{O}_l(t-1)$$

$$A'(t) = \frac{A(t) - \langle A \rangle_t}{\sqrt{\mathrm{var}_t(A)}}$$

- Input $\vec{I}$ in the previous slide can also be from the output $\vec{O}$ from a lower afferent level l, which is a running average of activation $\vec{A}_l$ of level l.
- This serves as "local memory."

Learning: Optimization Function

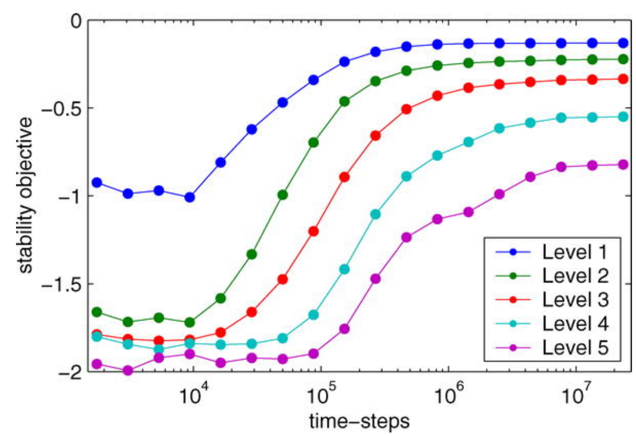
$$\psi_l = -\sum_i \frac{\langle (A_l^i(t) - A_l^i(t-\tau_l{}'))^2 \rangle_t}{\mathrm{var}_t(A_l^i)} - \beta \sum_{i \neq j} (\rho_l^{ij})^2 - \Gamma \sum_\iota \langle A_l^i \rangle_t$$

$$\rho_l^{ij} = \frac{\mathrm{cov}_t(A_l^i, A_l^j)}{\sqrt{\mathrm{var}_t(A_l^i)\mathrm{var}_t\,(A_l^j)}}$$

ψ maximized using standard gradient ascent on $\vec{W}$:

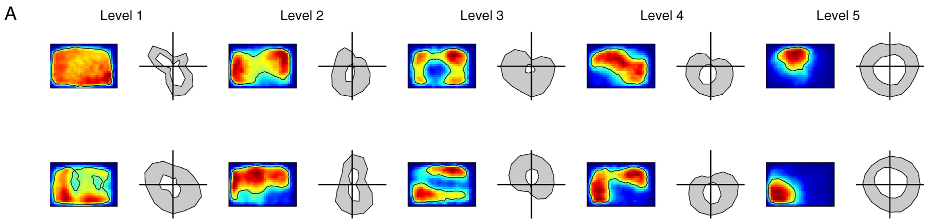
- Maximize smoothness in activation over time.
- Minimize correlation across neurons in the same level.
- Maximize sparseness of response.

Stability



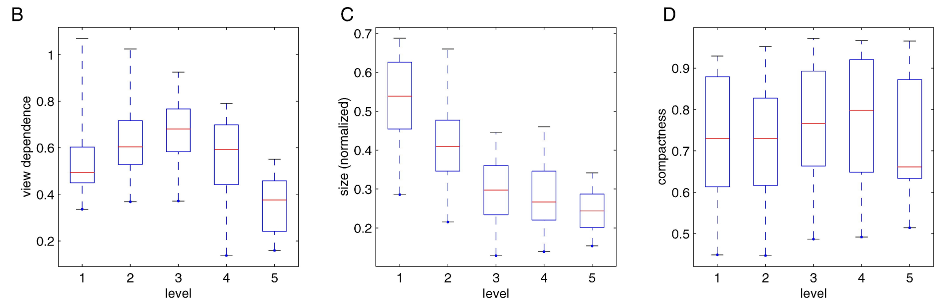
- Stability increases and converges in all layers.

Response Property over Regions in the Arena



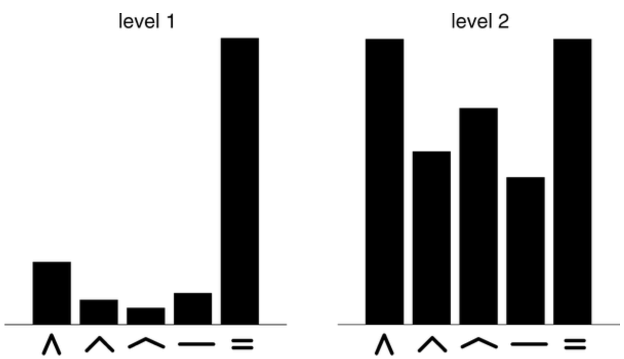
- Unit response within different locations in the arena (heat map).
- Orientation of robot when units are responding (polar plot).
- Highest level units respond only when the robot is in a small region within the arena.

Properties of Different Levels



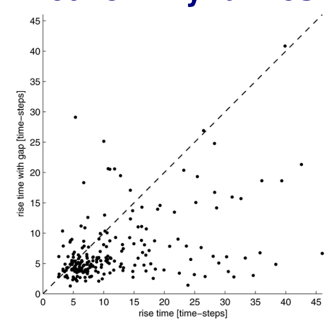
- View dependence: change in response dependent on viewing angle when the position of the robot is fixed.
- Size: size of region in which the robot is responsive, normalized by the total area of the arena.
- Compactness: perimeter of the responsive region divided by the perimenter of a disc of equal area.

Preference for Complex Inputs



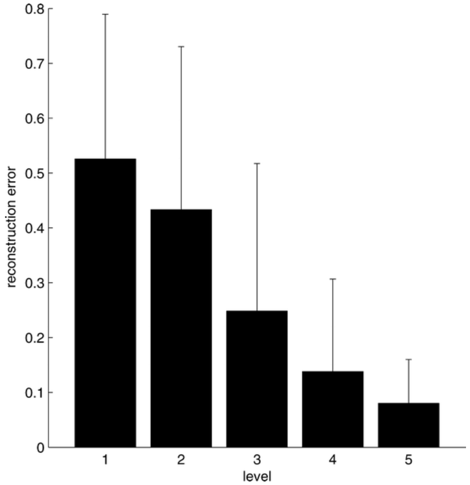
- Higher level prefers more complex objects.

Network Dynamics



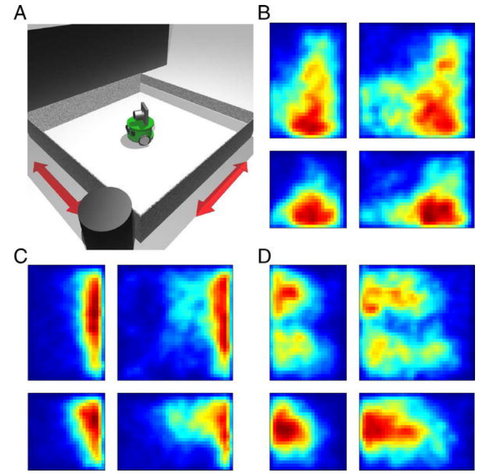
- Maximum stability may be counter-intuitive (it will result in slow change in activation). Does it mean the whole approach will give a slow robot? Use Bayes rule: $P(x|A) \propto P(A|x)P(x)$.
- Would the network respond fast to fast changing inputs?
- Response time with regular input : response time with fast input = 2.1 $\pm$ 1.9.

Position Reconstruction



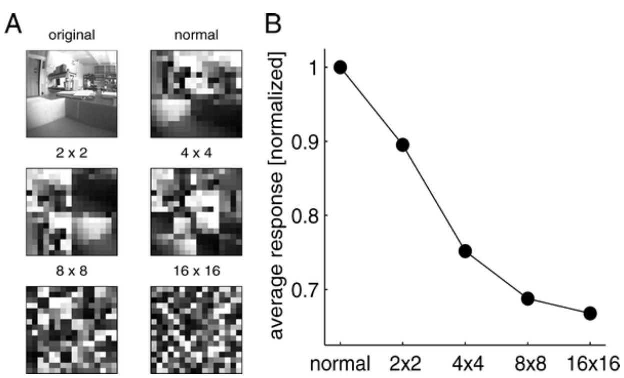
- Given the response, how well can you predict the position of the robot?
- Higher level becomes more and more position specific.

Environment Manipulation



- Adjusting the shape of the environment (the arena) largely preserves position preference (the position map gets stretched).

Input Scrambling



- Input scrambling causes gradual decrease in response (in level 3).

Discussion (YC)

- What are the relative contribution of the three terms in the optimization function? A systematic study with and without different combinations of these terms may have been helpful.
- How is the stability objective related to all of those resulting properties? Same question for local memory applies.
- Role of leaky-integrator memory is unclear.
- No link to the motor system.
- Motion is too restricted (camera cannot change pitch).
- Relation to Choe and Smith (2006)?

References

Choe, Y., and Smith, N. H. (2006). Motion-based autonomous grounding: Inferring external world properties from internal sensory states alone. In Gil, Y., and Mooney, R., editors, *Proceedings of the 21st National Conference on Artificial Intelligence*. 936–941.

Wyss, R., König, P., and Verschure, P. F. M. J. (2006). A model of the ventral visual system based on temporal stability and local memory. *PLoS Biology*, 4:e120.