A Neural Network for the Detection of Rotational Motion[1]

Irwin K. King king%rana@usc.edu
Jim-Shih Liaw liaw%rana@usc.edu
Michael A. Arbib arbib%pollux@usc.edu

Center for Neural Engineering
University of Southern California
Los Angeles, CA 90089-2520, USA

Abstract

This paper describes a biologically plausible neural network which detects rotational motion in the frontal-parallel plane (revolving around the Z-axis) produced by a multiple-dots stimulus against a stationary background. The network consists of multi-layered velocity sensitive sensory cells organized in a locally connected fashion. The network displays the ability to segment multiple rotational stimuli in time sequences as well as an object with both rotational and translational motion by extracting the center and the perimeter trace of the rotational stimulus. Different simulations have been performed and the results are displayed and discussed.

1. Introduction

Biological visual systems can analyze and interpret intensity changes impinging on our visual field produced by dynamic stimuli (see Nakayama 1985 for a review). Moreover, motion patterns alone elicit object interpretation responses without the presence of color, texture, and contour information. From psychophysical examinations, Wallach & O'Connell's (1953) experiment with the ***Kinetic Depth Effect*** (KDE) showed how viewers were able to infer the 3-D structure from the 2-D orthographic projected shadow of an object behind a translucent screen. Using dot stimuli, Johansson (1973) demonstrated with light sources attached to the major joints of a human subject conducting simple types of movement, the pure motion of the dots against a dark background alone elicited the correct motion interpretation from naive viewers usually within 200 ms, some even at 100 ms. Using a similar dynamic point-light display, Kozlowski & Cutting (1977) demonstrated the remarkable ability of the viewers to distinguish the sex of the walker. Furthermore, Börjesson & Von Hofsten (1975) developed a vector model to explain how relative motion between the dots and the common motion among the dots in the 2D image plane induced depth perception of rotation and translation motion in 3D from two-dot and three-dot stimuli. Others have also found how dot patterns give rise to depth perception (Braunstein 1976). These experiments illustrate how motion analyzed in isolation is sufficient for the correct interpretation of moving objects.

Aside from these psychophysical findings, physiological examinations have unearthed data substantiating the existence of neural mechanisms which are tailored to specific visual functions on the analysis and interpretation of motion in cortical regions. For instance, Hubel and Wiesel (1962, 1968) studied the diversity and the complexity of different sensory cell types and connections from extracellular microelectrode recordings in anesthetized cats and monkeys. The discovery of *simple*, *complex*, and *hyper-complex* cell types demonstrated orientation and velocity sensitivity characteristics of these units toward stationary and moving stimuli. In the cerebral cortex, Van Essen & Maunsell (1983) have mapped out a cortical hierarchy and in addition identified specific areas, e. g. V1, middle temporal (MT), and medial superior temporal (MST) as regions associated with the analysis of motion. These data are further supported by clinical studies of patients with lesions to the posterior right hemisphere. Vaina (1989) found deficits in 2D-Form-From-Motion tasks in patients having lesions in the right occipital-temporal (ROT) region. Hence, Vaina hypothesized MT's role is to conduct correlation and correspondence processes between sets of elements over time and space which results in more complex visual perception functionalities. Sakata et al. (1985) found passive visual (PV) neurons in area 7a of rhesus monkeys specifically sensitive to visual rotational stimuli. Detailed investigations further demonstrated that neurons differentiate between rotation in depth rather than in the frontoparallel plane, illustrating the complex functionality of these neurons (Saito et al. 1986). Through these studies, we are challenged to model the various perceptual functionalities using neural networks.

---

1 This research is supported in part by grant no. 1RO1 NS 24926 from the National Institutes of Health (M. A. Arbib principal investigator).

0-7803-0164-1/91/0000-0707$01.00©1991 IEEE

All the above psychophysical, clinical, and physiological evidence suggests specific neural mechanisms performing specialized motion perception tasks. Recent interest in neural networks has led researchers to investigate the potential use of these methodologies in motion detection, motion stereo, velocity estimation, and optical flow calculation problems (Zhou & Chellappa 1990, Courellis & Marmarelis 1990). The initial successes are encouraging; however, most models dealt with simple translational motion and left the more complex types of motion, e. g. rotational motion, unstudied. Therefore, in this paper we present a biologically plausible neural network model for the detection of rotational stimuli in the frontoparallel plane with a stationary background.

2. Neural Network Architecture

The neural network is constructed in a hierarchical fashion with three levels of abstraction. The basic building blocks of the network are leaky integrator neuronal units, which better approximate the biological data available than the commonly used bi-state neurons. The characteristic equation of the leaky integrator is described as,

$$\tau_m \frac{dm(t)}{dt} = -m(t) + S_m(t) + h_m \quad \text{and} \quad S_m(t) = \sum_i w_i X_i(t) \tag{1}$$

where m(t) is the membrane potential of the neuron m as a function of time, $S_m(t)$ is the sum of all inputs from other neurons to the neuron m, h_m is the resting potential, and τ_m is the time constant for decay. By using the leaky integrator model, more complex behaviors of the network can be generated but the difficulty of choosing the correct parameters also increases. To obtain the firing rate of the neuron, a non-linear function is applied to the membrane potential of the neuron. Instead of the sigmoid function, we select the RAMP function as shown in (2) to threshold the membrane potential x in computing the firing rate of the neuron in our model.

$$RAMP(x, A, x_1, y_1, y_2) = \begin{cases} y_1 & \text{if } x < x_1 \\ A - x_1 + y_2 & \text{if } x \geq x_1 \end{cases} \tag{2}$$

The first level of neurons are motion sensitive sensitive neurons (moving edge detectors) modelled after the R3 cells found in frogs and toads (Ewert & Tsui 1987). These units have transient response to leading illumination changes in the excitatory receptive field without motion direction preference (see Teeters 1989 for a detailed simulation of R3 cells). The next level of complex cells are the directional selective units using R3 neurons. Figure 1 illustrates a one dimensional directional selective units modelled after the asymmetrical lateral excitation and inhibition interactions reported by different investigators (Reichardt 1969; Barlow & Levick 1965). The combination of both circuitries elicits a response which excites (inhibits) activities in the neighboring cells when the input is in the preferred (null) direction.

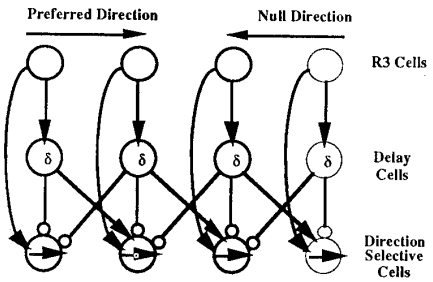


Figure 1. The architecture diagram for directional selective units. Each neuron excites (inhibits) the neighboring units when stimulus moves in the preferred (null) direction.

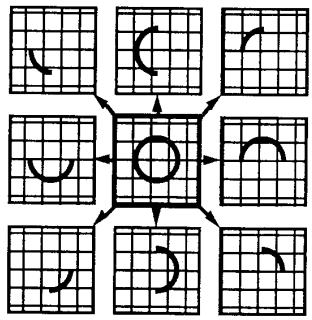


Figure 2. The distributed representation of a clockwise rotational stimulus in various layers of directional selective neuron units. The center grid is the input layer I and the eight remaining layers are the directional selective layers positioned according the directional preferences indicated by the arrow connections.

The last level of the neural network detects the rotational stimulus. It is made up of multiple layers of directional sensitive units. The network consists of $N_r \times N_c \times D$ directional selective units where N_r and N_c are the number of rows and columns of each layer of neurons. There are D layers of neurons where each layer corresponds to

a set of different direction selective cells. All neurons at every point (i, j) in the same layer where $1 \le i \le N_r$ and $1 \le j \le N_c$ have the same velocity sensitivity profile. At each layer k where $0 \le k \le D\text{-}1$, the preferred orientation θ_k in radians, of the layer is defined by $\frac{k\,2\pi}{D}$. Thus, the parameter D tunes the sampling resolution of the network.

Each layer L_k, receives input from the input layer I and the activity at (i,j) in the k^{th} layer is characterized as,

$$L_k(i,j,t+1) = \alpha I(i,j,t) - \sum_{i'=-m/2}^{m/2} \sum_{j'=-n/2}^{n/2} I(i+i', j+j', t-1) * M_k(i',j') \quad (3)$$

where α is a scaling factor for the input stimulus, * denotes the convolution operator which convolves the delay mask, M_k of size (m, n) in our case, against the input layer I. The mask is excitatory (inhibitory) towards the preferred (null) direction and is approximated by the following equation,

$$M_k = G_1(\alpha, x_1, y_1, \sigma_{x_1}, \sigma_{y_1}) - G_2(\beta, x_2, y_2, \sigma_{x_2}, \sigma_{y_2}) \quad (4)$$

where G_1 and G_2 are two Gaussian functions, β and γ are scaling factors of the excitation and the inhibition weight respectively, the σs control the width of the Gaussian windows, the point (x_1, y_1) gives the preferred direction in terms of the cartesian coordinate respect to point (x,y), and $(x_2, y_2) = (-x_1, -y_1)$.

As the rotational stimulus moves in the receptive field, an individual layer will respond to the stimulus when the path it traces matches coarsely to the preferential direction of that layer. Thus we have a very distributed representation in each layer to represent a rotating stimulus as shown in Figure 2.

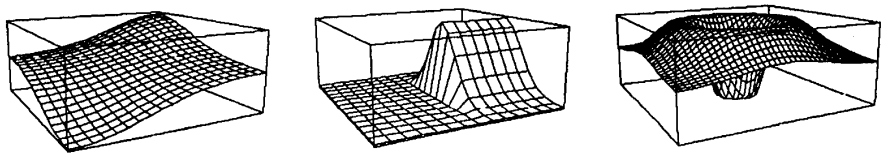
There are two output layers. One layer sums up all the distributed activities in all the layers resulting in a visual memory trace of the rotational motion. The activity in the output layer can be described as,

$$L_{out_1}(i,j,t+1) = L_{out_1}(i,j,t) - \sum_{k=0}^{D-1} L_k(i,j,t) * W_k \quad (5)$$

where W_k is a large mask as illustrated in Figure 3(b). The final output layer which corresponds to the center of a rotating stimulus is given by,

$$L_{out_2}(i,j,t+1) = L_{out_1}(i,j,t) * W_{out_2} \quad (6)$$

where W_{out2} is a on-surround and off-center DOG type mask illustrated in Figure 3 (c). It consists of a sharp inverse peak at the center and a broader but shallower surrounding. The spatial formation of this weight mask sums up activities in an approximated circular shape but inhibits activities near the center in order to avoid mishaps in ambiguous regions such as the overlapping areas of two rotational stimuli in our simulation.



(a) (b) (c)

Figure 3. These are qualitative representations of weight masks for the network. (a) This mask, M_k in equation (3), is used in the construction of directional sensitive units by having excitation (inhibition) in the preferred (null) direction for R3 cells in Figure 1. (b) We employ this type of mask when we sum up all directional sensitive layers, W_k in equation (5). The area we sum up is very acute (direction specific) as shown. (c) The inverted-hat, W_{out2} in equation (6), is utilized to extract the center of the rotational stimulus from the final layer by summing up the surround and inhibiting the center region.

2.1. Simulation Results

To facilitate implementation and testing, we have made some assumptions. (1) We have a featureless, uniform, and stationary background. This also implies no random noises in the input. (2) The ω of the rotation should not exceed the temporal characteristics of the receptive neuron. Hence, the stimuli are moving slowly such that there is no discontinuity of the motion, i.e., movement falls within the 8 adjacent neighboring pixels, during each simulation cycle. (3) Lastly, we have used two dots stimuli instead of a single dot stimuli for speedier result. The dots are offset by phase π. In this case, we achieve a quicker detection of the rotational motion. This is not surprising since single stimuli give less information of whether it will move in a circle than multiple stimuli; therefore, when more points are being introduced to the network, the quicker the response would be.

We used NSL, the Neural Simulation Language (Weitzenfeld 1990), to simulate our neural networks experimentally. The differential equation in (1) is approximated by the Eüler's method. In the simulation, we assume the input image plane size of $N_r \times N_c = (21 \times 21)$ for every layer. The stimulus on the image would be represented by an active pixel = 1 and the remaining pixels are set to 0. The rotation stimulus is centered at (x_c = 10, y_c = 10) and the radius has the following relation with respect to the receptive field size, $r < MIN(N_r/2, N_c/2)$. In our case, we set the radius to be 5 pixels wide. There are D = 8 layers of cells which correspond to 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° of directional sensitivity cells.

We have tested the network using several types of stimulus. First, we used a stationary stimulus to test against our model. We observed some initial activities but they died out quickly since the sensory cells in the network respond only to changes in intensity. Second, the network was tested using simple linear translation dot stimulus. Since the output layer collects the sum from many different layers, the sum of activities in only a few layers do not exceed the threshold of the output layer hence do not invoke any firing in the output. Nevertheless, the layers which have the approximate velocity preference as the input stimulus will fire to indicate a stimulus moving in the preferred direction. We then examine the effect of a rotational stimulus centered at (7,10) with a radius of 5 in the visual field. Figure 4 shows that the curve trace of the stimulus quickly invokes some firing activities in the output layer. As the rotational stimulus makes a complete rotation a population of neurons firing will result within the rotational boundary of the stimulus. By applying an On-Surround-Off-Center type Gaussian mask to the first output layer, we are able to locate the center of this rotation stimulus. The next set of experiments investigated the network under the condition when the stimulus possesses both rotational and translational motion. Even though the tangential velocity of the rotation is much greater than the translational motion, nonetheless, it demonstrates the ability of the network to track the center of the rotation as having the translational motion but at the same time also indicating the detection of the rotational motion as shown in Figure 5 (a), (b), and (c).

Aside from a single moving object, we also studied the neural circuit with multiple moving stimuli. Since the same procedure and weight mask are applied at each location of the output layer, the neural network is capable of detecting multiple instances of rotational stimulus. We have further divided experiments on multiple rotational objects into two sub-cases. The more interesting case involves two intersecting rotational dot stimulus objects since we were uncertain about the effect of disturbances and interferences in regions of intersection in the network. To our satisfaction, the result proved that the neural network is robust to perceive such overlapping formation by producing the correct answer as displayed in Figure 6 (a), (b), and (c). However, the time it took to separate overlapping stimuli is longer than the non-intersecting case. In the case of two non-intersecting rotational objects having the same ω, the network exhibited the ability to extract multiple rotational stimuli by yielding the centers of the stimuli accurately. The results are demonstrated in Figure 7 (a), (b), and (c).

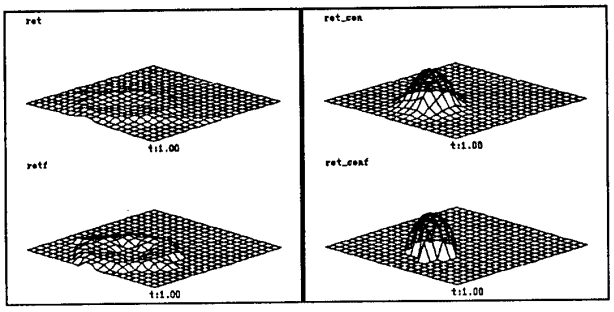
3. Discussion

There are several shortcomings in the rotational motion detection network. Currently, the direction of the rotation is indistinguishable. Moreover, the network fails in identifying one rotational stimulus within a larger rotational stimulus having a different center. Lastly, we would also like to detect a translating rotational object with greater speed and accuracy.

However, we have presented a perceptual neural network capable of detecting rotational dot-stimulus in this paper. The network is constructed from locally connected velocity sensitive cells in a multi-layer architecture composed of directional selectivity neuronal units. The detection of the rotational motion is ***scale*** and ***position*** invariant when the mask used is large enough to encompass the curved path trace of the stimulus. Furthermore, our simulation tests demonstrated the ability of the network to extract the rotational component of a rotational and

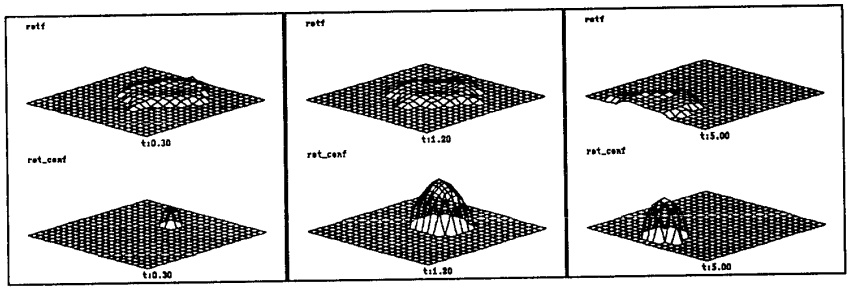
translational stimulus, detect multiple rotational objects, and overcome degraded stimulus in overlapping rotational stimuli.

In the future, we seek to develop a hierarchical structure of motion perception neural circuitries in obtaining higher visual perception functionalities by assembling different neural mechanisms having various perceptual capabilities, e.g. the looming detector by Liaw & Arbib (1991), to achieve machine perceptions rivaling that of biological systems. Our study will consist of (1) mapping out perceptual units through careful anatomical, physiological, and psychophysical studies and (2) simulating these componential functionalities for visual perception using neural networks.



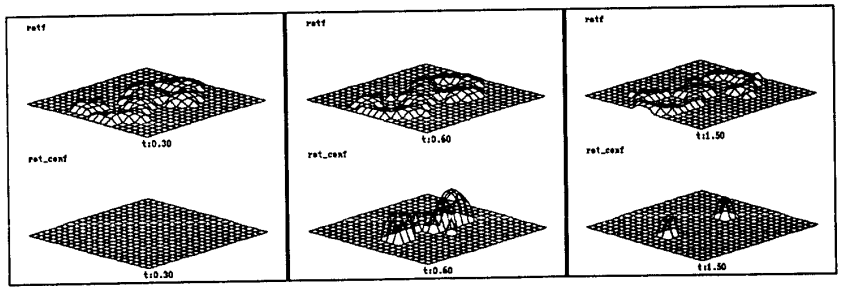
(a) (b)

Figure 4. After 50 simulation time steps, an off-centered 3-dots stimulus at (7, 10) gives an output at the center of rotational motion. **rot** is the membrane potential of cells in the output layer. **rotf** is the firing rate output from a non-linear transform function, the Ramp function in equation (2), using **rot** as the input. **rot_cen** is membrane potential layer indicating the center of the rotational stimulus.



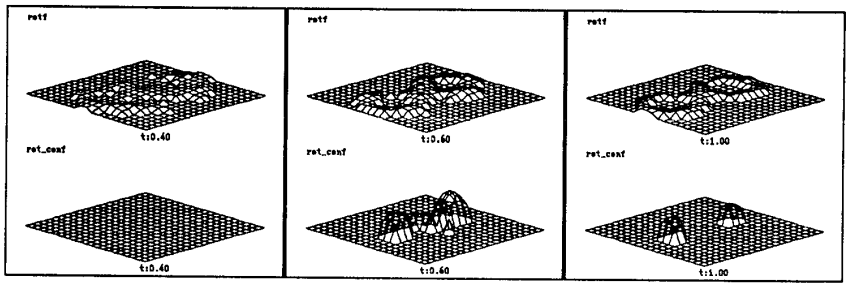
(a) (b) (c)

Figure 5. The stimuli in these figures are both rotating and translating simultaneously. However, there is no correlation between the speed of angular rotation and the translation since translation is relatively slow in this simulation. Nevertheless, we see that after 30 simulation time steps a rotational stimulus is detected as shown in (a). The stimulus has now translated across the X axis in (b) after 120 time steps and (c) after 500 time steps with an uninterrupted detection of the rotational stimulus as it moves across the receptive field.



(a) (b) (c)

Figure 6. This simulation uses two rotational stimuli which share some intersecting areas. There are some ambiguities around the areas where both rotational stimulus intersect but eventually the output became stable and the center of both rotational stimuli are detected after 150 stimulation steps.



(a) (b) (c)

Figure 7. Different from the previous figure, this simulation run used two nearby rotational stimuli in phase without any intersecting area. Notice that after 60 time steps, there is some interference caused by the local excitation bias from neighboring cells. Nevertheless, the network settled into a stable solution after 100 simulation time steps.

References

Barlow, H. B., & Levick, W. R. (1965). The Mechanism of Directionally Selective Units in Rabbit's Retina. *J. of Physiol.*, **178,** 477-504.

Börjesson, E., & Von Hofsten, C. v. (1975). A Vector Model for Perceived Object Rotation and Translation in Space. *Psychological Research*, **38,** 209-230.

Braunstein, M. L. (1976). *Depth Perception Through Motion.* New York: Academic Press.

Courellis, S. H., & Marmarelis, V. Z. (1990). An Artificial Neural Network for Motion Detection and Speed Estimation. in *Proceedings to IJCNN.* **1** San Diego, CA. (pp. 407-421). IEEE Neural Networks Council.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *J. of Physiol.*, **160,** 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive Fields and Functional Architecture of Monkey Striate Cortex. *J. of Physiol.*, **195,** 215-243.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**(2), 201-211.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, **21**(6), 575-580.

Liaw, J. S., Arbib, M. A. (1991). A Neural Network Model for Response to Looming Objects by Frog and Toad. in **Arbib, M. A., Ewert, J. P.,** *Visual Structures and Integrated Functions.* (Research Notes in Neural Computing). New York: Springer-Verlag.

Nakayama, K. (1985). Biol. Image Motion Processing: A Review. *Vision Research*, **25**(5), 625-660.

Reichardt, W. E. (1969). Movement Perception in Insects. in **Reichardt, W. E.,** *Processing of Optical Data by Organisms and Machines.* (pp. 465-493). New York: Academic.

Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., & Iwai, E. (1986). Integration of Direction Signals of Image Motion in the Superior Temporal Sulcus of the Macaque Monkey. *J. of Neurosci.*, **6**(1), 145-157.

Sakata, H., Shibutani, H., & Kawano, K. (1983). Functional Properties of Visual Tracking Neurons in Posterior Parietal Association Cortex of the Monkey. *J. of Neurophysiology*, **49**(6), 1364-1380.

Sakata, H., Shibutani, H., Kawano, K., & Harrington, T. L. (1985). Neural Mechanisms of Space Vision in the Parietal Association Cortex of the Monkey. *Vision Research*, **25**(3), 453-463.

Teeters, J. L. (1989). A Simulation System for Neural Networks and Model for the Anuran Retina. Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Tsai, H. J. & Ewert, J. P., (1987). Edge preference of retinal and tectal neurons in common toads (*Bufo bufo*) in response to worm-like moving stripes: the question of behaviorally relevant 'position indicators'. *J. Comp. Physiol.* **A 161**: 295-304.

Vaina, L. M. (1989). Selective Impairment of Visual Motion Interpretation Following Lesions of the Right Occipito-Parietal Area in Humans. *Biol. Cybern.*, **61,** 347-359.

Van Essen, D. C., & Maunsell, J. H. R. (September 1983). Hierarchical Organization and Functional Streams in the Visual Cortex. *Trends in Neurosciences*, 370-375.

Wallach, H., & O'Connell, D. N. (1953). The Kinetic Depth Effect. *J. of Experimental Psychology*, **45**(4), 205-217

Weitzenfeld, A. (1990). *NSL - Neural Simulation Language Version 2.0.1.* (TR 90-01). Los Angeles: Center for Neural Engineering, University of Southern California.

Zhou, Y. T., & Chellappa, R. (1990). A Network for Motion Perception. in *Proceedings to IJCNN.* **2** San Diego, CA. (pp. 875-884). IEEE Neural Networks Council.