A second criticism of the intersection-of-constraints model, and two stage models in general, comes from a paper by Victor and Conte [8]. They report a brain-injured patient with posterior cortical atrophy who is more severely impaired at perceiving the direction of motion of single gratings than that of a plaid. It is claimed that this undermines the two stage IOC model, which would predict that detection of plaid motion would be worse than component motion since detection of plaid motion depends on receiving accurate information about component motion.

The purpose of the present study was to examine the development of “neurons” in an artificial neural network to determine if the nature of the task for determining the direction of motion is enough to develop the type of receptive fields reported in the animal literature. Additionally, having an artificial network of this type allows for further examination of the difference between Type-I and Type-II plaids and a more detailed study of the patient recently reported by Victor et al. (1994) by making selective lesions in the network.

7. References


As reported in the animal literature, these nodes have broader tuning than the Layer-A nodes. However, unlike the nodes reported in the animal literature, the directional preference of these nodes to a moving grating was not the same as the directional preference to a moving plaid. As reported of neurons in MT, component motion-selective nodes were also found in Layer-B as well as several apparently untuned nodes.

During testing, the network was found to converge to the solution quite quickly, being near the final solution within one or two time steps. Type-II plaids were slightly slower at reaching their final solution. Network outputs are plotted as directions of perceived motion along with the intersection of constraints solution in Figure 6. Large arrows represent component grating direction-of-motion. Each input was presented for five time steps. The small arrows represent network outputs at each time step (number indicates which time step). Lines represent possible directions of motion for each component grating. The intersection of constraints solution occurs where these lines intersect.

6. Conclusions & Future Work

Our study demonstrated that a recurrent back-propagation network given moving plaid and grating input patterns and trained on the intersection-of-constraints solution, can generalize to novel plaid stimuli. Given this input/output relationship, the network develops representations similar to those seen in V1 and MT.

These results suggest that despite the existence of localized features which signal the veridical motion direction, there is a computational benefit for the processing of component motions. However they do not exclude a secondary role for local feature tracking. Both the physiological literature and the model presented in this paper contain ‘cells’ whose responses do not fall into the component or object motion dichotomy. These cells may be involved in a secondary motion processing system involving feature tracking.

Additionally, several recent studies have brought the intersection-of-constraints into question. The perceived direction of motion of Type-I plaids (whose perceived direction of motion lies between the direction of component grating motions) is accurately predicted by the IOC. However, for short presentation times, Type II plaids (whose perceived direction of motion lies outside the direction of component grating motion) appear to be moving in the direction of the vector sum of the component motions rather than that predicted by the IOC [7].
5. Network Representation

Several techniques were used to analyze the network. First, the receptive fields of the F-layer nodes were examined to determine what information they were extracting from the plaid pattern. The F-Layer nodes were found to be broadly tuned to the orientation of component gratings as reported of cells in primary visual cortex [6]. For example Node 48 is tuned for horizontal orientations and Node 52 is turned for vertical orientations as shown in Figure 3.

Other A-layer nodes were found to be relatively untuned or very widely tuned. Such cells have also been reported in the literature [2].

<table>
<thead>
<tr>
<th>Pattern Type</th>
<th>Number of Patterns</th>
<th>Average Network Error</th>
<th>Average Directional Error in Degrees</th>
<th>Average Speed Error in Pixels/Time Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plain Grating</td>
<td>162</td>
<td>0.021</td>
<td>10.79</td>
<td>0.11</td>
</tr>
<tr>
<td>Type I Plaid</td>
<td>456</td>
<td>0.032</td>
<td>9.60</td>
<td>0.21</td>
</tr>
<tr>
<td>Type II Plaid</td>
<td>1748</td>
<td>0.033</td>
<td>9.52</td>
<td>0.22</td>
</tr>
<tr>
<td>All Patterns</td>
<td>2366</td>
<td>0.032</td>
<td>9.62</td>
<td>0.21</td>
</tr>
<tr>
<td>Test Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plain Grating</td>
<td>174</td>
<td>0.023</td>
<td>8.54</td>
<td>0.11</td>
</tr>
<tr>
<td>Type I Plaid</td>
<td>460</td>
<td>0.058</td>
<td>17.19</td>
<td>0.41</td>
</tr>
<tr>
<td>Type II Plaid</td>
<td>1808</td>
<td>0.057</td>
<td>17.00</td>
<td>0.37</td>
</tr>
<tr>
<td>All Patterns</td>
<td>2442</td>
<td>0.054</td>
<td>16.44</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 1: Network Performance

As reported of V1 neurons in the animal literature [3], Layer-A nodes were found that were rather sharply tuned for the direction of motion of a moving grating (Figure 4). For each node, a predicted response to plaids composed of orthogonally oriented gratings was calculated using the method suggested by Adelson et al. [2]. A comparison of the predicted response and the actual node response indicates that the nodes are sensitive to the component motions of a moving plaid. The response of selected nodes to a moving grating is plotted in polar coordinates. The predicted response of the each node to a plaid pattern composed of orthogonally oriented gratings is very similar to that of the actual response, indicating that these nodes are sensitive to the component motion of a moving plaid. Note that the actual response is somewhat sharper than the predicted response due recurrent feedback.

Similar measurements were made the response of Layer-B nodes. Unlike, Layer-A nodes, the response of many Layer-B nodes to a moving plaid could not be accurately predicted from the response to component motion. These nodes appeared to be sensitive to the direction of the motion of the plaid as a whole rather than to the components of the plaid. The response of selected nodes to a moving grating is plotted in polar coordinates in Figure 5. The predicted response of each node to a plaid pattern composed of orthogonally-oriented gratings does not correspond to that of its actual response, indicating that these nodes are not sensitive to the component motion of a moving plaid. The nodes appear to be directionally tuned to the motion of the plaid as a whole, yet have a different preferred direction for the motion of a plaid than for the motion of a grating.
test set of 2593 patterns. Each input pattern was displayed to receptive fields of the network for 5 time units. The receptive fields of the network were constrained to be within a small aperture with a diameter of approximately 12 pixels. Figure 1 depicts five frames of a typical network input.

As an output, the network was trained to predict the perceived direction of motion of a plaid pattern (or single grating) as given by Adelson and Movshon’s [2] intersection-of-constraints calculation. Rather than a degree representation ($0^\circ - 360^\circ$) the output for the perceived direction was an $x$ and $y$ unit vector.

$$x = v \cos \theta \quad y = v \sin \theta \quad \text{(eq. 1)}$$

Where $v$ is the speed and $\theta$ is the direction of a moving plaid pattern as given by the intersection-of-constraints. This representation has the advantage that similar directions of motion require similar outputs from the network.

Additionally, the representation inherently contains information about the speed of the moving plaid, something that is not present in a degree representation:

$$v = \sqrt{x^2 + y^2} \quad \text{(eq. 2)}$$

To prevent difficulties with extreme values, the desired output for the network was scaled to keep all training values between 0.05 and 0.95. Sigmoidal, [0,1], output units require large input weights to reach extreme values and can never actually have values of 0 and 1.

Plaids moving at speeds faster that 5 pixels per time unit were eliminated from the training and testing set, constraining the values of $x$ and $y$ to be between -5 and 5. Therefore the actual scaled output values were defined as:

$$x = \left( \frac{v \cos \theta + 5}{10} \right) \quad y = \left( \frac{v \sin \theta + 5}{10} \right) \quad \text{(eq. 3)}$$

One disadvantage of this representation occurs when the perceived direction of motion is $90^\circ$. This requires the network to have an output of exactly 0.500. Nearby output values produce drastically different directions. For this reason, inputs resulting in perceived directions of $90^\circ$ were not included in calculating the performance of the network.

3. The Network

The network consisted of five layers, an input layer, $I$, a receptive field layer, $F$, intermediate hidden layers $A$ and $B$, and an output layer, $O$. Three additional layers, $C$, $D$ and $E$ hold recurrent copies of the intermediate layers and the output layer. Layers $A$, $B$, and the output layer $O$, receive input from the layer directly beneath and from copies of their activity during previous time steps (layers $C$, $D$, and $E$).

The intermediate layers and the output layer are fully connected by a set of learnable weights to their preceding layers and to their recurrent copies (See Figure 2). No weights are used for copying from the intermediate and output layers to the recurrent layers. Activations for the intermediate, ($A$ and $B$) and output ($O$) layers are directly copied into the recurrent copy layers ($C$, $D$ and $E$) at the end of each time step. All weights were initially set to small random values between -0.5 and 0.5.

Each node in the receptive field layer, $F$, is given a randomly chosen 5x5 pixel receptive field within the circular aperture when the network is first initialized (Figure 1). The $F$ nodes are connected to their respective inputs by a set of learnable weights.

4. Simulation Results

The network was trained using back-propagation with momentum. A learning rate of 0.1 and a momentum of 0.9 were used. Error was calculated for both the training and the test set and sorted according to the type of input stimuli (See...
A Neural Network Model of Motion Detection for Moving Plaid Stimuli

Lars Lidén
Department of Cognitive & Neural Systems, Boston University, Boston MA

Abstract

The perception of moving plaid stimuli is examined using a recurrent neural network trained on the intersection-of-constraints. It is found that the nodes in the network develop representations similar to neurons seen in the animal literature. Although motion signals from 2D-intersections are available, early layers in the network develop component directional-selectivity similar to neurons in V1 of the macaque monkey. Nodes higher up in the network show both object directional-selectivity and component directional-selectivity, similar to neurons found in MT of the macaque. These results support the notion of a two stage motion system which relies on component motions rather than feature tracking.

1. Introduction

The direction of motion of a single contour, such as a sinusoidally modulated grating, as seen through a small aperture is ambiguous. Only motion orthogonal to the grating is visible, parallel motion is undetectable. On the other hand, two non-parallel gratings perceived as belonging to a single moving object (a plaid) create an unambiguous directional percept [1]. Although a simple vector sum of the directions of motion of the component gratings will sometimes specify the perceived direction of plaid motion, this is frequently not the case. Adelson and Movshon [2] proposed the “intersection of constraints” (IOC) method for specifying the perceived direction of motion. Each component grating is thought to specify a range of directions which is consistent with the motion of the grating. The motion of the object as a whole is given by the intersection of the constraints imposed by each of the component gratings (See Figure 6).

Support was lent to this theory by the discovery of two different types of cells in the visual cortex. Neurons in V1 of the cat and macaque monkey, as well as about 40% of the cells in MT of the macaque, and most of the cells of lateral suprasylvian (LS) visual cortex of the cat, have been shown to have component directional selectivity [3]. They respond preferentially to a grating moving in a particular direction. When presented with a moving plaid, they show bi-lobed responses indicating that they are responding to the direction of motion of components of the plaid rather than the motion of the plaid as a whole. On the other hand about 25% of cells in MT and a small number of LS cells display pattern direction selectivity. Cells of this type show the same directional preference for the motion of a grating presented alone and for the motion of a plaid as a whole.

These findings have been used to suggest that there is a two stage process for determining motion direction, one sensitive to component motion, and another which reintegrates component motion and is sensitive to object motion as a whole [1]. Plaids, however, also contain localized features at the intersections of the maxima and minima of the component gratings. These features signal the veridical motion direction of plaids and have been suggested to play a role in the determination of motion direction [4], [5]. The role of such feature tracking in the perception of plaid motion has not been adequately resolved. The present study was designed to examine how an artificial neural network would learn to process moving plaid patterns. We examined whether such a network would break plaid motion into its component parts, as has been suggested of the human visual cortex, or whether such a network would use local feature tracking.

2. Network Input and Output

Sixteen sinusoidally modulated grating patterns of different orientations were used to create input stimuli for the neural network. Gratings were allowed to move orthogonally to their orientation at 1 or 2 pixels per time step.

Plaid training and testing sets were created by adding two grating patterns. All possible combinations of gratings at each speed were initially generated. Combinations in which only one of the gratings was moving were allowed, but combinations in which both gratings were stationary were eliminated. Inputs consisting of only a single moving grating were also included. This resulted in a total of 5,168 different possible input patterns, which were randomly divided into two sets, a training set of 2,575 patterns and a