시각적 선택에 대한 신경망 모형: FeatureGate 모형의 하향식 기제

A Neural Network Model for Visual Selection: Top-down mechanism of FeatureGate model

김 민 식*
(Min-Shik Kim)

요 약 시각적 선택에 대한 과거 정신물리학적, 신경 생리학적 연구결과를 토대로 FeatureGate 라는 신경망 모형을 제안하였다. 이 모형에는 공간 배치도가 위계적으로 구성되어 있으며, 정보의 흐름이 위계의 각 수준으로부터 그 다음 수준으로 넘어갈 때 주의 게이트에 의해 조절되도록 되어있다. 주의 게이트들은 독특한 세부 특징을 가진 위치에 반응하는 상향식 시스템과 표적 세부 특징이 있는 위치에 반응하는 하향식 기제 모두에 의해 조절된다. 본 연구는 FatureGate 모형의 하향식 기제에 초점을 맞추어 모형을 설명하고, 현재 다른 모형들이 설명하지 못하는 Moran & Desimone (1985)의 연구결과를 이 모형이 어떻게 설명하는지를 제시하고자 한다. FeatureGate 모형은 병렬적인 세부특징 검색, 계열적 접합표적 검색, 단서에 의한 주의의 점진적 감소 모형, 세부특징-주도적인 공간적 선택, 주의의 분할, 방해자극 위치의 억제, 주변 억제 등을 포함한 시각적 주의 연구의 여러 가지 많은 현상들을 설명하는데 하나의 일관적인 해석을 제공해 준다. 앞으로 이 모형을 더욱 확장, 발전시켜 세부특징의 조합된 배열에 반응하는 상위 수준의 유닛을 사용한다면 시각적 선택과정이 포함된 형태 재인 모형으로 개발될 수 있다.

주제어 Visual Attention, Spatial Attention, Top-down process, Neural Network Model (시각적 주의, 공간주의, 하향식 처리, 신경망 모형)

Abstract Based on known physiological and psychophysical results, a neural network model for visual selection, called FeaureGate, is proposed. The model consists of a hierarchy of spatial maps, and the flow of information from each level of the hierarchy to the next is controlled by attentional gates. The gates are jointly controlled by a bottom-up system favoring locations with unique features, and a top-down mechanism favoring locations with features designated as target features. The present study focuses on the top-down mechanism of the FeatureGate model that produces results similar to Moran and Desimone's (1985), which many current models have failed to explain. The FeatureGate model allows a consistent interpretation of many different experimental results in visual attention, including parallel feature searches and serial conjunction searches, attentional gradients triggered by cuing, feature-driven spatial selection, split attention, inhibition of distractor locations, and flanking inhibition. This framework can be extended to produce a model of shape recognition using upper-level units that respond to configurations of features.

---

* Department of Psychology, Yonsei University, Seoul, KOREA, 120-749
연세대학교 심리학과

연구세부분야: 인지 심리학
연락처: 02-361-2443
e-mail: kimm@yonsei.ac.kr

Our visual system is thought to consist of different processing stages, solving different computational problems and processing different representations at each stage (Marr, 1982). At any given moment the visual system receives more information than it can fully process. Thus, some portion of the input needs to be selected and processed more carefully than the rest. To understand when and how the selection occurs, a number of attention studies have tried to reveal what types of representations are selected. Most of the studies have suggested or assumed a two-stage theory of visual information processing in which the first stage is the preattentive, parallel processing and the second is the attentive, serial processing (e.g., Neisser, 1967). While the first stage processes simple features in parallel over the entire visual field, the second stage analyzes a complex form or identifies a visual object by selecting a particular location. Such a hypothesis has been supported by a number of visual search and spatial cuing studies, and can also be computationally justified by showing that parallel operations on the recognition of multiple objects over the whole visual scene result in a combinatorial explosion of computational resources (e.g., Ullman, 1984). From a computational viewpoint, a prime goal of the visual attention studies is to understand what is the most important problem in visual attention and how the problem can be solved. Based on our current, limited understanding of "how our visual system selects information or from what types of visual representations visual stimuli are selected," the selection is most likely based on location. If so, then how can the selection by location be implemented in biological visual systems? One approach to solve this problem will be to consider the known neurobiological and computational properties of our visual system, and then to formulate a model within these constraints. The present study will describe the general background for constructing our neural network model for visual selection, named the "Feature Gate (FG)" model (Kim & Cave, 1994; Cave, Kim, Bichot, & Sobel, 1999). Incidentally, although FG is still under development, the final version of the model will be constructed with enough detail so that it can generate many useful experimental predictions as well as new theoretical questions.

Neurophysiological and Psychophysical Constraints

Neurophysiological studies have shown that our visual system extracts progressively more complex forms of structure at each stage of visual processing. For example, numerous simple cells in area V1 code for the local orientation of contrast, each at a different position and scale in a massively parallel manner. At progressively higher levels, neurons in V2 and V4 can respond to more complex forms (e.g., hyperbolic patterns, Gallant, Braun, & Van Essen, 1993), and finally neurons in inferotemporal cortex (IT) and superior temporal sulcus (STS) appear to respond to complex objects (Gross, Rocha-Miranda, & Bender, 1972; Rolls & Baylis, 1986). Moreover, at the highest levels of visual processing (IT, STS), the neural response to the identity of the object ("what") is rarely affected by the location and size of the object ("where"). Many neurophysiological studies have shown that our visual system processes visual features mainly with two separate visual pathways, the "what" and "where" pathways (e.g., Mishkin, Ungerleider, & Macko, 1983; Livingstone & Hubel, 1988; Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). This implies that our visual system saves a great number of neural resources by encoding "what" separately from "where" (e.g., see Olshausen, 1994).

Besides these neurophysiological constraints, many attention models suggest that primary visual feature dimensions such as color, orientation, location and so on, are processed and represented with independent modules or separate feature maps (see, e.g., Treisman, 1988; Cave & Wolfe, 1990; Bundesen, 1990). In addition to the global constraints from separate coding structures, another important constraint came from a single-cell study by Moran and Desimone (1985), which most current attention models cannot explain.

Moran and Desimone recorded single cells in V4 and IT of monkeys who were trained to attend to stimuli at one location in the visual field and ignore stimuli at another location. When the two stimuli were both within the receptive field of a cell, the cell activity to the unattended stimulus was inhibited while the cell activity to the attended stimulus was not. Moreover, when one stimulus was inside the receptive field and another stimulus was outside, the cell activity to the stimulus inside the receptive field was not inhibited, whether it was attended or not. These physiological data give the following important implications in building a neural network model of visual selection: i) a target location (attended location) will be activated, regardless of whether distractors appear near the target or not, ii) distractor locations will be inhibited when the target is present nearby them, and iii) distractor locations will not be inhibited when target is not present or when the target is far from them.

Many psychophysical studies have shown that location plays an important role in the organization and selection of visual information even when location is irrelevant to correct responses in tasks (e.g., Tsal & Lavie, 1993; Cave & Pashler, 1995). Also, Kim and Cave (1995) measured spatial attention in visual search tasks using a probe technique and showed that both speed and accuracy for detecting probes are facilitated at the target location, and at the locations containing a distractor with one of the target's features, implying spatial attention driven by target features. Also, this study gave strong evidence that attention is used in a very easy search task with an "almost flat" slope which has been considered to be "parallel."

Based on these empirical results, a "location-specific" selection will be implemented with a neural network model. Like any other models, the model proposed here will have to explain a long line of empirical results from the previous studies noted until now. In particular, with the neural network model, the following properties of visual selection will be implemented: i) Selective processing will operate based on spatial attention, ii) Spatial attention will be allocated based on target features, iii) Selective spatial processing will occur at an early stage (e.g., see also Motter, 1993, for neurophysiological evidence that attentional effects were observed for V1 and V2 neurons as well as V4), iv) As suggested by the physiology of our visual system, the model will have a hierarchy of units (neurons), with larger receptive fields for the higher layers, v) To select locations quickly and efficiently, the model will rely on parallel distributed selection and local operations through the multiple layers, and vi) As suggested by Moran and Desimone (1985), activation of a location unit will be inhibited when a distractor feature occupies the location and at the same time a target feature is located within the same receptive filed.

Architecture of FG

In this section, a model architecture will be built up in steps to meet the above constraints. Although the model is designed to be consistent with known physiology, its particular components have not yet corresponded to particular cortical areas. For now, the model assumes that color and orientation are analyzed independently, just as

Feature Integration Theory (Treisman & Gelade, 1980; Treisman & Sato, 1988) and Guided Search (Cave & Wolfe, 1990) did. The model will explore basic architectures for visual selection, and as the model develops, it will reflect known physiological structures more and more. I first introduce some of the basic components of the model and then progressively add more components.

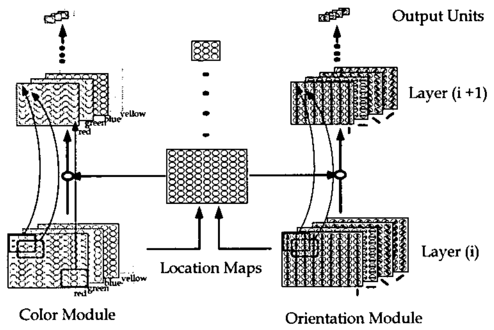

Figure 1. Color and Orientation Modules and Location Map.

Feature Modules and Location Map. Figure 1 illustrates a global framework of the model that consists of some of the basic components, such as Color Module, Orientation Module, and Location Maps. As shown in Figure 1, within each layer, different feature units respond to each color or orientation at each location. In the lower layer, all units have small receptive fields. As information travels up the hierarchy, the number of units in each layer decreases and the receptive field size relatively increases. At the top level, the receptive field for each output unit covers the entire visual field. Each unit in the layer (i+1) receives input from its receptive field ("neighborhood") that consists of a 3 x 3 array of units in the layer (i) via links that are dynamically gated by their location units.

Location Map. Figure 2 shows how the location units gate connections between Layer (i) and (i+1). For simplicity, only one

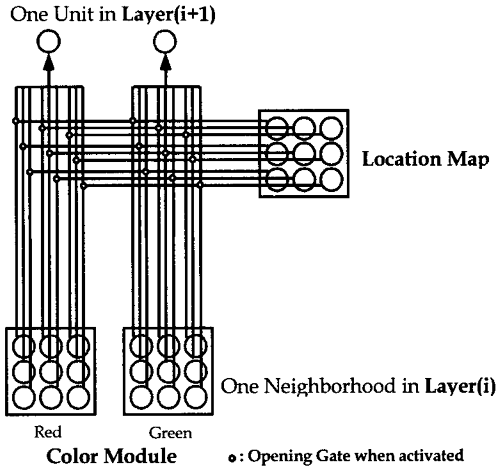

Figure 2. Location Map gates connections between two layers of Color Module.

neighborhood in Layer (i) was depicted using only red and green colors. Each unit in the location map controls the information flowing from its corresponding feature units to the next layer. Also, as shown in Figure 3, location units within each neighborhood interact competitively. It is a Winner-Take-All competition that allows the net to localize the most activated unit while inhibiting all others. Thus, only the information from that location will be transmitted to the next level.

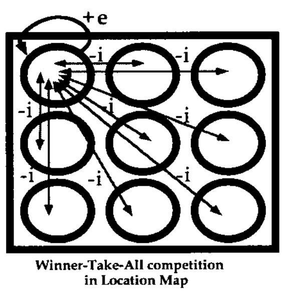

Figure 3. Competitive interaction within one Neighborhood in Location Map.

Selective Inhibitors. What determines the activations of location units? As shown in Figure 1, feature units in Layer (i) can activate their corresponding location units as well as a feature unit in the upper layer.

However, the connections between units of the feature map and the location map are also gated by an additional set of units, named Selective Inhibitors. As shown in Figure 4, each Selective Inhibitor unit receives inputs from three different sources. The first source is the Target Feature Unit. For example, if the target is a green item, then the Target Feature Unit for green would be active. The second source is their own feature units at their locations. The final source is other feature units at some other location within the same neighborhood. Activation of these three sources are combined by a set of Selective Inhibitors for each location, each receiving input from the current location, other locations feeding into the same units at the next level, and the Target Feature Units.

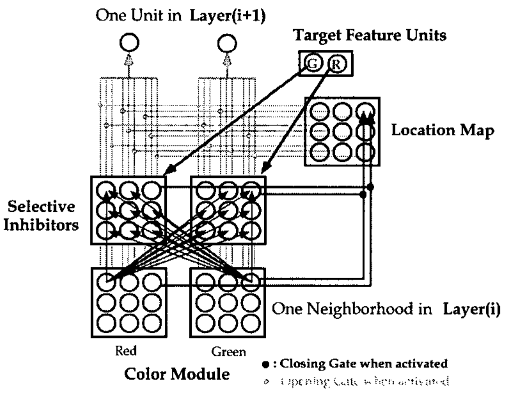

Figure 4. Selective Inhibitors gate links from Feature Units to Location Units.

Once these Selective Inhibitor units are activated, they in turn inhibit the link between its feature unit and location unit. Therefore, a location is inhibited when it contains a nontarget feature and another unit in the same neighborhood contains a target feature. In other words, a location is inhibited only when it carries a strong signal that will interfere with a selected signal at the next step up in the hierarchy. For example, when a target is a green item and red distractors appear with the green item, the Selective Inhibitor units in the red feature map will be activated, and thus will inhibit the activity transmitted from the red feature units to their corresponding location units. However, if both target and distractor features do not appear within the same neighborhood, then no Selective Inhibitors will be activated and thus no inhibition will occur. The way in which the Selective Inhibitors combine information illustrates a basic principle of the model. Any features will be gated or attenuated whenever they are at a location with a distractor feature that will interfere with a target feature at the next step up in the hierarchy. Under this principle, an object with the target feature will make it to the top of the hierarchy, no matter where it appears in the visual field, and what other distractors are present.

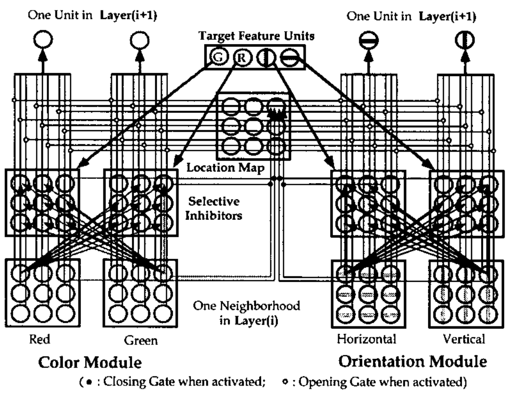

Figure 5. One portion of the model with Color and Orientation Modules

The same structure can be used for other feature dimensions such as orientation (Figure 5). If a unit in the Location Map is not activated, all information from that location will be blocked or attenuated, including color, orientation, and all other features. Moreover, the network shown in Figure 5 can be replicated may times to cover the entire visual field. This collection of nets would form one layer of a processing hierarchy. The output units at this level would be input units for the next level. All the units from one neighborhood feed into the same set of units at the next level. At each level of the hierarchy, the receptive fields are larger than the level below. Finally, at the top level, the receptive fields cover the entire visual field. Therefore, spatial attention determines what information makes it to the top of the hierarchy. Our visual system would require such a selection mechanism because without selection, all the features from all the visible objects would be on in the top level, and there would be no way to sort them out.

Example - with conjunction stimuli of color and orientation.

In this section, an example will be provided to show how the structure depicted in the previous section actually selects a particular location in steps. Here again, only one neighborhood is used for simplicity.

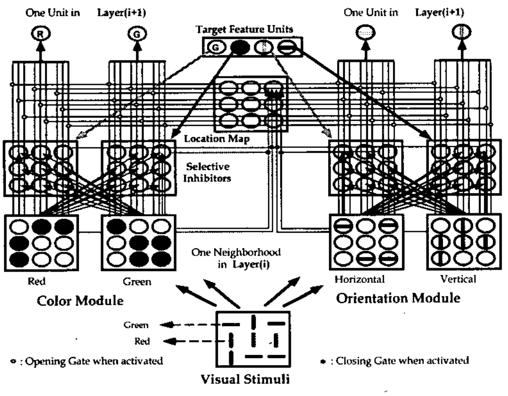

Figure 6. Stimuli are presented to the network through input units. Here, the target is a red horizontal bar.

As shown in Figure 6, visual stimuli are presented to the network through input units. The target is a red horizontal bar in this example. It is assumed that visual stimuli are encoded in Feature Maps preattentively according to the feature properties. If a target feature is present in the neighborhood, then Selective Inhibitors are activated at those locations that have distractor features (Figure 7). Each active Selective Inhibitor unit inhibits the connection between its corresponding feature and location units (Figure 8). Thus, distractor feature units cannot activate their corresponding location units when a neighboring location contains a target feature. Finally, the location unit whose activation is highest will be picked by the Winner-Take-All mechanism, and thus only information from that location flows freely to the next level (Figure 9).

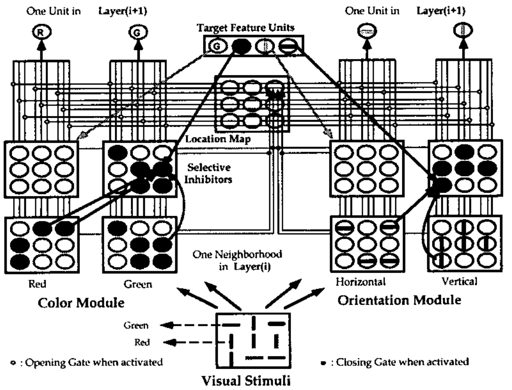

Figure 7. When a target feature is present in the neighborhood, Selective Inhibitors are activated at those locations that have distractor features. (Note: Black units are active. White units are inactive.)

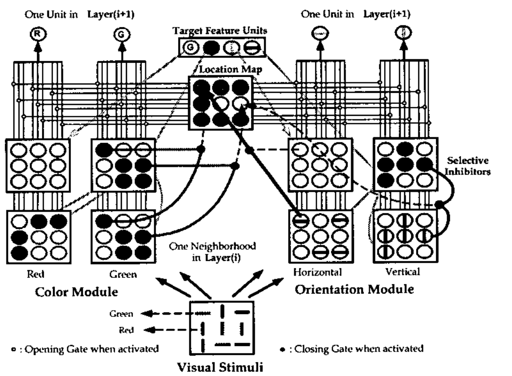

Figure 8. Each active Selective Inhibitor inhibits the link between its corresponding feature and location units. Here, distractor feature units cannot activate their corresponding location units, and the activation of the target location unit is largest.

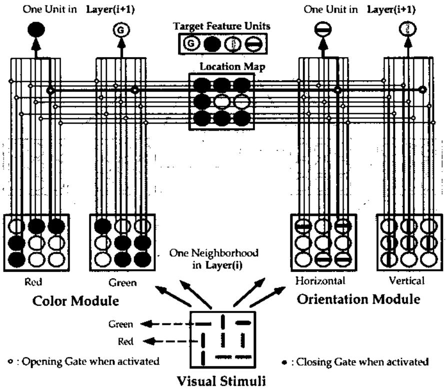

Figure 9. When a Location Map unit is active, activation from the feature units for that location flows freely to the next level. Here, the features at the target s location are selected while the features at the other locations are inhibited.

Computer Simulation

Up to now, I have described some basic structures and mechanisms of the top-down components of the model using only one neighborhood. Since the Feature Gate model is composed of network elements arranged in a hierarchy, its mechanism can route information from any location in the visual field to the top of the hierarchy while limiting the interference from distractor inputs. To demonstrate this, the top-down components of the model have been implemented as a computer simulation using two types of visual search, simple feature and conjunction searches. In the current simulation, the model structure contains five layers, and each neighborhood in each layer consists of a 3x3 region of locations and each neighborhood is partially overlaps nearby neighborhoods Layers 1 and 2.

Figure 10 shows a computer simulation of simple feature search. Here, the task is to search for a red item. The activities of all the units here are assumed to be analog values between 0 and 1 (the greater the activation, the darker it is). The net begins with the input layer (Layer 1) with each feature map composed of a 21x21 array of units, and the number of units decreases as the information feeds into the next layer. As shown in the Location Maps of Figure 11, the distractor locations sharing the same neighborhood with the target feature are inhibited at each Location Map. Also, random noise arranging -0.3 to +0.3 is presented in each location, as it was in Guided Search (Cave & Wolfe, 1990), to prevent search from being more efficient than that found in human subjects. Although it is assumed that the random noise could occur in the feature maps, Selective Inhibitors, the transfer of information between the different maps, or some combination of these, it was added only to each Location Map here for simplicity. With random noise distractor locations may obtain a high activation leading the wrong location to be selected. However, even with the random noise, the net can usually select the target feature successfully at the first run, as shown in the Output unit. The Combined Location Map illustrates an averaged activity for each location unit through all the four Location Maps. This map plays no role in the operation of the model, but displays how inhibition at different levels of the hierarchy combines for each location in the field. As you see, the target location has the highest value and distractor locations nearby the target are relatively more inhibited than those far from the target.

By adding orientation features, a simulation of conjunction search is performed(Figures 11 and 12). Without noise(Figure 11), the net selects the target features(in this example, red and horizontal) successfully on the first run. Also, the Combined Location Map shows the highest activation at the target location and gradually decreasing activation from that location, producing a "gradient" of spatial activation, as suggested by a spatial gradient model(Downing & Pinker, 1985;Mangun & Hillyard, 1988;LaBerge & Brown, 1989). However, with the same amount of random

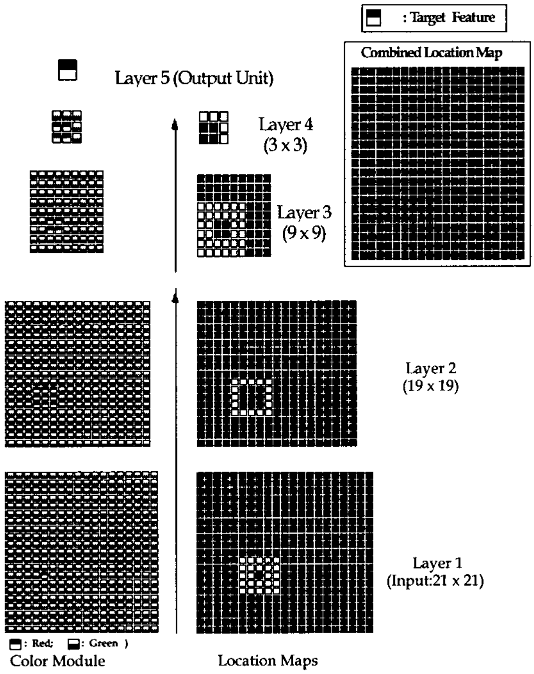

Figure 10. Computer simulation of simple feature search with ± 30 % noise in each Location Map.

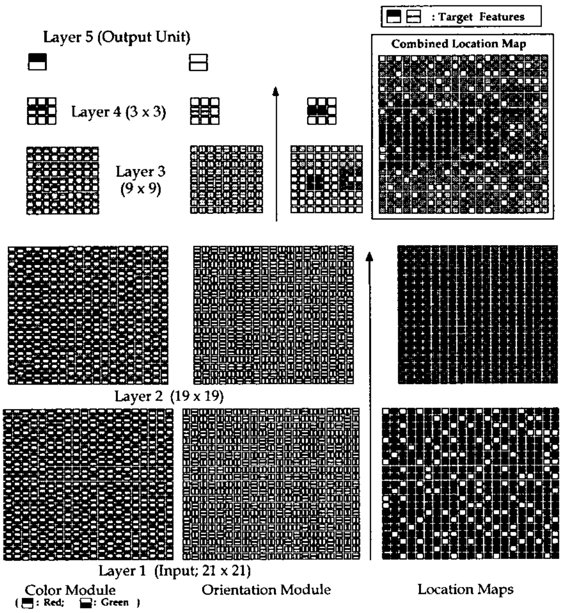

Figure 11. Computer simulation of conjunction search without noise in each Location Map. Without noise, the conjunction target is selected easily.

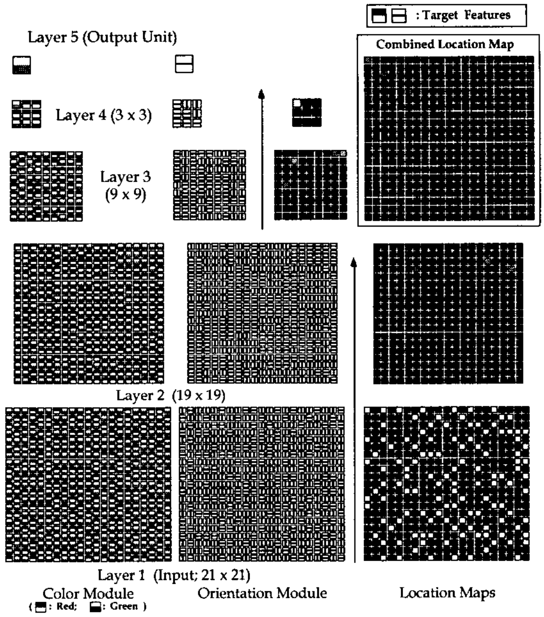

Figure 12. Computer simulation of conjunction search with ± 30 % noise in each Location Map. With noise, a distractor is selected over the target. The role of noise in this model is similar to that in Guided Search. If a nontarget is selected, the model inhibits the selected location and continues to search serially until it finds the target.

noise as before(+/- 0.3), a distractor is selected selected over the target (Figure 12). If a nontarget is selected, the model inhibits the selected location (inhibition of return) and continues to search serially until it finds the target.

Discussion

Predictions. First of all, as shown in the simulation of visual search, the Feature Gate (FG) model can predict different results between the simple feature search and conjunction search tasks (Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989). With the same amount of noise, the model can select a simple target feature very easily, independently of the number of distractors. However, in conjunction search, the model rarely selects the target as output at the first run. When the selected location is not the target, the net will inhibit that location (inhibition of return; Posner & Cohen, 1984; Klein, 1988), allowing selection of the next highly activated location. Thus, it is capable of ordered serial search similar to that in Guided Search (Cave & Wolfe, 1990; Wolfe, 1994). On the other hand, the amount of inhibition will be decreasing after some period of time, allowing items at the inhibited location to be revisited.

Also, FG can explain the spatial attention effects at the cued location, as shown in many cueing experiments (Eriksen & Hoffman, 1974; Posner, Snyder, & Davidson, 1980; Remington & Pierce, 1984). Although mechanisms concerned with bottom-up activation are not included in the current model, a peripheral cue will activate its location in the location map through bottom-up activation, allowing a later stimulus at that location to travel up the hierarchy more easily. Also, the model could explain central cuing experiments by including direct top-down connections to the Location Map.

The model can also explain why probes are detected more quickly or more accurately when they are at or near the primary target location (Hoffman & Nelson, 1981; Heinze, Luck, Münte, Gös, Mangun, & Hillyard, 1994; Kim & Cave, 1995, 1999a, 1999b; Cepeda, Cave, Bichot, & Kim, 1998). FG selects target features by activating their locations, or more specifically by inhibiting distractor locations. Once a Location Map unit has been inhibited, it takes time to adjust after a new stimulus appears. Thus, it will slow the detection of a probe appearing at the distractor location. On the other hand, since the model moves the locus of attention by inhibiting the currently selected location unit and activating a new one, rather than by sliding a spotlight across the visual field, it predicts shifts of visual attention that are independent of distance, as suggested by many psychologists (Remington & Pierce, 1984; Sagi & Julesz, 1985; Murphy & Eriksen, 1987; Eriksen & Webb, 1989; Kwak, Dagenbach, & Egeth, 1991). Also, the model can be implemented so that search displays with multiple targets are more readily responded to than those with a single target. That is, two target locations in the same neighborhood can activate each other, so that one of them can feed into the next level more quickly. This property of the model can explain Mordkoff, Yantis, and Egeth's (1990) results that most attention models based on serial search cannot explain. Mordkoff et al. (1990; also see Mordkoff & Yantis, 1993) showed that the fastest response to scenes including multiple targets is faster than the fastest response to scenes including single targets. These data are a striking contradiction of the serial binding hypothesis, which predicts that the fastest responses to scenes should not vary with the number of targets. In FG, all of the target locations can be selected simultaneously, so that they are all contributing to the activation of the units at the top level. Because these top-level units receive activation from multiple locations, they reach a threshold of activation more quickly, and a response is generated more quickly.

Incidentally, the model proposed here can explain many other previous studies. As mentioned earlier, the Combined Location Map in conjunction search (Figure 11) suggests an attentional gradient (Downing & Pinker, 1985; Mangun & Hillyard, 1988; LaBerge & Brown, 1989). However, with a popping-out target in simple feature search (Figure 10), it shows the highest activation at the target location, more inhibition nearby it, and gradually recovering activation, producing a 3-D symmetric Mexican-hat-shaped distribution of spatial attention. Also, the response of a single unit in this model is similar to that of a single cell in V4 and IT (Moran & Desimone, 1985). When both target (attended) and nontarget (unattended) features appear in the same neighborhood (receptive field), the unit (cell) activity to the nontarget features is inhibited while the activity to the target feature is not. However, when the target feature is outside of the neighborhood, nontarget features are not inhibited.

Comparison with other models. Most of attentional mechanisms suggested in cognitive psychology consist of unknown structures, whose characteristics are only conceptually described, such as a spotlight metaphor of spatial attention. Unlike those models, Feature Gate model contains a concrete network of simplified neuron-like units, and it can be simulated directly on computers. Moreover, one motivation for this visual attention model is to build the basis for a more complex model that will eventually explore how visual selection contributes to recognition. Most current attention models, such as Feature Integration Theory (Treisman & Gelade, 1980; Treisman & Sato, 1990) and Guided Search (Cave & Wolfe, 1990), cannot explain how the selected information is actually routed to recognition mechanisms. The current model could be expanded to identify shapes by including units in the higher levels of the hierarchy that respond to combinations of shape features. Mozer s (1991) model of letter and word recognition illustrates how such an architecture can be used to recognize complex shapes. Additional Target Feature Units for shapes would also be added. These units will allow the model to select locations containing particular shapes such as letters.

As mentioned earlier, one of the most surprising results from Moran and Desimone s (1985) study was that a unit was not inhibited when the attended area was outside the unit s receptive field. FG reproduces basically the same results as Moran and Desimone's by employing location-based operations within each receptive field. Most current attention models cannot explain this phenomenon. For example, Mozer (1991) set out to model object recognition and found that he had to include a mechanism for spatial attention. Although Mozer's attention mechanism serves similar functions to FG, such as gating the connections between two layers, it cannot explain Moran and Desimone's result because it suppresses all unattended stimuli at each level regardless their proximity to attended stimuli.

A neural network model by Olshausen, Anderson and Van Essen (1993) also attempts to describe how visual selection serves object recognition. In their model, which was developed from Anderson and Van Essen (1987), they tried to implement an idea that visual attention could be the key to forming position- and size-invariant object representations. Like FG and Mozer's model, this model has a hierarchy of units, with larger receptive fields in the higher levels. One of the basic features in their model is that it uses an attentional window in which interfering information is inhibited. Thus, in their model, any input beyond the attentional window or outside a cell's receptive field will not be inhibited, as shown in Moran and Desimone's results.

Although Mozer's model and Olshausen et al.'s model share many similar features, they differ in some aspects. For example, in Mozer s model, units in each layer encode combinations of the features at the layer below. Thus, spatial relations are not preserved but are re-encoded. In this type of system, selection and identification can occur simultaneously within the same set of units. In Olshausen et al.'s model, however, spatial relationships between different locations are preserved as information travels up to the hierarchy. That is, a region of visual input is selected and passed up through a hierarchy of levels without loss of its spatial relationships. Thus, the spatial layout of the selected portion of the visual field is represented at the top layer and then analyzed - therefore, identification occurs after selection.

Niebur, Koch, and Rosin (1993) offered a somewhat different account for Moran and Desimone's result based on the "temporal tagging" hypothesis (Crick and Koch, 1990). In their model, attention to a particular location initiates synchronous oscillations that are passed up the hierarchy. For example, at the lower layer (e.g., V1 pyramidal cells), each cell responds to red or green at a particular location. If a particular region of the visual field is selected, the firing rate of V1 cells within the selected region is adjusted to correspond to a 40 Hz range. This adjustment is detected by inhibitory cells at the next layer (V4 stellate cells) and is used to inhibit V4 pyramidal cells that respond to the distractor feature. As noted in their paper, the possibility of "tagged" input in attended area by such oscillations was suggested by some studies (e.g., Gray & Singer, 1989; Kreiter & Singer, 1992). However, other recent studies (e.g., Tovee & Rolls, 1992; Young, Tanaka, & Yamane, 1992) failed to find evidence for the oscillations in monkeys' areas V1 and IT. Also, this model cannot preserve spatial relationships within the selected area.

Cohen, Dunbar, and McClelland (1990) proposed another network model of attention. Their model provides an account of the Stroop effect which requires attentional shifts in a more general sort of attention between different tasks, and not spatial selection. That is, they are mainly concerned with shifts between different processing tasks such as word identification and color naming, and not shifts in processing between different locations in the visual field. Nonetheless, the basic mechanism in their model is similar to the gating in FG between the input and output units by the Location Map, and also the control over the Location Map from the Selective Inhibitors. Another network model that simulates the Stroop effect is proposed by Phaf, Van der Haijden, and Hudson (1990). While Cohen et al.'s model explains the Stroop effect based on weight differences as a consequence of differential learning, Phaf et al.'s model predicts the effect based on its particular architecture such as direct connections, which is similar to FG.

Humphreys and Müller (1993) presented a connectionist model that performs visual search in parallel across the visual field based on Duncan and Humphreys's (1989) visual search theory. Duncan and Humphreys claimed that search items are encoded in parallel, independently of whether they are simple features or conjunctions. The main point of their theory is that search difficulty increases as similarity of target to distractors increases and similarity between distractors decreases. Also, an efficient search can be accomplished by rejecting nontargets that are strongly grouped. Likewise, in the model of Humphreys and Müller (1993), similar items are grouped, and selected together. That is, once combined-features (e.g., L or T) are encoded in parallel, grouping is implemented by within-map facilitatory links and between-map inhibitory links. Then, areas with unambiguous grouping are rejected recursively until the target is found (positive response), or no items are left (negative response). Thus, this model predicts a flat search slope with a single distractor group, and steeper slopes with increased the number of distractor groups. This model was tested with several search tasks and showed results consistent with human subjects (see also, Müller, Humphreys, & Donnelly, 1994). One drawback in this type of model is that a number of connections are required for the grouping processes. Moreover, if grouping is assumed to operate on the basis of not only simple-form conjunctions as in their model, but also many different levels of representations, the number of required connections would increase explosively.

Conclusion

Examining the special role of spatial attention in visual selection raises a number of theoretical controversies in visual attention research. For example, one must consider the distinctions between early- versus late-selection, between object-based versus location-based selection, between parallel versus serial processing, between top-down versus bottom-up processing, and so on. Spatial attention undoubtedly plays a special role in visual selection and recognition, but the exact nature and extent of that role still remain unanswered. Moreover, If visual processing involves a number of separate stages, then each stage will have its own computational limits, and may have its own selection mechanism(s) rather than having a single unique attention mechanism across all stages. Assuming that such multiple attention systems exist, we should not interpret different results from different paradigms as being mutually exclusive. For example, both spatial attention and object-based attention occur in visual selection, one at an early stage, the other at a later stage, respectively, or both simultaneously. Yet, the FeatureGate model presented here is an attempt to provide a consistent interpretation of a great deal of empirical data based on a single selection mechanism, location-based selection, and this view is tenable and seems to provide the most parsimonious accounts for them. The Feature Gate model shows a single mechanism that can account for different attentional results from visual search, location cuing, and spatial probes. Its architecture is generally consistent with known physiology. It demonstrates how location-based selection can efficiently select targets defined by nonspatial properties such as color and orientation. It also shows that an attentional gradient can arise from a hierarchy of different scales. This general architecture can be developed into a model of visual recognition, and will help in exploring how spatial selection contributes to higher-level cognition.

The current model is presented as a starting point for future theorizing about visual selection, and it is still under developing. In the line of developing the model, Cave, Kim, Bichot, and Sobel (1999) added a bottom-up system to the model and elaborated the simulation so that it can reproduce results from different attention experiments, including visual search, cueing, and probe experiments. Although the model is meant to be primarily a model of spatial attention, eventually the model will be expanded to identify shapes, so that it can be used to explore how spatial selection contributes to object recognition in complex displays.

References

Anderson, C. H. & Van Essen, D. C. (1987). Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences*, 84, 6297-6301.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97, 523-547.

Cepeda, N. J., Cave, K.R., Bichot, N. P., & Kim, M.-S. (1998). Spatial selection via feature-driven inhibition of distractor locations. *Perception & Psychophysics, 60(5)*, 727-746.

Cave, K.R., Kim, M.-S., Bichot, N.P., & Sobel, K.V. (1999). *Visual selection within a hierarchical network: The FeatureGate model.* Manuscript in preparation.

Cave, K.R., & Pashler, H. (1995). Visual selection mediated by location: Selecting successive visual objects. *Perception and Psychophysics*, 57, 421-432.

Cave, K.R. & Wolfe, J.M. (1990) Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225-271.

Cohen, J. D., Dunbar, K., & McClelland J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332-361.

Crick, F., & Koch, C. (1990). *Towards a neurobiological theory of consciousness.* unpublished manuscript.

Downing, C., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI* (pp.171-187). Hillsdale, NJ: Erlbaum.

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433-458.

Eriksen, C.W. & Hoffman, J.E. (1974). Selective attention: Noise suppression or signal enhancement? *Bulletin of the Psychonomic Society, 4*, 587-589.

Eriksen, C., & Webb, J. (1989). Shifting of attentional focus within and about a visual display. *Perception & Psychophysics, 45*, 175-183.

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science, 259*, 100-103.

Goodale, M. A., & Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*, 20-25.

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Science, U.S.A., 86*, 1698-1702.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology, 35*, 96-111.

Heinze, H.-J., Luck, S.J., Münte, T.F., Gös, A., Mangun, G.R., & Hillyard, S.A. (1994). Attention to adjacent and separate positions in space: An electrophysiological analysis. *Perception & Psychophysics, 56*, 42-52.

Hoffman, J.E. & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics, 30*, 283-290.

Humphreys, G. W., & Müller, H. J. (1993). Search via recursive rejection (SERR): A connectionist model of visual search. *Cognitive Psychology, 25*, 43-110.

Kim, M.-S., & Cave, K.R. (1994). *A Neural network model for visual selection.* Poster presented at the meeting of the Psychonomic Society, St. Louis, Mo.

Kim, M.-S., & Cave, K. R.(1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science, 6*, 376-380.

Kim, M.-S., & Cave, K. R. (1999a). Top-down and bottom-up attentional control: On the nature of interference from a salient distractor. *Perception & Psychophysics, 61*, 1009-1023.

Kim, M.-S., & Cave, K. R. (1999b). Grouping effects on spatial attention in visual search. *Journal of General Psychology, 126*, 326-352

Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature, 334*, 430-431.

Kreiter, A. K., & Singer, W. (1992). Oscillatory neuronal responses in the visual-cortex of the awake macaque monkey. *European Journal of Neuroscience, 4*, 369-375.

Kwak, H.-W. Dagenbach, D., & Egeth, H. (1991) Further evidence for a time-independent shift of the focus of attention, *Perception & Psychophysics, 49*, 473-480.

LaBerge, D. & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96*, 101-124.

Livingstone, M. & Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, 240*, 740-749.

Mangun, G.R., & Hillyard, S.A. (1988). Spatial gradients of visual attention: behavioral and electrophysiological evidence. *Electroencephalography and Clinical Neurophysiology, 70*, 417-428.

Marr, D. (1982) Vision: *A Computational investigation into the human representation and processing of visual information*. Sanfransisco, CA: Freeman.

Mishkin, M., Ungerleider, L.G., & Macko, K.A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences 6*, 414-417.

Moran, J. & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782-784.

Mordkoff, J. T., Yantis, S., & Egeth, H. E. (1990). Detecting conjunctions of color and form in parallel. *Perception and Psychophysics, 48*, 157-168.

Mordkoff, J. T., Yantis, S. (1993). Dividing attention between color and shape: Evidence of coactivation. *Perception and Psychophysics, 53*, 357-366.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology, 70*, 909-919.

Mozer, M.C. (1991). *The Perception of Multiple Objects*. Cambridge, Mass., M.I.T. Press.

Müller, H. J., Humphreys, G. W., & Donnelly, N. (1994). Search via recursive rejection (SERR): Visual search for single and dual form-conjunction targets. *Journal of Experimental Psychology: Human Perception and Performance, 20, 235-258.*

Murphy, T., & Eriksen, C. (1987). Temporal changes in the distribution of attention in the visual field in response to precues. *Perception & Psychophysics, 42*, 576-586.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Niebur, E., Koch, C., & Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Research, 33*, 2789-2802.

Olshausen, B. A. (1994). *Neural routing circuits for forming invariant representations of visual objects*. Thesis for the Degree of Ph. D.

Olshausen B.A., Anderson C.H., Van Essen D.C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience, 13*, 4700-4719.

Phaf, R. H., Van Der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*, 273-341.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bowhuis (Eds.), *Attention and Performance X* (pp. 531-556). Hillsdale, NJ: Erlbaum.

Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160-174.

Remington, R., & Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics, 35*, 393-399.

Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex in the superior temporal sulcus of the monkey. 6*Experimental Brain Research, 65*, 38-48.

Sagi, D. & Julesz, B. (1985). Fast noninertial shifts of attention. *Spatial Vision, 1*, 141-149.

Tovee, M. J., & Rolls, E. T. (1992). Oscillatory activity is not evident in the primate temporal visual cortex with static stimuli. *NeuroReport, 3*, 369-372.

Treisman, A. M. (1988). Features and objects:

the Fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology, 40A*, 201-237.

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A.M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 456-478.

Tsal, Y. & Lavie, N. (1993) Location dominance in attending to color and shape. *Journal of Experimental Psychology: Human Perception & Performance, 19*, 131-139.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-159.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems: Separation of appearance and location of objects. In D. L. Ingle, M. A. Goodale, & R. J. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Wolfe, J. (1994). Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin and Review, 1, 202-238.Wolfe, J.M., Cave, K.R., & Franzel, S. (1989). Guided Search: An alternative to the feature integration model for visual search. Journal of Experimental *Psychology: Human Perception & Performance, 15*, 419-433.

Young, M. P., Tanaka, K., & Yamane, S. (1992). On oscillating neuronal responses in the visual cortex of the monkey. *Journal of Neurophysiology, 67*, 1464-1474.

This work was supported by BrainTech21 grant from the Ministry of Science and Technology of Korean government, and in part by Yonsei University Research Fund of 1998.