Artif Intell Rev (2013) 40:27–50
DOI 10.1007/s10462-011-9278-y

A critical review of selective attention: an interdisciplinary perspective

KangWoo Lee · Hyunseung Choo

Published online: 21 July 2011

Abstract During the last half century, significant efforts have been made to explore the underlying mechanisms of visual selective attention using a variety of approaches—psychology, neuroscience, and computational models. Among them, the computational approach emerged on the stage with the development of computer science and computer vision focusing researchers interests in this area. However, computer scientists often face the difficulty of how to construct a computational model of selective attention working on their own purpose. Here, we critically review studies of selective attention from a multidisciplinary perspective to take lessons from psychological and biological studies of attention. We consider how constraints from those studies can be imposed on computational models of selective attention.

Keywords Selective attention · Computational model · Multidisciplinary approach

1 Introduction

Due to the manifold nature of visual attention, many approaches—from psychological and biological studies, to computational modelling—have been used to understand, simulate and implement its functions. Marr's influential work distinguished three different levels involved in understanding complex visual systems—computational theory, algorithm and implementation (Marr 1982). In his terms, the computational level is the description of what the goal

This research was supported in part by MKE and MEST, Korean government, under ITRC NIPA-2010-(C1090-1021-0008) and WCU NRF (No. R31-2008-000-10062-0), respectively.

K. Lee
Department of Interaction Science, Sungkyunkwan University, 300, Chon-Chon Dong,
Jang-Ahn Ku, Suwon 440-746, Korea
e-mail: kangwooster@gmail.com

H. Choo (✉)
Department of Computer Engineering, Sungkyunkwan University, 300, Chon-Chon Dong,
Jang-Ahn Ku, Suwon 440-746, Korea
e-mail: choo@ece.skku.ac.kr

of the computation is, the algorithmic level is the operational description of how the computation is carried out, and the implementational level is the description of how the algorithm is realized physically in the brain or in a computer. The results of an interdisciplinary approach can extend our knowledge about visual attention. However, it simultaneously requires additional work to integrate the different stories from the various levels into a single and coherent account, as in a game where different pieces of a puzzle must be placed to make the overall picture.

The puzzle game, "what is visual attention", is not a random game to pick up a piece of a picture and put it in an arbitrary place. The rules of our game are constrained by the findings of studies from the different levels. That is, the constraints not only provide their own criteria—e.g. biological or psychological plausibility or computational efficiency that a study must achieve, but also require that we maintain the general consistency of an interdisciplinary approach through mutual constraints. This also leads researchers to investigate new and challenging problems that arise from conflict between studies from different levels.

This paper critically reviews studies of visual attention that have been carried out in three different areas—psychological, biological and computational studies—and compares them. In the first section, we introduce theoretical work developed in psychology. The theoretical work includes important, actively debated, issues such as 'why attention is necessary' and 'how it works'. This section is primarily motivated by Marr's first level, the goal of visual computation, but it is also motivated by Marr's second algorithmic level. In the second section, we introduce work on the biological foundations of visual attention. This section is closely linked to Marr's third level—how visual attention is implemented in the brain, and includes such issues as 'where the locus of attention is', 'what the underlying neural mechanism of visual attention might be' etc. In the third section, we review computational models that owe a debt to both psychological and biological studies. Even though we divide this paper into sections according to these areas, our reviews in each section are not strictly limited; rather, we cross boundaries between the areas to show the interdisciplinary constraints. Finally, a summary is given at the end of the paper.

2 Visual attention in psychological studies

2.1 Purpose of attention

First, we may ask *why visual attention is needed* as in Marr's metatheoretical conceptualization of the study of vision (Marr 1982). There may be many different reasons depending on the task that a person or a machine carries out. Among the many possible purposes of visual attention, some basic and important assumptions are introduced here.

The central assumption of the necessity of attention is the limited amount of perceptual resource available for a given task or process. This can vary with a number of factors such as motivation, difficulty of a task, alertness, etc. That is, the basic purpose of attention is to avoid possible information overload to protect a mechanism of *limited resource*. The resource limitation was originally conceptualized by Broadbent (Broadbent 1958). In his theory, known as filter theory, only a small portion of the incoming information is passed through the selective filter and identified, but other information is shut out from further analysis. The necessity of attention to overcome resource limitations becomes clear if we consider the analogy with a computer that has limited processing speed and limited memory, but a huge amount of input data, not all of which is relevant for a particular task. If the system can selectively process the small portion of information that is relevant to a current task, it can increase the efficiency of processing and prevent a breakdown caused by an overload.

From a different perspective, Treisman and her colleagues (Treisman and Gelade 1980; Treisman and Schmidt 1982) propose feature integration theory, in which attention is needed to solve the 'binding problem'. The binding problem is generally defined as the problem of how the visual system correctly links all the different features of complex objects. Correct combination of features belonging to different objects is essential for visual perception. Binding is also critically important for machine vision since objects in a scene are typically cluttered and occluded by others. Correct binding of parts or features of an object directly affects system performance. Moreover, selective binding from many possible feature combinations may avoid a combinational explosion.

Concerning invariant representation in terms of size, translation and rotation, Palmer (1998) argue that attention plays a role in the mapping of an object in a retinal-based spatial representation into a representation with a canonical reference frame. For example, size constancy could be achieved as an internal reference frame moves from a smaller to a larger size relative to the size of a stimulus, so that the stimulus is represented in object-centered coordinates not viewer-centered coordinates. The object-centered coordinate reference frame is seen as central to understanding spatial relationships. Neuropsychological evidence may also support this idea. For example, a patient, *JR*, who has a damaged bi-lateral parietal region (thought to have a role in attention), showed impaired responses to both within-object representation (to encode the spatial relations between parts of single objects) and between-object representation (to encode the spatial relations between separate perceptual objects) (Humphreys and Riddoch 1995).

Besides these purposes of visual attention, there are many other possible roles of attention in learning (Grossberg 2005), resolving ambiguity (Luck and Ford 1998), figure-ground separation (Qiu and von der Heydt 2007), multi-stable perception (Sterzer et al. 2009), consciousness (Koch and Tsuchiya 2007), and so on.

2.2 Selection process

To answer the question of how the goals described above can be achieved, it is essential to develop a system at the algorithmic level. The process of how a small portion of information is selected lies at the center of every theoretical explanation of visual attention. Three major issues in this selection process that concern the algorithmic level are reviewed here.

2.2.1 Early versus later selection

This is one of the controversial issues in the study of visual attention and concerns the stage at which the supposed bottleneck of limited capacity is located, i.e. where the selection process takes place. In the sequence of information processing stages (not time), we can roughly divide early and late selective attention. From the limited capacity assumption, Broadbent (Broadbent 1958) originally suggest that selection takes place at a relatively early stage of information processing, that is, before information reaches the semantic or identification stage. Various stimuli with different physical properties, such as intensity, color, orientation, location and so forth, are stored in short-term memory. The attention filter works on these physical or sensory attributes. Only a small portion of the information is selected, passed through an attentional filter and processed at a further stage. The remainder is blocked. That is, information is processed in parallel before a selection process occurs, but from then on, it is processed serially. This implies that we identify only one object at a time. The selected information is identified and becomes conscious.

In contrast to the early selection account, late selection theory assumes that selection takes place at a relatively late stage. The selection process works on semantic categorization, rather than physical attributes, so the limited capacity channel is located after the stage of semantic analysis (Norman 1968). That is, visual objects are processed massively and unselectively, and are passed without any capacity limitation up to the semantic filter. Even though both theories seem mutually exclusive and difficult to reconcile, a compromise may be reached between them. In attenuation theory, rejected visual information is attenuated rather than completely blocked (Treisman 1960). The two phases of selection work in both the early and late stages. In the first phase, early selection attenuates (or partially blocks) incoming signals. The second phase of attentional selection works on the process of identification. In some senses, the attenuation theory implies that as a human interacts with the surrounding environment, selective attention continuously and variably moves between two poles, a reflexive response to physical features and a goal-directed or knowledge-driven behavior, rather than at one or the other of the two discrete poles.

More recently, Lavie (1995, 2005) has suggested a hybrid model of attention in which the distinction between early and late selection rely on perceptual load. In a series of experiments, he found that distractors can be excluded from perception when the level of perceptual load in processing task-relevant stimuli is sufficiently high to exhaust perceptual capacity. However, in situations of low perceptual load, any spare capacity left over from the less demanding relevant processing may be allowed to deal with irrelevant distractors. That is, early selection is predicted for situations of high perceptual load, whereas late selection is predicted for situations of low perceptual load.

2.2.2 Space versus object based selection

Another issue is whether attention is deployed simply over space, or over an object that occupies a space. However, this does not seem a contentious issue; rather, it can be seen as a different aspect of selective attention since attention can work on both ‘where’ tasks and ‘what’ tasks.

One way to explain how attention works is to use an appropriate metaphor. According to the spotlight metaphor, attention can be characterized as an internal beam that throws light on the location where an object is placed (Posner et al. 1980). Therefore, the object in the light is highlighted and processed more effectively, but other objects out of the light are processed less effectively. The spotlight moves from one location to another, once an object has been processed. Posner and his colleagues (Posner et al. 1977; Posner 1980) carried out a series of cueing experiments that showed that the cued place where spotlight attention is allocated is more efficiently processed than an invalid cued place or the non-cued place. Early account of these cue validity effects is based on limited resource assumption. That is, processing resources are allocated to the location where the target is more likely to appear, and this improves the perceptibility, or quality of processing of the target at the cued location, relative to the uncued location (Posner 1980). Alternatively, cue validity effects can be simply explained by the Bayesian statistical model that assumes a weighted combination of noisy responses across the two locations without any resource limitation assumption (Eckstein et al. 2002; Shimozaki et al. 2003).

Another metaphor to explain the attention mechanism is the zoom lens metaphor (Eriksen and James 1986; Muller et al. 2003). In this account, attention initially covers a large area with low spatial resolution, and then zooms in on details with high spatial resolution. That is, attention moves from forest to trees. The idea of the zoom lens metaphor is related to 'global to local' or 'object and its parts' interaction.

In both metaphoric explanations, attention is directed to a spatial location, rather than to an object in a space. In contrast, an alternative approach to visual attention is based on objects. The main difference between these two approaches is that space-based attention enhances everything within the area that spatial attention illuminates regardless of objects or parts of objects, and consequently all objects in the area are equally processed. In contrast, in the object-based approach, since attention is allocated to an object in space, other objects that occupy the same space are not equally processed.

Experimental studies for object-based selection have focused on how attention is directed to process features within a corresponding object boundary and to organize them into a coherent object. For instance, Duncan (1984) carry out an experiment that supports the object-based approach. In his experiment, subjects were asked to report two features that belonged either to the same object or to different objects located in the same area. The stimulus consisted of two objects: a box with a gap and a line drawn over the box. The dimensions of the stimulus varied according to box length (short or long), gap location (right or left), line type (dashed or dotted) and line slant (tilted clockwise or tilted anticlockwise). The results of his experiments showed that subjects detected features that belonged to the same object more easily than features that belonged to different objects. This is known as the "same object advantage"

To demonstrate attentional benefits for processing visual objects, grouping principles such as continuation (Moore et al. 1988), collinearity (Lavie and Driver 1996), and similarity (Baylis and Driver 1992) were often adopted even though perceptual grouping had been considered as an early, image based and preattentive process (Julesz 1991). The studies reveal the close relation between attentional and perceptual organization processes, and show that various organizational processes constrain attentional selectivity. For example, some forms of grouping take place early, rapidly, and effortlessly depending on shape 'goodness' or 'simplicity' whereas other forms like figure-ground problems require controlled attentional processing (Kimchi and Razpurker-Apfeld 2004).

Different evidence for both space-based and object-based selection have led to a controversial debate about whether selection is object-based or space-based. However, this distinction could be elusive. The space-based and object-based strategies could be different aspects of visual attention resulting from tasks that require different visual pathways—the 'where' pathway and the 'what' pathway (Mishkin et al. 1983). For identification or recognition tasks, it is critical to understand detailed parts of an object to recognize it, regardless of where it is. However, to avoid bumping into others whilst walking through a large crowd, precise spatial representation is critical regardless of individual identities. Furthermore, recent studies showed that both spatial and object factors can simultaneously influence the allocation of attention (Egly et al. 1994), and can work in a mutually compensatory manner (Muller and Kleinschmidt 2003).

2.2.3 Object versus feature-based attention

We can voluntarily focus our attention on particular visual features of an object, for example on a particular color or shape. This ability is called 'feature-based attention'. Unlike bottom-up based attention, in which distinctive visual features automatically draw our attention, feature-based attention requires top-down feedback to guide a visual search. Even though both feature- and object-based attention is directed to features associated with a target, only object-based attention is confined to the features of a single object. Both attentional mechanisms support selection of the attended feature, but they differ strongly regarding the processing of non-attended features of the target object (Wegener et al. 2008). That is, irrelevant features of a target object are assumed to be activated in object-based attention, whereas the irrelevant features are assumed to be suppressed in feature-based attention.

There is experimental evidence that demonstrates how the fates of non-attended features are determined in both explanations. In an object-based experimental paradigm, conjunctive stimulus surfaces composed of more than two features, like motion and form, are often used to explore the fate of non-attended features (Schoenfeld et al. 2003). In such an experiment, directing attention to a particular feature belonging to one stimulus surface facilitates processing of other constituent features of that stimulus surface, even though these constituent features are task-irrelevant. A processing benefit for irrelevant features of an attended object has also been shown in more recent studies of binocular rivalry (Mitchell et al. 2004) and cross-modal attention (Turatto et al. 2005).

On the other hand, in a feature-based experimental paradigm, spatial cues were used to direct the subjects attention to individual features or an entire object between two gratings (Wegener et al. 2008). The cue in the experiment was associated with the indication of which feature (speed or color change) should be attended and which object (right or left object) should be attended. The results obtained from measured reaction times showed that incorrect cueing of the changing feature slowed down reaction times significantly, compared to object cueing. This indicates that selection of a specific feature may go along with suppressive mechanisms for unattended features. Moreover, recent studies also have challenged the same object advantage. The advantage can be considered in terms of a product of probabilistic and configural strategic prioritizations rather than of perceptual enhancement (Shomstein and Yantis 2004).

2.2.4 Bottom-up versus top-down based attention

Another distinction in visual attention is made by the direction of information processing. For bottom-up based attention, two major categories of stimulus properties that could, in principle, capture attention can be distinguished: ‘singletons’ refer to feature attributes such as color, orientation or motion that substantially differ from their backgrounds, and ‘abrupt-onset’ stimuli refer to visual targets or distractors that suddenly appear in a visual scene (Egeth and Yantis 1997). ‘Singleton’ attributes of a stimulus can easily attract our attention. For example, a red colored circle (a target) surrounded by blue colored triangles (distractors) is easily detected. Adding more distractors to a stimulus scene does not affect the time it takes to find a target.

The basic assumption for bottom-up based attention is that preattentive processing is driven by the bottom-up properties of the stimulus, prior to attentional allocation. This process occurs in multiple feature dimensions and a parallel manner. After the initial preattentive analysis of a scene, one object is selected on the basis of local feature contrast obtained from its relationship with respect to surroundings. Salience refers to the physical, bottom-up distinctiveness of an object. Attentional allocation is accompanied with ordering the most salient object to the least salient object.

In contrast, when properties of a target are not salient against its background, such as when we want to find a key on a messy desk, the task of finding the target requires the use of our knowledge of ‘what the key looks like’. Characteristically, in this case, searching is less efficient than in the case of finding a salient item. The response time to find a target increases when more distractors are added. This leads us to believe that top-down based attention uses serial processing .

Interestingly, the saliency that automatically attracts our attention may not necessarily be based on low-level features. For example, a toy car in a refrigerator is an unexpected and salient item among foodstuffs. Vanrullen et al. (2003) argued against the classical view that visual 'bottom-up' saliency automatically recruits the attentional system prior to object recognition, and proposed that saliency can be defined at multiple levels, such as luminance contrast, feature contrast, semantic discrepancy, and behavior discrepancy. Similarly, Fecteau and Munoz (2006) also point out that the terms salience and relevance are often treated as synonyms in neurophysiological literature since bottom-up and top-down sources of input converge to produce an amalgamated representation of priority.

3 Biological foundations for visual attention

So far, we have reviewed theoretical aspects of visual attention at the behavioral level. Now, we turn to biological foundations of visual attention that give some hint of how visual attention is implemented in the brain. We start from the two main visual pathways, briefly referred to above, to approach visual attention. This gives a general anatomical structure of the brain mechanisms involved in visual information processing. However, our review more specifically focuses on the attentional mechanism of the brain along these pathways.

3.1 Visual pathways

Anatomically, vision can be roughly divided into two main streams that correspond to 'what' and 'where' tasks, as shown in Fig. 1. The 'what' (or ventral) pathway runs from the occipital cortex to the temporal cortex, and it is thought to be specialized for object vision, while the 'where' (or dorsal) pathway runs from the occipital cortex to the parietal cortex, and is thought to be specialized for spatial vision. The original work of Mishkin et al. (1983) showed that a specific lesion in the temporal lobe produced severe impairment in an object recognition task, while a specific lesion to the parietal lobe produced severe impairment on a spatial selection task. More elaborate work on visual pathways has been done more recently by other researchers (DeYoe and Van Essen 1988; DeYoe et al. 1994; Kanwisher 2003). This shows that visual information is decomposed into different visual pathways that are specialized for perception of color, form, depth, motion, etc. at lower brain areas and fed into higher areas. These areas are connected in feed-forward and feed-backward manners, and also laterally interconnected. Attention studies on the brain focus to identify the loci of the brain and the underlying physiological mechanisms related to attentional tasks with various methods such as a single/multi cell recoding, brain imaging, etc (Fig. 2).

3.2 Mechanisms of visual attention

The diversity of visual attention suggests that attention is accomplished by many brain areas, rather than a single brain area. There are many different roles of brain areas in visual attention; we cannot describe all these different brain mechanisms here. Rather, we focus on the mechanisms in relation to the psychological issues introduced in the previous section.

3.2.1 Properties of receptive fields

Here, we introduce the concept of receptive fields (RFs) because many physiological studies of visual attention utilize them by estimating cells' responses to corresponding RFs in a given

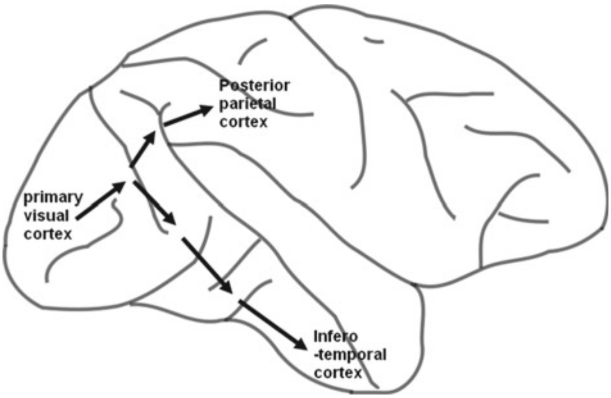

Fig. 1 The visual pathways in the macaque monkey. A schematic illustration of visual pathways for object recognition and spatial representation is shown

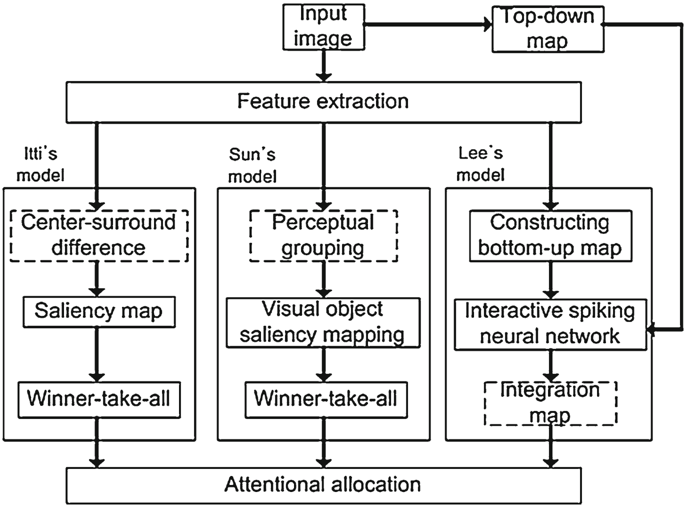

Fig. 2 Illustration of three computational models. From *left*, Itti et al. (1998), Sun and Fisher (2003), and Lee et al. (2005)'s model are presented with simplified characteristics. The major characteristics of the models are *boxed dashed lines*

attentional task. Classically, an RF is defined as the 'receptor area in which stimulation leads to response of a particular sensory neuron' (Levine and Shefner 1991). The RF has interesting properties in the visual hierarchy and is dynamically modulated by various factors, such as lateral interaction, as well as by attention. First, along a visual pathway from V1 to IT, the selectivity of neurons changes. Cells in early stages are sensitive to very low-level features, such as local contrast or edges. As one moves to a later stage, cells are sensitive to more complex patterns, such as shape. Second, the RF size of cells increases at later stages. This means that the response of a cell in a later stage becomes more independent of the location of one object; Thus, this can contribute to achieving translation invariant recognition. Third, the cell's activity is degraded if more than one stimulus falls into an RF, in comparison to the activity evoked by a single stimulus presented within the RF.

3.2.2 Biased competition

Desimone and Duncan (1995) proposed a biased competition hypothesis that is logically derived from the limited resource assumption. According to them, the RF can be viewed as a critical visual processing resource, for which objects in the visual field must compete because the resource is limited. The experimental paradigm so-called 'sensory suppression' is often adopted to show supportive evidences of the hypothesis (Kastner and Ungerleider 2000; Beck and Kastner 2008). In this paradigm, the responses of a cell to a single effective stimulus presented within its RF are compared to the responses to the same stimulus paired with an ineffective stimulus within the RF. With only the effective stimulus, cells in V4, MT and IT of monkeys produced a high firing rate, whereas the paired stimulus elicited relatively lower firing rates of the cells. This result implies that multiple objects in a restricted RF compete for limited processing resources in a mutually suppressive way. The degraded response of the cell can be enhanced by selectively attending an object and ignoring the others.

Desimone and his colleagues investigated attentional effects at a neuronal level along visual pathways (Moran and Desimone 1985; Luck et al. 1997; Reynolds and Desimone 2003) . In a typical neurophysiological experiment (Moran and Desimone 1985), neural activities from a monkey were recorded during an attention task that required it to attend to one object and to ignore the other. The objects were presented inside or outside the RF of the cell. When two stimuli were presented inside the RF, the neuron's response to the attended stimulus appeared to be of normal magnitude, while its response to the ignored stimulus was suppressed. However, they reported that no attentional modulation effect was obtained when only one stimulus was presented inside the RF.

Similar results were also obtained using a functional magnetic resonance imaging (fMRI) technique (Kastner et al. 1998). Comparing two conditions in which stimuli were presented simultaneously or sequentially, brain activation in the sequential condition is higher than that in the simultaneous condition. The activation in V4 and TEO in the simultaneous condition is profoundly reduced, but not in V1. In addition, the attentional effects that increase brain activation in the simultaneous condition are larger than those in the sequential condition. It is worth noting that attentional modulation effects on V1 are not consistently identified. There are conflicting reports as to whether attention can affect processing here (Luck et al. 1997).

Even though competition among neurons is an important brain mechanism that helps achieve stability of the brain, with converging neural responses activated by multiple stimuli producing a dominant and coherent perceptual experience, and with attentional selectivity enhancing relevant information and suppressing irrelevant information, the emphasis on competition may neglect another important aspect of selective visual attention, cooperation. As noted earlier, attention involves many brain areas. The dynamic and cooperative interaction between those areas may help to reduce the processing burden caused by limited resources (Deco and Rolls 2005).

3.2.3 Early versus late selection in neural mechanism

The attentional effect on an early stage of visual processing seems controversial, as noted above. Some experimental results from Event-Related Potentials (ERPs) did not reveal any difference in the responses of the primary cortex to attended and unattended items, whereas other brain imaging studies did find attentional modulation effects in that brain area. Kanwisher and Wojciulik (2000) interpreted these conflicting results based on perceptual load hypothesis (Lavie 1995). That is, the stage of selection depends on the processing load of a task. Early selection may occur when the processing load is high, whereas late selection may occur when the processing load is low. Even though both behavioral and neuroimaging studies support the hypothesis (Beck and Lavie 2005), it seems not to be compatible with the arguments derived from the biased competition account. According to the biased competition account, cells in higher processing stages have larger RFs and have a greater chance that more objects will fall inside their RFs, so that higher processing loads would be required for the cells in higher stages. Recently, an attempt was made to link biased competition to the perceptual load hypothesis (Torralbo and Beck 2008). That is, suppressive interactions among stimuli in the visual cortex are reduced when the target and non-targets are easily distinguishable. In contrast, a strong top-down bias from higher brain areas will be needed to bias the competition in favor of the target when those are difficult to be distinguished.

Another aspect of the issue 'early versus late' selection is concerned with that level of information—physical features or more meaningful objects. Evidence for feature-based selection was provided by a brain image study in which subjects were asked to attend to different features of the same visual array—color, shape or speed of motion of the elements in the array (Corbetta et al. 1990). Different regions in the extrastriate cortex were activated during the tasks according to which features were being attended. Conversely, neurological evidence also supports object-based selection. For example, in a brain imaging study, subjects were asked to attend to one of two stimuli (face or house) that were transparently superimposed (O'Craven et al. 1999). Attending to the face highly activated the fusiform face area that is believed to be specialized for face representation.

To summarize these findings, attention seems to be involved in multiple stages of visual information processing in a continuous manner rather than in a discrete or unipolar manner.

3.2.4 Space versus object based attention in neural mechanism

The brain areas linked to both space and object-based attention are identified along the two major visual pathways. Some areas, including parts of the parietal lobe and prefrontal cortex, are activated by both space and object-based attention (Wojciulik and Kanwisher 1999; Roelfserna et al. 1998). Some areas are differentially activated by either space or object-based attention (Colby and Goldberg 1999). Moreover, a hemispheric difference between these attention mechanisms has been reported (Fink et al. 1997). Concerning space representation, different parts of the parietal lobe are involved in different spatial relations. That is, we can possibly construct multiple spatial maps to establish a relationship between a viewer and objects, among objects, and within an object. For example, in order to grasp an object, the relationship between my hand and the object should be represented and it should send continuous feedback signals from my hand's trajectory. In contrast, for identification of an object, parts of the object may need to be represented in terms of the object itself, not a viewer, since a viewer can be in many different positions, and at many different angles, distances, etc. The reference frames for spatial representation include retinocentric, body-centric, and object-centric.

Colby and Goldberg (1999) summarized specific areas of the parietal lobe and their cognitive functions in space representation. The ventral intraparietal area may contribute to representing space in retinocentric and head-centric terms, and specify goals for movements of the head, lips, and tongue, and facilitate reaching with the mouth. The medial intraparietal area is specialized to respond to stimuli within reaching distance and is thought to guide reaching movement. The anterior intraparietal area is linked to the control of hand shape and grip. Importantly, the lateral intraparietal area responds to sudden abrupt-onset stimulus and maintains its activity before a saccadic eye movement. It is considered that this area is responsible for the representation of attended or salient spatial location. In addition, the area transforms retinocentric coordinates into motor coordinates for saccadic eye movement by tracing memory of stimuli.

Neuropsychological studies on the parietal lobe reveal its link with attentive functions. Patients with damage to the parietal lobe showed a variety of symptoms, such as somatosensory agnosia (a disorder of tactile appreciation of one's body), Balint syndrome (a disorder of recognition for more than one object at the same time), neglect (a disorder of perception of, or action on, the side of a space or object opposite that of a lesioned parietal lobe), and so on (Ellis and Young 1996). These studies suggest that the parietal lobe is involved in both spatial attention and non-spatial attention for object recognition.

In general, object-based and space-based attention share common neural mechanisms in the parietal lobes, but differentially activate brain areas depending on the related tasks. Early segmentation or grouping tasks may take place in the primary visual cortex (Lamme 1995), whereas more complex object representation tasks may take place in higher cortical areas such as lateral occipital (LO) cortex (de-Wit et al. 2008) or temporal cortex (Serences et al. 2004).

Research efforts to extend the behavioral study of the 'same object advantage' to brain study were also made. However, those efforts do not provide conclusive evidence that supports the psychological hypothesis. For instance, an fMRI/ERP study investigates the dynamics of object-based attentional selection using objects defined by color and motion features (Schoenfeld et al. 2003). The study showed that processing of an irrelevant color feature in the ventral occipital cortex was enhanced when it belonged to the attended surface of a moving dot array. In contrast, another study showed that ERPs were elicited by a distractor that shares the same color feature of a target when the task-irrelevant colored distractor was presented with a target object in a search task (Boehler et al. 2010). That is, object-based selection of task-irrelevant features may spread globally in space, not limited to within the target object.

3.2.5 Bottom-up and top-down based attention in neural mechanisms

Unlike the segregated visual pathways for object and spatial vision, less clearly distinguished systems have been identified for bottom-up and top-down based attention. Many areas in the brain are commonly involved in both stimulus driven and goal driven attention. Nevertheless, some evidence suggests that partially segregated networks of the brain carry out these two different attentional functions (Hopfinger et al. 2000; Buschman and Miller 2007). Corbetta and Shulman (2002) argue that a network, including parts of the intraparietal cortex and superior frontal cortex, is involved in preparing and applying goal-directed selection for stimuli and response, whereas a network including the temporoparietal cortex and inferior frontal cortex, which is lateralized to the right hemisphere, is involved in stimulus driven attentional tasks. The frontoparietal areas, such as the lateral intraparietal and frontal eye field, not only respond to a distinctive stimulus, but also show task-driven responses. Moreover, these areas are activated during search and detection tasks. Corbetta and Shulman (2002) assert that these areas may be involved in constructing a saliency map, combining both bottom-up and top-down information.

Concerning the saliency map that originally introduced in a computational model of selective attention, an important question raised in biological studies is about where the saliency map is located in the brain. There seems to be no particular brain area that may be considered as a saliency map. V1, for instance, has been often considered as a major candidate for salience map location because of the topological arrangement and tuning patterns of V1 cells, but not necessarily in a separate and explicit form (Nothdurft 2006; Zhaoping 2008).

More higher areas along the two visual pathways—parietal and V4 areas—are also considered as salience maps. The posterior superior parietal cortex in the dorsal pathway is thought to be a space-based saliency map because of its winner-take-all like activation patterns toward stimulus set size (e.g. higher lateral inhibition if set sizes increase) (Roggeman et al. 2010; Gottlieb et al. 1998). V4 in the ventral pathway is also thought of as a saliency map because of its retinotopic representation that guides exploratory eye movements. There are many possible candidates for saliency maps that emphasize different aspects of selective attention. This may imply that the saliency map is not necessarily a neural substance, rather it could be emergent properties of neural systems from various processing stages.

One way to investigate bottom-up versus top-down based attention is to measure neural activities over time in response to a cue, during an attention task. That is, in order to provide precued information to a current attentional task, neural activities in the brain area in response to the cue signal need to be sustained during the task. These sustained responses to a cue can be considered as a top-down control, and can be distinguished from transient responses that may be thought of as purely sensory driven responses. Some brain imaging studies show these distinctive responses to a given cue. For example, parts of the parietal lobe (intraparietal cortex) and frontal cortex are activated during both spatial and object based attentional tasks. Unlike the areas in the occipital lobe that transiently respond to a cue, these areas show a sustained response during attention (Bar 2003). Possibly, there are several different ways to provide top-down influence to an attended stimulus: enhancement of a relevant stimulus, suppression of an irrelevant stimulus, priming an expected location or object, etc.

3.2.6 Neuronal activities in attentional selection

The debate between object and feature-based attention has focused on the neuronal processing for task-irrelevant features underlying the selection mechanisms as predicted by ‘integrated competition theory’ (Duncan et al. 1997). According to the theory, directing attention to an object feature allows non-relevant features of the object to be coselected and results in binding them to the object. In contrast, explanations of feature-based attention assert that processing of irrelevant features are suppressed.

At the neuronal level, this issue concerns whether the neural responses of the non-relevant object feature are activated or suppressed. Some empirical evidence supports object-based selection, but other evidence does not. For instance, a fMRI experiment measured neural activities of cortical areas involved in both color processing and word reading during Stroop tasks (Wuhr and Frings 2008; Polk et al. 2008). The results of this experiment show that activation in functionally defined color areas increases while activation in functionally defined word areas decreases. That is, the processing of ignored object features is suppressed. Furthermore, an ERP study showed that object-based selection of irrelevant features is not confined to the attended object, rather it is spatially global over the visual field (Boehler et al. 2010).

Also, the neuronal responses of feature-based attention are often compared to those of space-based attention. Two aspects, gain and tuning, of neuronal responses reveal modulation effect led by different attentional mechanisms in visual pathways. The gain effect means that overall population responses to a stimulus are increased or decreased by a multiplicative factor across all feature detectors, and thus the amplified response is effective when external noise is low. On the other hand, the tuning effect means that the population responses profile to a stimulus are sharpened or broadened, and thus the narrowed responses would benefit when the external noise is high and should be suppressed (Maunsell and Treue 2006; Ling et al. 2009). The results from the modulation effects led by space-based attention indicate that spatial attention boosts the gain of population response and strengthens the representation of attended locations without changing neuronal tuning curves. In contrast, feature-based attention exerts a multiplicative gain upon neuronal response and sharpens the tuning curve of population response. This implies that attentional benefits from both mechanisms can be differentiated when external noise is high.

4 Computational approaches to selective attention

In this section, computational models of selective attention are introduced in terms of information processing and computational purpose. These two aspects give insight as to how an algorithm is chosen to transform inputs to outputs, and what computational goals of selective attention are achieved with the algorithm. Even though computational modeling is closely related to the psychological and neuroscientific studies discussed above, we will not refer to them in detail to avoid repetition, but take up general lessons in the summary.

Roughly speaking, any computational model for visual attention has two distinctive processing stages. This distinction is due to the assumption of capacity limitation that divides information processing stages into the preattentive and attentive stages. At the preattentive stage, the visual system works in a parallel manner without capacity limitation, whereas at the attentive stage, the visual system deals with only one item at a time. Except for a few computational models explaining the attentive process in terms of psychological phenomenon, after selection occurs, most current models do not refer to it. Therefore, our review focuses on rather specific algorithmic stages of computational models, such as feature extraction, selection process, and their applications.

4.1 The feature extraction process

Prior to a selection process, there is a processing stage in which visual stimuli are transformed into a preattentive representation. The goal of preattentive representation is to mark conspicuous image locations and make them more salient for perceptual pop-out. At this stage, various physical features, such as orientation, color, intensity, and size, are extracted in parallel and are composed of feature maps as in Itti's saliency based model (Itti et al. 1998; Itti and Koch 2000, 2001). Different processing schemes are taken up for their own computational application to utilize those features to achieve computational goals.

Basically, a center-surround scheme is commonly used to obtain salient features in saliency based models. The difference in comparison to the surrounding visual input at a location is calculated by this scheme (Itti et al. 1998). In fact, a region of homogenous features nullifies the response of a center-surround filter, whereas any discontinuous region of features can be localized by the response of the filter. However, the linear property of this center-surround scheme is criticized, since psychophysical evidence shows nonlinear and asymmetric response (e.g. a Q surrounded by Os versus an O surrounded by Qs).

Alternatively, the center-surround saliency is considered as a classification problem in the sense of how distinct a stimulus at a location is from the stimuli in its surroundings (Gao et al. 2008). Thus, the saliency can be formulated by the mutual information between features and its two classes (center and surround) that provides an optimal solution in a decision theoretic sense. Another criticism is based on the semantic analysis of the saliency. That is, a salient location obtained from the center-surround scheme does not correspond to an object or a part of an object. Rather, it simply corresponds to a pixel of an image scene that has higher contrast (Vanrullen et al. 2003).

In response to the criticism above, the salient points of images derived from the center-surround scheme were recently investigated using the '*LabelMe*' database in which objects of an image were manually annotated (Elazary and Itti 2008). Assuming that the labeled regions are interesting, a few of the most salient locations of the images are compared to the marked regions of the images. It was found that one or more of the top three salient locations, in 76% of all images, fell on a labeled region. On this basis, it is argued that selecting interesting objects in a scene is largely constrained by low level visual properties.

Second, a multi-scale mechanism is used to obtain an image representation from a coarse spatial scale to a finer spatial scale, with the zoom lens metaphor accomplished with the mechanism (Olshausen et al. 1993; Deco and Schurmann 2000; Sun and Fisher 2003; Sun et al. 2008). The information carried by multiple spatial scales is different. The general structure of the visual object is conveyed with a large-scale spatial resolution, while details of the object can be conveyed with a small-scale spatial resolution. In Deco's model, for instance, the coarsest level of spatial resolution is utilized to find the location of an interesting object in a priority map (Deco and Schurmann 2000). Once an object is located, the object is identified by increasing spatial resolution to a finer level until it is confirmed to be a target. For object recognition, detailed spatial representation is not necessary to locate an object, but it is necessary to identify what the object is. Similarly, global scene information can be used to guide attentional allocation at the local salient parts of the scene obtained from bottom-up saliency (Torralba et al. 2006; Oliva and Torralba 2007).

Interestingly, Sun and Fisher (2003); Sun et al. (2008) attempt to combine both the center-surround and multiscale decomposition schemes in a similar way to that of the zoom lens models. In the model, the attentional scan passes from a global saliency map to its local saliency maps based on global-to-local interaction. That is, if we admit the homogeneity of 'within-object' and discontinuity of 'between-object' at a feature level including intensity, color, texture, etc., the center-surround contrast with multiresolution provides not only a salient location in a visual scene, but also the boundary between visual objects or their parts.

4.2 Selection process

This part reviews the selection process implemented in various computational models. Some selection algorithms are more explicitly implemented in a psychological and biological context, while others are not. We first raise an issue of competitive and cooperative aspects of selection, then some issues of selection process introduced in previous sections.

4.2.1 Selection—cooperative or competitive?

The biased competition hypothesis, referred to above, implies that neurons at a given processing stage take part in an inevitable war for limited resources. In computational models, the concept of competition is embedded in the WTA network in which units are mutually interconnected and are inhibited by each other (Itti and Koch 2000; Indiveri 2008). Only the one unit surviving in the competition is selected, and consequently, the limited resource problem can be solved as the winner takes all the resource. Ironically, Lee (2008) argues that the logic of inevitable competition can be applied to the necessity of cooperation. That is, the limited resource assumption may also require the cooperation of independent brain areas or neural channels that help reduce the burden of processing in various ways. In his model, selection is accomplished by the integration of bottom-up information with cooperative cues. For instance, the task of 'finding a man in red t-shirt' deals with two types of information driven from bottom-up features, such as skin color and facial shape, and a cooperative cue feature, such as red color. Face candidates near the cue feature are more likely to be selected, and thus the whole image is not exhaustively searched in a face-by-face manner.

4.2.2 Which stage of selection—early or late?

Models of selective attention concerned with the locus of selective attention, can be classified into early or late selection models. In particular, the issue debated in psychology is related to whether the selection process occurs before object recognition or after. Even though some computational models insist that a late selection mechanism is utilized in the models, 'late' does not mean 'after object recognition' and the selection process does not work on semantically meaningful information. In this sense, most current computational models are based on early selection.

First, in the bottom-up saliency based model the selection is accomplished by saliency calculated from the center-surround feature contrasts. The selected location does not meaningfully correspond to the location of an object as noted above. Rather, it simply corresponds to the location where it gives the strongest contrast, e.g., possible edges. It seems that not many things can be done with the selected point for further processing even though Itti (Itti and Koch 2001) argues that it can be the front end to object recognition. Some variations adopting Itti's model use additional region segmentation methods for further processing after a saliency map was constructed (Mendi and Milanova 2010). Even though the selection algorithm of the model is intuitive and easily applied to the feature level of a visual image, little benefits for later processing stages can be obtained from this selection mechanism. So, one may ask if it is too early to select meaningful features.

Little bit late, but still early, Sun and Fisher's computational model groups the feature elements such as color, intensity and orientation into more meaningful perceptual units (objects) before the selection process operates (Sun and Fisher 2003). The selection process in Lee's model (Lee et al. 2005) occurs at a more abstract representational level that combines facial features, such as facial color, shape, and symmetry, using a dynamic neural network. Possible benefits from the computational model based on this 'bit late' selection are: (1) selection can be accomplished in a meaningful way by allocating an attentional window at the object location, not at a salient point; (2) bottom-up information can be easily coordinated with top-down information, since a smaller number of objects (not points) are activated at a time compared to those models using early selection.

Recently, more meaningful features such as similarity, familiarity, and symmetry have been introduced. For instance, Lee et al. (2010) use the feature 'familiarity' that is a measure of the resemblance of local features extracted from the input image to features of trained object models stored in a database. This measure provides the degree of evidence whether a task-relevant target object exists or not. They argue that an advantage of using familiarity is that it does not require additional information other than the object database available in an object recognition system.

In general, possible benefits of early and late selection can be considered in terms of computational efficiency and robustness. If selective attention takes place at an early stage of information processing, the computational load for later stages will become lighter. However, it is difficult to decide which features are important since objects are likely to share the same physical features. Conversely, selection would be relatively easy if attention takes place at a late stage, because selection works on a small number of units that correspond to more abstract factors such as concept or category. However, it also implies that more computational resource is required to process all objects in a visual scene up to the late stage. The tradeoff between ease of selection and the amount of required computational resource is clearly problematic if one wants to develop a computational model that has efficiency and robustness.

4.2.3 Employing object knowledge or not

The distinction between bottom-up and top-down is based on whether or not knowledge of a target object is employed for a selection process. For a bottom-up based model, selection is determined by physical properties of image features including intensity, orientation, color, and motion. A key issue for bottom-up based selection concerns how regions of interests are selected on the basis of the physical properties.

The center-surround difference scheme has been dominantly employed for the selection criteria of the early bottom-up based models. Recently, various methods are being developed to obtain the saliency. For instance, entropy based saliency estimates the local unpredictability of an image region that may correspond to an informative (or distinctive) part of the image (Itti and Baldi 2006). The spatio-temporal saliency can be calculated from density estimation of pixels both in spatial and temporal domain (Mahadevan and Vasconcelos 2010). This spatio-temporal saliency is closely related to Shannon's information which the quantity of an event is inversely proportional to the probability of the observation of the event. The spectral residual, which can be obtained from the difference between log-spectrum and averaged spectrum of an image, is used to detect saliency (Hou and Zhang 2007) Since all these methods rely on the intrinsic nature of the stimulus itself, no top-down knowledge about a target is necessary and thus they are independent from the tasks requiring knowledge.

In contrast, top-down based selection is based on the knowledge that involves a task. Therefore, a key issue for top-down based selection concerns how selection criteria are set from knowledge related to a task. It should be noted that top-down knowledge implies both knowledge as directly related to a target object (such as color of the target object), and the contextual knowledge explicitly or implicitly given.

The 'Guided Search' model developed by Wolfe (1994) is a milestone that has influenced many computational models of visual attention. In the model, different primitive features such as color and orientation are extracted in parallel and construct a set of feature maps. Next, the feature maps are passed through a differencing mechanism that yields the bottom-up activation. This activation can be used to guide attention toward distinctive items in the visual field. However, with only the bottom-up activation, a desired item can not be located if the item is not salient. Top-down information is need to enhance relevant feature properties and to suppress irrelevant feature properties. This is accomplished by top-down feedback directed to each feature map.

More recently, Hamker (2006) proposed a computational model that shares many aspects with the 'Guide Search' model, but it is different in details. The model, that is biologically inspired, also considers the influence of top-down information about a target template in the prefrontal cortex (PFC). The target template encodes features of the target object by a population of sustained activated neuronal units. Feedback from higher levels such as PFC and TE in the ventral pathway transfers the target template to lower levels such as V4. As a result, units corresponding to the features of interest different levels of the hierarchy are enhanced by the feedback.

Recent extensions of Itti's basic model uses task knowledge or feedback for object selection (Navalpakkam and Itti 2005, 2006). In these models, top-down knowledge about a target object is directed to an early stage of processing. The top-down knowledge is utilized by imposing weighting parameters on the stage before or after constructing feature maps. That is, the top-down influence works on the early stage by enhancing wanted features or suppressing unwanted features, so that certain features relevant to a target object can easily pop-out.

Torralba et al. utilize contextual knowledge to select target location (Torralba et al. 2006; Oliva and Torralba 2007). In their approach, the probability of the presence of a target at a location is described in terms of the conditional probability of the target object at the location over jointly given target features and contextual features. For simplicity, the conditional probability can be decomposed into three terms—object likelihood, local saliency and contextual priors. Contextual priors provide a strong bias about the location of a target.

4.2.4 Selection in space

Even though the issue of whether attention is directed to a location, feature, or object has been extensively studied in both psychology and neuroscience, there are only few computational models that implement these selective properties.

In behavioral studies, space-based attention has been studied with Posner's cueing paradigm (Posner 1980). Basic findings of the paradigm are straight forward. Attention benefits performance in detecting a target at a cued location, whereas attention costs performance in detecting a target at an uncued location. This validity effect is the core of computational modeling of spatial attention. In Mozer and Sitton's model, a winner-take-all network was used since only one attentional unit should be active at a time (Mozer and Sitton 1998). The attentional units received both exogenous inputs coming from sensory data and endogenous inputs coming from previous learning or cueing. Thus, exogenous and endogenous inputs take a role to activate an attentional unit, but inputs from the other competitive units take a role to suppress it. In this scheme, the validity effect results from competition among units, rather than interaction between the exogenous and endogenous inputs.

Another explanation of the validity effect is based on Bayesian statistic (Eckstein et al. 2002; Chikkerur et al. 2010). In this explanation, the likelihood of the responses given a signal at the cued location, the likelihood of the responses given a signal at the uncued location, and the likelihood of the responses given a noise at the both cued and uncued locations are separately calculated, weighted by prior probabilities, and then the likelihood ratio between signal and noise can be obtained. A decision can be made by comparing the likelihood ratio with the decision criteria. Basically, the validity effect is caused by both the weights obtained from prior probabilities and the signal-to-noise ratio. Therefore, the validity effect can be explained in terms of an optimal decision boundary that maximizes benefit and minimizes cost.

4.2.5 Selecting proto-object

The concept 'object based attention' in some literatures of computer vision is confusingly different from that in behavioral and biological studies. For instance, Borji et al. (2010) use the phrase to select a node of a U-TREE algorithm in which objects are represented in those nodes. Thus, object-based selection simply means to select a node corresponding an object, rather than to select parts or features belonging to the same object.

More carefully, 'region based selection' in computer vision seems closely linked to object-based attention (Hu et al. 2008; Avraham and Lindenbaum 2010). In this approach, segmentation process, that partitions a digital image into multiple segments that may be more meaningful or corresponding to an object, is accomplished to find regions of interest. These segmented regions are also known as 'proto-objects or 'pre-attentive objects (Orabona et al. 2007).

Interestingly, a polar transformation of features is often adopted in this approach. The advantages of polar transformation are that feature homogeneity and spatial proximity are taken into account when analyzing regions.

In Sun's model, the hierarchical relationship between segmented parts is established to guide the attentional movements shifting from one locus of attention to another that may belong to the same object. In the model, the Gestalt cues such as spatial proximity and similarity are considered to bias the allocation of attention. Furthermore, Sun argued that object-based attention has interesting properties that space-based models do not have: (1) more efficient visual search; (2) less chance to select a nonsensical or empty location; and (3) naturally hierarchical selectivity.

However, although their model showed more interesting and dynamic properties not yet exhibited by current computational models, the segmented regions still remains loosely separated, rather than concretely combined. Moreover, if we apply more strict criteria shown in psychological studies, in which objects overlap at the same location and attention is switched from one object to another, it is still questionable if the model can carry out the task.

Interestingly, Kim and Lee (2004) attempted to recognize superimposed patterns with top-down selective attention. In the model, an attentional layer (filter) is located between the input and the multi-layer classifier (MLP), and its unit has one-to-one connection with input units of MLP. Attentional gain is adjusted by the gradient descent rule that suppresses the irrelevant and noisy features in the input or enhances the relevant features in the input using the knowledge of trained patterns. With the attentional switch algorithm that changes attentional gains to 0 if they are greater or equal to 1, and to 1 otherwise, the recognition of the superimposed pattern is changed from one class to another by attentional switching.

4.3 What and where tasks

Researchers want to tackle two main types of tasks—search or detection ('where') tasks and object recognition ('what') tasks. Even though the labels of 'spatial-based' and 'object-based' models are closely associated with what they can perform, it does not necessarily mean that a spatial-based model can only work for a 'where' task and an object-based model can only work for a 'what' task. Rather, as pointed out by Miau et al. (2001), spatial-based attention may serve as the front-end to object recognition by eliminating many irrelevant locations. Conversely, object-based attention may serve as the front-end to a search by eliminating many irrelevant objects.

4.3.1 Object recognition

An important aspect of the saliency based approaches that is also shared by many other models is that the saliency map is retino-topographically organized. This is a useful coordinate system as long as the task for a model is limited to a searching or detection task. However, as pointed out, when attention engages a specific object, retinocentric spatial representation needs to be transformed to an object-centered spatial representation for invariant object recognition. Sun and Fisher (2003) also criticize the saliency-based attention models since they fail to perform in cluttered scenes or where objects overlap or share some common properties. For object recognition, selecting a location does not guarantee that the location is meaningful, since local saliency could be due to noise or unimportant parts of objects. For some structured objects, properties that are more functional are required to select locations, objects, features, and their groupings.

To overcome this problem, various approaches have been proposed. Walther et al. (2005) use region growing and adaptive thresholding methods to work on salient points. The model combines the scale-invariant feature transform (SIFT) algorithm for recognition with the saliency based model and performs better with attention. Won et al. (2006) use a scale selection algorithm based on Kadir's (Kadir and Brady 2001) approach. Therefore, the number of salient points in an image can be dramatically reduced to the number of salient regions. The saliency based model proposed by Itti et al is also used for object recognition by combining an additional recognition system called HMAX (Riesenhuber 2005).

The scanpaths (or sequences of attentional focus) can be used for object recognition (Rybak et al. 2005). Through the scaning of an image, primary features (e.g. edges) are extracted and transformed into invariant second order features using a reference frame and are then stored in 'what' memory. The sequence of attentional focus is stored in 'where' memory. The model learns these 'what' and 'where' pieces of information for a complex scene such as a face image. When a new image is presented, the model operates 'search mode' to scan the image until an input retinal image similar to one of the stored retinal images is found. Once a similar retinal image is found, a recognition mode is executed to compare the stored local contents to the local contents of the new image along the stored scanpath. The model shows the ability to recognize complex images invariantly with respect to shift, rotation and scale.

In a different approach, object recognition is considered as a reconstruction problem in which several attended pieces of an image are required to combine into an object (Fu et al. 2009; Gouet-Brunet and Lameyre 2008). In this approach, a segmentation process is adopted to break an image into perceptually homogeneous regions. The pieces may correspond to a part of an object or background noise. The attentional task of this approaches is to select the pieces and to reconstruct a whole object that maximizes a global attention function.

4.3.2 Search

Saliency in most current models is defined at the feature level, such as feature contrasts. Saliency is calculated by summing up all contrast values obtained from center-surround differences on each feature map. In those models, a location of the saliency map does not correspond to an object or a part of object. This means that the model exhaustively searches a possible target location in a point-by-point (or pixel-by-pixel) way, even though far fewer objects are contained in an image scene.

The task 'search' is intrinsically knowledge-driven, since we have to know what we want to find regardless of whether or not a target is salient. A later version of the saliency-based models proposed by Itti (Navalpakkam and Itti 2005), and the guided search model proposed by Wolfe (1994), used top-down knowledge that provided weighting parameters on preattentive stages. Similarly, in Hamker's model, feature values and saliency values are combined into a population code (Hamker 2006). This coded information is sent further upwards to V4 and IT cells. A target template is stored in prefrontal memory and is compared to the activation pattern of IT. The feedback from higher stages and lower stages guides searching behaviors by enhancing task-relevant information and suppressing task-irrelevant information.

In Lee's model in which attention is also guided by a cue feature, the location of a target candidate is represented with respect to the spatial reference frame of a cue feature (Lee et al. 2005). For instance, the task 'finding a man in red t-shirt' is intrinsically required to represent the location of a target candidate relative to the location of a red colored region. Even though the model does not provide an explicit explanation of how the validity effect is produced, facilitatory and interference effects result from the multiplicative interaction between cued locations and target candidates. That is, a unit corresponding to a target above cued locations (e.g. a face above a colored t-shirt) is positively gained by multiplicative interaction, while a unit corresponding to a target candidate at uncued locations (e.g. hands below a colored t-shirt) can be negatively gained.

Other models also have tackled a search task. One interesting approach is based on a learning mechanism. According to Balkenius (2000), attention can be controlled in the same way as actions using similar learning mechanisms. In particular, two learning mechanisms, habituation and conditioning, may take an important role in attentional control. Habituation, the ability to adapt to an environment, can be considered as a learning process in which an animal learns to ignore a stimulus carrying less informative values. In conditioning, an animal learns a pairing relation between a conditional stimulus and an unconditional stimulus and produces a response based on the relationship, or an animal learns the relationship between an action and its results.

5 Summary

This paper reviewed visual attention in psychological, neuroscientific and computational studies and introduced some of the fundamental issues in these areas. Computational goals were introduced to explain why visual attention is necessary and leads to selectivity of processing. The distinctions between early and late, space-based and object-based, and bottom-up and top-down processing highlight how selection may work. We introduced underlying mechanisms of brain areas that are linked to visual attention and correspond to psychological studies. We described the visual pathways, properties of RF, locus of visual attention, and attentional networks that are being investigated to uncover the mechanisms of visual attention at the biological level. Computational models owe a debt to studies in both areas. We introduced a number of computational models for the processing of visual attention. Even though current models focus on bottom-up spatial attention, there is a significant and experimental attempt to integrate top-down knowledge and combine spatial mechanisms with object recognition, and to reconcile early and late selection.

Computational modelers clearly have a purpose when they construct a model. However, regardless of their purpose, the difficulty they face is how they can achieve their goal. This review aims to provide guidance on how to construct a computational model of selective attention. The lessons from psychological and neuroscientific studies can be helpful to guide which kinds of properties are required in the model and what constraints should be imposed on the performance of a model. These constraints may work on different levels, such as model architecture, feature extraction, selection, learning, and attentional modulation algorithms. Conversely, computational models can contribute to extend our understanding of selective attention by implementing and testing ideas from psychological and biological studies. Interweaving all the efforts together, a piece of the puzzle can be placed in a more appropriate location.

References

Avraham T, Lindenbaum M (2010) Esaliency (extended saliency): meaningful attention using stochastic image modeling. IEEE Trans Pattern Anal Mach Intell 32(4):693–708

Bar MA (2003) Cortical mechanism for triggering top-down facilitation in visual object recognition. J Cogn Neurosci 15:600–609

Balkenius C (2000) Attention, habituation and conditioning: toward a computational model. Cogn Sci Q 1(2):171–214

Baylis GC, Driver J (1992) Visual parsing and response competition: the effect of grouping factors. Percept Psychophys 51:145–162

Beck DM, Kastner S (2008) Top-down and bottom-up mechanisms in biasing competition in the human brain. Vis Res 49(10):1154–1165

Beck DM, Lavie N (2005) Look here but ignore what you see: effects of distractors at fixation. J Exp Psychol Hum Percept Perform 31:592–607

Boehler CN, Schoenfeld MA, Heinze HJ, Hopf JM (2010) Object-based selection of irrelevant features is not confined to the attended object. J Cogn Neurosci. doi:10.1162

Borji A, Ahmadabadi MN, Araabi BN, Hamidi M (2010) Online learning of task-driven object-based visual attention control. Image Vis Comput 28(7):1130–1145

Broadbent DE (1958) Perception and communication. Pergamon, London

Buschman RJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Sci 315:1860–1862

Chikkerur S, Serre T, Tan C, Poggio T (2010) What and where: a Bayesian inference theory of attention. Vis Res 50(22):2233–2247

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1990) Attentional modulation of neural processing of shape, color, and velocity in humans. Science 248:1556–1559

Colby CL, Goldberg ME (1999) Space and attention in parietal cortex. Annu Rev Neurosci 22:319–349

Corbetta M, Shulman GL (2002) Control of goal-direct and stimulus-driven attention in the brain. Nat Neurosci Rev 3:201–215

Deco G, Schurmann B (2000) A hierarchical neural system with attentional top-down enhancement of the spatial resolution for object recognition. Vis Res 40(20):2845–2859

Deco G, Rolls ET (2005) Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. J Neurophysiol 94(1):295–313

Desimone R, Duncan J (1995) Neural mechanism of selective attention. Annu Rev Neurosci 18:193–222

de-Wit LH, Kentridge RW, Milner AD (2008) Object-based attention and visual area LO. Neuropsychologia 47:1483–1490

DeYoe EA, Van Essen DC (1988) Concurrent processing streams in monkey visual cortex. Trends Neurosci 11:219–226

DeYoe EA, Felleman DJ, Van Essen DC, McClendon E (1994) Processing streams in visual area v4 and inferotemporal cortex of the macaque monkey. Nature 37:1151–1154

Duncan J (1984) Selective attention and the organization of visual information. J Exp Psychol Gen 113: 501–517

Duncan J, Humphreys GW, Ward R (1997) Competitive brain activity in visual attention. Curr Opin Neurobiol 7:255–261

Eckstein MP, Shimozaki SS, Abbey CK (2002) The footprints of visual attention in the Posner cueing paradigm revealed by classification images. J Vis 2:25–45

Egeth HE, Yantis S (1997) Visual attention: control, representation, and time course. Annu Rev Psychol 48:269–297

Egly R, Driver J, Rafal RD (1994) Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. J Exp Psychol Gen 123:161–177

Elazary L, Itti L (2008) Interesting objects are visually salient. J Vis 8(3):1–15

Ellis A.W., Young A.W. (1996) Human cognitive neuropsychology: a textbook with readings. Psychology Press, Hove

Eriksen CW, James JD (1986) Visual attention within and around the field of local attention: a zoom lens model. Percept Psychophys 40(4):225–240

Fecteau JH, Munoz DP (2006) Salience, relevance, and firing: a priority map for target selection. Trends Cogn Sci 10(8):382–390

Fink GR, Dolan RJ, Halligan PW, Marshall JC, Frith CD (1997) Space-based and object-based visual attention: shared and specific neural mechanism. Brain 120:2013–2028

Fu H, Chi Z, Feng D (2009) An efficient algorithm for attention-driven image interpretation from segments. Pattern Recogn 42(1):126–140

Gao D, Mahadevan V, Vasconcelos N (2008) On the plausibility of the discriminant center-surround hypothesis for visual saliency. J Vis 8(7):1–18

Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481–484

Gouet-Brunet V, Lameyre B (2008) Object recognition and segmentation in videos by connecting heterogeneous visual features. Comput Vis Image Understand 111(1):86–109

Grossberg S (2005) Linking attention to learning, expectation, competition, and consciousness. In: Itti L, Rees G, Tsotsos J (eds) Neurobiology of attention. Academic Press, Elsevier pp 652–662

Hamker FH (2006) Modeling feature-based attention as an active top-down inference process. BioSystems 86:91–99

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3:284–291

Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Minneapolis, MN, pp 1–8

Hu Y, Rajan D, Chia L-T (2008) Detection of visual attention regions in images using robust subspace analysis. J Vis Commun Image Represent 19(3):199–216

Humphreys GW, Riddoch MJ (1995) Separate coding of space within and between perceptual objects: evidence from unilateral visual neglect. Cogn Neuropsychol 12(3):283–311

Indiveri G (2008) Neuromorphic VLSI models of selective attention: from single chip vision sensors to multichip systems. Sensors 8(9):5352–5375

Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259

Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vis Res 40:1489–1506

Itti L, Koch C (2001) Computational modeling of visual attention. Nat Neurosci Rev 21:314–329

Itti L, Baldi P (2006) Bayesian surprise attracts human attention. In: Weiss Y, Scholkopf B, Platt J (eds) Advances in neural information processing systems, vol 18. MIT press, MA pp 1–8

Julesz B (1991) Early vision and focal attention. Rev Mod Phys 63:735–772

Kadir T, Brady M (2001) Scale, saliency and image description. Int J Comput Vis 30(2):77–116

Kanwisher N, Wojciulik E (2000) Visual attention: insights from brain imaging. Nat Neurosci Rev 1:91–100

Kanwisher N (2003) The ventral visual object pathway in humans: evidence from fMRI. In: Chalupa L, Werner J (eds) The visual neurosciences. MIT Press, Cambridge, MA pp 1179–1189

Kastner S, Ungerleider LG (2000) Mechanism of visual attention in the human cortex. Annu Rev Neurosci 23:315–341

Kastner S, DeWeerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in ventral extrastriate cortex as revealed by functional MRI. Science 282:108–111

Kim BK, Lee S-Y (2004) Sequential recognition of superimposed patterns with top-down selective attention. Neurocomputing 58(60):633–640

Kimchi R, Razpurker-Apfeld I (2004) Perceptual grouping and attention: not all groupings are equal. Psychon Bull Rev 11(4):687–696

Koch C, Tsuchiya N (2007) Attention and consciousness: two distinct brain processes. Trends Cogn Sci 11:16–22

Lamme VAF (1995) The neurophysiology of figure-ground segregation in primary visual cortex. J Neurosci 15:1605–1615

Lavie N (1995) Perceptual load as a necessary condition for selective attention. Exp Psychol Hum Percept Perform 21:451–468

Lavie N. (1995) Perceptual load as a necessary condition for selective attention. J Exp Psychol Hum Percept Perform 21:451–468

Lavie N (2005) Distracted and confused?: selective attention under load. Trends Cogn Sci 9:75–82

Lavie N, Driver J (1996) On the spatial extent of attention in object based visual selection. Percept Psychophys 58(8):1238–1251

Lee KW (2008) Guiding attention by cooperative cues. J Comput Sci Technol 23(5):874–884

Lee KW, Feng J, Buxton H (2005) Cued search: a computational model of selective attention. IEEE Trans Neural Netw 16(4):910–924

Lee S, Kim K, Kim J-J, Kim M, Yoo H-J (2010) Familiarity based unified visual attention model for fast and robust object recognition. Pattern Recogn 43:1116–1128

Levine MW, Shefner JM (1991) Fundamentals of sensation and perception. Brooks/Cole, CA

Ling S, Liu T, Carrasco M (2009) How spatial and feature-based attention affect the gain and tuning of population responses. Vis Res 49:1194–1204

Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex. J Neurophysiol 77:24–42

Luck SJ, Ford MA (1998) On the role of selective attention in visual perception. Proc Natil Acad Sci USA 95:825–830

Mahadevan V, Vasconcelos N (2010) Spatiotemporal saliency in dynamic scenes. IEEE Trans Pattern Anal Mach Intell 32(1):171–177

Marr D (1982) Vision. W. H. Freeman and Company, San Francisco

Maunsell JHR, Treue S (2006) Feature-based attention in visual cortex. Trends Neurosci 29:317–322

Mendi E, Milanova M (2010) Contour-based image segmentation using selective visual attention. J Softw Eng Appl 3:796–802

Miau F, Papageorgiou C, Itti L (2001) Neuromorphic algorithms for computer vision and attention. In: Proceedings of the SPIE 46 annual international symposium on optical science and technology, vol 4479, pp 12–23

Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. Trends Neurosci 6:414–417

Mitchell J, Stoner G, Reynolds J (2004) Object-based attention determines dominance in binocular rivalry. Nature 429:410–413

Moore CM, Yantis S, Vauchan B (1988) Object-based visual selection: evidence from perceptual completion. Psychol Sci 9:104–110

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

Mozer MC, Sitton M (1998) Computational modeling of spatial attention. In: Pashler H (ed) Attention. Psychology Press, London pp 341–393

Muller NG, Bartelt OA, Donner TH, Villringer A, Brandt SA (2003) A physiological correlate of the “Zoom Lens” of visual attention. J Neurosci 23:3561–3565

Muller NG, Kleinschmidt A (2003) Dynamic interaction of object- and space-based attention in retinotopic visual areas. J Neurosci 23:9812–9816

Navalpakkam V, Itti L (2005) Modeling the influence of task on attention. Vis Res 45(2):205–231

Navalpakkam V, Itti L (2006) An integrated model of top-down and bottom-up attention for optimal object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. New York, NY, pp 2049–2056

Norman DA (1968) Towards a theory of memory and attention. Psychol Rev 75:522–536

Nothdurft HC (2006) Salience and target selection in visual search. Vis Cogn 14(4-8):514–542

O’Craven K, Kansisher N, Downing P (1999) fMRI evidence for objects as the units of attentional selection. Nature 401:584–587

Oliva A, Torralba A (2007) The role of context in object recognition. Trends Cogn Sci 11(12):520–527

Olshausen BA, Anderson CH, Van Essen DC (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J Neurosci 13(11):4700–4719

Orabona F, Metta G, Sandini G (2007) A Proto-object based visual attention model. In: Paletta L, Rome E (eds) Lecture notes in artificial intelligence, vol 4840. Berlin, Heidelberg, pp 198–215

Palmer SE (1998) The psychology of perceptual organization: a transformational approach. In: Beck J, Hope B, Rosenfeld A (eds) Human and machine vision. Academic, Orlando pp 269–339

Polk TA, Drake RM, Jonides JJ, Smith MR, Smith EE (2008) Attention enhances the neural processing of relevant features and suppresses the processing of irrelevant features in humans: a functional magnetic resonance imaging study of the Stroop task. J Neurosci 28:13786–13792

Posner MI, Snyder CRR, Davidson BJ (1980) Attention and the detection of signals. J Exp Psychol Gen 109:160–174

Posner MI, Nissen MJ, Ogden WC (1977) Attended and unattended processing modes: the role of set for spatial location. In: Pick HL, Saltzman IJ (eds) Modes of perceiving and processing information. Erlbaum, Hillsdale pp 160–174

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25

Qiu FT, von der Heydt R (2007) Neural representation of transparent overlay. Nat Neurosci 10:283–284

Reynolds JH, Desimone R (2003) Interacting roles of attention and visual salience in V4. Neuron 37: 853–863

Riesenhuber M (2005) Object recognition in cortex: neural mechanisms and possible roles for attention. In: Itti L, Rees G, Tsotsos J (eds) Neurobiology of Attention. Academic Press, Elsevier pp 279–287

Roelfserna Lamme PR, Lamme VAF, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. Nature 395:376–381

Roggeman C, Fias W, Verguts T (2010) Salience maps in parietal cortex: imaging and computational modeling. Neuroimage 52(3):1005–1014

Rybak IA, Gusakova VI, Golovan AV, Podladchikova LN, Shevtsova NAA (2005) Attention-guided recognition based on what and where representations: a behavioral model. In: Itti L, Rees G, Tsotsos J (eds) Neurobiology of Attention. Academic Press, Elsevier pp 2387–2400

Schoenfeld MA, Tempelmann C, Martinez A, Hopf JM, Sattler C, Heinze HJ, Hillyard SA (2003) Dynamics of feature binding during object-selective attention. Proc Natl Acad Sci 100:11806–11811

Serences JT, Schwarzbach J, Courtney SM, Golay X, Yantis S (2004) Control of object-based attention in human cortex. Cereb Cortex 14:1346–1357

Shimozaki SS, Eckstein MP, Abbey CK (2003) Comparison of two weighted integration models for the cueing task: Linear and likelihood. J Vis 3(3):209–229

Shomstein S, Yantis S (2004) Configural and contextual prioritization in object-based attention. Psychon Bull Rev 11:247–253

Sterzer P, Kleinschmidt A, Rees G (2009) The neural bases of multistable perception. Trends Cogn Sci 13(7):310–318

Sun Y, Fisher R (2003) Object-based visual attention for computer vision. Artif Intell 146(1):77–123

Sun Y, Fisher R, Wang F, Gomes HM (2008) A computer vision model for visual object based attention and eye movements. Comput Vis Image Understand 112(2):126–142

Torralbo A, Beck DM (2008) Perceptual load-induced selection as a result of local competitive interactions in visual cortex. Psychol Sci 19(10):1045–1050

Torralba A, Oliva A, Castelhano M, Henderson JM (2006) Contextual guidance of attention in natural scenes: the role of global features on object search. Psychol Rev 113(4):766–786

Treisman A, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97–136

Treisman A, Schmidt H (1982) Illusory conjunctions in perception of objects. Cogn Psychol 14:107–141

Treisman A (1960) Contextual cues in selective listening. Q J Exp Psychol 12:242–248

Turatto M, Mazza V, Umilta C (2005) Crossmodal object-based attention: auditory objects affect visual processing. Cognition 96:B55–B64

Vanrullen R (2003) Visual saliency and spike timing in the ventral visual pathway. J Physiol (Paris) 97(2): 365–377

Walther D, Rutishauser U, Koch C, Perona P (2005) Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. Comput Vis Image Understand 100:41–63

Wegener D, Ehn F, Aurich MK, Galashan FO, Kreiter AK (2008) Feature-based attention and the suppression of non-relevant object features. Vis Res 48:2696–2707

Wojciulik E, Kanwisher N (1999) The generality of parietal involvement in visual attention. Neuron 23: 747–764

Wolfe JM (1994) Guided search 2.0: a revised model of visual search. Psychon Bull Rev 1:202–238

Won W-J, Ban S-W, Moon J (2006) Biologically motivated face selective attention system. In: Proceedings of the international joint conference on neural networks, Vancouver, Canada, pp 4292–4297

Wuhr P, Frings C (2008) A case for inhibition: visual attention suppresses the processing of irrelevant objects. J Exp Psychol Gen 137:116–130

Zhaoping L (2008) Attention capture by eye of origin singletons even without awareness–a hallmark of a bottom-up saliency map in the primary visual cortex. J Vis 8(1):1–18