Different Binding Strategies for the Different Stages of Visual Recognition

John K. Tsotsos[1], Antonio Jose Rodriguez-Sanchez,
Albert L. Rothenstein, and Eugene Simine

Dept. of Computer Science and Engineering, and Centre for Vision Research
York University, Toronto, Ontario Canada

Abstract. Many think attention needs an executive to allocate resources. Although the cortex exhibits substantial plasticity, dynamic allocation of neurons seems outside its capability. Suppose instead that the processing structure is fixed, but can be 'tuned' to task needs. The only resource that can be allocated is time. How can this fixed structure be used over periods of time longer than one feed-forward pass? Can the Selective Tuning model provide the answer? This short paper has one goal, that of explaining a single figure (Fig.1), that puts forward the proposal that by using multiple passes of the visual processing hierarchy, both bottom-up and top-down, and using task information to tune the processing prior to each pass, we can explain the different recognition behaviors that human vision exhibits. To accomplish this, four different kinds of binding processes are introduced and are tied directly to specific recognition tasks and their time course.

1 Introduction

Topics like visual attention, recognition, or binding command a large, conflicting literature. For example, the nature of the attentional influence has been debated for a long time. Among the more interesting observations are those of James (1980) "everyone knows what attention is..." juxtaposed with that of Pillsbury (1908) "attention is in disarray" and Sutherland's (1998) "after many thousands of experiments, we know only marginally more about attention than about the interior of a black hole". Even Marr, basically discounted the importance of attention by not considering the time intervals of perception where attentive effects appear. When describing grouping processes and the full primal sketch, he says, "our approach requires that the discrimination be made quickly - to be safe, in less than 160ms - and that a clear psychophysical boundary be present" (Marr 1982, p.96). Not only is the number of experimental investigations enormous, but also the number of different models, theories and perspectives is large. Attention has been viewed as early selection (Broadbent 1958), using attenuator theory (Treisman 1964), as a late selection process (Norman 1968, Deutsch & Deutsch 1963), as a result of neural

---

[1] JKT is also with the Depts. of Computer Science and of Ophthalmology at the University of Toronto, and holds the Canada Research Chair in Computational Vision.

synchrony (Milner 1974), using the metaphor of a spotlight (Shulman et al. 1979), within the feature integration theory (Treisman & Gelade, 1980), as an object-based phenomenon (Duncan 1984), using the zoom lens metaphor (Eriksen & St. James 1986), as a pre-motor theory subserving eye movements (Rizzollati et al. 1987), as biased competition (Duncan & Desimone 1995), as feature similarity gain (Treue & Martinez-Trujillo 1999), and more.

Within all of these different viewpoints, the only real constant seems to be that attentional phenomena seem to be due to inherent limits in processing capacity in the brain (Tsotsos 1990). But even this does not constrain a solution. Even if we all agree that there is a processing limit, what is its nature? How does it lead to the mechanisms in the brain that produce the phenomena observed experimentally?

We suggest that the terms attention, recognition and binding have become so loaded that they mask the true problems; each may be decomposed into smaller, easier problems. For example, consider the observations that different recognition tasks require different processing times. Detection and categorization seem to take about 150ms, identification takes about 65ms longer, localization of a stimulus so that detection can be expressed through a saccade or pointing action takes 200ms or more, and harder tasks such as detection in clutter, transparent motion or difficult conjunctions take even longer.

We propose that the process of binding visual features to objects in each of these tasks differs and that different sorts of binding actions take different amounts of processing time. Some require attention, others do not. We introduce a novel set of four binding processes: convergence, partial and full recurrence, and iterative binding. These are tied to different recognition tasks: detection or categorization, identification, localization and hard detection. The Selective Tuning model (Tsotsos 1990; Tsotsos et al. 1995), through its execution time course and due to its inherent tuning functionality, provides much of the computational substrate for these types of binding, recognition and attentive modulation.

2   The Stages of Recognition

Which knife can one use to carve 'recognition' into manageable slices? There are many possibilities. Should those slices be different brain areas, each responsible for different sub-tasks? Should those slices be different tasks? What about varying feature, object or scene complexity? The argument made by this paper is that the same neural machinery of the visual cortex is used for any of these dimensions (admittedly, some areas perhaps more involved than others) and that the most effective way of carving up the problem is to cut along the dimension of time. That is, different tasks are known to take different amounts of processing time even though they require the same neural machinery for that processing.

Much past experimental research has already provided what is needed. Consider the time course of events during a typical visual search experiment. An abstraction of this appears in the next paragraph. For each step, it is the same pair of eyes, the same retinal cells, the same, LGN, V1, V2 and so forth, that process the incoming stimuli. Each step in the processing pathway requires processing time; no step is instantaneous

or can be assumed so. In such experiments, the timings for each of the input arrays is manipulated presumably in order to investigate different phenomena.

Consider the following characterization of the typical course of activities for an experiment investigating attention or recognition behavior:

1. provide the subject with task information, including
   - what are the cues if any
   - what is the task and what criteria are used to judge a successful trial
   - what sequence of events will the subject see
2. attend fixation stimulus
3. onset of stimulus array
4. process stimulus array, perhaps including
   - detect items in array
   - attend to one or more items, re-applying or modifying task guidance in order to solve the task
   - interpret item's characteristics as required for the task
5. respond to stimulus array using one of the following
   - key press while continuing to fixate
   - saccade to perceived location
6. subject feedback on response
7. onset of next stimulus array, using one or more of
   - mask
   - blank
   - new stimuli to relate to previous

There are many, many variations on this basic theme and this is where the ingenuity of the best experimentalists can shine. However, for those wishing to explain the experimental observations the sequence of actual events plays a more important role than has been acknowledged. A modeler cannot simply take the conclusion of the experiment as the basis of a model without also including the spatial and temporal environment of the experiment into account. This would only lead to models that do not reflect the reality of the experiment or do not generalize and thus produce useful predictions.

3   Different Stages of Recognition

If models are to be sensibly compared to results from human experimentation, the models must consider the same sequence of events as in the experiment and examples of such a sequence appeared in the previous section.

Most models assume that a hierarchical sequence of computations defines the selectivity of a neuron. A feed-forward pass through the hierarchy would yield the strongest responding neurons if stimuli match existing neurons, or the strongest responding component neurons if stimuli are novel.

But, the first set of computations to be performed, following the sequence presented in the previous section, is related to priming of the hierarchy of processing

areas. Task knowledge, such as fixation point, target/cue location, task success criteria, is applied to 'tune' the hierarchy (Posner et al. 1978). In experiments, it has been shown that such task guidance must be applied 300 to 100ms before stimulus onset to be effective (Müller & Rabbitt 1989). This informs us that significant processing time is required for this step alone. It is a sufficient amount of time to complete a top-down traversal of the full processing hierarchy. Figure 1 shows a proposed sequence of processing stages in visual recognition tasks. The first stage, the leftmost element of the figure, shows the priming stage. Once complete, the stimulus can be presented (the second element of the figure from the left).

The third element of Fig. 1 represents the *Detection/Categorization Task*. Detecting whether or not a particular object is present in an image seems to take about 150ms (Thorpe et al. 1996). This kind of 'yes-no' response can be called 'pop-out' in visual search with the added condition that the speed of response is the same regardless of number of distractors (Treisman & Gelade 1980). To name the object, or to categorize, also seems to take the same amount of time (Grill-Spector & Kanwisher 2005; Evans & Treisman 2005). Interestingly, the median time required for a single feed-forward pass through the visual system is about 150ms (Bullier 2001). Thus, many conclude that a single feed-forward pass suffices for this visual task. This first feed-forward pass is shown in the figure emphasizing the feed-forward divergence of neural connections and thus stimulus elements are 'blurred' progressively more in higher areas of the hierarchy. This task does not include location or location judgments, the need to manipulate, point, or other motor commands specific to the object and usually, all objects can be easily segmented. These are the kinds of stimuli Marr had in mind for his work as mentioned previously.

To provide details about an object, such as identity (within-category identification) or type, requires additional processing time, 65ms or so (Grill-Spector & Kanwisher 2005; Evans & Treisman 2005); this is the *Identification Task* and is represented by the fourth from the left element of Figure 1. If the highest levels of the hierarchy can provide the basic category of the stimulus, such as 'bird', where are the details that allow one to determine the type of bird? The sort of detail required would be size, color, shape, and so forth. These are clearly lower level visual features and thus they can only be found in earlier levels of the visual hierarchy. They can be accessed by looking at which feature neurons feed into those neurons that provided the category information. One way to achieve this is to traverse the hierarchy downwards, beginning with the category neuron and moving downwards through the needed feature maps. This downwards traversal is what requires the additional time observed. The extent of downward traversal is determined by the task, that is, the aspects of identification that are required. It is interesting to consider an additional impact of a partial downwards traversal. This traversal may be partial not only because of the task definition but also because the full traversal is interrupted and not allowed to complete either because new stimuli enter the system before there is enough time for completion or because not enough time is permitted due to other tasks. The result is that there is the potential for errors in localization and these may lead to the well-known illusory conjunction phenomenon (Treisman & Schmidt 1982).

If additional localization is required for description or a motor task, (pointing, saccade, etc..), then the top-down traversal process must be allowed to complete and
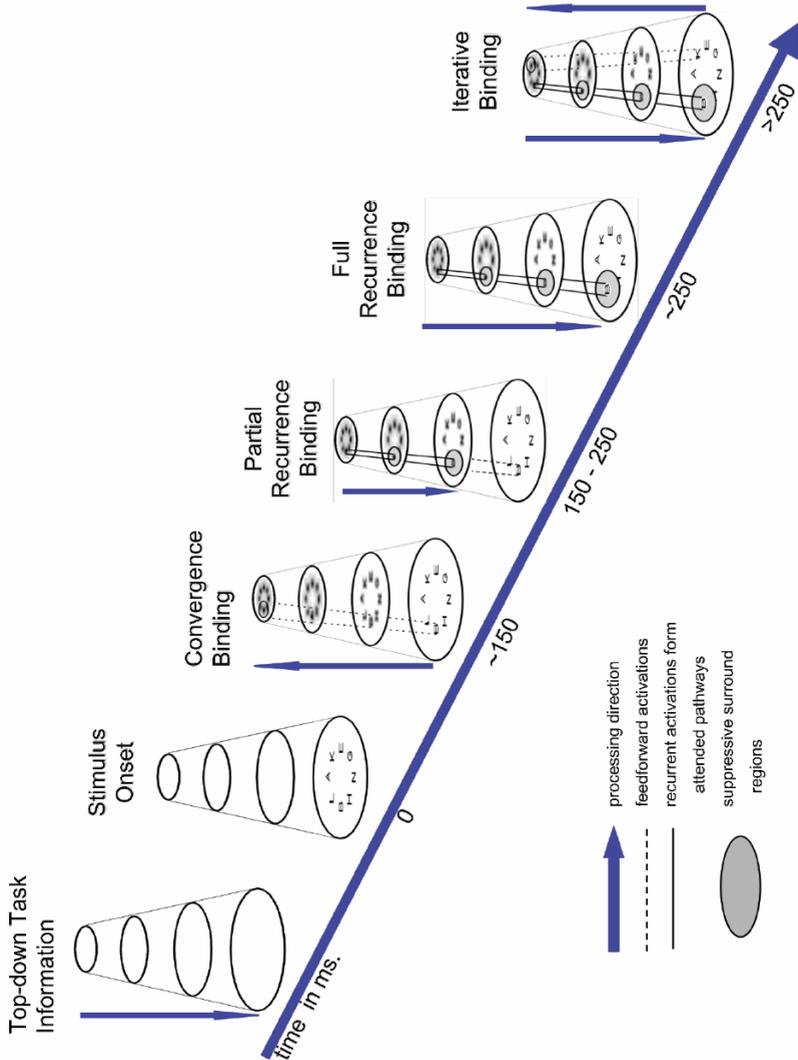
Fig. 1. The time course of visual recognition stages with types of visual binding required for each

thus additional time is required. This is called the *Localization Task*. How much time? A single saccade seems to require 200ms (with a range of 100-300ms) of processing time (Becker 1991). A lever press response seems to need 250-450ms in monkey (Mehta et al. 2000). During this task, the temporal pattern of attentional modulation shows a distinct top-down pattern over a period of 35 - 350ms post-stimulus. The 'attentional dwell time' needed for relevant objects to become available to influence behavior seems to be about 250ms (Duncan et al. 1994). Pointing to a target in

humans seems to need anywhere from 230 to 360ms (Gueye et al. 2002; Lünenburger & Hoffman 2003). Still, none of these experiments cleanly separate visual processing time from motor processing time; as a result, these results can only provide an encouraging guide for the basic claim of our model and further experimental work is needed.

Behavior, i.e., an action relevant to the stimulus, requires localization. The location details are available only in the earliest layers of the visual processing hierarchy because that is where the finest spatial resolution of neural representation can be found. As a result, the top-down traversal initiated for the Identification Task must complete so that is reaches these earliest layers as shown in the figural element second from the right in Fig. 1.

All of the above tasks as described can be characterized by stimuli that are well separated, can be easily segregated from the background, and are in an important sense, simple. In most real world scenes and many more complex experimental displays, even more time is needed. The *Hard Recognition Task* includes difficult conjunction searches, resolving illusory conjunctions, determining transparency, recognizing objects in cluttered scenes, and more (Treisman & Gelade 1980; Treisman & Schmidt 1982; Wolfe 1998; Schoenfeld et al. 2003). The final element of the figure, the rightmost element, depicts the start of a second feed-forward pass to illustrate this. The idea is that it is likely that several iterations of the entire process, feed-forward and feedback, may be required to solve difficult tasks.

4   The Visual Feature Binding Problem

Following Roskies (1999), the canonical example of binding is the one suggested by Rosenblatt in which one sort of visual feature, such as an object's shape, must be correctly associated with another feature, such as its location, to provide a unified representation of that object. Such explicit association ("binding") is particularly important when more than one visual object is present, in order to avoid incorrect combinations of features belonging to different objects, otherwise known as "illusory conjunctions" (Treisman & Schmidt 1982). Binding is a broad problem: visual binding, auditory binding, binding across time, cross-modal binding, cognitive binding of a percept to a concept, cross-modal identification and memory reconstruction.

Classical demonstrations of binding seem to rely on two things: the existence of representations in the brain that have no location information, and, representations of pure location for all stimuli. However, there is no evidence for a representation of location independent of any other information. Similarly, there is no evidence for a representation of feature without a receptive field. Nevertheless, location is *partially* abstracted away within a hierarchical representation as part of the solution to complexity (Tsotsos 1990). A single neuron receives converging inputs from many receptors and each receptor provides input for many neurons. Precise location is lost in such a network of diverging feed-forward paths yet increasingly larger convergence onto single neurons. How can location be recovered and connected to the right features and objects as binding seems to require?

We might begin by developing requirements for the solution of the binding problem. Define the binding task as requiring the solution of three sub-problems: Detection (is a given object/event present in the display?), Localization (location and spatial extent of detected object/event) and Attachment (explicit object/event links to its constituent components). We will be able to recognize a solution when an algorithm can correctly provide correct answers to the above, and this occurs in images that: a) contain more than one copy of a given feature each at different locations; b) contain more than one object/event each at different locations; and, c) contain objects/events that are composed of multiple features and share at least one feature type. These constraints provide us with a way of designing solutions and testing them with well-defined success criteria. They also provide constraints on what kinds of stimuli and tasks actually require binding in the first place. They will be used in the next section to suggest solutions to the kinds of binding that the different stages of recognition require. Previous proposals for the binding problem (see Roskies 1999) have not dealt with such constraints on the definition of the problem and this points to the uniqueness of the present proposal.

5   The Kinds of Binding Needed for the Stages of Recognition

A novel set of four different binding processes are introduced that are claimed to suffice for solving the recognition tasks described above.

Convergence Binding achieves the *Detection/Categorization Task* via hierarchical neural convergence, layer by layer, in order to determine the strongest responding neural representations at the highest layers of the processing hierarchy. This feed-forward traversal follows the task-modulated neural pathways through the 'tuned' visual processing hierarchy. This is consistent with previous views on this problem (Treisman 1999; Reynolds & Desimone 1999). This type of binding will suffice only when the abstraction achieved as a result of neural convergence does not obscure location or feature information that may be needed, and if stimulus elements do not lead to ambiguity at the higher levels due to the large receptive fields. That is, stimulus elements that fall within the larger receptive fields must not be too similar or otherwise interfere with the response of the neuron to its ideal tuning properties. Such interference may be thought of as 'noise' with the target stimulus being 'signal'. Convergence binding provides neither method for reducing this noise nor a method for recovering precise location. According to the requirements for a binding solution, this is not strictly an example of binding; it is named so here for continuity with past literature.

Full Recurrence Binding achieves the *Localization Task*. If Convergence Binding is followed by a complete top-down traversal, attended stimuli in each feature map of the hierarchical representation can be localized. Recurrent (or feedback) traversals through the visual processing hierarchy 'trace' the pathways of neural activity that lead to the strongest responding neurons at the top of the hierarchy resulting from the feed-forward traversal. Even for the strongest responding neurons with very large receptive fields where a number of stimulus elements fall within that receptive field, the reason why that response is strong lies within that receptive field and can be found.

There is one more critical component of the top-down traversal, appearing on the figures as gray regions indicating areas of neural suppression or inhibition in the area surrounding the attended stimulus. This area is defined by the projection of the receptive field of the neuron that best describes the stimulus through the processing hierarchy. That is, it is the set of neural pathways that feed that neuron. The reason for this particular definition stems from the previous discussion on signal versus noise in the input scene. Inputs corresponding to the stimulus most closely matching the tuning characteristics of the neuron form the signal while the remainder of the inputs is the noise. Any lateral connections are also considered as noise for this purpose. Thus, if it can be determined what those signal elements are, the remainder of the receptive field is suppressed, enhancing the overall signal-to-noise ratio of processing for that neuron. The method for achieving this was first described in (Tsotsos 1993) and fully detailed together with proofs of convergence and other properties in (Tsotsos et al. 1995). It is based on the assumption that the signal is defined by the strongest responses in each layer and within the receptive field of the neuron or neurons selected at the top of the hierarchy (again by strongest response).

However, the top-down process is complicated by the fact that each neuron within any layer may receive input from more than one feature representation. How do the different representations contribute to the selection? Different features may have different roles. For example, there are differing representations for many different values of object velocity however an object can only exhibit one velocity. These different representations can be considered as mutually exclusive, so the top-down search process must select one, the strongest. On the other hand, there are features that cooperate, such as the features that make up a face (nose, eyes, etc.). These contribute to the neuron in a weighted sum manner and the top-down search process much select appropriate elements from each. There may be other roles as well. The key here is that each neuron may have a complex set of inputs, specific to its tuning properties, and the top-down traversal must be specific to each. This is accomplished by allowing the choices to be made locally, at each level, as if there were a localized saliency representation for each neuron (Tsotsos et al. 2005). There is no global representation of saliency in this model.

If the full recurrence binding process does not complete for any reason, this is called *Partial Recurrence Binding*. Partial recurrence binding can find the additional information needed to solve the *Identification Task* if it is represented in intermediate layers of the processing hierarchy. If this is not deployed directly due to task needs but is due to interruption, then this results in illusory conjunctions. A variety of different effects may be observed depending on when during the top-down traversal the process is interrupted.

Iterative Binding is needed for the *Hard Detection Task*, i.e., discrimination, description, search, etc. Iterative Binding is defined as one of more Convergence Binding-Full Recurrence Binding cycles. The processing hierarchy may be tuned for the task before each traversal as appropriate. The iteration terminates when the task is satisfied. This iterative feed-forward-feedback cycle was first described in Tsotsos 1990.

Simulations of this strategy show strong agreement with a variety of psychophysical and neurophysiologic experiments such as static visual searches of

varying difficulties, segregation of transparent dot pattern motions, surround inhibition, and more (Rothenstein & Tsotsos 2006; Rodriguez-Sanchez et al. 2006; Tsotsos et al. 2005; Tsotsos et al. 1995). In particular the surround inhibition prediction seems well supported by a variety of experimental studies (Cutzu & Tsotsos 2003: Hopf et al. 2005: Tombu & Tsotsos 2007). The top-down attentional modulation hypothesis also proposed by Selective Tuning has strong support (Mehta et al. 2000; O'Connor et al. 2002).

6   Conclusion

A novel view of how attention, visual feature binding, and recognition are inter-related has been presented. It differs from any of those presented previously (Roskies 1999). The greatest point of departure is that it provides a way to integrate binding by convergence with binding depending on attention. The visual binding problem is decomposed into four kinds of processes each being tied to one of the classes of recognition behaviors that have been investigated experimentally over the past decades that are defined by task and time course. This view differs from conventional wisdom that considers both binding and recognition as monolithic tasks. The decomposition has the promise of dividing and conquering these problems, and the Selective Tuning strategy previously presented is proposed as the computational substrate for their solution. There are three basic ideas behind this solution:

- top-down task directed priming before processing;
- feed-forward traversal through the 'tuned' visual processing hierarchy following the task-modulated neural pathways;
- recurrent (or feedback) traversals through the visual processing hierarchy that 'trace' the pathways of neural activity that lead to the strongest responding neurons at the top of the hierarchy that result form the feed-forward traversal.

These three basic steps are used in combination, and repeated, as needed to solve the given visual task. The details of how exactly these processes may be accomplished are detailed elsewhere (Tsotsos 1990; Tsotsos et al. 1995; Tsotsos et al. 2005). In simulation with artificial as well as real images as input, the model exhibits good agreement with a wide variety of experimental observations.

The model has a number of important characteristics: a particular time course of events during the recognition process covering the simplest to complex stimuli that can be directly compared with experimental time courses; an iterative use of the same visual processing hierarchy in order to deal with the most complex stimuli; iterative tuning of the same visual processing hierarchy specific to task requirements; suppressive surrounds due to attention that assist with difficult segmentations; a particular time course of events for recognition ranging from simple to complex recognition tasks; a top-down localization process for attended stimuli based on tracing feed-forward activations guided by localized saliency computations. Each of these may be considered a prediction for human or non-human primate vision. It would be very interesting to explore each.

Acknowledgments

References

Becker, W.: Saccades. In: Carpenter, R.H.S. (ed.) Eye Movements, pp. 95–137. CRC Press, Boca Raton, USA (1991)

Broadbent, D.: Perception and communication. Pergamon Press, NY (1958)

Bullier, J.: Integrated Model of Visual Processing. Brain Research Reviews 36, 96–107 (2001)

Cutzu, F., Tsotsos, J.K.: The selective tuning model of visual attention: Testing the predictions arising from the inhibitory surround mechanism. Vision Research 43, 205–219 (2003)

Deutsch, J., Deutsch, D.: Attention: Some theoretical considerations. Psych. Review 70, 80–90 (1963)

Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention. Ann. Rev. of Neuroscience 18, 193–222 (1995)

Duncan, J.: Selective attention and the organization of visual information. J. Exp. Psychol. Gen. 113(4), 501–517 (1984)

Duncan, J., Ward, J., Shapiro, K.: Direct measurement of attentional dwell time in human vision. Nature 369, 313–315 (1994)

Eriksen, C., James, J.S.: Visual attention within and around the field of focal attention: a zoom lens model. Percept. Psychophys. (4), 225–240 (1986)

Evans, K., Treisman, A.: Perception of Objects in Natural Scenes: Is It Really Attention Free? J. Experimental Psychology: Human Perception and Performance 31(6), 1476–1492 (2005)

Grill-Spector, K., Kanwisher, N.: Visual recognition: As soon as you know it is there, you know what it is. Psychological Science 16, 152–160 (2005)

Gueye, L., Legalett, E., Viallet, F., Trouche, E., Farnarier, G.: Spatial Orienting of Attention: a study of reaction time during pointing movement. Neurophysiologie Clinique 32, 361–368 (2002)

Hopf, J.-M., Boehler, C.N., Luck, S.J., Tsotsos, J.K., Heinze, H.-J., Schoenfeld, M.A.: Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. PNAS 103(4), 1053–1058 (2006)

James, W.: The Principles of Psychology, H. Holt (1890)

Lünenburger, L., Hoffman, K.-P.: Arm movement and gap as factors influencing the reaction time of the second saccade in a double-step task. European J. Neuroscience 17, 2481–2491 (2003)

Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Henry Holt and Co., New York (1982)

Treue, S., Martinez-Trujillo, J.: Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399(6736), 575–579 (1999)

Mehta, A., Ulbert, I., Schroeder, C.: Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. Cerebral Cortex 10(4), 343–358 (2000)

Milner, P.: A model for visual shape recognition. Psych. Rev. 81, 521–535 (1974)

Müller, H., Rabbitt, P.: Reflexive and Voluntary Orienting of Visual Attention: Time course of activation and resistance to interruption. J. Exp. Psychology: Human Perception and Performance 15, 315–330 (1989)

Norman, D.: Toward a theory of memory and attention. Psych. Review 75, 522–536 (1968)

O'Connor, D., Fukui, M., Pinsk, M., Kastner, S.: Attention modulates responses in the human lateral geniculate nucleus. Nature Neuroscience 5(11), 1203–1209 (2002)

Posner, M.I., Nissen, M., Ogden, W.: Attended and unattended processing modes: The role of set for spatial locations. In: Pick, Saltzmann (eds.) Modes of Perceiving and Processing Information, pp. 137–158. Erlbaum, Hillsdale, NJ (1978)

Reynolds, J., Desimone, R.: The Role of Neural Mechanisms of Attention in Solving the Binding Problem. Neuron 24, 19–29 (1999)

Rizzolatti, G., Riggio, L., Dascola, I., Umilta, C.: Reorienting attention across the horizontal and vertical meridians - evidence in favor of a premotor theory of attention. Neuropsychologia 25, 31–40 (1987)

Rodriguez Sanchez, A., Simine, E., Tsotsos, J.K.: Feature Conjunctions in Serial Visual Search. In: Int. Conf. Artificial Neural Networks, Athens, Greece (September 10-14, 2006)

Roskies, A.: The Binding Problem - Introduction. Neuron 24, 7–9 (1999, 1999b)

Rothenstein, A., Tsotsos, J.K.: Selective Tuning: Feature Binding Through Selective Attention. In: Int. Conf. Artificial Neural Networks, Athens, Greece (September 10-14, 2006)

Schoenfeld, M., Tempelmann, C., Martinez, A., Hopf, J.-M., Sattler, C., Heinze, H.-J., Hillyard, S.: Dynamics of feature binding during object-selective attention. Proc. National Academy of Sciences 100(20), 11806–11811 (2003)

Shulman, G., Remington, R., McLean, J.: Moving Attention through Visual Space. J. Experimental Psychology 92, 428–431 (1979)

Sutherland, S.: Feature Selection. Nature 392, 350 (1998)

Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. Nature 381, 520–522 (1996)

Tombu, M., Tsotsos, J.K.: Attending to orientation results in an inhibitory surround in orientation space. Perception & Psychophysics (to appear, 2007)

Treisman, A.: The effect of irrelevant material on the efficiency of selective listening. American J. Psychology 77, 533–546 (1964)

Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology 12(1), 97–136 (1980)

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

Treisman, A.: Solutions to the Binding Problem: Progress through Controversy and Convergence. Neuron 24(1), 105–125 (1999)

Tsotsos, J.K.: A Complexity Level Analysis of Vision. Behavioral and Brain Sciences 13, 423–455 (1990)

Tsotsos, J.K.: An Inhibitory Beam for Attentional Selection. In: Harris, L., Jenkin, M. (eds.) Spatial Vision in Humans and Robots, pp. 313–331. Cambridge University Press, Cambridge (1993)

Tsotsos, J.K., Culhane, S., Wai, W., Lai, Y., Davis, N., Nuflo, F.: Modeling visual attention via selective tuning. Artificial Intelligence 78(1-2), 507–547 (1995)

Tsotsos, J.K., Liu, Y., Martinez-Trujillo, J., Pomplun, M., Simine, E., Zhou, K.: Attending to Motion. Computer Vision and Image Understanding 100(1-2), 3–40 (2005)

Wolfe, J.M.: Visual Search. In: Pashler, H. (ed.) Attention, pp. 13–74. Psychology Press Ltd., Hove, UK (1998)