

Modeling the Dynamics of Feature Binding During Object-Selective Attention

Albert L. Rothenstein and John K. Tsotsos

Centre for Vision Research and Dept. of Computer Science & Engineering
York University, Toronto, Canada
albertlr, tsotsos@cs.yorku.ca

Abstract. We present a biologically plausible computational model for solving the visual feature binding problem. The binding problem appears to be due to the distributed nature of visual processing in the primate brain, and the gradual loss of spatial information along the processing hierarchy. The model relies on the reentrant connections so ubiquitous in the primate brain to recover spatial information, and thus allows features represented in different parts of the brain to be integrated in a unitary conscious percept. We demonstrate the ability of the Selective Tuning model of visual attention [1] to recover spatial information, and based on this we propose a general solution to the feature binding problem. The solution is used to simulate the results of a recent neurophysiology study on the binding of motion and color. The example demonstrates how the method is able to handle the difficult case of transparency.

1 Introduction

Convergent evidence from many different kinds of studies suggests the visual cortex is divided into specialized areas for color, orientation, depth, movement, etc., organized into two main processing streams, a dorsal pathway, responsible for encoding motion, space, and spatial relations for guiding actions, and a ventral pathway, associated with object recognition and classification, conclusions supported by functional imaging, neurophysiology, and by strikingly selective localized lesions [2]. This high selectivity of the various cortical areas has led researchers to ask the obvious questions how, despite this specialization, the visual percept is unitary, and what are the mechanisms responsible for, in effect, putting all this distributed information together. Following Roskies [3], “the canonical example of binding is the one suggested by Rosenblatt [4] 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 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” [5]. Limiting the resources available for visual processing through increased loads and/or reduced time leads observers to erroneously associate basic features present in the image into objects that do not exist, e.g. a red X and a blue O are sometimes reported

as a blue O and a red X. Studies have shown that these are real conjunction errors, and can not be attributed to guessing or memory [6]. Several other examples of the varieties of binding problems in the literature appear in the special issue of neuron on the binding problem - Neuron 24(1) (1999).

1.1 Overview of Approach

This paper presents a solution to the feature binding problem in the context of the Selective Tuning (ST) model of visual attention, and demonstrate it by comparing its behavior and temporal characteristics with recent experimental results. After introducing ST, we will highlight the special role of spatial information in binding, and demonstrate STs ability to recover spatial information from abstract, high-level representations. This ability will be used to model the binding of motion and color, in a setup similar to that used by Schoenfeld et al. [7], a study that investigated the timing and cortical localization of feature-integration mechanisms in object-based attention. In this study, recordings of event-related potentials and magnetic fields were combined with functional MRI while subjects attended to one of two superimposed transparent surfaces formed by arrays of dots moving in opposite directions. A spatiotemporal analysis revealed an increase in neural activity localized to a color-selective region of the fusiform gyrus when the surface moving in the attended direction displayed an irrelevant color feature. The delay of this increase is on the order of $\sim 40-60$ ms, indicating that attentional modulation of color stimuli require a serial processing of information, first in the motion analysis pathway from V1 to MT/ V5 and thereafter via the ventral stream to IT.

An analysis of the model flow of information will reveal a pattern of temporal activity that is very similar to the one argued for by the experiments. Similar timing results have been reported for the binding of task-relevant features [8,9,10] and for form-from-motion experiments [11], indicating that the model results could be extended to other feature domains.

2 Selective Tuning

Selective Tuning (ST) [1] is a computational model of visual attention characterized by the integration of feedforward and feedback pathways into a network that is able to take high level decisions, and, through a series of winner-take-all processes, identify all the neurons that have participated in that decision. This identification satisfies the key requirement for a kind of visual feature binding that ST was demonstrated to solve [12], despite the loss of spatial information inherent in a pyramidal system. The ST feedback process does not need collocation if neural convergence is guaranteed, so ST is able to select all parts of a stimulus, even if they do not share a location (e.g. stimuli with discontinuities due to overlap, or stimuli that are separated spatially due to the nature of the cortical feature maps). The partial solution to binding proposed in [12] is able to correctly bind all the activations that have contributed to a high level decision (*convergence*) and even non-convergent representations if the problem can be

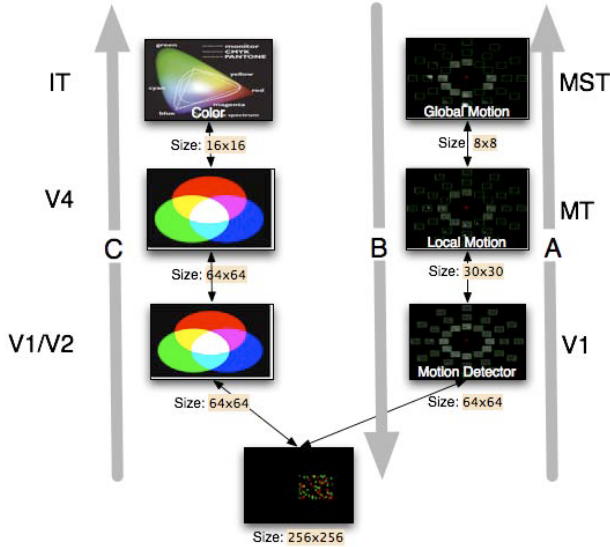


Fig. 1. Diagram of the network. On the left side is the color pathway, while on the right side is the motion pathway. The arrows show the flow of information. See the text for details.

solved at the spatial resolution of the top level of the pyramid (a weak form of *collocation*) - i.e. there is sufficient spatial separation between the target and the distractors (see [13,14] for the importance of spatial separation in attention and recognition). It is important to note that the feedback process will select only the units responding to the selected stimulus, and not units that just happen to share locations with it, thus ensuring that overlapping and transparent stimuli will be handled correctly.

The general structure of the neural network used in this paper is presented in Fig. 1, consisting of two biologically inspired processing pathways, corresponding to the ventral and dorsal visual pathways. The pathways are pyramidal, meaning that successive layers represent more and more abstract concepts, and location, size and (direct) shape information is lost. The dorsal pathway recognizes affine motions, and is described in detail in [12]. The ventral pathway is a simple Gaussian blurring pyramid that detects the presence of color in the input image, while gradually losing spatial specificity. All the weights in the neural network are preset, no learning was used. The network operates in parallel at all times, i.e. all neurons are evaluated at every step, and the serial behavior described is a result of the their interaction and of the structure of the network.

Note that all the segmentation results are based purely on feedback from higher (and much more abstract) levels and not on simple color thresholding or other image manipulation techniques.

2.1 Recovering Spatial Information

In Feature Integration Theory (FIT) [15], location is considered a feature that is faithfully represented in a “master map” of locations but, as Taraborelli [16] points out: “the idea of binding itself is nothing but a spatial conjunction of information concerning visual attributes of the same item.” Tsotsos et al. [12] note that considering location as a feature can not be valid as location precision (resolution) changes layer to layer in any pyramid representation, and propose that location should be considered as the anchor that permits features to be bound together. At the same time, Robertson [17] lists three phenomena that demonstrate the special role of spatial attention in binding: illusory conjunctions under divided attention, dependence on number of distractors for conjunction searches, and the elimination of the influence of distractors with spatial cueing. In effect, a solution to the binding problem must address this seemingly incompatible requirement: binding is ultimately only a spatial conjunction, but at the same time it must be based on high-level information, allowing for object and feature-based selection.

While the special role of spatial information in feature binding has been highlighted by many researchers [15,16,12,17], the two experiments reviewed in this paper indicate that this spatial information needs to be as high-resolution as possible, to accurately recover shape and to distinguish between the different layers of transparent motion. This high-resolution spatial information is only available in the early stages of visual processing, and both the lateral geniculate nucleus [18,19] and the striate cortex [20,21] have been proposed as sites of visual feature binding. For object- and high-level feature- based attention to allow binding to take place, a mechanism is needed to recover the exact information in the input image that has contributed to the high-level decision, and this mechanism is provided by Selective Tuning.

As mentioned above, ST uses a series of winner-take-all processes to identify all the neurons that have participated in a high level activation. At the same time, this process inhibits all the connections that have not participated in this decision, creating an inhibitory surround that increases the signal-to-noise ratio for the selected stimulus, generating the observed attentional modulation of neuronal responses [22,23]. The inhibitory surround prediction of ST [23] has received significant experimental support [24,25,26,27,28,29,30,31,32,33,13,34,35,36,14].

This selection combined with surround suppression allows ST to perform feedback-directed image segmentation, as illustrated in Fig. 2.

Fig. 2(a) and Fig. 2(b) illustrates the segmentation of a face in a face detection system using ST (from [37], using an image from the Yale Face Database [38]). The system detects faces as spatial configurations of circular features at different scales in a pyramidal neural network, and the attentional system is used to recover the spatial extent of the detected face. See [37] for details on the system and additional examples (including the detection and segmentation of partially occluded faces).

Transparent stimuli pose a difficult problem for any segmentation system. Fig. 2(c) contains a stimulus composed of two overlapping grids of colored points. The

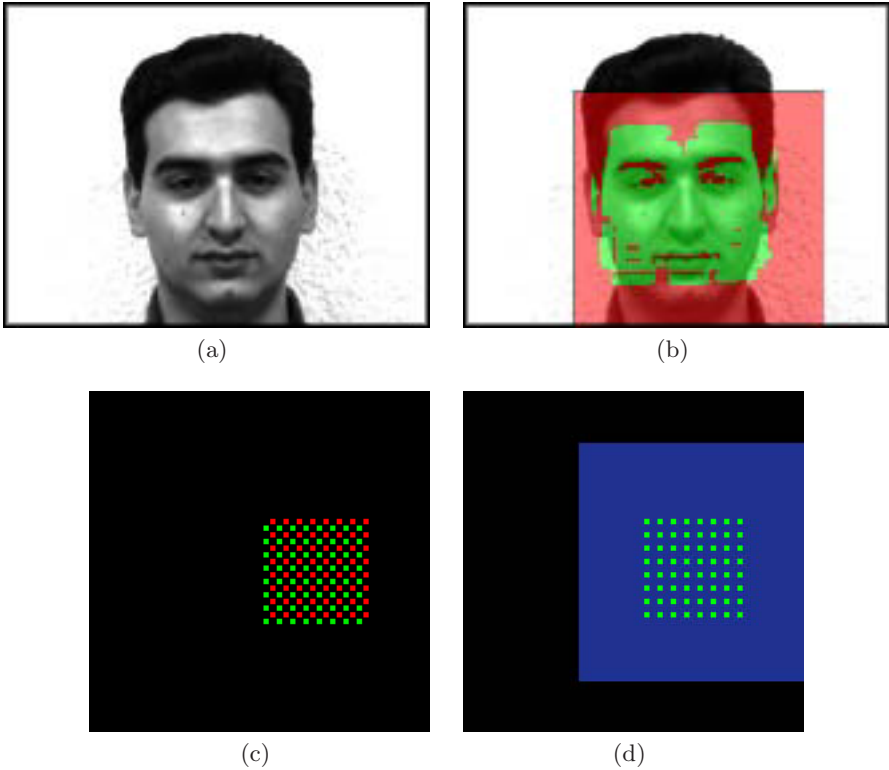


Fig. 2. Examples of Selective Tuning image segmentation. In the segmentation images, green depicts the selected region/pixels, red the inhibitory surround. (a) Face image. (b) Segmentation of the face in the input layer using the attentional beam. (c) Red/green color transparency grid. (d) Segmentation of the red stimulus.

attentional system detects the presence of the two colors, selects one, and through feedback isolates out the points of the corresponding color, while inhibiting the other color Fig. 2(d).

Other examples are presented in [12,39], including real video sequences.

3 Proposed Feature Binding Solution

This section motivates and introduces an original approach to the feature binding problem. The solution proposed is based on the general *collocation* principle of FIT, but using Selective Tuning to integrate high-level representations in the spatial selection process, and performing the spatial selection without requiring a “master map” of locations. The proposal, illustrated in Fig. 1 is to allow one feedforward pass through the network (arrow A in the figure). Similar to the primate visual system, this first pass occurs in parallel in all visual areas. Task information is used to select one high-level representation (in this case, the most

salient motion representation), and proceed backwards through the system in Selective Tuning manner (arrow B), selecting compatible representations that have contributed to the winning units, and inhibiting all the activations that are incompatible. As this feedback proceeds, lower level representations that participated in the salient activation are selected, and distractors inhibited, all the way to the first layer of processing. This allows further feedforward processing to be selectively directed towards the selected object and/or color (arrow C), in effect eliminating the influence of spatially near competing stimuli and allowing the ventral pathway to detect the shapes defined by the motion signal. When processing ends, the remaining active high-level representations all describe the selected stimulus in a sparse, distributed fashion ideal for maximizing the capacity of associative memories [40]. At the same time all the components of the attended stimulus will be selected throughout the visual system for recognition, and the location information can be used for the planning of actions towards the selected stimulus.

4 Binding Motion and Color

Our system, described above and represented in Fig. 1, will process the image in parallel, along the independent processing pathways, detecting the presence of the different colors and motion patterns. The attentional process will select one top-level representation for further analysis, and the ST process will localize the corresponding pixels in the input image through feedback. ST will also inhibit pixels in the surround of the attended item, thus enhancing the relative saliency of the attended stimulus and improving the signal-to-noise ratio in the system. A second feedforward pass through the pyramids will refine the representation of the selected object, and at the same time select all the (distributed) representations that belong to it, thus achieving binding. The process can be repeated until the desired target is found, implementing a visual search mechanism.

As mentioned in the Introduction, an extreme example of transparency is the completely overlapping motion and color class of stimuli used by [7], and illustrated in Fig. 3(a).

The attentional process will select one top-level motion representation for further analysis (the rightward motion, in this case), and the ST process will localize the corresponding pixels in the input image, and inhibit all nearby pixels, thus enhancing the saliency of the attended stimulus. Fig. 3(b) shows the selected and inhibited input areas in green and red, respectively, and 3(c) overlaps the selected and inhibited areas onto a frame of the input sequence.

Based on the ST selection and inhibition model, the red color representation is strongly inhibited, while the green representation is unchanged (its input has not changed, but if the input image was noisy, or if the system implemented lateral inhibition between representations, these effects would be diminished, making the representation more salient).

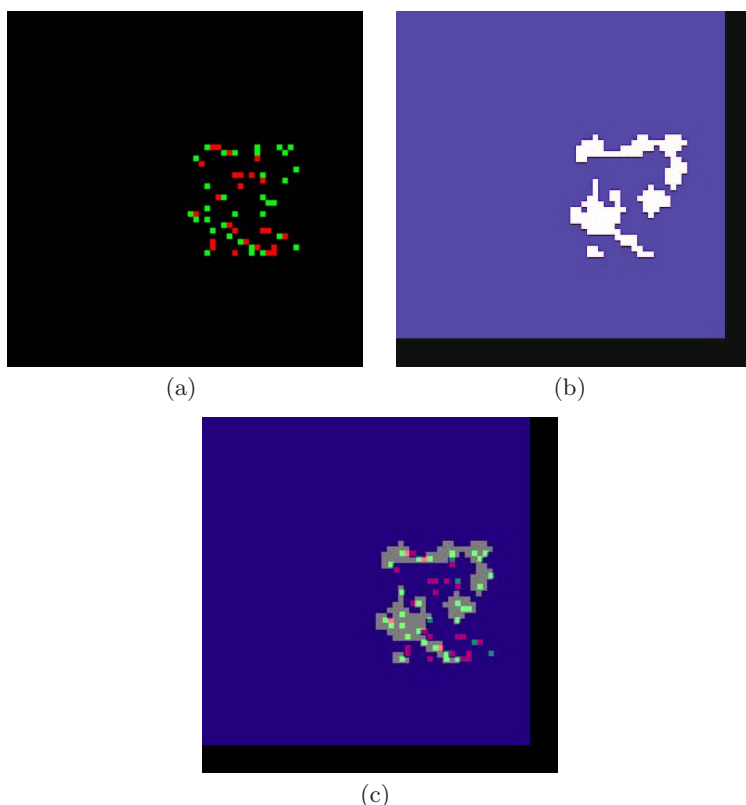


Fig. 3. Selective Tuning segmentation of a transparent motion stimulus. In the segmentation images, green depicts the selected region/pixels, red the inhibitory surround. (a) One frame in the transparent motion stimulus. The red dots translate to the right, the green dots to the left. The color of the dots is irrelevant, segmentation is based on the motion of the dots. (b) Selection of the green dots in the input layer using the attentional beam. (c) Segmentation results overlapped on top of the input frame. The green dots are selected, while the red dots are inhibited.

Thus, the ratio between the red and green representations is increased in the attended condition, duplicating the results presented by [7].

Note that the selection is not perfect, errors occur due to the absence of motion signals in the case of points that have just entered the frame or to the presence of illusory motion when points emerge after being occluded by other points. Analysis of the selection performance reveals that, on average, the system correctly selects 85% of the desired points, while allowing for 16% false positives.

4.1 Temporal Analysis

Fig. 4 illustrates the sequence of steps that accomplishes the task. After the stimulus appears, its information is processed along the two pathways – Fig. 4(a). The

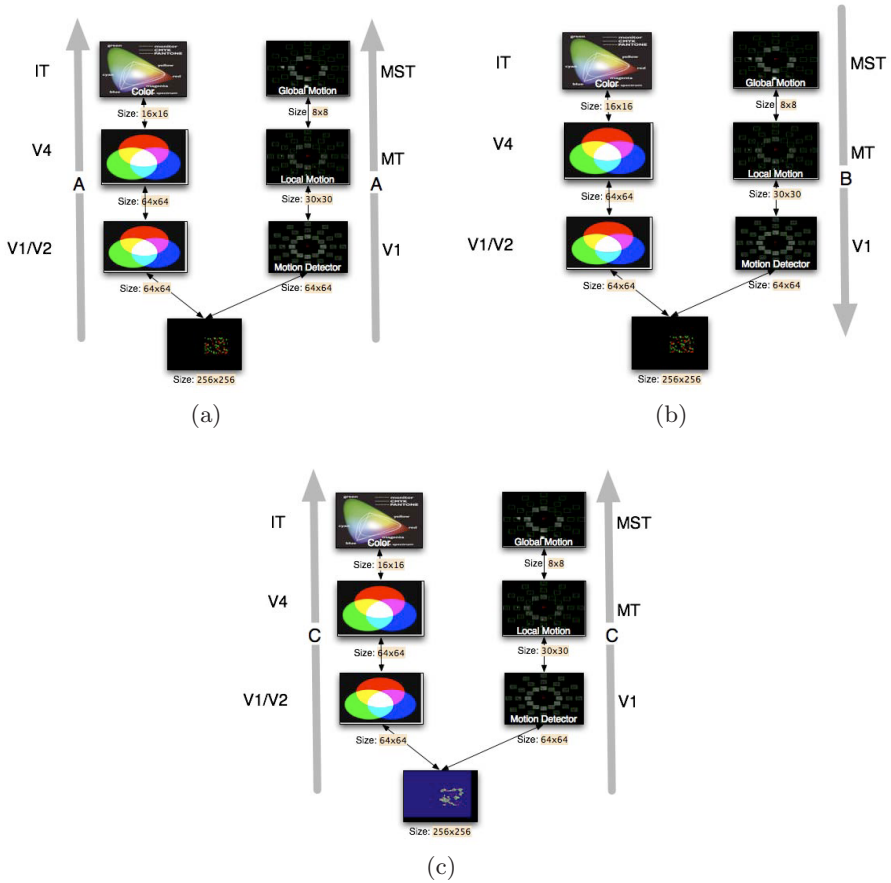


Fig. 4. Temporal flow of information. The arrows show the flow of information. See text for discussion. a) Stimulus is processed in parallel along both pathways. Attention, under task guidance, selects one direction of motion. b) Selective Tuning feedback localizes the corresponding moving dots in the input. c) A second feedforward pass along both pathways reevaluates the attended stimulus.

task is to attend to the leftward motion, so the corresponding high-level representation is selected, and the ST feedback process is applied to localize it in the input – Fig. 4(b). The attentional modulation in the dorsal pathway proceeds from higher level areas to lower level ones [41]. After the leftward motion is localized in the input and the rightward motion is inhibited, information again flows up the network, resulting in a reevaluation of the representations – Fig. 4(c). This reevaluation consists of refined representations for leftward motion and its color, green, and of inhibited representations for rightward motion and its color, red. This temporal pattern of activation corresponds in attentional modulation of the ventral

color representations following the dorsal attentional selection by a time interval proportional to the number of visual areas that the signals have to traverse. The experimentally observed time interval of $\sim 40\text{-}60\text{ms}$ corresponds to information passing through 3-5 visual areas [42] – MT–V1–V4 or MST–MT–V1–V2–V4, respectively.

5 Discussion

While the importance of space in binding is captured in the Feature Integration Theory, high level representations, object- and feature-based attention mechanisms are not easily integrated into FIT. In this paper we have presented an original solution to the feature binding problem in visual perception, by recovering spatial information from high level representations through Selective Tuning style feedback. Another important contribution of this research is a process of recovering spatial information that does not require a “master map” of locations or any other separate representation of space. We have demonstrated this solution by modeling experimental results. While these preliminary results are encouraging, the representations used are very simplistic, and significant work needs to be done to prove the generality of the solution.

A series of studies regarding the time course and processing sequence of motion and form/color [8,9,10,7,11] conclude that dorsal activation (area V5/MT) precedes ventral activation (areas LO and IT) by approximately 50ms (the actual delay depending on the task, stimuli, measuring methodology, etc).

These and other results are inconsistent with models that propose binding through convergence in higher-level areas [43], mainly due to the progressive loss of spatial specificity of neural responses.

Another explanation for the results could rely on direct connections between the processing streams at some intermediate level. This proposal seems highly unlikely, for at least the following reasons: The progressive loss of spatial information does not allow information to be passed around with the necessary resolution, such an interaction would require dense interconnections between the dorsal and ventral pathways, at a density that is not observed in practice, and the observed delay is not consistent with direct connections.

It is argued that the various cortical representations of a stimulus are bound together due to the quasi-synchronous firing of all the neurons that represent it [44,45,46] (but also see [47] for a critical evaluation of this hypothesis). The experiments presented here can not discount synchrony as an indicator of successful binding, but the observed delays clearly demonstrate that some process is needed to perform the binding before it can be signaled in such a manner.

The fact that precise spatial information needs to be recovered based on high-level perceptual decisions to accomplish the tasks presented here argues strongly for a feedback model similar to the one proposed in this paper. Also, the timing of the observed effects is consistent with information passing through 3-5 visual areas [42]. Our model is also supported by evidence that attentional modulation appears sooner in higher level areas than in lower level ones [41].

It has been shown that the extraction of visual information can occur very rapidly, and based on very brief exposures to stimuli [48,49], but the amount of information extracted is very limited. In particular, localization of stimuli is poor, and detailed recognition beyond the categoric level is severely impaired [50]. The proposed mechanism is consistent with this fast detection/slow localization and recognition dichotomy.

The idea that attention binds together distributed cortical activations at multiple levels of the visual hierarchy involved in processing attended stimuli has recently received significant experimental support [51], and reentrant connections between extrastriate areas and V1 are gaining support as the substrate for attention and conscious perception (see [52] for a review).

6 Methods

All simulations were implemented in the TarzaNN neural network simulator [53]. The simulator, instructions and all additional files needed to reproduce the results presented here are available online at <http://www.TarzaNN.org>. The simulations were performed on a Macintosh PowerMac G5. Note that the simulator will also work on other platforms, including Windows, Linux and Solaris.

References

1. Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y.H., Davis, N., Nuflo, F.: Modeling visual-attention via selective tuning. *Artif. Intell.* 78(1-2), 507–545 (1995)
2. Goodale, M.A., Milner, D.A.: Separate visual pathways for perception and action. *Trends in Neuroscience* 15(1), 20–25 (1992)
3. Roskies, A.L.: The binding problem. *Neuron* 24(1), 7–9 (1999)
4. Rosenblatt, F.: *Principles of Neurodynamics: Perceptions and the Theory of Brain Mechanisms*. Spartan Books (1961)
5. Treisman, A.M., Schmidt, H.: Illusory conjunctions in the perception of objects. *Cognit Psychol.* 14(1), 107–141 (1982)
6. Wolfe, J.M., Cave, K.R.: The psychophysical evidence review for a binding problem in human vision. *Neuron* 24(1), 11–17 (1999)
7. Schoenfeld, M.A., Tempelmann, C., Martinez, A., Hopf, J.M., Sattler, C., Heinze, H.J., Hillyard, S.A.: Dynamics of feature binding during object-selective attention. *Proceedings of the National Academy of Sciences of the United States of America* 100(20), 11806–11811 (2003)
8. Karayanidis, F., Michie, P.T.: Evidence of visual processing negativity with attention to orientation and color in central space. *Electroencephalography and Clinical Neurophysiology* 103(2), 282–297 (1997)
9. Cortese, F., Bernstein, L.J., Alain, C.: Binding visual features during high-rate serial presentation. *Neuroreport* 10(7), 1565–1570 (1999)
10. Smid, H.G.O.M., Jakob, A., Heinze, H.J.: An event-related brain potential study of visual selective attention to conjunctions of color and shape. *Psychophysiology* 36(2), 264–279 (1999)

11. Schoenfeld, M.A., Woldorff, M., Duzel, E., Scheich, H., Heinze, H.J., Mangun, G.R.: Form-from-motion: MEG evidence for time course and processing sequence. *Journal of Cognitive Neuroscience* 15(2), 157–172 (2003)
12. Tsotsos, J.K., Liu, Y., Martinez-Trujillo, J.C., Pomplun, M., Simine, E., Zhou, K.: Attending to visual motion. *Comput. Vis. Image Und.* 100(1-2), 3–40 (2005)
13. Cutzu, F., Tsotsos, J.K.: The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vision Research* 43(2), 205–219 (2003)
14. 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. *Proceedings of the National Academy of Sciences of the United States of America* 103(4), 1053–1058 (2006)
15. Treisman, A.M., Gelade, G.: Feature-integration theory of attention. *Cognitive Psychology* 12(1), 97–136 (1980)
16. Taraborelli, D.: Feature binding and object perception. Does object awareness require feature conjunction? In: *ESPP 2002. 10th Annual Meeting of the European Society for Philosophy and Psychology*, Lyon (2002)
17. Robertson, L.: *Space, Objects, Brains and Minds. Essays in Cognitive Psychology*. Psychology Press (2004)
18. Koch, C.: A theoretical analysis of the electrical properties of an x-cell in the cat's LGN: does the spine-triad circuit subserve selective visual attention? Technical report, MIT, Artificial Intelligence Laboratory (1984)
19. Sherman, S.M., Koch, C.: The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. *Experimental Brain Research* 63, 1–20 (1986)
20. Li, Z.: A saliency map in primary visual cortex. *Trends in Cognitive Sciences* 6(1), 9–16 (2002)
21. Lee, T., Yang, C., Romero, R., Mumford, D.: Neural activity in early visual cortex reflects behavioral experience and higher order perceptual saliency. *Nature Neuroscience* 5(6), 589–597 (2002)
22. Moran, J., Desimone, R.: Selective attention gates visual processing in the extrastriate cortex. *Science* 229(4715), 782–784 (1985)
23. Tsotsos, J.K.: Analyzing vision at the complexity level. *Behavioral and Brain Sciences* 13(3), 423–444 (1990)
24. Schall, J.D.: Neural basis of saccade target selection. *Reviews in Neuroscience* 6(1), 63–85 (1995)
25. Steinman, B.A., Steinman, S.B., Lehmkuhle, S.: Visual attention mechanisms show a center-surround organization. *Vision Research* 35(13), 1859–1869 (1995)
26. Caputo, G., Guerra, S.: Attentional selection by distracter suppression. *Vision Research* 38(5), 669–689 (1998)
27. Bahcall, D.O., Kowler, E.: Attentional interference at small spatial separations. *Vision Research* 39(1), 71–86 (1999)
28. Mounts, J.R., Melara, R.D.: Attentional selection of objects or features: Evidence from a modified search task. *Perception & Psychophysics* 61(2), 322–341 (1999)
29. Mounts, J.R.: Evidence for suppressive mechanisms in attentional selection: feature singletons produce inhibitory surrounds. *Perception & Psychophysics* 62(5), 969–983 (2000)

30. Vanduffel, W., Tootell, R.B.H., Orban, G.A.: Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. *Cerebral Cortex* 10(2), 109–126 (2000)
31. Kristjansson, A., Nakayama, K.: The attentional blink in space and time. *Vision Res.* 42(17), 2039–2050 (2002)
32. Slotnick, S.D., Hopfinger, J.B., Klein, S.A., Sutter, E.E.: Darkness beyond the light: attentional inhibition surrounding the classic spotlight. *Neuroreport* 13(6), 773–778 (2002)
33. Slotnick, S.D., Schwarzbach, J., Yantis, S.: Attentional inhibition of visual processing in human striate and extrastriate cortex. *Neuroimage* 19(4), 1602–1611 (2003)
34. Müller, N.G., Kleinschmidt, A.: The attentional ‘spotlight’s’ penumbra: center-surround modulation in striate cortex. *Neuroreport* 15(6), 977–980 (2004)
35. Schall, J.D.: On the role of frontal eye field in guiding attention and saccades. *Vision Research* 44(12), 1453–1467 (2004)
36. Müller, N.G., Mollenhauer, M., Rösler, A., Kleinschmidt, A.: The attentional field has a Mexican hat distribution. *Vision Research* 45(9), 1129–1137 (2005)
37. Rothenstein, A.L., Zaharescu, A., Tsotsos, J.K.: Second-order (non-Fourier) attention-based face detection. In: Kollias, S., Stafylopatis, A., Duch, W., Oja, E. (eds.) *ICANN 2006. LNCS*, vol. 4132, pp. 518–527. Springer, Heidelberg (2006)
38. Bellhumer, P.N., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(7), 711–720 (1997)
39. Rothenstein, A.L., Tsotsos, J.K.: Selective Tuning: Feature Binding Through Selective Attention. In: Kollias, S., Stafylopatis, A., Duch, W., Oja, E. (eds.) *ICANN 2006. LNCS*, vol. 4132, pp. 548–557. Springer, Heidelberg (2006)
40. Földiák, P., Young, M.: Sparse coding in the primate cortex. In: Arbib, M.A. (ed.) *The Handbook of Brain Theory and Neural Networks*, pp. 895–898. MIT Press, Cambridge (1995)
41. Mehta, A.D., Ulbert, I., Schroeder, C.E.: Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. *Cerebral Cortex* 10(4), 343–358 (2000)
42. Bullier, J.: Integrated model of visual processing. *Brain Research Reviews* 36(2-3), 96–107 (2001)
43. Barlow, H.B.: Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* 1(4), 371–394 (1972)
44. Milner, P.: A model for visual shape recognition. *Psychol. Rev.* 81, 521–535 (1974)
45. von der Malsburg, C.: Nervous structures with dynamical links. *Ber. Bunsenges. Phys. Chem.* 89, 703–710 (1985)
46. Singer, W.: Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65 (1999)
47. Shadlen, M.N., Movshon, J.A.: Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24, 67–77 (1999)
48. Biederman, I.: Perceiving real-world scenes. *Science* 177(43), 77–80 (1972)
49. Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. *Nature* 381(6582), 520–522 (1996)
50. Evans, K.K., Treisman, A.M.: Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance* 31(6), 1476–1492 (2005)

51. Haynes, J.D., Tregellas, J., Rees, G.: Attentional integration between anatomically distinct stimulus representations in early visual cortex. *Proc. Natl. Acad. Sci.* 102(41), 14925–14930 (2005)
52. Pollen, D.A.: Explicit neural representations, recursive neural networks and conscious visual perception. *Cerebral Cortex* 13(8), 807–814 (2003)
53. Rothenstein, A.L., Zaharescu, A., Tsotsos, J.K.: TarzaNN: A general purpose neural network simulator for visual attention modeling. In: Paletta, L., Tsotsos, J.K., Rome, E., Humphreys, G.W. (eds.) *WAPCV 2004*. LNCS, vol. 3368, pp. 159–167. Springer, Heidelberg (2005)