Attention in Stereo Vision: Implications for Computational Models of Attention

Neil D. B. Bruce and John K. Tsotsos

Department of Computer Science, York University, Toronto, Canada, M3J 1P3

Centre for Vision Research, York University, Toronto, Canada, M3J 1P3

ABSTRACT

The stereo correspondence problem is a topic that has been the subject of considerable research effort. What has not yet been considered is an analogue of stereo correspondence in the domain of attention. In this chapter we bring this problem to light revealing important implications for computational models of attention and in particular how these implications constrain the problem of computational modeling of attention. A model is described which addresses attention in the stereo domain, and it is revealed that a variety of behaviors observed in binocular rivalry experiments are consistent with the model's behavior. Finally, we consider how constraints imposed by stereo vision may suggest analogous constraints in other non-stereo feature domains with significant consequence to computational models of attention.

1 INTRODUCTION

An important problem faced by the primate brain, is that of understanding the 3D structure of one's environment based on the two dimensional view received by each eye. Having two slightly different perspectives of the scene allows the possibility of estimating scene structure based on small differences between the images captured by the left and right eyes. Differences between the location of points appearing in the left and right eye are referred to as stereo disparities, and the perception of depth based on such disparities is referred to as stereopsis. An important problem then becomes that of deciding which features observed in one eye correspond to features observed in the other; this is referred to in the literature as the stereo correspondence problem.

In light of the problems posed by stereo vision in the domain of attention modeling, we demonstrate that the Selective Tuning model is able to accommodate for stereo vision with no additional assumptions or requirements imposed on the model. This is in contrast to other classes of models for which we highlight the pitfalls posed by stereo correspondence in the attention domain.

An additional consideration is the relationship between binocular rivalry and attention. In recent years it has become increasingly evident that attention plays a significant role in the perceptual alternation observed when viewing a rivalry stimulus. In section 7, we reveal that when applied to stereo vision, behavior consistent with psychophysical results in this domain emerge from the Selective Tuning model (Tsotsos et al., 1995).

As a whole, the body of work demonstrates that stereo vision in itself imposes strict requirements on computational models of attention, Selective Tuning is able to accommodate for stereo vision with a variety of rivalry behaviors emerging directly from the model, and some of the problems posed by stereo vision may have non-stereo analogues.

The balance of this chapter is structured as follows: In section 2, we motivate the need for attention in biological systems, appealing to issues tied to interference among competing signals within a neural representation, and also to the inherent complexity of visual search. In section 3, we summarize evidence tied to the relationship between attention, and deployment in 3D space. In particular, we show that while this relationship is complicated, there does appear to be the basic ability to attend in depth. We further introduce a novel constraint on systems that achieve attention in depth deemed the Attentional Stereo Correspondence Problem (ASCP) in this work. In section 4, we describe a basic hierarchical computational model that realizes stereo correspondence, inspired by energy models in visual computation. This is followed in section 5 by a more generic description of a mechanism of attentional selection that acts upon the interpretive network that simulates stereo computation. Section 6 describes how the attentional computation outlined in section 5 acts upon the putative visual representation in section 4 towards allowing attention to be deployed within 3-dimensional space. This is followed in section 7 by considering some of the possible implications of such a marriage including possible explanations for observed binocular and pattern rivalry behavior. In section 8, we comment on the manner in which the ASCP constrains models of attention as a whole, and highlight certain classes of models for which this domain may pose a challenge. Finally, we close the chapter by summarizing some of the important points, and highlight the elements most relevant as a guide to computational modeling within this area.

2 THE NEED FOR ATTENTION

Attention provides a mechanism for selection of particular aspects of a scene for subsequent processing while eliminating interference from competing visual events. A common misperception is that attention and fixation are one and the same phenomenon. Attention focuses processing on a selected region of the visual field that needn't coincide with the centre of fixation. This is perhaps exemplified by the perceived ability to look out of the corner of ones eye. There exist numerous formal arguments demonstrating the necessity of attention to solve the visual search problem (Tsotsos, 1988; Burt, 1988). In lieu of exhaustively describing each of these arguments, we instead summarize some of the more important elements and comment specifically on implications in the domain of stereo vision.

A question that frequently arises with regards to attention, is that of why attention is necessary. Many arguments for attention appeal to reducing the complexity of visual search. The intention of this section is to motivate why this is not the entire story, since the issue at hand is greater in

scope than simply reducing computational complexity. One of the primary goals of attention, unrelated to complexity, concerns interference between signals generated by unrelated visual events: In a feedforward network, crossover between signals and blurring may result in a response at the output level that is highly confused.

Tsotsos examined the problem of visual search as derived from first principles (Tsotsos, 1988) within a well defined framework including images, a model base of objects and events, and an objective function that affords a metric of closeness between an image subset and an element of the model base. On the basis of this formulation, it may be shown that visual search in the general case (i.e. when no explicit target is given) is NP-complete. One conclusion that emerges on the basis of this analysis and other complexity arguments (Uhr, 1972; Burt, 1988; Anderson and Essen, 1987; Nakayama and Silverman, 1991), is that the computational complexity of vision demands a pyramidal processing architecture. Such an architecture is observed in the primate brain on the basis of increasing receptive field size and the observed connectivity between neurons as one ascends visual pathways (Palmer, 1999). Pyramidal processing may greatly reduce the computation required to accomplish a particular task by reducing the size of instances to be processed. Tsotsos et al. outline four major issues that arise in a pyramid processing architecture, all of which result in corruption of information as input flows from the earliest to later layers (Tsotsos et al., 1995). The four cases are depicted in figure 1. The pyramid depicted in the top left (Fig. 1a) demonstrates the context effect. The response of any given unit at the top of the pyramid results from input from a very large portion of the input image. As such, the response of a given unit at the top of the pyramid may result from a variety of different objects or events in the image. On the basis of this observation, it is clear that the response at the output layer with regards to a particular event depends significantly on the context of that event. The top right pyramid (Fig. 1b) demonstrates the blurring effect. A small localized event in the input layer eventually impacts on the response of a large number of neurons at the output layer. This may result in issues in localizing the source of the response at the output layer, as a localized event may be represented by a large portion of the highest layer. In the stereo domain this also implies the inability to determine which eye the winning activation derives from. The pyramid on the bottom left (Fig. 1c) displays the cross-talk effect. Cross-talk refers to the overlap of two image events in the pyramid which results in interference between signals in higher layers of the pyramid. This issue is of particular importance in the context of stereo vision since there exist many neurons in the human visual system that respond to input from the two eyes. Often the input from each eye is in agreement in which case interference is not an issue. However, in the event that the two eyes receive disparate input, an appropriate mechanism is required to resolve such interference. Finally, the pyramid on the bottom right (Fig. 1d) displays the boundary effect. Units at the outer edges connect to fewer units in higher layers of the pyramid. As a result, a significantly stronger response may result from the same stimulus centered in the visual field relative to near the boundaries. Means of overcoming this difficulty are discussed in detail in (Tsotsos et al., 1995; Culhane, 1992).

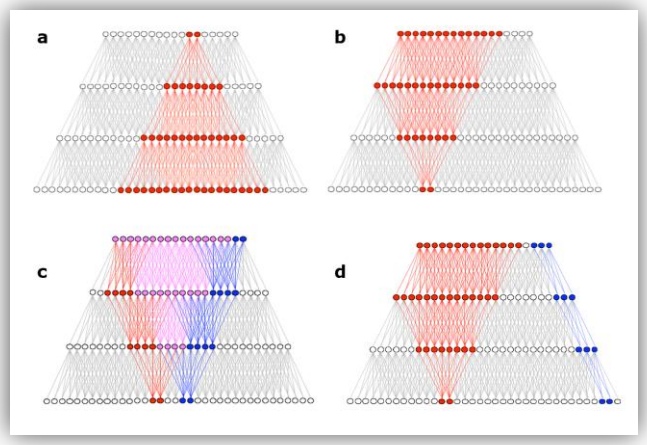

Fig. 1. Four major issues in pyramid information flow: a. The context effect, b. The blurring effect, c. The cross-talk effect, d. The boundary-effect. Adapted from (Tsotsos, 2003).

At this point, the rationale of the preceding discussion may not yet be apparent. The motive for addressing such issues is that an appropriate attentional mechanism may overcome the aforementioned interference issues inherent in pyramid processing. In particular, the Selective Tuning Model (Tsotsos et al., 1995) was designed with these issues in mind. Attenuation of appropriate connections in the network allows each of the aforesaid issues to be overcome. The exact mechanism by which such issues are handled becomes evident in the description of the Selective Tuning Model presented in section 5.

3 ATTENTION AND DEPTH

Perhaps the most crucial question in this discussion is that of whether attention operates in three-dimensional space. Discussion of attention in depth may become very convoluted owing to apparent differences between viewer centred, object centred, or action centred frames of reference in the context of allocating attention in three dimensions. The following section intentionally avoids this distinction, instead presenting evidence in favour of and against a three-dimensional focus of attention. This basic element is the sole consideration that has any bearing on the discussion presented in the sections that follow. In contrast to results on 2D deployment of attention, the literature involving 3D attention is far more contemporary. There are a great deal of conflicting results found in the current literature, but the following establishes that the proverbial spotlight of attention appears to reside in 3D space.

In a three-dimensional analogue of the Posner paradigm, Hoffman and Mueller show that a cue (brightening a dot) can produce cueing effects associated with a 3D location (Hoffman and Mueller, 1994).

In a pair of studies investigating the role of interference from distractors, it was observed that attending to a specific location defined by disparity, eliminated interference from distractors in other depth planes (Arnott and Shedden, 2000; Theeuwes et al., 1986).

Atchley et al. conducted a set of experiments in which observers were cued to one of four positions (left-right and near-far) in a stereoscopic display (Atchley et al., 1997). They found a larger reaction time in shifting attention in x,y and depth than switching attention in the frontoparallel plane alone suggesting a ”depth-aware” spotlight.

The results of Atchley and colleagues receive support from the Event-Related-Potential (ERP) study of Kasai et al (Kasai et al., 2003). Kasai et al. observed ERP responses when subjects directed attention to locations appearing on the left and right of the display and either near or far relative to fixation. A response to a target at the attended location was required. Previous findings observed that attending to a location modulates incoming sensory signals, reflected by P1 and N1 ERP components. Kasai et al. observed a greater effect of P1 amplitude for left- right selection in the near condition, and an N1 amplitude increase for the combination of location- and depth.

Viswanathan and Mingolla considered the allocation of attention in depth in a multi-element tracking paradigm (Viswanathan and Mingolla, 2002). The task required tracking a subset of 2-8 elements moving around the display. They demonstrated that depth cues improve performance in a multi-element tracking task and establish through control experiments that such improvement derives from the spatial separation in three dimensions.

Theeuwes and Pratt consider inhibition of return in the stereo domain. Their results suggest that attentional cueing happens in three-dimensional space while inhibition of return appears to spread across depth planes (Theeuwes and Pratt, 2003). The authors suggest that this result may be explained by an inhibition of return (IOR) mechanism that avoids returning to any part of a previously attended object. While this result begs additional questions about the role of IOR in stereo attention, it further supports the notion that attention operates in three-dimensional space.

Andersen and Kramer carried out an experiment involving a response compatibility task. Subjects were instructed to respond to a central target while ignoring distractors (Andersen and Kramer, 1993). Flanking distractors were presented at 7 different depths as well as 3 different horizontal and vertical distances. The findings were that response-compatibility varied in x-y directions as well as in depth. Interestingly, the effect was stronger for horizontal shifts than for vertical shifts and stronger for crossed versus uncrossed disparities.

Marrara and Moore describe a set of experiments aimed at discerning specific conditions on observing attention in depth (Marrara and Moore, 2000). They demonstrate that some previous failures to observe attention in depth relate to issues of timing. Their results as a whole strongly suggest a 3D focus of attention, for which perceptual organization is an important influence.

There also exist a handful of studies that fail to find any effect of cueing in depth (Ghirardelli and Folk, 1996; Iavecchia and Folk, 1995; Theeuwes et al., 1998). Most authors seem to attribute inconsistencies in results to either the attentional requirements of the tasks involved, or issues pertaining to frame of reference. While these failures to observe attention in depth do not invalidate the body of literature as a whole, they do highlight the apparent complexity of the mechanisms involved.

It seems fair to conclude that the bulk of the literature is in favor of a representation of attention that resides in three-dimensional space. We will demonstrate that this consideration has important implications when cast into the domain of attention modeling in the sections that follow.

The Attentional Stereo Correspondence Problem (ASCP)

In section 1, the stereo correspondence problem was briefly described. In this section, we will establish why stereo correspondence presents an additional problem in the domain of attention. Recall that the stereo correspondence problem refers to matching points appearing in the left eye to the same visual stimuli appearing in the right eye. The additional problem posed by attention, is that the locus of attention need not correspond to the locus of fixation. Therefore in addition to determining which points in the two eyes correspond for the purpose of inferring depth, one must also ascertain that the locus of attention falls on a single visual event, even in the case that this event falls on different coordinates in the left and right eye. Although the geometry of stereo correspondence may be rather involved in the general case, which may involve vertical disparities and torsional eye movements, for the purposes of exposition the discussion in this chapter considers the simpler case of horizontal disparities such that a stimulus to be attended falls on coordinates (x, y) in the left eye and (x + δ, y) in the right eye.

Consider the situation presented in figure 2. While the eyes are fixated on the background target, the focus of attention as indicated by the highlighted area falls on the closer hexahedron. This requires that the system that deploys attention holds some knowledge of which points in the left eye correspond to the same item or visual event in the right eye. If this constraint is not satisfied, one might imagine undesirable movements of the two eyes, or a signal that is comprised of interference between two unrelated items in an image as activation corresponding to the two attended items converges on binocular neurons later on in the visual system. In the later sections, we further discuss how this problem constrains the type of model that might achieve intuitively appropriate deployment of attention.

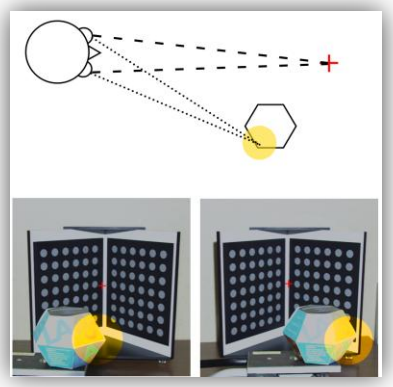

Fig. 2. An illustration of the attentional stereo correspondence problem. The two eyes fixate on the red cross located on the background target. Attention is deployed covertly to the yellow region residing on the hexahedron, which requires knowledge of stereo correspondence to deploy attention appropriately.

4 A SUFFICIENT MODEL

In this section we present a model of stereo computation in primates with specificity sufficient to demonstrate constraints on achieving attentional selection in depth. The proposed framework is based largely on the Ohzawa and colleagues model of early primate stereo vision (Ohzawa, 1998). Properties of the model include disparity selective monocular neurons at early layers, binocular neurons at higher layers, pooling across orientation, and spatial frequency, and increasing receptive field size as once ascends the visual hierarchy. All of these elements are consistent with the organization of the primate visual system and also consistent with more general properties of neural organization. The organization of the stereo framework is as follows:

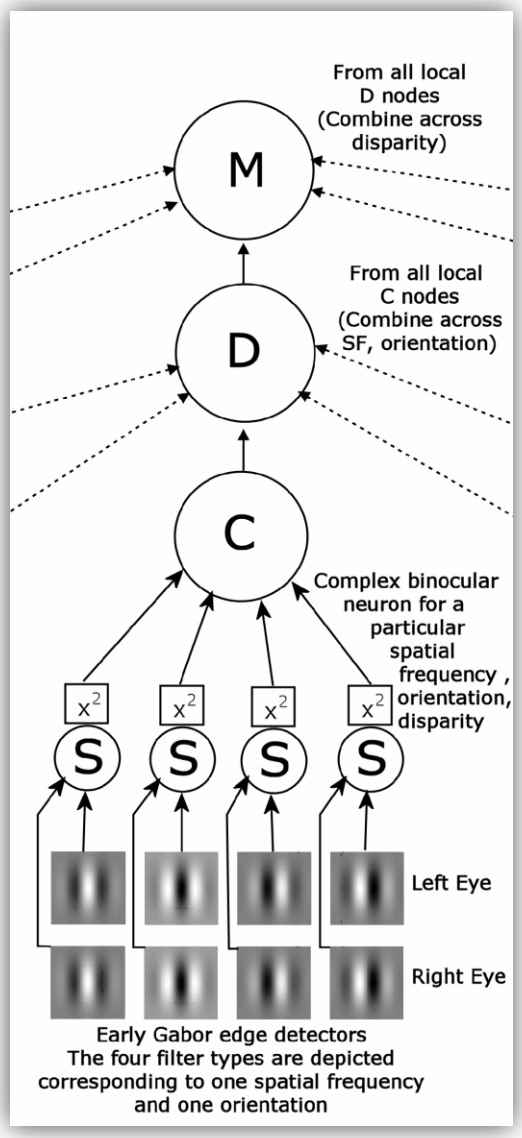

Figure 3. The computational architecture underlying disparity based computation.

1. Layer 1: Gabor maps of the form:

$$G_\beta(x, y, \theta, \sigma, \phi, \psi) = \psi \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) cos\left(2\pi \frac{x'}{\lambda} + \phi\right)$$

with $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$. $\beta \in \{l, r\}$ indicates the eye from which input to G is derived (left or right). Four different orientations were included in the implementation corresponding to $\theta \in \{0, \pi/4, \pi/2, 3\pi/4\}$. γ is fixed at 1 yielding a circular region of support for all feature maps, and feature maps include values for σ corresponding to 23 by 4 to 39 cycles per 100 pixels. $\phi \in \{0, \pi/2\}$ and $\psi \in \{-1, 1\}$ yielding 0 or $\pi/2$ radians phase positive and negative filters. These are depicted in layer 1 of figure 3. As a whole, this results in 2*4*5*2*2=160 feature maps at layer 1 corresponding to 2 eyes, 4 orientations, 5 spatial frequencies and 4 combinations of phase and sign.

2. Layer 2: Binocular simple cells tuned to various disparities (δ = 0 to 12 pixels in increments of 2) are derived by summing the output of a Gabor filter at a particular spatial frequency, and orientation acting on each of the left and right eye input, and shifted by the degree of disparity δ:

$$s_\delta(x, y, \theta, \sigma, \phi, \psi) = G_l(x, y, \theta, \sigma, \phi, \psi) + G_r(x + \delta, y, \theta, \sigma, \phi, \psi)$$

This gives rise to 560 feature maps (160*7/2) corresponding to 7 disparities δ for each feature set in layer 1 and combined across the two eyes. Binocular simple cells of this type are found among early visual areas, and often involve a differential contribution from each of the two eyes according to variation in ocular dominance along neural columns. In the context of our implementation, including this consideration would give rise to an appreciably larger number of feature maps and no generality is lost in our argument in assuming equal weighted inputs from each eye.

3. Layer 3: Complex binocular cells are produced by summing the squared output of 4 simple binocular neurons corresponding to the 4 different Gabor filter types for a particular orientation, spatial frequency and disparity.

$$C_\delta(x, y, \theta, \sigma) = \sum_{\phi \in \{0, \pi/2\}} \sum_{\psi \in \{-1, 1\}} s_\delta(x, y, \theta, \sigma, \phi, \psi)^2$$

The choice of this operation is biologically motivated and means that output is not sensitive to contrast polarity, and disparity sensitivity is constant for all stimulus positions in the receptive field. The output of the above operation to compute C_δ is convolved with a gaussian with σ=5 pixels to simulate the pooling of simple cell responses by complex cells as in (Chen and Qian, 2004).

4. Layer 4: Responses are combined across orientation, and spatial frequency giving rise to 9 feature maps corresponding to the 9 disparities considered.

$$D_\delta = \sum_\theta \sum_\alpha C(\theta, \alpha)$$

This operation is suggested in the stereo model of Fleet et al. (Fleet et al., 1996) which is also inspired by the energy model of Ohzawa (Ohzawa, 1998). Combining across orientation and spatial frequency reduces false peaks inherent in narrow band signals. This operation is also appropriate since most models of stereo vision will presumably rely on some form of pooling at a later stage of processing.

5. Layer 5: The 7 disparity maps are averaged to produce a single representation of

disparity related activity:

$$M = \sum_{\delta} D_{\delta}^{2}$$

M is convolved with a Gaussian of $\sigma = 5$ pixels to produce a smooth master activation map at the highest layer.

The overall architecture is depicted in figure 3.

5 THE SELECTIVE TUNING (ST) MODEL

In this paper, selective attention in depth is achieved in the context of the Selective Tuning model. In the later discussion, it should become clear that this model is consistent with a broad range of psychophysical results pertaining to attention in depth, while other efforts encounter difficulties when tested on the same conditions.

Many design choices in the Selective Tuning Model are formed on the basis of overcoming the issues of complexity and problems inherent in pyramid processing discussed earlier. Selective Tuning simultaneously handles the issues of spatial selection of relevant stimulus and features. Spatial selection is accomplished by way of inhibition of appropriate connections in the network. Feature selection is accomplished through bias units which allow inhibition of responses to irrelevant features. The Selective Tuning Model is characterized by a multi-scale pyramid architecture with feedforward and feedback connections between units of each layer. A high level schematic of the model is depicted in figure 4. Details concerning the connectivity between adjacent layers are displayed in figure 5.

Variables shown in figure 5 are as follows (Also refer to (Tsotsos et al., 1995) for a more detailed description):

- $\hat{I}_{l,k}$: interpretive unit in layer l and assembly k
- $\hat{G}_l, k, j$: jth gating unit in the Winner-Take-All (WTA) network in layer l, assembly k which links $\hat{I}_{l,k}$ to $\hat{I}_{l-1,j}$
- $\hat{g}_{l,k}$: gating control unit for the WTA over inputs to $\hat{I}_{l,k}$
- $\hat{b}_{l,k}$: bias unit for $\hat{I}_{l,k}$
- $q_{l.j.i}$: weight corresponding to $\hat{I}_{l-1,i}$ in computing $\hat{I}_{l,j}$
- $n_{l,x}$: scale normalization factor
- $M_{l,k}$: set of gating units corresponding to $\hat{I}_{l,k}$

Selection is accomplished through two traversals of the pyramid. First, the responses of interpretive units are computed from the lowest level to the highest level of the pyramid in a feedforward manner. Next, WTA competition takes place between all units at the highest layer to select a single winning unit. In subsequent layers, units in layer *l* that connect to the winning unit in layer *l +1* compete for selection. This ultimately leads to selection of a localized response in the input layer. Note that interference between competing elements is eliminated by way of selection. Bias is handled through a connected network of bias units that impact on the response of interpretive units they are tied to in a multiplicative manner. Bias units may be used to modify the response of interpretive units that correspond to a particular stimulus type, such as

bright items, horizontal edges, or blue objects. Bias values less than one might be assigned to the response of non-blue units to bias selection in favor of blue pixels. The exact circuitry for initiating bias is left unspecified in this implementation except to assume that there is some circuitry that allows the response of units of any type and at any layer to be modulated. The WTA process employed in Selective Tuning differs from that of Koch and Ullman (Koch and Ullman, 1985) in a number of aspects. The effect of unit i in the WTA network on unit j is quantified by the following expression:

$$y = \begin{cases} q_{l,k,i}G_{l,k,i}^{t-1} - q_{l,k,j}G_{l,k,j}^{t-1}, if & 0 < \zeta < q_{l,k,i}G_{l,k,i}^{t-1} - q_{l,k,j}G_{l,k,j}^{t-1} \\ 0, & otherwise \end{cases} \quad (1)$$

with $\zeta = \frac{Z}{2^{\gamma}+1}$, γ a parameter that controls the convergence rate of the WTA network (converges within γ iterations) and $G_{l,k,j}^{t_0} = b_{l-1}n_{l-1}I_{l-1,j}$. The choice of this particular scheme is tied to provable guarantees associated with convergence, and the desire to preserve the nature of the winning signal. A more detailed version of the preceding description concerning the WTA scheme, and in particular parameters tied to the interpretive network, may be found in (Tsotsos et al., 1995). Appropriate deployment of attention in 3D space is achieved via Selective Tuning combined with the interpretive network comprised of the basic biologically inspired model of primate binocular vision described in section 4.

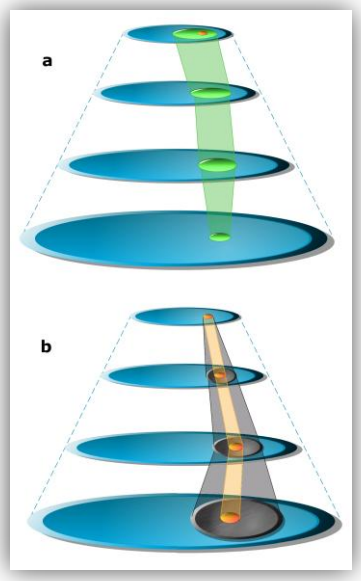

Fig. 4. A high-level schematic of the selective tuning model. a. Bottom up feedforward computation. Stimulus at the input level (green) causes a spread in activity in successively higher layers. Winner selected at the highest layer is shown by the orange oval. b. Top-down WTA selection. WTA selection happens in a top-down manner with the winning unit at each level indicated by the orange region. A suppressive annulus around the attended item caused by inhibition of connections is depicted by the grey region.

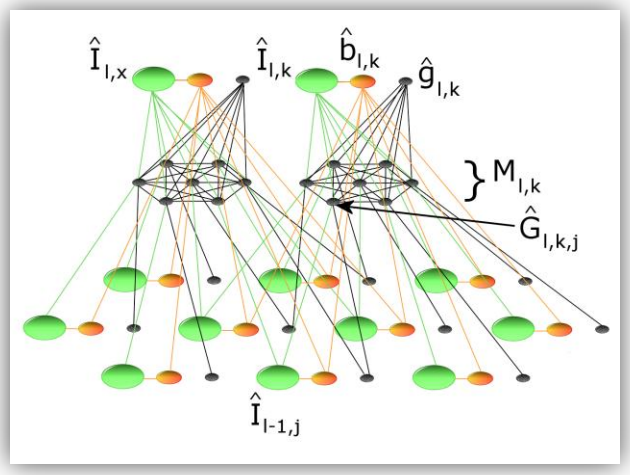

Fig. 5. A detailed depiction of connectivity between units and layers in the Selective Tuning model. (From (Tsotsos et al., 1995)).

6 SHIFTING ATTENTION IN DEPTH

It is straightforward to describe the manner in which selection of appropriate units in each eye is achieved given the principles of Selective Tuning, and the computational circuitry involved in the implementation presented here. Figure 6 demonstrates 4 distinct stages important for demonstrating how appropriate attentional selection is achieved.

In figure 6 i. A vertically oriented bar appears in front of the background surface. The receptive fields of a unit at a single position responding to vertical and horizontal stimuli respectively is shown for stimuli appearing in the left eye for position (x, y), and corresponding to three positions (x − δ, y), (x, y), (x + δ, y) in the right eye. The network hierarchy above this is shown in a manner similar to that of figure 3. Each of the lowercase letters s/c, d, m are used to emphasize that these are merely single neurons of thousands in the feature maps. It should be noted that only 2 of the 4 orientations are shown, and only one of the 4 filter types (0 phase positive) is shown. The computation which takes place from the inputs to the binocular simple cells, to the output of the binocular complex cells is compressed into a single operation shown as (s/c) for the purpose of exposition.

In figure 6 ii., activation flows upwards through the hierarchy. The saturation of green indicates the intensity of firing of the neurons involved. Note that some of the binocular cells are weakly activated by a stimulus appearing in only one of the two eyes, but the strongest response derives from a true correspondence.

In figure 6 iii., winner-take-all competition selects the winning units at the highest layer which includes the single neuron belonging to layer M shown in the diagram. Subsequently, WTA competition is initiated at the next layer down for units that contribute to the winner selected at layer M. This process propagates down the pyramid with flow along connections that do not contribute appreciably to the winning response attenuated (selected units appear as orange).

The cascade of WTA activation is shown up to the point where monocular units from the two eyes converge on a binocular unit.

In figure 6 iv., a variety of outcomes may arise at this stage. In most cases, the winning binocular unit will elicit a strong response which derives from a stimulus of appropriate disparity appearing in both eyes. This is the case in the example shown, and the focus of attention is directed to region (x, y) in the left eye, and (x + δ, y) in the right eye. An alternative possibility is that the winning binocular response derives from a very strong stimulus appearing in only one of the two eyes. This case is considered in later sections.

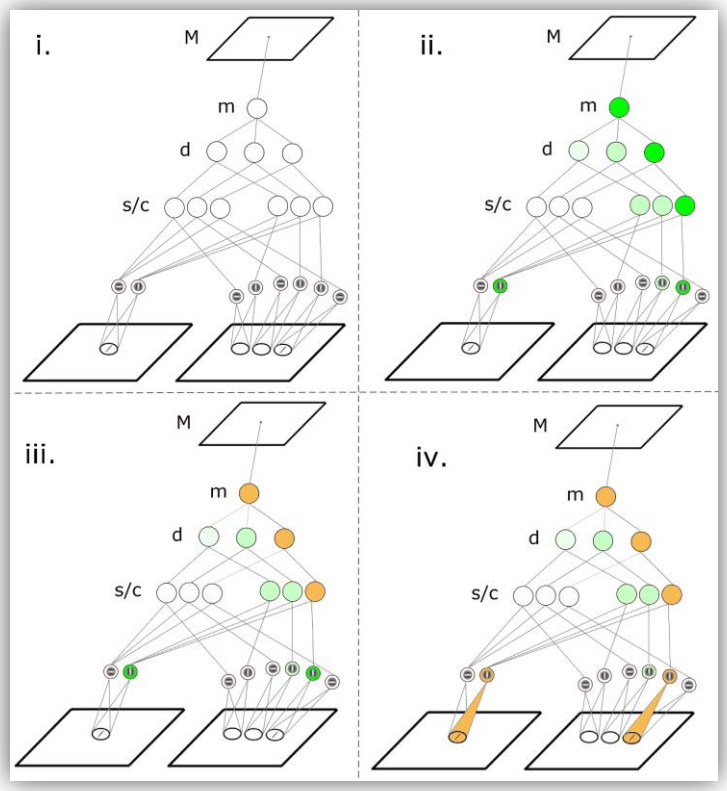

Fig. 6. Four stages in the feedforward interpretive computation, and feedback selection that achieves appropriate deployment of attention in depth. Refer to the text for further details.

At this stage, we have shown that a selection mechanism that acts on the same interpretive network that realizes stereo correspondence is a sufficient condition on achieving appropriate deployment of attention in three dimensions. The simulation results produced by the model also demonstrate that top-down selection on such a hierarchy provides sufficient conditions for appropriate attentional selection. In figure 7: Top row: A pair of random dot stereograms is presented to the two eyes (right eye on left for cross fusing). Middle Row: The resulting selected regions in the right and left eye respectively. Bottom row: Ground truth and selected regions superimposed on ground truth, white corresponds to 12 pixels and black to 0.

Figure 8 demonstrates the result of two successive runs of Selective Tuning. Top Row: The random dot stereogram on which results in figure 8 is based. Second Row: Bias is initiated in favor of near disparity with the bias values corresponding to disparities of 0,2,4,6,8,10, and 12 pixels given by multiplicative modulation by the square root of 0, 1, 2, 4, 8, 16 and 32 respectively. Third Row: Bias is initiated in favor of far disparity (non-zero) with bias values of the square root of 0, 32, 16, 8, 4, and 2 respectively. Fourth Row: The ground truth disparity map is shown for comparison.

It is perhaps worth emphasizing again that fixation and attention are not the same. This point is sometimes overlooked in work that deals with 2D attention because changes in image correspond to changes in fixation. In human vision, there is variable spatial resolution and so the distinction is important. In three-dimensions, the problem becomes even larger as fixation introduces an additional dimension of complexity corresponding to the additional geometric constraints.

This section established that Selective Tuning is capable of attentional deployment in three dimensions. The implementation results are included largely for proof of concept from a practical standpoint. In the sections that follow, less common cases than the standard stereo correspondence problem are considered, with implications of such cases discussed.

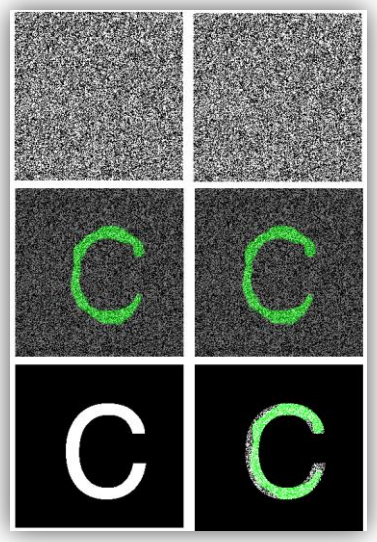

Fig. 7. Selection results based on a letter C that appears in front of the background. (From left to right) Top row: View given as input to right eye and left eye respectively. A central "C"-shaped region containing random dots is offset horizontally from one eye to the other, yielding a stereo view when each image is presented to one of the two eyes. This can also be observed using the cross-eyed, or parallel viewing techniques for stereo image pairs. 2nd row: Region selected by algorithm for right and left eyes respectively superimposed on original image at half contrast. Bottom row: Ground truth, and selection superimposed on ground truth aligned with right eye view.

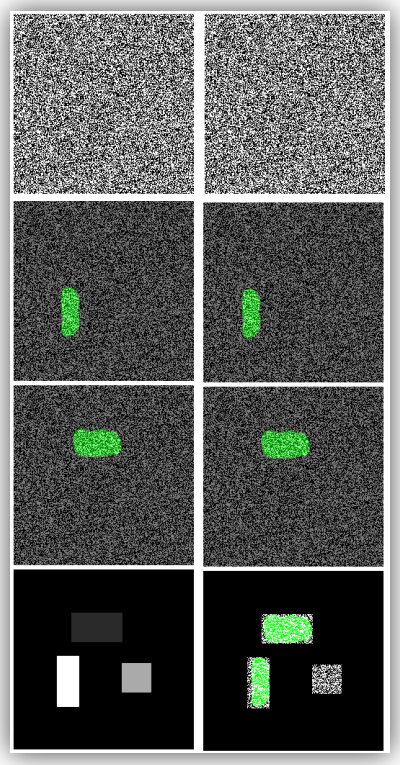

Fig. 8. Selection results for a random dot stereogram containing three boxes at different depths. (From left to right) Top row: View given as input to right and left eyes respectively. Three box shaped regions containing random dots are offset by a different horizontal extent yielding a stereo view when each image is presented to one of the two eyes. This can also be observed using the cross-eyed, or parallel viewing techniques for stereo image pairs. from 2nd row: Regions selected for right and left eyes in the attend near condition. 3rd row: Regions selected for right and left eyes in the attend far condition. Bottom row: Ground truth and selected regions superimposed on ground truth aligned with right eye view.

7 ATTENTION AND BINOCULAR RIVALRY

Binocular rivalry is a phenomenon that occurs when the two eyes are presented with very different images. The resulting percept involves one of the images appearing for a brief period of time, then the other, then the first and so on. The following section considers the role of attention in binocular rivalry with specific reference to the model we have described. Attention has recently been shown to play an important role in determining the nature of perceptual alternation observed during binocular rivalry. The following section establishes that: i. Perceptual alternation of the kind observed during binocular rivalry is predicted by the model proposed in section 4. ii. The behavior is consistent with a wide array of psychophysical data with regard to the role of top-down attention and saliency in the perceptual alternation that is observed. iii. The psychophysical literature is suggestive of a hierarchical structure underlying binocular rivalry, consistent with the representation assumed by the model. As a whole, this section provides insight into the computational architecture underlying binocular rivalry and

additionally, further establishes the plausibility of Selective Tuning as an accurate account of attentional function in primates.

Rivalry and Winner-take-all Competition

One existing effort at modeling binocular rivalry is complementary to our proposal (Wilson, 1999). In (Wilson, 1999) it is revealed that perceptual reversals of the form observed in binocular rivalry paradigms may be achieved via the combination of WTA behavior of the form assumed by ST in combination with neural decay in the form of spike frequency adaptation. Selection of one of the rivalry patterns causes spike frequency adaptation to the extent that WTA competition eventually selects the competing rivalry stimulus once sufficient adaptation has transpired. A hypothetical example of this appears in figure 9: Feedforward activation leads to selection of a unit at the highest layer of the processing hierarchy by ST. Subsequently, a cascade of winner-take-all competitions proceeds downward from the highest layer to the lowest, selecting units that contribute strongly to observed activation, while attenuating pathways that do not contribute appreciably to the observed activation at each layer. At the level of binocular complex cells, one of the alternatives among the 2 competing stimuli is selected, with the input from the competing stimulus suppressed by virtue of gating. Subsequently, following a period of adaptation, this cascade of competition switches to selecting the alternative rivalrous stimulus. Both are not selected simultaneously owing to the fact that there are no units higher up in the pyramid that respond to orthogonal orientations in the two eyes (i.e. a horizontal stimulus in one eye along with a vertical stimulus in the other).

Additionally, units driven by different patterns exert mutually inhibitory influence, while collinear facilitation is driven by recurrent excitation among units attuned to the same pattern type. Such a configuration is easily realized within the proposed model by adding appropriate local lateral inhibitory connections between neurons that respond to different features, and excitatory connections within feature. The effect of such connections, is nonlinear wave propagation resulting in a percept that changes in sweeping waves from one rivalry stimulus to the other. It is important to note that at the core of this behavior, is the presence of WTA competition on units at multiple levels of a hierarchical visual processing framework.

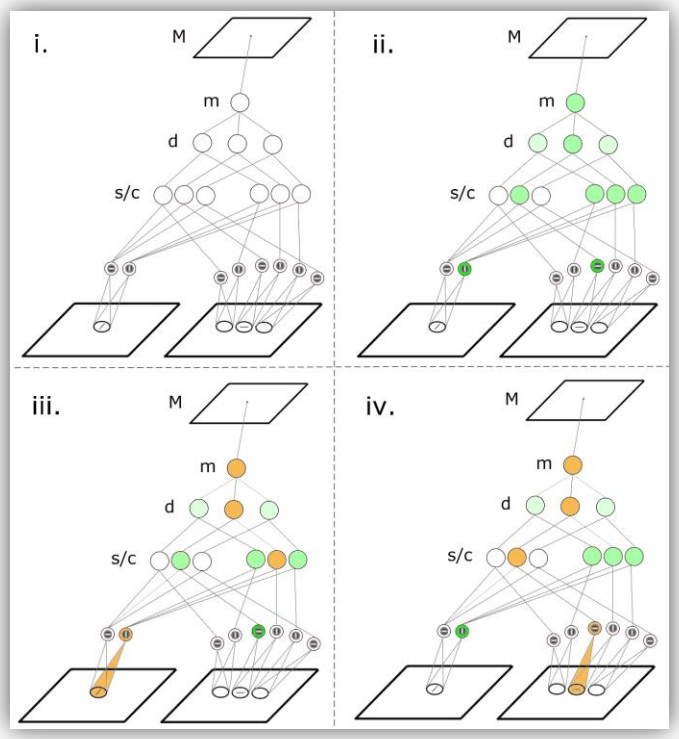

Fig. 9. A demonstration of network behavior for the rivalry stimuli. Four stages in the feedforward interpretive computation, and feedback selection that achieves appropriate deployment of attention in depth. Refer to the text for further details.

The role of saliency

It has long been debated whether rivalry happens by virtue of direct interocular competition, or competition between high-level pattern representations. The latter would be suggestive of a greater role of attention in determining dominance. A natural question to begin with is: What is the role of saliency in determining pattern dominance? One might expect that if a pop-out cue accompanies the rivalry stimulus appearing in one eye, this may bias the dominant image in favor of the eye containing the popout stimulus.

There are a few studies that shed some light on this consideration. Kanai et al. investigated the effect of visual transients on four types of perceptual alternation, including binocular rivalry (Kanai et al., 2005). A variety of experiments were performed each of which included a bistable stimulus followed by a flash. In each case, the influence of the flash on perceptual dominance was observed. As a whole the experiments establish that a flash may trigger perceptual reversals for all cases provided the flash is in the vicinity of the bistable stimulus.

Ooi and He carried out a study involving rivalry stimuli, with the image appearing in one of the eyes accompanied by a pop-out cue in some cases (Ooi and He, 1999). They found that there is a greater likelihood of a rivalry stimulus becoming dominant in the event that it is accompanied by a pop-out cue provided the cue is proximal to the rivalry stimulus.

The main conclusion that may be drawn from the above discussion is that it appears possible to selectively bias bistable perception in favor of one eye via exogenous cues provided such cues are proximal to the rivalry stimulus appearing in one eye. It is interesting to note that the aforementioned considerations agree with the expected behaviour of ST as is established in the example that follows.

A pop-out or transient cue in the vicinity of the rivalry stimulus in one eye tends to result in the dominance of that eye. Intuitively, we would expect that the cue will result in an increase in activation associated with stimulus appearing in one of the two eyes increasing its likelihood of becoming dominant. With respect to the behavior of the model, a salient cue proximal to the rivalry stimulus in one eye will elicit a strong response from neurons within the vicinity of the salient cue in one of the two eyes. Subsequent WTA selection is likely to then select the stimulus in this eye, which should then trigger nonlinear wave propagation associated with the stimulus properties appearing in the target eye via inter-feature inhibition and intra-feature facilitation (e.g. collinear facilitation for simple edges and inhibition among orthogonal orientations).

Top-down influences on bistable perception

The preceding section establishes that there exist bottom-up influences on dominance in binocular rivalry. Another important consideration is the extent to which top-down influences play a role in binocular rivalry. In this case, it is perhaps worthwhile to consider predictions of the model in this context. One might posit that as it is possible to attend to a particular feature type (e.g. a red circle) and effectively modulate the response of neurons associated with this feature, the same kind of volitional control should allow the extension of the time taken for a dominant image to be suppressed as activation associated with the dominant image decays. One might also expect that selective attention to the suppressed stimulus might lower the required threshold sufficiently to cause a perceptual reversal.

The psychophysical literature concerning volitional control in binocular rivalry is largely in agreement with the aforementioned prediction. Mitchell et al. carried out an experiment in which attention was directed to one of two overlapping transparent surfaces and subsequently the image of one surface was deleted from each eye (Mitchell et al., 2004). In most cases, only the cued surface was perceived following deletion.

Meng and Tong presented a rivalry stimulus consisting of a face in one eye and a house in the other (Meng and Tong, 2004). They found a significant effect on time of exclusive dominance of the face or house when attention was cued to the face or the house respectively.

Ooi and He employed a covert attention cueing paradigm to consider the effect of voluntary attention to a rivalry stimulus (Ooi and He, 1999). They found that voluntary attention to a grating appearing in one eye decreased the incidence of suppression of the grating by a moving target appearing in the other eye. This was only true if the attended patch was in the proximity of the moving target. In cases where the moving target passed over a non-attended patch, the patch was suppressed with the same incidence as the control case. The conclusion drawn by the authors is that voluntary attention may extend perception of the dominant image in a rivalry

stimulus. An important companion question is that of whether the dominant percept may be suppressed by selectively attending to the non-dominant stimulus. Ooi and He claim that this is probably not the case as attention is directed to the grating in cases that it is suppressed and should revert to dominance if such reversal is possible. This explanation is hardly satisfactory since their paradigm precludes considering this possibility on account of the short time course involved.

The claim of Ooi and He concerning perceptual reversal is in conflict with the predicted behavior of our model. As noted, one would expect that lowering the threshold associated with the suppressed stimulus might allow it to regain dominance. It is worth noting that the model predicts that this result should be more difficult to observe than that of whether or not it is possible to extend the period of the dominant percept. Since driving the neurons associated with the dominant percept or attenuating the response of the non-dominant units is guaranteed to extend the period of dominance assuming decay of the signal, the result that Ooi and He have demonstrated should be observed without fail. However, there is no guarantee that voluntary attention may allow the attenuation of the dominant percept to the extent that the suppressed stimulus regains dominance. That said, we predict there should be instances where voluntary attention expedites perceptual reversal such that a suppressed stimulus regains dominance. Verification of this claim would prove invaluable in validating the model, and might be realized in an experimental design that looks specifically at triggering perceptual reversal and allows a longer time course for such reversal to occur.

Hierarchical organization and dynamics

The preceding sections assume the notion of competition within a hierarchical neural framework. There are a variety of issues related to this assumption as follows: 1. To what extent is there support for top down hierarchical cortical competition as opposed to direct competition of inputs at a low level? 2. What implications does a hierarchical organization hold with respect to binocular rivalry and other forms of pattern rivalry? 3. Do imaging studies support hierarchical competition in binocular rivalry? 4. What are the dynamics associated with binocular rivalry and are such dynamics consistent with the proposed model?

It is reasonable to assume that competition associated with rivalry might take place prior to the convergence of visual pathways that are purely monocular.

There are a variety of models that make this assumption, most of which assume competition at the level of the LGN or within layer IV of V1 (Matsuoka, 1984; Lehky and Maunsell, 1996; Blake, 1989; Mueller, 1990).

More recent evidence suggests that rivalry has no impact on LGN neurons (Lehky and Maunsell, 1996). Further, there now exist some rather convincing arguments against such simple competitive models and in favor of a hierarchical configuration. In one of the more prominent studies, Logothetis and colleagues performed single cell recordings in trained monkeys while the rivalry stimulus was rapidly swapped between the two eyes (Logothetis et al.,

1996). Their results reveal a number of surprising results: i. The majority of neurons whose activity correlates with perceptual alternation are binocular and are found in higher visual areas. ii. Perceptual alternation in this paradigm exhibits dynamics identical to the static (non-swapping) stimulus case under certain conditions. iii. Competition reflects a high-level representation of the stimulus and is eye independent. iv. The perceptual alternation observed in binocular rivalry is much closer to other forms of bistable perception than previously thought.

An even stronger result appears in (Wilson et al., 2001) demonstrating that a hierarchical model consisting of at least two layers is necessary to account for rivalry data. Wilson presents a two-stage model in which orthogonal gratings compete via strong reciprocal inhibition at both monocular and binocular levels of processing.

There also exists rivalry of a slightly different nature termed pattern rivalry (or monocular rivalry). Pattern rivalry refers to the wavering percept which results from superimposing two transparent patterns and results in periods of exclusive visibility of each pattern. Maier et al. demonstrate that multistable perception in pattern rivalry appears to be driven by a global holistic interpretation of the stimuli (Maier et al., 2005). Typically when a stimulus of sufficient contrast appears in one eye, with no superimposed stimulus in the other eye, the monocular stimulus is perceived with little or no suppression. Such a configuration is oft referred to as a rivalry-free region. Maier et al. demonstrated that the stimulus in a rivalry-free region may be suppressed on the basis of properties shared with the more global surround. This result indicates that competition relies on resolution of a more global representation of the scene rather than merely local spatial or ocular conflict.

The above discussion supports a variety of important considerations from a modeling perspective: i. Rivalry exists within a hierarchy and is present at several levels of representation. ii. There is a significant role of saliency in determining perceptual dominance. iii. Primates may exert some degree of volitional control on the rate of perceptual alternation. iv. There is significant similarity between binocular rivalry and other forms of perceptual rivalry.

This type of configuration may also account for the behavior described by Maier et al. Global selection of a pattern that is distinct from the pattern constituting the rivalry-free region might prompt its suppression via lateral inhibitory interaction.

Discussion

Binocular rivalry is a useful domain to consider owing to its apparent ties to attention, allowing the relationship between attention and binocular vision to be examined. It is interesting to note that Selective Tuning in conjunction with the simplistic model of primate binocular vision is able to account for a wide array of behavior associated with binocular rivalry.

Overall, the discussion included in this section provides a strong case for selection in the binocular domain within a hierarchical configuration of the form that Selective Tuning assumes. A variety of aspects of the model are in strong agreement with the psychophysical literature including competition within a hierarchy, competition at every level of the hierarchy, a role of attention (bottom-up and top-down) in all forms of pattern rivalry, and with more global

competition mediating local suppression. Further, the dynamics associated with rivalrous perception may be produced through the addition of simple lateral connections. Each of these elements provides a case not only for Selective Tuning, but additionally implies a set of constraints on how attentive behavior is achieved. What is not yet apparent is how such behavior might be achieved in the context of other classes of models. This consideration is dealt with in some detail in the section that follows.

8 ATTENTIONAL STEREO CORRESPONDENCE REVISITED

In the preceding section, the discussion of binocular rivalry in the context of attention alluded to a variety of issues that may deny explanation by particular types of attention models. In this section, we further explore this notion by considering conditions on achieving appropriate attentive behavior under the assumption of binocularity. For the sake of discussion, let us begin by assuming that the sole constraint that stereo attention need satisfy is the selection of a region, feature, or object that corresponds to a true stereo correspondence. We will refer to this as the attentional stereo correspondence problem (ASCP).

The ASCP evidently poses a challenge for computational models of attention. The Selective Tuning model might be described as a model that includes Top-Down selection on a feedforward interpretive network. Other models fall predominantly in one of two categories: i. Saliency Based models, a category of models that derive from Treisman's feature integration theory. (Treisman and Gelade, 1980) ii. Models with feedforward selection mechanisms wherein flow through the network is gated as it proceeds up the feedforward interpretive network. Each of these classes encounters problems when faced with the problem of stereo vision and require further thought in light of the observations presented in this paper.

Let us first consider implications of stereo vision for models based on the notion of a saliency map; a ubiquitous element of attention models in the literature (Treisman and Gelade, 1980; Koch and Ullman, 1985; Sandon, 1989; Wolfe and Cave, 1989; Mozer, 1991; Itti et al., 1998; Bruce and Tsotsos, 2006, 2009). There appears to be inherent limitations to such a representation which preclude the ability to localize a corresponding region, feature or object in the binocular domain.

Consider the saliency map in its current form. A variety of feature maps are derived from the retinal input, and subsequently converge on a single unique topographical representation of saliency. A selection mechanism then acts on this representation to select an attended location. The problem with such a representation is that the saliency map retains no memory of what gave rise to the observed activation. Although the selection mechanism knows where to attend, it has no knowledge of what is being attended.

The most obvious issue with this kind of representation is that each eye has a slightly different view, and a feature may fall on different retinal coordinates in each eye. Consequently, binocular attention could require the selection of two locations from a single saliency map. Since no knowledge of what gives rise to the resulting activation is maintained, there is no hope of selecting a true stereo correspondence on this basis. It would in principle be possible to have disparity selective neurons project onto the saliency map with attention acting in a cyclopean

reference frame. This however is in disagreement with the psychophysical data summarized in section 3.

A second possibility that proponents of a saliency-based architecture might suggest is that each eye may have its own saliency map. This claim is also inherently flawed in that it would require solving the correspondence problem on a representation devoid of information concerning local structures or features. Although it is conceivable that matching could proceed on the basis of the saliency landscapes, there are also additional complications inherent in making this assumption. For example, there exist neurons tuned to near, far and at fixation binocular disparity. Amalgamating the representation carried by these different units into a saliency map would imply blurring in depth. It would seem that such a configuration would be prone to errors in attending in 3D. As a whole, a computational saliency based mechanism based on a pair of saliency maps invites numerous problems when attending in depth is considered.

As a whole, the issue at hand is that a saliency map does not retain the information required to solve the stereo correspondence problem. In the section that follows, we consider the possibility that this is simply a special case of a more general problem. The issue of what information is available, and when, is important in understanding attention. In the section that follows, we explore how such a consideration constrains computational models of attention.

On the basis of the basic problem posed by ASCP, one might also challenge the possibility that selection proceeds as activation ascends a feedforward hierarchical interpretive network. In this case, selection at the monocular level would occur first, and may preclude attending to true correspondences. That said, the section that follows considers the possibility that the ASCP simply serves to highlight a more general problem in the domain of attentional selection which raises questions for models that assume feedforward interpretive computation.

A more general problem?

The preceding section is suggestive of a more general problem with how saliency based models operate. In the case of stereo vision, it is evident that a saliency map lacks the information necessary to attend to a true stereo correspondence. This case is especially convincing since it relies on spatial coordinates but hints at a much more important issue: that of how non-spatial content is selected. A saliency map includes no memory of which features gave rise to the winning activation in the same manner that the information necessary to establish stereo correspondence is lacking. A consequence of this consideration is that selection in space must precede selection of features in an exogenous cueing paradigm for any model based on a saliency map.

This last consideration runs contrary to the experimental literature on the subject. In perhaps the strongest result in opposition to the possibility that spatial attention may precede feature based attention, Hopf et al. demonstrate that attention to features precedes attention to locations (Hopf et al., 2004) in an ERP/ERMF paradigm. In the experiment of Hopf et al. a red and green C were presented to left and right visual field with the position of each determined randomly. In half the blocks the subjects responded to the red C and in the other half, responded to the green C. The response required indicating the orientation of the target C. The red and green C's

appearing in each visual field were both flanked by six distracting blue C's on all trials. Four conditions included distractors in both visual fields sharing the same orientation as the target C, distractors in neither visual field sharing the same orientation as the target C, or target compatible C's appearing in only the target or non-target hemifield with non-compatible oriented C's appearing in the non-target or target hemifield. Their chief finding was that modulation occurred primarily within the hemisphere contralateral to distractors sharing the same orientation as the target, independent of the hemifield in which the target letter appeared. Approximately 30 ms following such modulation, the neural response reflecting focused attention on the target began (N2pc component). This strongly suggests that knowledge of the target properties is present prior to localization, a consideration strongly inconsistent with saliency based models. It is also shown in (Grill-Spector and Kanwisher, 2005) that the categorization of a presented object happens with a time course at least as short as detection of the object.

These studies raise questions for models that assume feedforward WTA selection. Given a feedforward WTA configuration, one might expect activity corre-sponding to localization to precede or at least coincide with activity pertaining to target features. This is a consideration inconsistent with the experimental literature.

Rivalry revisited

In addition to the issues highlighted earlier in this section, saliency based models hold little hope for explaining any of the considerations pertaining to binocular rivalry as discussed in section 7. Consideration in regard to specific issues would require making brazen assumptions on how saliency maps might be extended to handle binocular vision.

There is however one fundamental issue which is worth discussing. Rivalry is clearly cortical and not eye-based. Rivalry happens at every layer of the visual cortex. It is distributed and with global competition directing local suppression (Maier et al., 2005). These considerations paint a picture that bears little if any connection to a model based on a unique topographical feature blind saliency map.

With respect to feedforward WTA models, the assumption of feedforward gating is inconsistent with experiments indicating that perceptual alternation observed during binocular rivalry appears to be based on a global interpretation of the scene which presumably requires first the involvement of higher visual areas. This consideration holds even stronger when one considers that the same argument may be made of pattern rivalry in general. Monocular rivalry may be observed when pairs of transparent stimuli are presented overlapping, with temporal windows of exclusive visibility corresponding to one pattern or the other. Such alternation has little to do with local competition among competing stimuli, but rather the entire scene alternates on the basis of global competition (Maier et al., 2005). Attention is undeniably involved in the perceptual alternation observed during monocular rivalry and should provide an additional constraint on attention modeling at large.

9 GENERAL DISCUSSION

The problem of visual attention within a stereo framework has received relatively little consideration. This is especially true on the modeling side with few efforts considering how stereo vision informs the problem of visual attention. Previous computational efforts that give any consideration to binocularity generally treat stereo as a feature that contributes to some representation of global saliency in a cyclopean frame of reference (See. (Itti and Koch, 2001)). As we have pointed out, this is a consideration that runs contrary to contemporary psychophysics literature.

In this paper, we argue that the ASCP is an important consideration for attention models to address, and highlighted difficulties for certain classes of models when faced with the ASCP. This hints at some potential problems that are more general, and in particular, the necessity to consider what kinds of information are available at different levels of visual processing including the relative order of distractor suppression, detection, localization and recognition. A few recent imaging studies shed some light on some of these details, and provide greater insight on the order and timing of these various attention related events.

We also argue, that the problem of binocular rivalry is one that cannot be disentangled from binocular vision, and the presence of rivalry behaviors that may be explained by a model based on top-down gating is encouraging. The last ten years has provided mounting evidence that attention plays an important role in binocular rivalry and also pattern rivalry, and one might argue that these are serious considerations for any computational proposal relating to attention to address. The fact that a global interpretation of a scene appears to factor appreciably in the type of modulation that takes place makes the distinction between models that assume feedforward gating versus top-down gating an important one.

It has been stressed that an important role of attention is in resolving interference within a hierarchical network. Resolution of cross-talk, and precise localization are important elements of an attention model, and we have established additional situations in the domain of stereo vision for which these issues are important.

We have put forth a proposal for attentional selection that appears to afford intuitively appropriate behavior with respect to simple selection in depth and various forms of rivalry. The significant agreement between Selective Tuning and behavioral observations associated with stereo attention provides a strong case for ST as a description of the computational hierarchy underlying visual attention. It is our hope that the variety of issues highlighted in this paper will provoke useful discussion on the strengths and shortcomings of existing models in the hope that an understanding closer to a consensus might be reached.

Consideration of stereo vision poses questions for any model of attention that does not include selection on the same interpretive network that resolves stereo correspondence, or, does not first factor in more global attributes of the scene to implement more local modulation. This consideration poses a challenge for the attention community at large, and should invoke important discussion on elements that constrain the space of possible models of attention

drawing knowledge from connectionist arguments, known neurophysiology, timing, and the body of imaging data that will emerge in the next several decades.

REFERENCES

Andersen, G. J. and Kramer, A. F. (1993). Limits of focused attention in three-dimensional space. Perception and Psychophysics, 53(6):658–667.

Anderson, C. and Essen, D. V. (1987). Shifter circuits: a computational strat egy for dynamic aspects of visual processing. Proceedings of the National Academy of Science, 84:6297–6301.

Arnott, S. R. and Shedden, J. M. (2000). Attention switching in depth using random-dot autostereograms: Attention gradient asymmetries. Perception and Psychophysics, 62(7):1459–1473.

Atchley, P., Kramer, A. F., Andersen, G. J., and Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a depth-aware attentional focus. Psychonomic Bulletin and Review, 4(4):524–529.

Blake, R. (1989). A neural theory of binocular rivalry. Psychological Review, 96:145–167.

Bruce, N.D.B. and Tsotsos, J.K. (2006). Saliency based on information maximization. Advances in Neural Information Processing Systems, 18:155–162.

Bruce, N.D.B., and Tsotsos, J.K. (2009). Saliency, attention and visual search: An information theoretic approach. Journal of Vision, 9:3:1–24.

Burt, P. (1988). Attention mechanisms for vision in a dynamic world. Proceedings Ninth International Conference on Pattern Recognition, pages 977–987.

Chen, Y. and Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. Neural Computation, 16:1545–1577.

Culhane, S. (1992). Implementation of an Attentional Prototype for Early Vision. University of Toronto, M.Sc. Thesis.

Fleet, D., Wagner, H., and Heeger, D. (1996). Neural encoding of binocular disparity: Energy models. Vision Research, 36(12):1839–1857.

Ghirardelli, T. G. and Folk, C. L. (1996). Spatial cueing in a stereoscopic display: Evidence for a ”depth-blind” attentional spotlight. Psychonomic Bulletin and Review, 3:81–86.

Grill-Spector, K. and Kanwisher, N. (2005). As soon as you know it is there, you know what it is. Psychological Science, 16:2:152–160.

Hoffman, J. and Mueller, S. (1994). An in depth look at attention. Annual Meeting of the Psychonomic Society, St. Louis, MO.

Hopf, J. M., Boelmans, K., Schoenfeld, M. A., Luck, S. J., and Heinze, H. J. (2004). Attention to features precedes attention to locations in visual search: Evidence from electromagnetic brain responses in humans. Journal of Neuroscience, 24(8):1822–1832.

Iavecchia, H. P. and Folk, C. L. (1995). Shifting visual attention stereographic displays: A time course analysis. Human Factors, 36:606–618.

Itti, L. and Koch, C. (2001). Computational modeling of visual attention. Nature Reviews Neuroscience, 2(3):194–203.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11):1254–1259.

Kanai, R., Moradi, F., Shimojo, S., and Verstraten, F. A. J. (2005). Perceptual alternation induced by visual transients. Perception, 34:803–822.

Kasai, T., Morotomi, T., Katayama, J., and Kumada, T. (2003). Attending to a location in three-dimensional space modulates early erps. Cognitive Brain Research, 17:273–285.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology, 4:219–227.

Lehky, S. R. and Maunsell, J. H. (1996). No binocular rivalry in the lgn of alert macaque monkeys. Vision Research, 36:1225–1234.

Logothetis, N. K., Leopold, D. A., and Sheinberg, D. L. (1996). What is rivalling during binocular rivalry. Nature, 380:621–624.

Maier, A., Logothetis, N. K., and Leopold, D. A. (2005). Global competition dictates local suppression in pattern rivalry. Journal of Vision, 5:668–677.

Marrara, M. T. and Moore, C. M. (2000). Role of perceptual organization while attending in depth. Perception and Psychophysics, 62(4)):786–799.

Matsuoka, K. (1984). The dynamic model of binocular rivalry. Biological Cybernetics, 49:201–208.

Meng, M. and Tong, F. (2004). Can attention selectively bias bistable perception? differences between binocular rivalry and ambiguous figures. Journal of Vision, 4:539–551.

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

Mozer, M. (1991). The Perception of Multiple Objects. MIT Press.

Mueller, T. J. (1990). A physiological model of binocular rivalry. Visual Neuroscience, 4:63–73.

Nakayama, K. and Silverman, G. (1991). Serial and parallel processing of visual feature conjunctions. Nature, 320:264–265.

Ohzawa, I. (1998). Mechanisms of stereoscopic vision: the disparity energy model. Current Opinion in Biology, 8:509–515.

Ooi, T. L. and He, Z. J. (1999). Binocular rivalry and visual awareness. Perception, 28:551–574.

Palmer, D. (1999). Vision Science: Photons to Phenomenology. Cambridge Massachusetts: MIT Press.

Sandon, P. (1989). Simulating visual attention. Journal of Cognitive Neuroscience, 2(3):213–231.

Theeuwes, J., Atchley, P., and Kramer, A. F. (1986). Serial and parallel processing of visual feature conjunctions. Nature, 320:264–265.

Theeuwes, J., Atchley, P., and Kramer, A. F. (1998). Attentional control within 3-d space. Journal of Experimental Psychology: Human Perception and Performance, 24(5):1476–1485.

Theeuwes, J. and Pratt, J. (2003). Inhibition of return spreads across 3-d space. Psychonomic Bulletin and Review, 10(3):616–620.

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

Tsotsos, J. (1988). A complexity level analysis of immediate vision. International Journal of Computer Vision, 2:303–320.

Tsotsos, J. (2003). Visual Attention Mechanisms; The Selective Tuning Model. Kluwer Academic.

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

Uhr, L. (1972). Layered recognition cone networks that preprocess, classify, and describe. IEEE Transactions on Computing, 21:758–768.

Viswanathan, L. and Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. Perception, 31:1415–1437.

Wilson, H. R. (1999). Spikes, decisions and actions. Oxford University Press. Wilson, H. R.,

Blake, R., and Lee, S. (2001). Dynamics of travelling waves in visual perception. Nature, 412:907–910.

Wolfe, J. and Cave, R. (1989). Guided search: An alternative to feature integration theory for visual search. Journal of Experimental Psychology: Human Perception and Performance, 15:419–443.