Swimming in the Underlying Stream: Computational Models of Gaze in a Comparative Behavioral Analysis of Autism

Frederick Shic (frederick.shic@yale.edu)
Department of Computer Science, 51 Prospect St
New Haven, CT 06511 USA

Warren Jones (warren.jones@yale.edu)
Yale Child Study Center, 230 South Frontage Road
New Haven, CT 06520 USA

Ami Klin (ami.klin@yale.edu)
Yale Child Study Center, 230 South Frontage Road
New Haven, CT 06520 USA

Brian Scassellati (scaz@cs.yale.edu)
Department of Computer Science, 51 Prospect St
New Haven, CT 06511 USA

Abstract

Focal visual attention is a scarce resource. In order to best utilize this resource, the brain allocates movements of the eye to focus upon locations in the spatiotemporal visual scene that are maximally informative. How informative a location in the visual field is, however, depends on the dynamic internal goals and intrinsic preferences of the observer himself. This interplay suggests that, by tracking, recording, and modeling the movements of subjects watching some visual scene, we can tap into the underlying stream of human motivation. In this paper we present a framework for the computational modeling of human gaze and, by instantiating this framework, demonstrate how the visual strategies of human subjects can be quantified and compared. This comparison is formulated in terms of the subsequent implications of shared and unshared strategies in a population of adolescents with autism and with matched controls.

Introduction

The eye can only fixate upon one point in the spatiotemporal scene at a time. Consequently, the dynamics of foveal fixation represent the allocation of a scarce resource that reflects the changing internal processes, goals, and motivations of the human observer (Luck, Hillyard, Mouloua & Hawkins, 1996). Though the exact mechanisms, purpose, and utility of eye movements is a subject of some debate, that eye movements are in some way affected by internal mental processes is not (e.g. Hayhoe & Ballard, 2005). Furthermore, though it is possible to allocate visual attention covertly without an overt shift in gaze fixation, when an overt shift in gaze does occur, it seems to simultaneously demand attention (Deubel & Schneider, 1995). Thus, the movement of the eyes can serve as a window into the internal motivations of the mind.

Where a task is given explicitly, models of eye dynamics can be formulated. For instance, models used to describe saccadic actions in visual search paradigms can be computationally framed in terms of bottom-up features (Itti, Koch & Niebur, 1998; Wolfe, 1994; Wolfe & Gancarz, 1996). Models that seek to describe higher level tasks, such as those tasks that require synchronization of physical motor skills, can be framed in terms of reinforcement learning and uncertainty propagation (Sprague & Ballard, 2003). In fields such as marketing, where the goal is maximization of brand recall or user retention, Markov models may be used to describe advertisement exploration (Wedel & Pieters, 2000). Naturally, highly specialized visual activities, such as reading, may exhibit very specific patterns of transitional gaze activity that can be explicitly coded (Rayner, 1998). While these various models are capable of describing the behaviors observed in specific situations, in natural settings, or free-viewing tasks, such models typically have much less utility- a model for reading, for example, would be useless in describing the search patterns of an individual who is viewing a face.

The goal of this paper is to examine computational models of gaze in natural situations where a goal or task is not given. This is not to say that these are situations in which internal representations do not come into play; on the contrary, the difficulty of this domain is that the individual motivations governing eye fixation transitions *exist* but are not known *a priori*. To this end we begin this paper by formulating a general framework for computational models of visual attention. We instantiate this model with one particular implementation and use it to analyze the gaze patterns of individuals with autism as compared to matched controls. In contrast to previous work that examined this same dataset at the highly semantic level of social interaction and associated face processing (Klin, Jones, Schultz, Volkmar & Cohen, 2002), our aim is to examine eye patterns from the level of elementary features. We show that such analysis leads to results compatible with high-level behavioral interpretations, suggesting that differences between typical and atypical gaze patterns can be captured along a continuum reflecting common underlying behavioral currents.

A Framework for Computational Models of Visual Attention

Given a representation of the visual scene, computational models of visual attention determine a point in that scene to which focal attention is directed. One common framework for these models is shown in Figure 1. Models within this framework begin with a representation of the spatiotemporal scene $I(s,t)$ as a function of some spatial coordinate s and temporal index t. This representation is then decomposed, by feature extraction, into a set of features $F(s,t)$ that maps in many-to-one fashion onto the original spatiotemporal coordinate system. Operating over these features, an

attentional system converts these features into a saliency map, $S(s,t)$. Finally, a gaze policy is applied to the saliency map in order to extract a point, $g(t)$, that corresponds to a location that will actually be fixated upon. Many computational models of visual attention (Itti, Niebur, & Koch, 1998; Brezeal & Scassellati, 1999; Wolfe & Gancarz, 1996) obey this formulation.
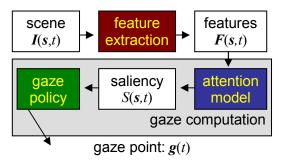

Figure 1: A generic framework for computational models of visual attention.

Feature Extraction

In computational models of attention, feature extraction is the process of extracting from the input stream abstract representations or key characteristics relevant to the final attentional decision. What exactly comprises the best set of features for guiding visual attention is an open question, though much progress has been made, especially in areas pertaining to visual search (Wolfe & Horowitz, 2004). Most feature extraction modules, however, choose their attributes based on a combination of biological inspiration and practical concerns. For example, the model of Itti et al. (1998) uses separate channels for image intensity, edge orientation, and color, where each channel is in turn composed of even more elementary channels, such as the "redness" or "brightness" of points.

Note that though the chosen features are processed early in the visual pathway, their computational formulation or characterization can be arbitrarily simple or complex. For example, by considering an augmented set of features that depend upon previously computed internal variables, we can account for models of selective attention, such as the selective tuning model of Tsotsos et al. (1995), which incorporates bidirectional excitation and inhibition between the feature extraction module and the attention model. This is an important feature, as strictly bottom-up models of visual attention adequately represent neither the true neurophysiological underpinnings of visual attention (Desimone and Duncan, 1995; Posner & Petersen, 1990) nor its computational capabilities and limitations (Tsotsos. 1988).

The Attention Model

The role of the attention model is to convert the elementary features into a saliency map, an intermediate representation first proposed by Koch and Ullman (1985) that associates with each point in the spatiotemporal scene a specific value

representing that point's visual prominence. While there exists some evidence for a saliency map in the brain (Li, 2000), computational models of visual attention typically employ saliency maps for computational and organizational reasons and do not assume a direct biological correlate.

Currently, many different strategies are available for the computation of saliency. Most strategies rely upon the feature integration theory of Treisman and Gelade (1980) which views saliency as the integration of multiple input modality maps, often by linearly weighted summation or nonlinear transfer of linearly weighted summation (Balkenius, Astrom & Eriksson, 2004; Brezeal & Scassellati, 1999; Itti et al, 1998; Wolfe & Gancarz, 1996). Others view salience in more theoretical terms. For instance, Itti & Baldi (2006) view the salience of spatiotemporal locations in terms of Bayesian "surprise", and Torralba (2003) characterizes global contextual factors affecting salience in information-theoretic terms. Later in this paper we will present another perspective on saliency maps by framing salience as a classification problem on points attended-to by observers and points that are not attended-to.

Gaze Policy

A gaze policy takes the saliency map as input and from it derives the location where attention should be next directed. Formally, if the salience at each point in the saliency map is real-valued, we can simply define this point as:

$$g(t) = \arg\max{}_s (S(s,t))$$

As with the other steps in our framework, the actual implementation of a gaze policy can be more involved, incorporating higher order interactions such as inhibition of return (as in Itti et al., 1998). Furthermore, the actual action of fixating the eye can involve a change in visual input as the high-resolution fovea rotates to sample the area at a chosen point non-linearly (as in Wolfe & Gancarz, 1996). Thus there may exist some level of interaction between the gaze policy and the scene input to the system, completing a circuit describing this framework for visual attention.

Comparing Gaze Patterns

One natural metric for judging how well a model performs is to compare it against human subjects (Ouerhani, von Wartburg, Hugli & Muri, 2004; Parkhurst, Law & Niebur, 2002; Shic & Scassellati, 2006). To do this, the free parameters of a model can be tuned so that the model best describes the gaze allocation behavior of, say, one particular individual.

How should the similarity between simulated gaze patterns generated by a computational model and the actual gaze actions of subjects be measured? A simple measure would be to consider the time-varying Euclidean distance between two gaze trajectories. Figure 2 demonstrates one of the problems with this naïve approach.
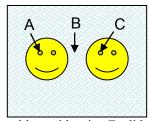
Figure 2: The problem with using Euclidean distance as a measure for gaze similarity. An individual who focuses on B is most likely not using the same scanning strategy as the person who is focusing on A. In contrast, an individual who focuses on C might be using a very similar gaze strategy.

A better solution is to phrase the distance between gaze patterns in terms of the similarity of the features underlying the points of fixation. To do this we require some way of comparing features. This is easily accomplished by viewing the generation of saliency within the attention model as a classification problem that separates attended-to locations in the visual stream from those locations that are not-attended-to. A Bayesian formulation is:

$$p(c_i \mid \boldsymbol{f}) = \frac{p(\boldsymbol{f} \mid c_i)p(c_i)}{p(\boldsymbol{f})}$$

Where $\boldsymbol{f}$ is the underlying set of features associated with a particular location, c_0 is the attended-to class and $c_1 = \neg c_0$ is the not-attended-to class. By transforming this to a classifier that would choose class c_0 if $p(c_0|\boldsymbol{f}) > \Theta\, p(c_1|\boldsymbol{f})$ for some threshold Θ, and would choose class c_1 otherwise, we can define the saliency associated with features $\boldsymbol{f}$ to be:

$$S(\boldsymbol{f}) = \frac{p(\boldsymbol{f} \mid c_0)}{p(\boldsymbol{f} \mid \neg c_0)}$$

If we further relax the strict probabilistic interpretation of salience, however, we can access a much larger set of dimensionality reduction techniques. For instance, by maximizing the Fisher criterion function we can find a projection w that in some sense represents an optimal 1D projection for discriminating between attended-to and not-attended-to locations. That is we can obtain the solution:

$$w = S_W^{-1}(m_1 - m_2)$$

$$\text{where } m_i = \frac{1}{|c_i|}\sum_{x \in C_i} x, \; S_W = S_1 + S_2, \text{ and}$$

$$S_i = \sum_{x \in C_i}(x - m_i)(x - m_i)^t = (|c_i|-1)\Sigma_i = k_i \Sigma_i$$

(for reference see Duda and Hart, 2001). In this manner we avoid having to estimate $p(\boldsymbol{f} \mid c)$, a task that can be quite difficult in high dimensions, and can require more complicated approximation techniques even in lower dimensions.

Both the pure probabilistic formulation using Bayesian inference and the dimensionality reduction strategy employing Fisher's linear discriminant are natural methods for tuning computational models of visual attention to the gaze patterns of an individual. Once the model is tuned, the corresponding maps of salience at every point in time and space for that individual are easily generated. We can obtain a measure of how well the model fits by examining the salience at locations where the individual actually looks in comparison to the salience of the locations that the individual does not look.

Once we have a tuned model, however, we are not limited to model-individual comparisons. We can also take this same model and apply it to *other* individuals. That is, we can evaluate how well a particular model, tuned to one particular individual, explains the gaze patterns of other individuals. Furthermore, the results of model cross-application can be aggregated in order to investigate population specific trends.

Experiment

As a test of our framework and comparative techniques, we apply our methods to the analysis of a population of individuals with autism and matched controls. We know that differences in gaze patterns exist between these two groups both qualitatively (Figure 3) and as a result of the high-level analysis conducted by Klin et al. (2002) which showed that individuals with autism, in comparison to controls, focused more on mouths and objects than on eyes. In this work, we are primarily interested in the implications of cross-population and inter-population statistics upon the developmental and cognitive deficits inherent in autism.

Subjects and Data

The data and subjects used in this study were drawn from a subset of the data obtained in Klin et al. (2002). In this experiment, adolescents and young adults diagnosed with autism (N=10) were matched with a control group (N=10) on the basis of age and verbal-IQ. These individuals watched two different one-minute clips of the 1966 black-and-white movie "Who's Afraid of Virginia Woolf?" in a controlled environment while their eye movements were tracked via a head mounted eye-tracker. The movie occupied a width of approximately 34º in the visual field, and the eye tracker was accurate to ±0.3º over a horizontal and vertical range of ±20º. For further information regarding the parameters of data acquisition, subject statistics, and diagnostic criteria, see Klin et al (2002).

As a control against computational bias, several synthetic gaze trajectories were also incorporated into the experiment. These gaze trajectories were uncorrelated with the visual scene and included (i) random filters (random weight matrices in the Fisher's linear discriminant formulation) (ii) random saccades (a sequence of fast jumps across the screen triggered probabilistically) and (iii) random walks (small movements across the screen every frame).
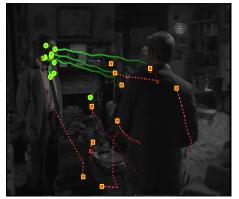
Figure 3: Eye scanning paths of controls (solid lines from circles) as compared to individuals with autism (dotted lines from squares) on a scene from the 1966 movie "Who's Afraid of Virginia Woolf?" (Klin et al., 2002). The instantaneous fixation point is the circle or square and each path stretches 250 ms into the future. The gaze locations of controls are clustered on the left-most face; the gaze locations of individuals with autism are scattered.

Computational Model

Feature Extraction – The features used in this experiment consisted of a linearization of raw patch features drawn from points in history. That is, points of eye fixation corresponding to attended-to locations (and 15 randomly selected points at least 2.9º distant from the actual gaze point for not-attended-to locations) were considered the center of a square area which was further subdivided spatially into a uniform grid of subblocks. Each subblock within the grid was taken to be representative of the underlying spatial content by averaging (i.e. the subblock represented the corresponding region by a single average intensity), and the set of all subblocks associated with selected points in time prior to the fixation constituted the features associated with an attended-to location. The entire grid spanned approximately 9.3º and was divided into 11x11 subblocks, sampled at 100ms and 300ms in the past. Temporal sampling was necessary to allow for motion encoding, as the scene was time-varying. Though this feature set was not completely physiological, being coarser in sampling and larger in extent than the fovea, its simple expression struck a useful tradeoff between spatiotemporal extent and computational expedience. Several other feature sets were also tested, including both a multiscale representation as well as the more complex biologically-inspired model of Itti et al. (1998). Neither the use of these other feature sets, nor the variation of their associated parameters within a wide range, impacted the nature of our final results.   Further details on varying feature representations can be found in Shic & Scassellati (2006).

Attention Model – Saliency maps were generated by using the method of dimensionality reduction via projection of features upon Fisher's linear discriminant. To compensate for the much larger sampling of non-attended-to locations versus attended-to locations, the coefficient k associated

with the covariance matrix was taken to be equal for all classes. Training of models occurred over odd frames of one particular clip, allowing for testing over the highly-correlated even frames of the same clip, as well as an independent comparison on a completely different clip.

Comparative Method

Our computational framework provides a method for determining, for some particular individual, the saliency of every spatiotemporal point in the visual scene. If we thus generate a model for an individual A, we can see how well our techniques work by examining the reported saliencies at the locations of A's gaze (Figure 4). If our techniques are good, the average saliency at the locations where A fixates should be high. Furthermore, we can take A's model and look at the locations where another individual, B, looks. This gives us a measure of how well the model of A describes the gaze trajectories of B, leading to a natural measure for the distance between the two individuals.
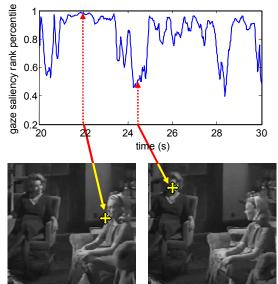

Figure 4: Time-varying salience record representing how well a particular model, tuned to one individual, describes the gaze behavior of another individual. In this example the yellow crosses correspond to locations actually fixated upon by an individual. These locations are associated with particular salience values. When the actual fixation is adequately represented by the model, salience is high (close to 1.0); when not, salience is close to chance (0.5).

In order to maintain consistency and comparability across all frames in the movies and all individuals we first normalized the saliency values in each frame to a rank percentile. That is, if a particular spatial location in some given frame was the 95th highest percentile (most salient) location, it was normalized to a value of 0.95 regardless of what its actual projected value was. This reflected the fact that fixation is a relative and not an absolute decision process. Next, the gaze patterns of a particular individual were indexed into the salience map generated by another

individual. From this we were able to obtain time-varying salience records (Figure 4). Finally, in order to obtain an overall score representing how well a model matched an individual, the median salience value from the time-varying salience record was taken as representative. This provided a robust measure of average model effectiveness.

Results

By applying models tuned for each trajectory (both human and synthetic operating over two movie clips) to every other trajectory in our data set, we were able to obtain a large number of cross-trajectory comparisons. By aggregating the data into groups we obtained the statistics of Figure 5 & 6.

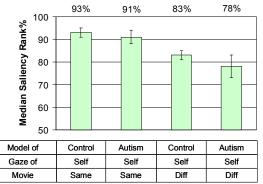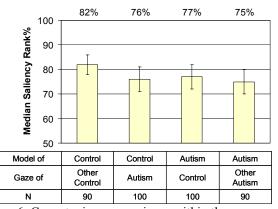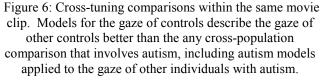| Model of	Control	Autism	Control	Autism
Gaze of	Self	Self	Self	Self
Movie	Same	Same	Diff	Diff

Figure 5: Self-tuning comparisons across movies. Results are aggregated (N=10 in each condition) for models trained on one individual (control or autism) and tested on the gaze patterns of that same individual (watching either the same movie or a different movie). When a model is trained on one movie and applied to another movie, we get a drop in performance. However, in all cases, human models describe the gaze of other humans much better than random as determined theoretically (50%) and empirically (52±13%, N=600). Error bars span two standard deviations.

Model of	Control	Control	Autism	Autism
Gaze of	Other Control	Autism	Control	Other Autism
N	90	100	100	90

Figure 6: Cross-tuning comparisons within the same movie clip. Models for the gaze of controls describe the gaze of other controls better than the any cross-population comparison that involves autism, including autism models applied to the gaze of other individuals with autism.

The application of our framework leads to several results. First, all applications of a human's model to a human's gaze trajectory lead to performance much better than those obtained by random chance, as developed by synthetic gaze trajectories (Figure 5 & 6; p<0.01). This suggests that both individuals with autism and control individuals rely on some common scanning approach, implying the existence some core human strategy. Furthermore, this result suggests that it is unlikely that a methodological bias exists in either the learning technique or the feature representation.

Second, the extremely high matched-application (control on self and autism on self groupings) within-movie scores (Figure 5) suggest that each subject relies upon some specific individual strategy. This specific individual strategy does not seem to transfer across scenes, as demonstrated by matched comparison score drops as we move from within-movie comparisons to across-movie comparisons, suggesting that top-down or contextual influences on gaze strategy are significant.

Third, as highlighted by Figure 6, control individuals, who are taken to be socially more typical than individuals with autism, exhibit much greater coherence (p<0.01) in terms of attraction to underlying features than cross-application cases that involve individuals with autism. This suggests that the strategies of controls transfer well to other controls, but that the strategies of individuals with autism do not transfer to the same degree to either normal individuals or even other individuals with autism.

Discussion

The original Klin et al. (2002) study found that individuals with autism spent more time focusing on mouths, bodies, and objects, whereas controls spent significantly more time looking at eyes. In terms of elementary features, eyes vary the least; objects vary the most. Thus our results in this paper could derive specifically from this disparity. If eyes vary the least, and controls focus on eyes much more often than individuals with autism (the difference between eye fixation time fractions between the two populations exceeds 40%), we would expect a higher correspondence among control individuals. Similarly, if features associated with bodies and objects vary most, we would expect individuals with autism to exhibit fine tuned strategies specific to particular objects or image characteristics not generally found elsewhere. If these strategies are extremely fine tuned, they cannot transfer to other individuals.

The disadvantage of a featural level analysis, compared to higher-level considerations, is that much of the internal circuitry of low-level models is impenetrable. For instance, we can frame the results obtained in our results in terms of semantic labels associated with subject fixation. However, the converse, predicting high level implications from low level aggregate effects, could prove very difficult. On the other hand, since we do have as many time-varying salience records as we have comparisons, it is possible that by pinpointing locations of mutually high salience we could discover classes of highly correlated specific gaze behavior. The use of our comparative techniques as an exploratory tool in this manner remains to be investigated.

The advantage of featural level analysis is that preexisting labels with associated semantic implications are not assumed. If the underlying featural representation associated with a particular computational model of visual attention is sufficient to represent some common underlying strategy within a population, our techniques should uncover this fact. In this investigation we have uncovered two tiers of shared strategies. The first tier represents the underlying gaze patterns associated with the scanning behavior of all humans, mechanisms likely hardwired into the early visual system. The second tier is found between controls, likely representing typical development versus early derailment as predicted by enactive mind theory (Klin, Jones, Schultz, and Volkmar, 2003). Finally, the ability for models to match specific individual preferences suggests that order *does* exist in the gaze patterns of individuals with autism, suggesting that when early derailment of social skill development occurs it is replaced by some other set of visual behavior that likely reflects a unique cascading specialization.

Conclusions

We have presented a general framework for visual attention, realized one particular implementation of this framework, and applied the resulting model to analyze the eye scanning trajectories of individuals with autism and matched controls. We show that such feature level analysis offers a wealth of insight into the fundamental behaviors and preferences of subject populations, and that these uncovered insights are consistent with higher-level analysis. Future avenues for investigation include augmented sets of features such as scanpath memory via Markov model, application of various dimensionality reduction techniques, and the use of our comparative method as an exploratory tool for decomposing the global scanning behavior into component elements.

References

Balkenius, C., Eriksson, A. P. & Astrom, K. (2004). Learning in Visual Attention. In *Proceedings of LAVS '04*. St Catharine's College, Cambridge, UK.

Breazeal, C. and Scassellati, B. (1999). A Context-Dependent Attention System for a Social Robot. In T. Dean, (Ed.), *Proc. 16th Intl. Conf. Artificial Intel.*, 1146-1153. Morgan Kaufmann Publishers, San Francisco, CA.

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

Deubel, H. & Schneider, W.X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vis. Research,* 36(12),1827–1837.

Duda, R., Hart, P., & Stork, D. (2001). Pattern Classification. New York: Wiley.

Hayhoe, M. & Ballard, D. (2005). Eye Movements in natural behavior. *Trends in Cog. Sci.* 9(4), 188-194.

Luck, S. J., Hillyard, S. A., Mouloua, M., & Hawkins, H. L. (1996). Mechanisms of visual-spatial attention: Resource allocation or uncertainty reduction? *Journal Exp. Psych.: Human Perception and Performance*, 22, 725-737.

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

Itti, L. & Baldi, P. (2006). Bayesian Surprise Attracts Human Attention. In *Adv. Neural Information Processing Systems* (pp. 1-8). Cambridge, MA: MIT Press.

Klin, A., Jones, W., Schultz, R., Volkmar, F. & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry,* 59(9), 809-816.

Klin, A., Jones, W., Schultz, R. & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Phil. Trans. of the Royal Society of London, Series B, Biological Sciences,* 358(1430), 345–360.

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

Li, Z. (2002). A salience map in primary visual cortex. *Trends Cogn. Sci.,* 6, 9–16.

Ouerhani, N., von Wartburg, R., Hughli, H., Muri, R. (2004). Empirical Validation of the Saliency-based model of Visual Attention. *Electronic Letters on Computer Vision and Image Analysis,* 3(1), 13-24.

Parkhurst, D., Law, K. & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research,* 42(1), 107-123.

Posner, M. I. & Petersen, S.E. (1990). The attention system of the human brain., *Ann. Rev Neurosci.* 13, 25-42.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin,* 124(3), 372-422.

Shic, F. & Scassellati, B. (2006). A behavioral analysis of robotic models of visual attention. Under review.

Sprague, N. and Ballard, D. (2003) Eye movements for reward maximization. In *Advances in Neural Information Processing Systems,* Vol. 16, MIT Press.

Torralba, A. (2003) Modeling global scene factors in attention. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1407–1418.

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

Tsotsos, J. K. (1988). A 'complexity level' analysis of immediate vision. *Intl. Journal Comp. Vis.*, 1(4):303–320.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78, 507-545.

Wedel, M., Rik, P. (2000), Eye fixations on advertisementsand memory for brands: A model and findings. *Marketing Science,* 19(4), 297–312.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bull. and Rev.*, 1(2), 202-238.

Wolfe, J. M., & Gancarz, G. (1996). Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminarayanan (Ed.) *Basic and Clinical Applications of Vision Science*, Dordrecht, Netherlands: Kluwer Academic.

Wolfe, J.M. & Horowitz, T.S. (2004). Opinion: What attributes guide the deployment of visual attention and how do they do it? *Nature rev., Neuroscience,* 5(6), 495.