Spatiotemporal Saliency: Towards a Hierarchical Representation of Visual Saliency

Neil D.B. Bruce and John K. Tsotsos

Department of Computer Science and Engineering and
Centre for Vision Research
York University, Toronto, ON, Canada
{neil,tsotsos}@cse.yorku.ca
http://www.cse.yorku.ca/~neil

Abstract. In prior work, we put forth a model of visual saliency motivated by information theoretic considerations [1]. In this effort we consider how this proposal extends to explain saliency in the spatiotemporal domain and further, propose a distributed representation for visual saliency comprised of localized hierarchical saliency computation. Evidence for the efficacy of the proposal in capturing aspects of human behavior is achieved via comparison with eye tracking data and a discussion of the role of neural coding in the determination of saliency suggests avenues for future research.

Keywords: Attention, Saliency, Spatiotemporal, Information Theory, Fixation, Hierarchical.

1 Introduction

Certain visual search experiments demonstrate in dramatic fashion the immediate and automatic deployment of attention to unique stimulus elements in a display. This phenomenon no doubt factors appreciably into visual sampling in general influencing fixational eye movements and our visual experience as a whole. Some success has been had in emulating these mechanisms [2], reproducing certain behavioral observations related to visual search, but the precise nature of the principles underlying such behaviors remains unknown.

One recent proposal deemed Attention by Information Maximization (AIM) is grounded in a principled definition for what constitutes visually salient content derived from information theory, and has had some success in explaining certain aspects of behavior including the deployment of eye movements [1] and other visual search behaviors [3]. In this paper we further explore support for this proposal through consideration of spatiotemporal visual stimuli. This includes a comparison of the proposal against the state of the art in this domain. The following discussion reveals the efficacy of the proposal put forth in AIM to explain eye movements for spatiotemporal data and also describes how the model

L. Paletta and J.K. Tsotsos (Eds.): WAPCV 2008, LNAI 5395, pp. 98–111, 2009.

fits in with the *big picture*. Specifically, we address how the proposal fits with distributed hierarchical attentional architectures of the sort put forth by Tsotsos [4] for which favorable evidence has appeared in recent years.

2 AIM: Information Maximizing Saliency

In the following section, we briefly review the proposal put forth in [1], which is applied to a set of neurons that code for content in space-time within the evaluation included in this work. The following offers only a brief overview; for a detailed account, readers should refer to [1].

The central premise of AIM is that saliency computation should serve to maximize information sampled from one's environment from a stimulus driven perspective. Specifically, given an ensemble of neurons $C_{i,j}$ that code for content at spatial coordinates i, j with $C_{i,j,k}$, $k = 1...N$ corresponding to the different types of cells with receptive fields centered at i, j the self-information or surprisal associated with $C_{i,j}$ is given by $-log(p(C_{i,j}))$ with the likelihood determined by observing the response of cells in the surround of $C_{i,j}$. Given the assumption of independence on the response of different types of cells (an assumption made reasonable by sparsity as discussed in the section that follows), this quantity may be computed as $\sum_{k=1}^{N} -log(C_{i,j,k})$. Saliency in this context then amounts to the surprisal or self-information of the response associated with a cell as defined by its surround. In other words, saliency is inversely proportional to the likelihood of predicting the response of any given neuron in observing the response of neurons in its surrounding spatiotemporal context. For any given cell type it is straightforward to derive a likelihood estimate by constructing a probability density estimate based on cells of the same type in the surround. An overview of the model with reference to the specifics of the implementation for spatiotemporal stimuli is presented in the section that follows.

3 Extension to Space-Time

The general nature of the original proposal implies that it may be applied to any set of neurons that constitute a sparse basis. For this reason, extension to space-time is straightforward assuming the early coding of spatiotemporal content observed in the cortex satisfies these criteria. There exist many efforts documenting the relationship between early visual cortical neurons and coding strategies that demonstrate that learning a sparse code for local grey-level image content yields V1 like receptive fields similar to oriented Gabor filters [5,6]. Further efforts have demonstrated this same strategy yields color-opponent coding for spatiochromatic content [7] and also cells with properties akin to V1 for spatiotemporal data [8]. We have employed the same data and strategy put forth in [8] to learn a basis set of cells coding for spatiotemporal content. The data described in [8] was subsampled taking every second frame to yield data at 25

frames per second. The data set consists of a variety of natural spatiotemporal sequences taken from various angles of a moving vehicle traveling in a typical urban environment. Spatiotemporal volumes were then randomly sampled from the videos to yield 11x11x6 (x,y,t) localized spatiotemporal volumes that served as training data. Infomax ICA [9] was applied to the training set resulting in a spatiotemporal basis consisting of cells that respond to various frequencies and velocities of motion and for which the correlation between cell firing rates is minimized. The basis resulting from dimensionality reduction via PCA retaining 95% variance followed by ICA yields a set of 60 spatiotemporal cells. A subsample of these (corresponding to 1st, 3rd and 6th frame of the volume) are shown in figure 1. Note the response to various angular and radial frequencies and selectivity for different velocities of motion. Aside from the application to spatiotemporal data and the different basis set, the saliency computation proceeds according to the description put forth in [1].

An overall schematic of the model based on the learned spatiotemporal basis appears in figure 2. A localized region from adjacent frames (3 of 6 shown) are projected onto the learned basis. This yields a set of coefficients for the local region that describes the extent to which various types of motion are observed at the given location. The likelihood of each response is then evaluated by observing the response of cells of the same type in the surround or in this implementation, over the entire image. A sum of the negative log likelihood associated with all of the coefficients corresponding to the given coordinate (pixel) location yields a local measure of saliency.

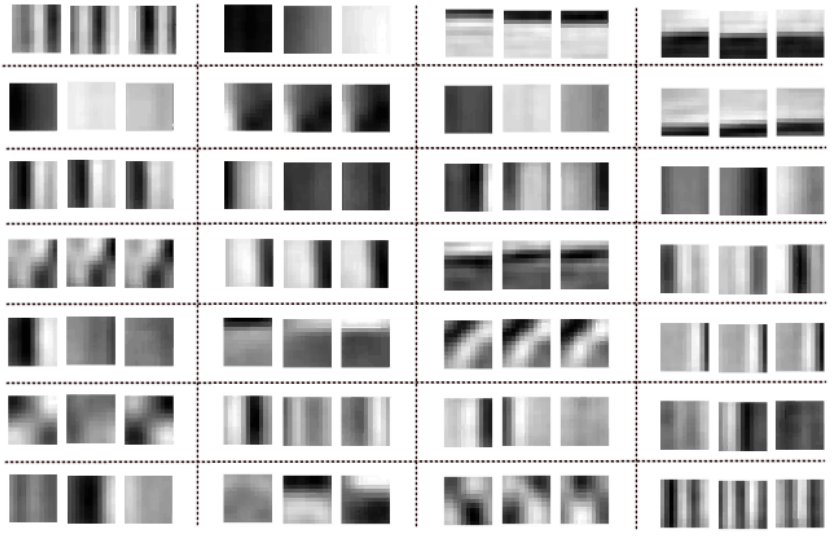

Fig. 1. The receptive field profile of a subsample of the learned basis corresponding to frames 1, 3 and 6 of the spatiotemporal volume. Note the selectivity for various angular and radial frequencies and velocities and directions of motion.

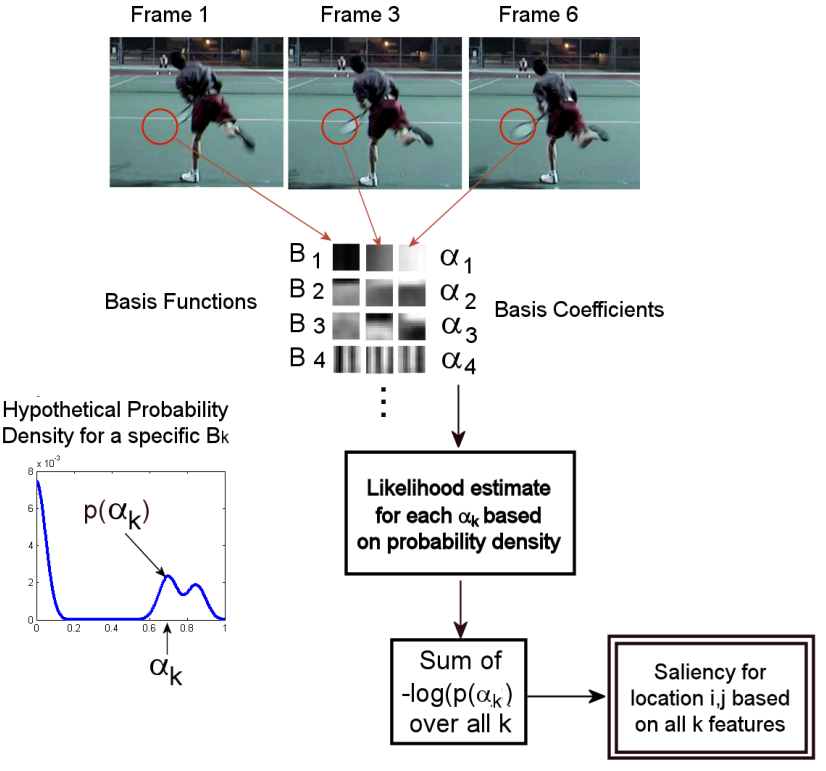

Fig. 2. An overview of the computation performed by AIM. A spatiotemporal volume is projected onto a learned basis based on independent component analysis. The likelihood of any given cells firing rate may be estimated by observing the distribution of responses associated with cells of the same type in the surround or over the entire image. A summation of these likelihoods subjected to a log transform then yields a local measure of information. For a complete description the reader should refer to [1].

4 Evaluation

An evaluation of the efficacy of the model in predicting spatiotemporal fixation patterns is achieved via comparison with eye tracking data collected for video stimuli. The eye tracking data employed for this study was that used in [10] and performance evaluation was carried out according to the same performance metric described in the aforementioned work.

The data consists of eye tracking data for a total of 50 video clips and from 8 subjects aged 22-32 with normal or corrected to normal vision. Videos consist of indoor and outdoor scenes, news and television clips and video games. Videos were presented at a resolution of 640x480 and at 60 Hz and consist of over 25 minutes of playtime. The total number of saccades included in the analysis is 12,211.

For any given algorithm, one may compare the saliency at fixated locations with randomly sampled locations. The Kullback-Leibler divergence of two distributions corresponding to these quantities is given by

$$D_{KL}(P,Q) = \sum P(i) log \frac{P(i)}{Q(i)}$$

where P and Q correspond to the distribution of randomly sampled and atfixation sampled saliency values respectively based on 10 bin histogram estimates. The KL-divergence offers a performance metric allowing comparison of

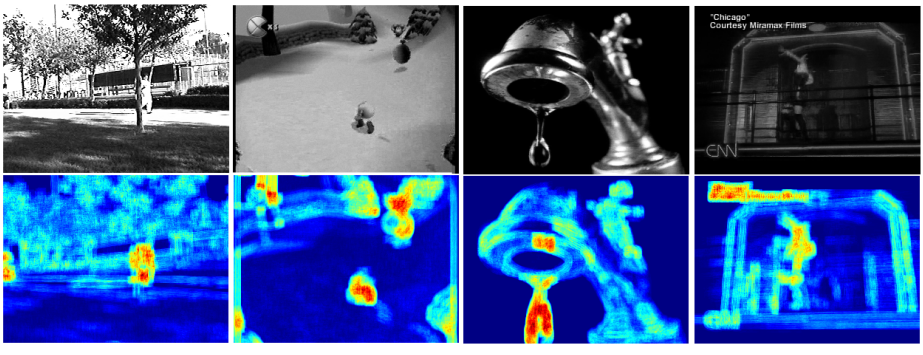

Fig. 3. Relative saliency of each pixel for a variety of frames from different videos allowing a qualitative assessment of model performance

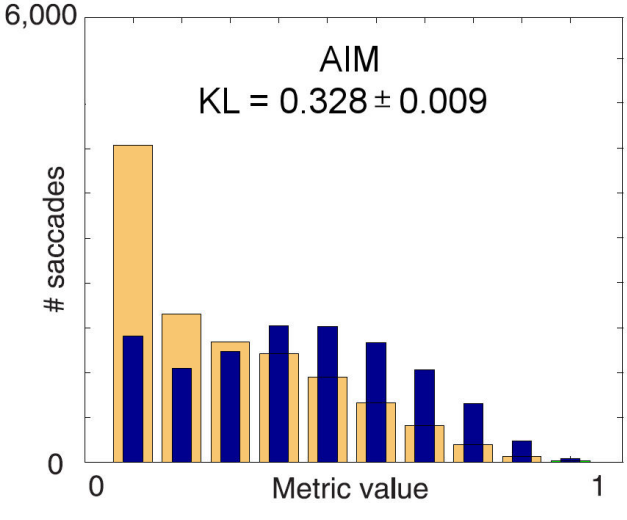

Fig. 4. A histogram representation comparing saliency values at fixated versus randomly located display locations. KL-divergence is 0.328 as compared with 0.241 for the algorithm presented in [10] and 0.205 for that appearing in [2].

various algorithms. Results are compared against those put forth in [10] and proceeds according to the same performance evaluation strategy.

Figure 3 demonstrates the relative saliency of pixel locations for a variety of single frames from a number of videos. Note the inherent tradeoff between moving and stationary content as observed for the running tap, and park scene as well as the ability to detect salient patterns on a relatively low contrast background (rightmost frame).

Figure 4 demonstrates a histogram of the saliency associated with the fixated locations as compared with those from uniformly randomly sampled regions. Of note is the shift of the distribution towards higher saliency values for the distribution associated with fixated relative to random locations. The KL-divergence of the two distributions shown is 0.328. This compares favorably with the Surprise metric of Itti and Baldi [10] which gives rise to a KL-divergence score of 0.241 and the saliency evaluation of Itti and Koch [2] which yields a KL-divergence score of 0.205. This result demonstrates that relative to competing proposals the saliency associated with fixated relative to random locations is greatest for AIM.

5 Surround Suppression, Gain Control and Redundancy

An important consideration in any model that posits a specific proposal for how saliency computation is achieved, is that of a possible neural implementation. Perhaps the foremost consideration pertaining to neural circuitry, is the extent to which the proposal agrees with observations concerning cortical circuitry and neurophysiology. To this end, this section reviews a variety of classic and recent results derived from psychophysics and imaging experiments on the nature of surround suppression within the cortex. Necessary conditions of an architecture that seeks to maximize information in its control of neural gain are weighed against the experimental literature in order to evaluate the plausibility of AIM from the perspective of a possible neural basis for its implementation. As a whole, the discussion establishes that a variety of peculiar and very specific constraints imposed by the implementation show considerable agreement with the computation implicated in surround suppression further providing support for AIM, and also offering some insight on the nature of computation responsible for iso-orientation surround suppression in early visual cortex. Debate concerning the specific nature and form of surround suppression has rekindled in recent years, which has resulted in a large body of interesting results that further elucidate the details of this process. The following discussion reviews these results and offers further insight through a meta-analysis of recent studies. In each case, experimental findings are contrasted against the computational constraints on AIM to establish plausibility of the proposed computation.

5.1 Types of Features

A great deal of research has focused specifically on the suppression that arises from introducing a stimulus in the surround of a localized oriented Gabor target. The specific nature of iso-orientation (iso-feature) surround suppression as

dictated by the details of AIM includes two key considerations: 1. Suppression of a cell whose receptive field lies at the target location should occur only for a surround stimulus that is the effective stimulus for this cell. For example, for a vertically oriented Gabor target, suppression of a cell that elicits a response to the target will occur only by way of a similar stimulus appearing in the surround. Recall that a fundamental assumption is that the responses of different types of cells at a given location are such that the correlation between their responses is minimal and this is a phenomenon that is observed cortically. In the domain of studies pertaining to surround suppression, the literature is undivided in its agreement with this assumption. When considering the cell response or psychometric threshold associated with a target patch, suppression from a surround stimulus is highly stimulus specific and is at a maximum for a surround matching the target orientation, with suppression observed only for a narrow orientation band centered around the target orientation [11,12,13,14,15,16]. This is consistent with a local likelihood estimate in which the independence assumption is implicit. 2. Suppression should be observed for all feature types, and the nature of, and parameters associated with suppression should not differ across feature type. This is an important consideration since studies of this type have largely focused on oriented sinusoidal stimuli but nevertheless similar suppression associated with color, or velocity of motion for example, should also be observed and the nature of such suppression should be consistent with that observed in studies involving oriented sinusoidal target and surrounds. One recent effort provides strong evidence that this is the case through single cell recording on macaque monkeys [14]. Shen et al. demonstrate that centre-surround fields defined by a variety of features including color, velocity and oriented gratings all elicit suppression and with suppression at a maximum for matching centre and surround stimuli.

5.2 Relative Contrast

Given a cell with firing rate $N_{i,j}$ that codes for a specific quantity at coordinates i,j in the visual field (e.g. a cell selective for a specific angular and radial frequency as part of a basis representation with its centre at location i,j), a density estimate on the observation likelihood of the firing rate associated with $N_{i,j}$ as discussed earlier in this section is given by:

$$p(N_{i,j}) = \sum_{\forall s,t \in \Omega} f(N_{i,j} - N_{s,t}) \tag{1}$$

Where f is a monotonic symmetric kernel with its maximum at f(0) and Ω the region over which the surround has any significant impact. For further ease of exposition in observing the behavior of equation 1, assume without loss of generality that f comprises a Gaussian kernel. Then equation 1 becomes:

$$\frac{1}{\sigma\sqrt{2\pi}} \sum_{\forall s,t \in \Omega} e^{-(N_{j,k} - N_{s,t})^2 / 2\sigma^2} \tag{2}$$

As there also exists a spatial component to this estimate, it may be more appropriate to also include a parameter that reflects the effect of distance on the

contribution of any given cell to the estimate of $N_{i,j}$ which might appear as follows:

$$\frac{1}{\sigma\sqrt{2\pi}} \sum_{\forall s,t \in \Omega} \Psi(s,t) e^{-(N_{j,k}-N_{s,t})^2/2\sigma^2} \tag{3}$$

Ψ drops off according to the distance of any given cell from the target location, reflecting the decreasing correlation between responses. Assuming that surround suppression is the basis for the computation involved in AIM equation 1 demands a very specific form for the suppressive influence of a surrounding stimulus on the target item. According to the form of equation 3, suppression depends on the relative response of centre and surround stimuli and should be at a maximum for equal contrast centre and surround stimuli: Raising or lowering the contrast of a stimulus pattern will generally result in a concomitant increase in the response of a cell for which the pattern in question is the effective stimulus. There is therefore a direct monotonic (nonlinear) relationship between the firing rate attributed to centre or surround, and their respective contrasts. Support for suppression as a function of relative centre versus surround contrast is ubiquitous in the literature [17,18,14,11,19,20,15,21] although there is as of yet no consensus on why this should be the specific form for the suppressive influence of a surround stimulus. There also exists a large body of prominent studies revealing that this suppression is indeed at a maximum for equal contrast centre and surround stimuli [17,18,14,11,15]. Note that this implies mathematical equivalence between surround suppression and a likelihood estimate on a given cell's response as defined by the response of neighboring cells and implies divisive modulation of a cells response by a function of its likelihood. This is an important consideration as it offers insight on the role of surround suppression which has recently become an issue of considerable dispute [16] and implicates surround suppression as the machinery underlying the implementation of AIM. It is also worth noting that the suppressive impact of cells in the surround is observed to drop off exponentially with distance from the target giving the specific form of Ψ [16].

5.3 Spatial Configuration

For the sake of exposition, let us assume that the computation under discussion is restricted to V1. From the perspective of efficient coding, no knowledge of structure is available at V1 beyond that which lies within a region the size of single V1 receptive field. A pure information theoretic interpretation of the surprisal associated with a local observation as determined at the level of V1 should reflect this implying an isotropic contribution to any likelihood estimate in the vicinity of the target cell, regardless of the pattern that forms an effective stimulus for the cell in question. That is, for a unit whose effective stimulus is a horizontal Gabor pattern, equidistant patterns of the same type in the vicinity of the target should result in equal suppression regardless of where they appear with respect to the target and this is reflected in the implementation put forth in [1]. It is also expected that likelihoods associated with higher order structure over larger receptive fields are mediated by higher visual areas either

implicitly at the single cell level or explicitly via recurrent connections. In line with the assumption that computation is on the observation likelihood of a pattern within a given region, and that structures are limited to an aperture no larger than a V1 receptive field, it is indeed the case that suppression from the surround is isotropic with respect to the location of a pattern appearing in the surround independent of target and surround orientations [16]. By virtue of the same consideration, one would also expect the spatial extent of surround suppression to be invariant to the spatial frequency of a target item. This is also a consideration that is evident in the literature [16]. In consideration of observation likelihoods associated with more complex patterns, it is interesting to consider the nature of surround suppression among higher visual areas. Recent studies are discovering more and more examples of suppressive surround inhibition among higher visual areas with the same properties and divisive influence as those that are well established in V1. Extrastriate surround inhibition of this form has been observed at least among areas V2 [22,23], V4 [24,25], MT [26,27,28], and MST [29]. This is suggestive of the possibility that saliency is represented within a distributed hierarchy, with local saliency computation mediated by surround suppression at various layers of the visual cortex.

5.4 Fovea versus Periphery

If the role of local surround suppression is in attenuating neural activation associated with unimportant visual input and/or redirecting the eyes via fixational eye movements one would expect the influence of such a mechanism to be prominent within the periphery of the visual field. Petrov and McKee demonstrated that surround suppression is in fact strong in the periphery and absent in the fovea [16]. This is consistent, as Petrov and McKee point out, with a role of this mechanism in the control of saccadic eye movements. Furthermore, there are additional points they highlight that support this possibility, including the fact that the extent of suppression is invariant to stimulus spatial frequency. Also of note, is the fact that the inaccuracy of a first saccade is proportional to target eccentricity and this correlates with the extent of surround suppression as a function of eccentricity [16]. Note that the cortical region over which surround suppression is observed does not vary with eccentricity implying that computationally, an equal number of neurons contribute to any given likelihood estimate of the form appearing in equation 1. All of these considerations are in line with a role of this mechanism in the deployment of saccades.

5.5 Summary

We have put forth the proposal that the implementation of AIM is achieved via local surround circuitry throughout the visual cortex. As a whole, there appears to be considerable agreement with the proposal and the specific form of surround suppression. The demonstration of equivalence of a likelihood estimate on the surround of a cell with the apparent form of suppressive inhibition implies modulation of cell responses at a single cell level through divisive gain as

a function of the likelihood associated with that cell's response. This provides a more specific explanation for the nature of computation appearing in suppressive surround circuitry and further bolsters the claim that saliency computation proceeds according to a strategy of optimizing information transmission.

6 On the Role of Neural Encoding

As discussed, probability density estimation, or any sort of neural probabilistic inference, requires an efficient representation of the statistics of the natural world in order to meet computational demands. The specific nature of this representation within many biological brains seems to be an encoding of natural stimuli in a manner that minimizes the correlation or mutual dependence between neurons [30,31,32,33,34]. A consequence of this computationally is that likelihoods in regard to a neural firing rate can be considered independent of the firing rates of neurons that code for different features. In this regard, the pop-out versus serial search distinction may be seen as an emergent property of this coding strategy. Since likelihoods associated with orientation statistics are considered independently of those that represent chromatic information, the conjunction of these features fails to elicit pop-out [3]. It is also interesting to note in support of this line of reasoning, that as radial and angular frequency are coded jointly within the cortex, a unique item defined by a conjunction of spatial frequency and orientation does result in a pop-out stimulus [35]. In light of this observation, it may be said more generally, that the specific nature of neuron properties has a considerable influence on the behavior that manifests. It is well established that search efficiency is more involved than a simple dichotomy of serial versus parallel searches [36]. It has been demonstrated that one can observe a wide range of behaviors from very efficient to very inefficient depending on the chosen stimuli. One might suggest that the extent to which a search may be carried out efficiently reflects the complexity of the neural code corresponding to target and distractor elements. For stimuli that are highly natural and may be represented by the response of a small number of neurons, one might expect a far more efficient search than that associated with a highly unnatural stimulus that gives rise to a widely distributed neural representation. This may also extend beyond simple V1-like features to explain the surprising efficiency with which some search tasks involving complex stimuli are completed, such as search tasks involving 3D-shape [37], depth from shading [38] and even very complex forms such as faces [39] which are known to have a highly efficient cortical representation within the primate cortex [40,41,42]. Considerations pertaining to coding may also shed some light on the role of novelty in determining search efficiency. Inter-element suppression of stimulus items may occur more strongly for those representations that are relatively efficient and carried by only a small number of cells. Behaviorally this is consistent with visual search paradigms in which familiarity with distractors yields a relatively efficient search [43,44] assuming familiarity with target items leads to a more efficient or even template like representation of the relevant stimuli. As a whole, it may be said that the role that principles underlying coding

within the visual cortex play within attention and visual search is an aspect of the problem that has been underemphasized. Many behaviors, in particular the specific efficiency with which a search is conducted, may be seen as properties that surface from very basic principles underlying the neural representation of visual patterns, and consideration of the specific role of coding in attention and visual search should serve as a target for further investigation.

7 Towards a Hierarchical Representation of Saliency

The preceding results demonstrate that the proposal originally tested on spatiochromatic data extends well to explain spatiotemporal data. A question that naturally follows from this, is the extent to which the proposal may extend to capture more high-level behaviors associated with neurons coding for more complex stimuli and appearing higher in the cortex. As the saliency associated with a pixel location is a simple summation of the individual saliency attributed to each cell for each location, it is evident that saliency may be evaluated at the level of a single cell. It follows that the same proposal that has been depicted in a form more akin to the traditional saliency map style representation may also reside within a distributed hierarchical representation in which the representation of saliency is implicit and computed via local modulation as opposed to a single explicit topographical representation of saliency. Such a proposal is in line with models of attention that posit a distributed hierarchical selection strategy [4]. Additionally, as the constraints on the cells involved are satisfied among higher visual areas, one might propose that the proposal put forth in AIM extends to higher visual areas to explain some of the apparent *high-level* effects documented in the previous section. For example, a hierarchical coding structure combined with AIM should afford some of the pop-out effects associated with high-level features such as depth from shading assuming an appropriate code for such features among higher visual areas.

8 Conclusion

We have considered how AIM extends to capture behaviors associated with visual patterns distributed over space and time. The plausibility of the proposal as a description of human behavior is validated through a comparison with eye tracking data on a wide range of qualitatively different videos. The proposal emerges as very effective in explaining the behavioral data as was demonstrated for the spatiochromatic case. We have also described how the proposal put forth in AIM is compatible with distributed architectures for attentional selection [4] including related details pertaining to coding and neural implementation. This is an important contribution as the topic of saliency [4] is seldom discussed in a context independent of the assumption of an explicit topographical saliency map. Future work will aim to further explore saliency computation as a process involving attention acting on a distributed hierarchical representation with saliency realized via localized modulation throughout the cortex.

Acknowledgments. The authors wish to thank Dr. Laurent Itti for sharing the eye tracking data employed in the evaluation of spatiotemporal saliency. The authors gratefully acknowledge the support of NSERC in funding this work. John Tsotsos is the NSERC Canada Research Chair in Computational Vision.

References

1. Bruce, N.D.B., Tsotsos, J.K.: Saliency Based on Information Maximization. In: Advances in Neural Information Processing Systems, vol. 18, pp. 155–162 (June 2006)
2. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)
3. Bruce, N.D.B., Tsotsos, J.K.: An information theoretic model of saliency and visual search. In: Paletta, L., Rome, E. (eds.) WAPCV 2007. LNCS, vol. 4840, pp. 171–183. Springer, Heidelberg (2007)
4. Tsotsos, J.K., Culhane, S., Yan Kei Wai, W., Lai, Y., Davis, N., Nuflo, F.: Modeling visual attention via selective tuning. Artificial intelligence 78, 507–545 (1995)
5. Bell, A.J., Sejnowski, T.J.: The 'Independent Components' of Natural Scenes are Edge Filters. Vision Research 37(23), 3327–3338 (1997)
6. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609 (1996)
7. Wachtler, T., Lee, T.-W., Sejnowski, T.J.: The chromatic structure of natural scenes. J. Opt. Soc. Amer. A 18(1), 65–77 (2001)
8. van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. Lond. B 265, 359–366 (1998)
9. Lee, T.W., Girolami, M., Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Computation 11(2), 417–441 (1999)
10. Itti, L., Baldi, P.: Bayesian Surprise Attracts Human Attention. In: Advances in Neural Information Processing Systems, vol. 19, pp. 547–554 (2006)
11. Yu, C., Levi, D.M.: Surround modulation in human vision unmasked by masking experiments. Nature 3(7), 724–728 (2000)
12. Williams, A.L., Singh, K.D., Smith, A.T.: Surround modulation measured with fMRI in the visual cortex. Journal of Neurophysiology 89(1), 525–533 (2003)
13. Xing, J., Heeger, D.J.: Measurement and Modeling of Centre-Surround Suppression and Enhancement. Vision Research 41, 571–583 (2001)
14. Shen, Z.M., Xu, W.F., Li, C.Y.: Cue-invariant detection of centre surround discontinuity by V1 neurons in awake macaque monkey. Journal of Physiology 583, 581–592 (2007)
15. Yu, C., Klein, A.K., Levi, D.M.: Cross-and Iso-oriented surrounds modulate the contrast response function: The effect of surround contrast. Journal of Vision 3, 527–540 (2003)
16. Petrov, Y., McKee, S.P.: The effect of spatial configuration on surround suppression of contrast sensitivity. Journal of Vision 6(3), 224–238 (2006)
17. Adini, Y., Sagi, D.: Recurrent networks in human visual cortex: psychophysical evidence. Journal of the Optical Society of America A 18(8), 2228–2236 (2001)

18. Olzak, L.A., Laurinen, P.I.: Contextual Effects in fine spatial discriminations. Nature 381(6583), 607–609 (2005)
19. Cannon, M.W., Fullencamp, S.C.: A model for inhibitory lateral interaction effects in perceived contrast. Vision Research 36(8), 1115–1125 (1996)
20. Xing, J., Heeger, D.J.: Centre-surround interactions in foveal and peripheral vision. Vision Research 40, 3065–3072 (2000)
21. Yu, C., Klein, A.K., Levi, D.M.: Surround modulation of perceived contrast and the role of brightness induction. Journal of Vision 1, 18–31 (2001)
22. Zhang, B., Zheng, J., Watanabe, I., Maruko, I., Bi, H., Smith, E.L., Chino, Y.: Delayed maturation of receptive field centre/surround mechanisms in V2. Proceedings of the National Academy of Sciences 102(16), 5862–5867 (2005)
23. Solomon, S.G., Pierce, J.W., Lennie, P.: The impact of suppressive surrounds on chromatic properties of cortical neurons. Journal of Neuroscience 24(1), 148–160 (2004)
24. Schein, S.J., Desimone, R.: Spectral properties of V4 Neurons in the macaque. Journal of Neuroscience 10(10), 3369–3389 (1990)
25. Kondo, H., Komatsu, H.: Suppression on neuronal responses by a metacontrast masking stimulus. Neuroscience Research 36(1), 27–33 (2000)
26. Tadin, D., Lappin, J.S.: Optimal Size for perceiving motion decreases with contrast. Vision Research 45, 2059–2064 (2005)
27. Born, R.T., Bradley, D.C.: Structure and Function of Visual Area MT. Annual Review of Neuroscience 28, 157–189 (2005)
28. Huang, X., Albright, T.D., Stoner, G.R.: Adaptive Surround Modulation in Cortical Area MT. Neuron 53(5), 761–770 (2007)
29. Eifuku, S., Wurtz, R.H.: Response to Motion in Extrastriate Area MSTI: Centre-Surround Interactions. Journal of Neurophysiology 80(11), 282–296 (1998)
30. Foldiak, P., Young, M.: Sparse coding in the primate cortex. In: Arbib, M.A. (ed.) The Handbook of Brain Theory and Neural Networks, pp. 895–898 (1995)
31. David, S.V., Vinje, W.E., Gallant, J.L.: Natural stimulus statistics alter the receptive field structure of v1 neurons. Journal of Neuroscience 24(31), 6991–7006 (2004)
32. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. Annual Review Neuroscience 24, 1193–1216 (2001)
33. Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., Fried, I.: Invariant visual representation by single neurons in the human brain. Proceedings of the National Academy of Science 102(16), 5862–5867 (2005)
34. Kreiman, G.: Neural coding: computational and biophysical perspectives. Physics of Life Reviews 2, 71–102 (2004)
35. Sagi, D.: The combination of spatial frequency and orientation is effortlessly perceived. Perception and Psychophysics 43, 601–603 (1988)
36. Wolfe, J.M., Horowitz, T.S.: What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience 5, 1–7 (2004)
37. Enns, J.T., Rensink, R.A.: Sensitivity to three-dimensional orientation in visual search. Psychological Science 1, 323–326 (1990)
38. Ramachandran, V.S.: Perception of Shape from Shading. Nature, 163–166 (1988)
39. Hershler, O., Hochstein, S.: At first sight: a high-level pop out effect for faces. Vision Research 45(13), 1707–1724 (2005)
40. Sergent, J., Ohta, S., MacDonald, B.: Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain 115(1), 15–36 (1992)

41. Kanwisher, N., McDermott, J., Chun, M.M.: The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience 17(11), 4302–4311 (2006)
42. Grill-Spector, K., Sayres, R., Ress, D.: High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nature Neuroscience 9(9), 1177–1185 (2006)
43. Wang, Q., Cavanagh, P., Green, M.: Familiarity and pop-out in visual search. Perception and Psychophysics 56(5), 495–500 (1994)
44. Shen, J., Reingold, E.M.: Visual search asymmetry: the influence of stimulus familiarity and low-level features. Perception and Psychophysics 63(3), 464–475 (2001)