Perceptual Salience as Novelty Detection in Cortical Pinwheel Space

Paul Sajda and Feng Han
Department of Biomedical Engineering, Columbia University, NY, USA

***Abstract*— We describe a filter-based model of orientation processing in primary visual cortex (V1) and demonstrate that novelty in cortical "pinwheel" space can be used as a measure of perceptual salience. In the model, novelty is computed as the negative log likelihood of a pinwheel's activity relative to the population response. The population response is modeled using a mixture of Gaussians, enabling the representation of complex, multi-modal distributions. Hidden variables that are inferred in the mixture model can be viewed as grouping or "binding" pinwheels which have similar responses within the distribution space. Results are shown for several stimuli that illustrate well-known contextual effects related to perceptual salience, as well as results for a natural image.**

***Keywords*— perceptual salience, novelty detection, orientation pinwheel, mixture of Gaussians, primary visual cortex**

I. Introduction

Perceptual salience is often defined as the degree to which a part of a visual scene "pops-out" [1] from the background. It is considered the bottom-up component of visual attention and is thus a major contributor to how the visual system parses a scene. In addition to its significance for visual neuroscience, a theory/model of perceptual salience could have implications for a number of applications, from improved cueing in computer-assisted medical image analysis to perceptually optimized progressive video transmission/compression.

The phenomenon of visual perceptual salience has been well-studied psychophysically, computationally and physiologically. Several models have emerged which attempt to explain the phenomenon. Many of these models are mechanistic, for example attempting to account for perceptual salience via well-characterized neural mechanisms. Itti and Koch [2] for example, build an attention model which is based on the construction of a saliency map. The saliency map represents the magnitude of filter responses, and is constructed as a combination of low-level feature responses (color, orientation, intensity) coupled with various degrees of facilitation and inhibition. The magnitude of the saliency map responses (i.e. firing rate in a neural network model) is the metric for perceptual salience. A winner-take-all network determines the most salient location in the scene. Yen and Finkel [3] and Li [4] focus on the role long-range horizontal connections play in perceptual salience. Yen and Finkel, for example, argue that facilitatory interactions between cortical neurons, mediated by long-range connections, will increase the levels of synchronized activity. In their model this level of synchronized activity is seen as a metric for salience.

An alternative approach is to define a probabilistic metric for salience. For example, we might define salient structure as something that is "novel" given the other structure in the image. A natural probabilistic metric for detecting novelty is the negative log likelihood, $-log(p(f|P))$, where f represents some "response value" for a given element in the scene and P represents the population over which one models the distribution of those values. Large values of the negative log likelihood indicate novel structure. A probabilistic definition of salience enables seamless incorporation of top-down components, via Bayes rule and inclusion of priors. It is less clear how to incorporate such information into a metric which is based on firing rate or synchronization. A key question, however, is how does one model $p(f|P)$.

One approach is to consider the population distribution of the response values to be Gaussian. In this case the negative log likelihood is given by the Mahalanobis distance,

$$-log(p(f|P)) = (f-\mu)^T \Sigma^{-1} (f-\mu) \qquad (1)$$

where μ is the estimated population mean and Σ the estimated population covariance. If f is vector valued, then we have a multi-variate Gaussian, with vector valued means and a covariance matrix. Rosenholtz showed that using the Mahalanobis distance as a simple salience metric, one can predict a number of motion pop-out phenomena [5]. However, critical in this analysis is that the stimuli are composed of elements that oscillate back and forth across a region of the image. This implies that, in velocity space, the distribution of element velocities is well-approximated by a Gaussian. In general, however, one would not expect a Gaussian distribution to adequately model the population response.

In this paper we describe a probabilistic model of perceptual salience which is based on detecting novel (i.e. large negative log likelihood) structure in a scene relative to the distribution of the population response. We construct the population distribution in orientation "pinwheel" [6] space. Pinwheel responses are estimated using a filter-based model which includes both classical receptive fields and long-range horizontal cortical interactions. We construct a distribution in pinwheel space using a mixture of Gaussians, providing a more flexible framework for modeling of complex, multi-modal, distributions. We compute perceptual salience as the negative log likelihood of a pinwheel's responses under the modeled population distribution. In the following we describe the details of the model and present simulation results demonstrating performance.

II. The Model

We construct a simple filter-based model of primary visual cortex (V1). Our model of V1 is tessellated with an array of orientation pinwheels, which receive input retinotopically from the visual scene. Filters are organized into orientation pinwheels, with each wedge of the pinwheel indicating a different preferred orientation. Input to a given wedge comes

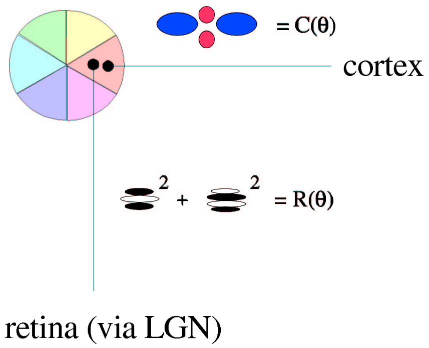

Fig. 1. Inputs to the orientation pinwheel. Each wedge of the pinwheel represents filters (i.e. neurons) tuned to a given orientation θ. Input directly from the scene (from the Retina via the LGN) is oriented energy, $R(\theta)$, computed via quadrature mirror filters. Input from other cortical pinwheels is orientation specific, with the cortical interaction filter, $C(\theta)$, applied to only wedges of other cortical pinwheels having the same orientation preference.

both directly from the visual scene (i.e. from retina via the lateral geniculate nucleus (LGN)) and from other cortical pinwheels, via long-range lateral connections. Figure 1 illustrates these two sets of inputs. The component of the pinwheel response due to retinal input, $f_r(\theta)$ is given by the convolution,

$$f_r(\theta) = (R_e(\theta) \otimes I)^2 + (R_o(\theta) \otimes I)^2 + n \qquad (2)$$

where θ determines the orientation tuning of the filter and therefore the wedge of the pinwheel, I is the image (intensities) and $R_e(\theta)$ and $R_o(\theta)$ are a pair quadrature mirror filters (QMF) [7] and n is Gaussian noise. $f_r(\theta)$ can be considered the localized oriented energy in the scene. The pinwheel response due to input from other cortical pinwheels, $f_c(\theta)$, is computed by convolving a cortical interaction function, $C(\theta_i)$ with $f_r(\theta_j)$,

$$f_c(\theta) = \begin{cases} C(\theta_i) \otimes f_r(\theta_j) + n & \theta_i = \theta_j \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Figure 2 illustrates the cortical interaction function for a pinwheel wedge with a preferred vertical orientation. This function is in good agreement with the empirically determined non-linear contextual maps of neurons in V1 [8]. The effect of the cortical interaction function is to force, via facilitation, collinear elements to have similar firing rates, which in turn forces their clustering in pinwheel space.

The total response of a pinwheel is given by the product of the retinal and cortical components,

$$f(\theta) = f_r(\theta)(1 + f_c(\theta)) \qquad (4)$$

Note that this equation implies that the retinal component must be non-zero for there to be a response, while the cortical component acts more or less as a modulation term. This is in agreement with the neurophysiology [9].

A pinwheel's response can be represented as a coordinate in an N dimensional space, where N indicates the number of preferred orientations (and therefore the number of wedges in a pinwheel). Figure 3A shows an example of this space for $N = 3$. Note that proximity in this space implies similar responses/firing rates.

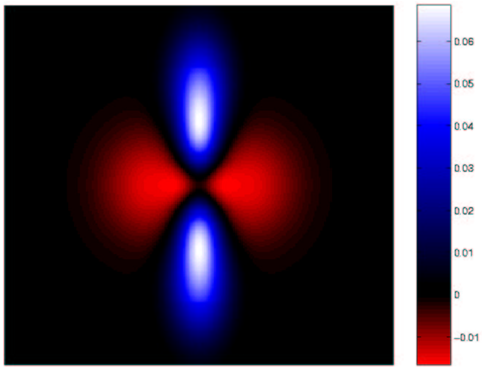

Fig. 2. The cortical interaction filter, $C(\theta)$. The filter is modeled after the non-linear context maps, established physiologically and psychophysically, by Gilbert et al [8]. There are two primary zones of facilitation, oriented collinear, and two zones of inhibition, oriented orthogonal to the preferred orientation. For this example, the cortical integration filter is shown for preferred orientation of 90^o (vertical).

Perceptual salience of a given pinwheel, S_i, is computed as the novelty of a pinwheel response relative to the population, P. Novelty is quantified as the negative log likelihood of a pinwheel response given the population distribution,

$$S_i = -log(p(f_i|P)). \qquad (5)$$

We estimate the population distribution using a mixture of multi-variate Gaussians [10]. The distribution of the population is therefore given by,

$$p(f|P) = \sum_c p(f|c, P)p(c|P) \qquad (6)$$

where $p(f|c, P)$ is modeled as a spherically symmetric multivariate Gaussian, $N(\mu, \mathbf{I}\sigma)$, and $p(c|P)$ is modeled as a normalized probability $\frac{\pi_c}{\sum_c \pi_c}$ and represents the probability of the pinwheel response under Gaussian component c. Estimation of the parameters in the model (μs, σs and πs) is accomplished using the expectation-maximization (EM) algorithm [11].

Figure 3 illustrates the significance of modeling the population response as a mixture of Gaussians (Figure 3C) compared to to a single Gaussian (Figure 3B). A single Gaussian is a poor model for the multi-modal population distribution. Computing the novelty of a pinwheel response will result in incorrect estimates of salience. However, approximating the population distribution as a mixture of Gaussians enables the population distribution to be more accurately modeled.

III. Results

To demonstrate our model we first compute perceptual salience for the four images shown in Figure 4. We model pinwheels as having six orientations (wedges). The population distribution is modeled as a mixture of Gaussians, with

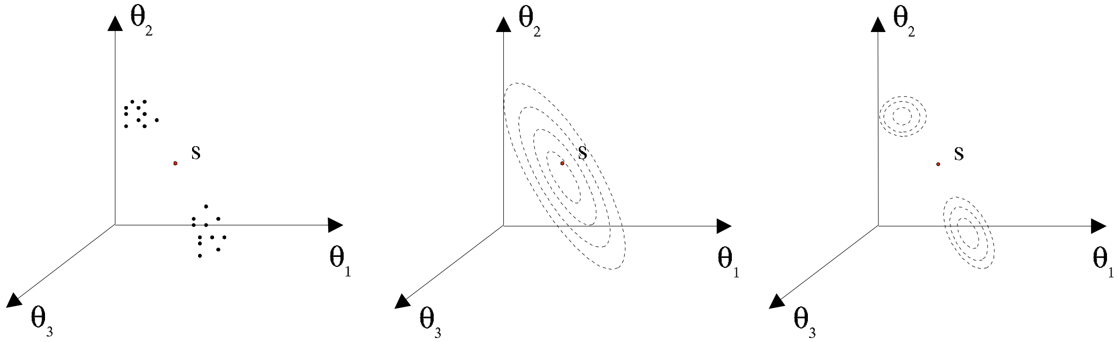

Fig. 3. Illustration of salience as novelty in cortical pinwheel space. Only three dimensions of the space are shown. (A) Clustering of pinwheel responses. (B) Responses modeled as as a Gaussian pdf (i.e. negative log likelihood given by the Mahalanobis distance). Note that point s would not be salient. (C) Responses modeled as mixture of Gaussians. In this case s should have a large negative log likelihood and would be salient.

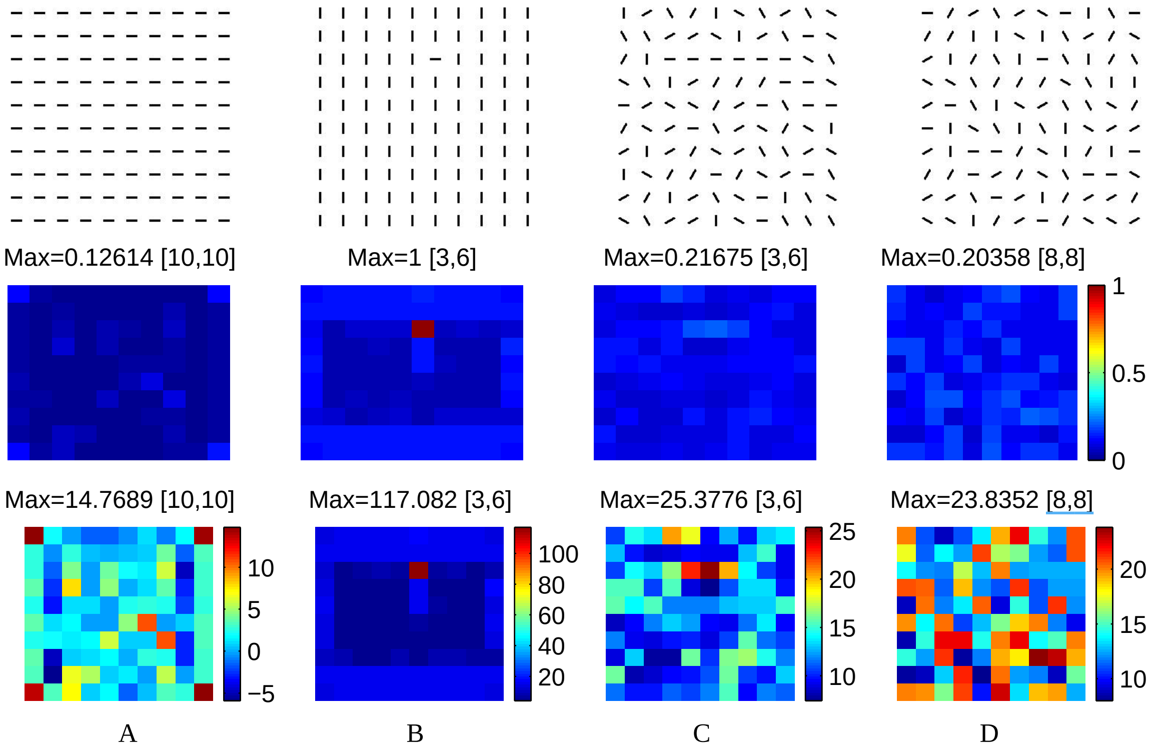

Fig. 4. Results for four stimuli demonstrating contextual effects on perceptual salience. The distribution in pinwheel space is computed using a mixture of Gaussians. (top) Stimuli used in the experiments. (middle) Normalized (across all stimuli) negative log likelihood. Also shown for each case is the maximum value of the normalized negative log likelihood, and its spatial coordinate. Normalized values can be used to qualitatively determine the relative salience of structure in different images. (bottom) Unnormalized negative log likelihoods. Also shown for each case is the maximum value and its spatial coordinate.

2-6 components.[1] The fours images consist of oriented bars placed in a 10-by-10 grid. Each of the four images has a horizontal bar at $[3, 6]$ (we will refer to this as the target). [2] In spite of having the same orientation at $[3, 6]$, the perceptual salience at location $[3, 6]$ is very different for the four images, which demonstrates the well-known contextual effects that modulate salience.

Results show that the model is in qualitative agreement with perceived salience. In case (A) all data are well modelled by the mixture distribution and there is no pop-out. In case (B) the target bar pops out from the iso-orientation background, as indicated by the large negative log likelihood. In case (C) flankers on each side of the target bar induce contextual effects which make the target salient compared with the randomly oriented bars in the background. In case (D) all bars have a randomly distributed orientation. The relative salient locations indicate colinear or co-circular bars,

[1] Note that the number of components is a parameter and can be estimated using a number of model selection criteria. However we do not address the issue of component selection in this paper.

[2] We index the images as a matrix [rows, columns] with $[1, 1]$ being the top left corner.

which occur by chance.

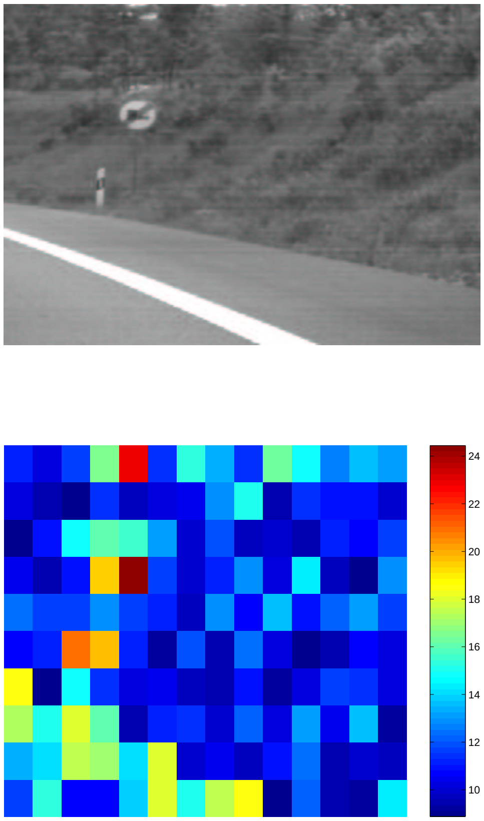

Fig. 5. Perceptual salience results for a natural image.

Results applying the model to a real image are shown in Figure 5 (in this case we use a 10-by-14 pinwheel array). We see that the model determines that the traffic sign is most novel under the population distribution, and therefore the most salient. Also seen as novel are the road post and lane demarcation (white line). Finally, novel structure is seen near the top edge of the image (location $[1, 6]$ in pinwheel space). Though this structure is salient relative to the rest of the background trees, its novelty is also likely due to edge effects in the model.

IV. Conclusion

In this paper we have argued that perceptual salience can be viewed as novelty detection in an appropriate feature space. We consider responses of orientation pinwheels as a feature space and construct a simple filter-based model of V1 to estimate those responses. The model includes input from what is considered the classical receptive field as well as contextual inputs mediated by long-range horizontal connections. The pinwheel responses are represented as points in a high dimensional space and the population distribution is estimated using a mixture of multi-variate Gaussians. The negative log likelihood is computed for each pinwheel response and used as an estimate of an element's salience. We demonstrate that the model is in qualitative agreement with perceived salience.

It is interesting to consider how the nervous system might compute the population distribution. It is unlikely that an EM algorithm is directly implemented in the cortex. However critical is determining which cluster each pinwheel should be assigned. In our approach we compute the likelihood of a pinwheel being assigned to a given cluster u, (i.e. $p(f_i|c = u, P)$). Since clusters of pinwheels tend to have similar responses, and therefore firing rates, it may be that pinwheels are assigned to clusters via synchronization of their responses. It has been demonstrated that synchronization of neural responses occurs when firing rates are nearly equal [12] [13]. Synchronization has been implicated as a possible mechanism for binding/grouping elements in a scene. In the case of our model, we would argue that synchronization is a mechanism for estimating the probability of a response, given an underlying hidden variable (cluster) and therefore it is directly linked to a probabilistic representation of the scene. Future work will consider more biologically-plausible methods for estimating the population distribution.

Acknowledgments

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-01-1-0625 as well as a grant from the National Imagery and Mapping Agency, NMA201-02-C0012.

References

[1] A. Treisman, "Features and objects: The fourteenth Bartlett memorial lecture." *Quarterly Journal of Experimental Psychology*, vol. 40A, pp. 201–236, 1988.

[2] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol. 40, no. 10–12, pp. 1489–1506, 2000.

[3] S.-C. Yen and L. Finkel, "Extraction of perceptually salient contours by striate cortical networks," *Vision Research*, vol. 38, no. 5, pp. 719–741, 1998.

[4] Z. Li, "Contextual influences in V1 as a basis for pop out and asymmetry in visual research." *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, pp. 10 530–10 535, 1999.

[5] R. Rosenholtz, "A simple saliency model predicts a number of motion popout phenomena," *Vision Research*, vol. 39, pp. 3157–3163, 1999.

[6] T. Bonhoeffer and A. Grinvald, "Iso-orientation domains in cat visual cortex are arranged in pinwheel like patterns," *Nature*, vol. 353, pp. 429–431, 1991.

[7] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-13, no. 9, pp. 891–906, 1991.

[8] M. Kapadia, G. Westheimer, and C. Gilbert, "Spatial distribution of contextual interactions in primary visual cortex and in visual perception," *The American Physiological Society*, pp. 2048–2062, 2000.

[9] M. Kapadia, M. Ito, C. Gilbert, and G. Westheimer, "Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys," *Neuron*, vol. 15, pp. 843–856, 1995.

[10] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. New York: John Wiley and Sons, 2001.

[11] M. I. Jordan and R. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, vol. 6, pp. 181–214, 1994.

[12] P. Matthews, R. Mirollo, and S. Strogatz, "Dynamics of a large system of coupled nonlinear oscillators," *Physica D*, vol. 52, pp. 293–331, 1991.

[13] M. Tsodyks, I. Mitkov, and H. Sompolinsky, "Pattern of synchrony in inhomogeneous networks of oscillators with pulse interactions," *Phys. Rev. Lett.*, vol. 71, pp. 1280–1283, 1993.