Appears in *IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007

Bottom-up saliency is a discriminant process

Dashan Gao Nuno Vasconcelos
Department of Electrical and Computer Engineering
University of California, San Diego
{dgao, nuno}@ucsd.edu

Abstract

A bottom-up visual saliency detector is proposed, following a decision-theoretic formulation of saliency, previously developed for top-down processing (object recognition) [5]. The saliency of a given location of the visual field is defined as the power of a Gabor-like feature set to discriminate between the visual appearance of 1) a neighborhood centered at that location (the center) and 2) a neighborhood that surrounds it (the surround). Discrimination is defined in an information-theoretic sense and the optimal saliency detector derived for a class of stimuli that complies with known statistical properties of natural images, so as to achieve a computationally efficient solution. The resulting saliency detector is shown to replicate the fundamental properties of the psychophysics of pre-attentive vision, including stimulus pop-out, inability to detect feature conjunctions, asymmetries with respect to feature presence vs. absence, and compliance with Weber's law. It is also shown that the detector produces better predictions of human eye fixations than two previously proposed bottom-up saliency detectors.

1. Introduction

It has long been known that mechanisms of selective visual attention play an important role in biological vision [28]. By identifying certain regions of the visual field as more important, or *salient*, than others they enable a non-uniform allocation of perceptual resources that eases the computational burden posed, to an observer, by pattern recognition or other visual tasks. The deployment of visual attention has long been believed to be driven by the interaction of two complementary components: a *bottom-up*, fast, *stimulus-driven mechanism*, and a *top-down*, slower, *goal-driven mechanism.* While various bottom-up [11, 2, 13] and top-down [27, 5, 17] saliency algorithms have been proposed in the computer and biological vision literatures, little has been achieved in what concerns the development of a unified framework for the two saliency components. In fact, very little has been proposed in terms of generic principles that could drive the design of both bottom-up and top-down saliency detectors.

One exception is the principle of *discriminant saliency*, initially proposed in [5] for visual recognition problems. It defines as salient the features whose response best distinguishes a visual concept (e.g. object) to recognize from all others that may be of possible interest (e.g. the set of all other object classes that compose the recognition problem). Discriminant saliency has, so far, been applied to the design of object recognition systems, a task where the resulting saliency detector has been shown to perform well [5, 7]. While this poses the principle as a purely top-down approach to saliency, the idea of equating saliency with discrimination applies equally well to the problem of bottom-up saliency. This is due to the ubiquity of "center-surround" mechanisms in the early stages of biological visual systems [9]. Such mechanisms could be naturally interpreted as detectors of features that are distinct from the surrounding background.

It should be noted that, although top-down saliency may have greater immediate value for computer vision (given the connections to object recognition), it is important to study saliency in the bottom-up context. The reason is that the bottom-up visual pathway of biological vision is much better understood than its top-down counterpart. Comparing the predictions of bottom-up discriminant saliency with the vast repository of results available in the psychophysics literature is a natural strategy to test the underlying assertion that saliency is, in general, a discriminant process. This has motivated us to study the effectiveness of discriminant saliency as a driving principle for bottom-up saliency.

The contributions of this study, which we present here, are four-fold. First, we derive the bottom-up detector itself. Like its top-down counterpart [5], it is optimal in a decision-theoretic sense, but with respect to center-surround discrimination, not object recognition. Second, in the spirit of Barlow and others [1], we show that by exploiting the regularities of the visual world, it is possible to implement the optimal detector with computational efficiency. In particular, we show that by exploiting 1) previously observed properties of feature dependencies, and 2) a widely used model of the statistics of natural image features, a generalized Gaussian distribution, the optimal detector can be implemented with great computational simplicity. Third, we show that the proposed model is compatible with the psychophysics of human pre-attentive vision. In particular, it is shown that discriminant saliency replicates various fundamental properties of human pre-attentive vision, including stimulus pop-out, disregard of feature conjunctions, and saliency asymmetries. Finally, in a more engineering oriented vein, the performance of the discriminant detector is compared with those of two other bottom-up saliency detectors previously proposed in the literature. This experimental comparison addresses both the ability to replicate classical psychophysics results, and the ability to predict human eye fixations on natural scenes. It is shown that, in both cases, discriminant saliency achieves superior performance.

2. Previous work

Computational modeling of bottom-up visual saliency has been a subject of interest, for a few decades, in both computer and biological vision.

2.1. Previous work

Saliency has a long history in computer vision. A substantial amount of work aims to model mechanisms of perceptual organization, such as contour saliency [19], illusory contours [26], or more general Gestalt phenomena. This goes beyond the problem addressed here, the detection of salient locations in the visual field, which is closer to what is often called "interest point" detection in computer vision. While interest point detectors have been successful in various applications, e.g. object tracking [6] or recognition [12, 16], they are hardly related to biological attention. They are also designed with respect to cost functions (e.g., scale and affine invariance [16]), which are less central to perception than discrimination, the essential component of optimal decision making. While, in the near future, we intend to evaluate the performance of discriminant saliency with respect to properties such as stability, this goes beyond what is addressed here.

A large corpus of existing saliency detectors has been inspired by, or aims to replicate, known properties of the psychophysics and physiology of pre-attentive vision [11, 13]. Many of these detectors do not propose a clear unifying computational principle for their various steps. Others have justified all computations as optimal under generic saliency principles, such as the maximization of self-information [2] or "surprise" [10]. Our comparisons focus on this body of biologically plausible detectors, namely [11], which is arguably the most popular detector in current use [25], and [2], one of the most recent attempts at deriving detectors that are optimal under generic principles.

Finally, to the best of our knowledge, there has been no previous effort, in the literature, to develop a unified formulation for both bottom-up and top-down saliency. In this context, the discriminant saliency principle, first proposed for top-down processing in [5], is of particular interest.

2.2. Discriminant saliency

Discriminant saliency is rooted on a decision-theoretic interpretation of perception. Saliency is defined with respect to a stimulus of interest and a *null* hypothesis, composed of stimuli that are not salient. Once this null hypothesis is available, the locations of the visual field that can be classified, with *lowest probability of error*, as not belonging to it are classified as salient. Mathematically, this is accomplished by 1) defining a binary classification problem that opposes the stimulus of interest to the null hypothesis, 2) finding the visual features that are most discriminating for this problem, and 3) equating the saliency of each location in the visual field to the strength of response of these features, at the location.

This definition has, at least, two interesting properties. First, different specifications of the stimulus of interest and the null hypothesis enable its specialization to top-down or bottom-up saliency. Second, the search for discriminant features is a well-defined, and computationally tractable, problem that has been widely studied in the literature. In [5], a top-down saliency detector has been derived by equating the stimulus of interest to an object class and the null hypothesis to the set of objects in all other classes. In this work, we consider the problem of bottom-up saliency.

3. Bottom-up discriminant saliency

A common formulation for bottom-up saliency is that the saliency of each location is a function of how distinct it is from the surrounding background [11]. This formulation is supported by the ubiquity of "center-surround" mechanisms in the early stages of biological vision [9]. We next introduce a discriminant solution for center-surround saliency.

3.1. Mathematical formulation

Center-surround saliency can be formulated in decision-theoretic terms by 1) defining the stimulus of interest, at location l, as the visual appearance within a neighborhood $\mathcal{W}_l^1$ of l (the *center*), 2) the null hypothesis as the visual appearance within a surrounding window $\mathcal{W}_l^0$ (the *surround*), and 3) searching for the location l^* where the responses of a pre-defined feature set are most discriminant for the decision between *center* and *surround*.

The feature responses within the two windows are interpreted as observations from a random process $\mathbf{X}(l) =$

$(X_1(l), \dots, X_d(l))$, of dimension d, conditioned on the state of a hidden random variable $Y(l)$. Observations in $\mathcal{W}_l^1$ are drawn when $Y(l) = 1$ while surround observations ($\mathcal{W}_l^0$) are drawn when $Y(l) = 0$. The feature vectors observed in each region are, therefore, drawn according to the conditional densities $P_{\mathbf{X}(l)|Y(l)}(\mathbf{x}|c), c \in \{0, 1\}$. Observations drawn with $Y(l) = c$ are referred to as belonging to class c. The observed vector at any location j is denoted by $\mathbf{x}(j) = (x_1(j), \dots, x_d(j))$, and the saliency of the feature set at location l, $\mathbf{X}(l)$, is defined by how discriminant the feature responses are for the classification of the observations $\mathbf{x}(j), \forall j \in \mathcal{W}_l = \mathcal{W}_l^0 \cup \mathcal{W}_l^1$, into center and surround. In particular, the saliency of location l, $S(l)$, is quantified by the discriminant power of the entire feature set at l, as measured by the mutual information between features, $\mathbf{X}$, and class label, Y,

$$I_l(\mathbf{X}; Y) = \sum_c \int p_{\mathbf{X}(l),Y(l)}(\mathbf{x}, c) \log \frac{p_{\mathbf{X}(l),Y(l)}(\mathbf{x}, c)}{p_{\mathbf{X}(l)}(\mathbf{x}) p_{Y(l)}(c)} d\mathbf{x}. \tag{1}$$

The l subscript emphasizes the fact that the mutual information is defined locally, within $\mathcal{W}_l$, and saliency detection consists of identifying the locations where (1) is maximal.

3.2. Computational parsimony

The exact maximization of (1) is usually impractical, since it requires density estimates on a potentially high-dimensional feature space. Computational efficiency can be achieved by exploiting a known property of the statistics of band-pass natural image features, e.g. Gabor or wavelet coefficients: that features in this class exhibit strongly *consistent* patterns of dependencies across a wide range of imagery [3, 8]. These regularities are illustrated by Figure 1, which presents three images, the histograms of one coefficient of their wavelet decomposition, and the histograms of that coefficient conditioned on its parent. Although the drastically different visual appearance of the images affects the scale (variance) of the marginal distributions, *their shape, or that of the conditional distributions between coefficients,* is quite stable. The observation that these distributions follow a canonical (bow-tie) pattern, is remarkably consistent over the set of natural images. This consistency indicates that, even though the fine details of feature dependence may vary from scene to scene, its coarse structure follows a universal statistical law that appears to hold for all natural scenes. This, in turn, suggests that feature dependencies are not greatly informative about the image class [24, 3] or, in the particular case of saliency, about whether observations originate in the center or surround. The following theorem (see [23] for a proof) shows that, when this is the case, (1) can be drastically simplified.

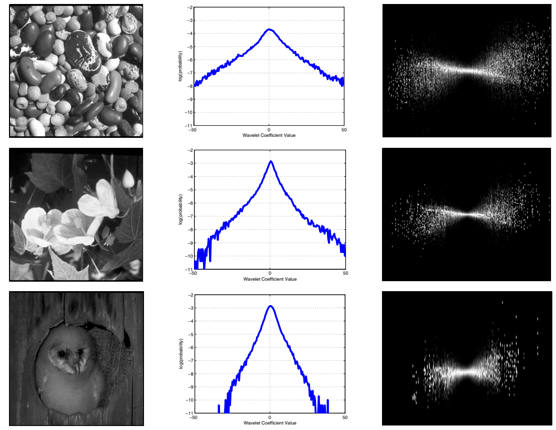


Figure 1. Constancy of natural image statistics. Left: three images. Center: each plot presents the histogram of the same coefficient from a wavelet decomposition of the image on the left. Right: conditional histogram of the same coefficient, conditioned on the value of its parent.

Theorem 1. *Let $\mathbf{X} = \{X_1, \dots, X_d\}$ be a collection of features, and Y the class label. If*

$$\frac{\sum_{i=1}^d [I(X_i; \mathbf{X}_{1,i-1}) - I(X_i; \mathbf{X}_{1,i-1}|Y)]}{\sum_{i=1}^d I(X_i; Y)} = 0, \tag{2}$$

where $\mathbf{X}_{1,i} = \{X_1, \dots, X_i\}$, then

$$I(\mathbf{X}; Y) = \sum_{i=1}^d I(X_i; Y). \tag{3}$$

The left hand side of (2) is a measure of the ratio between the information for discrimination contained in feature dependencies and that contained in the individual features. While this ratio is usually non-zero, it is generally small for band-pass natural image features [24]. Hence, the approximation of saliency, (1), by

$$S(l) \quad = \quad \sum_{i=1}^d I_l(X_i; Y), \tag{4}$$

is a sensible compromise between decision theoretic optimality and computational parsimony. Note that this approximation *does not* assume that the features are independently distributed, but simply that their dependencies are not informative about the class. The function $S(l)$ is referred to as the *saliency map*.

3.3. Leveraging well known statistical properties

Extensive research on the statistics of natural images has shown that, for band-pass features, these densities are well approximated by a generalized Gaussian distribution (GGD) [14],

$$P_X(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left\{-\left(\frac{|x|}{\alpha}\right)^\beta\right\}, \tag{5}$$

where $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} \mathrm{d}t$, $t > 0$, is the Gamma function, α a *scale* parameter, and β a *shape* parameter. The parameter β controls the decay rate from the peak value, and defines a sub-family of the GGD (e.g., the Laplacian family when $\beta = 1$ or the Gaussian family when $\beta = 2$).

Whenever the class conditional densities, $P_{X|Y}(x|c)$, and the marginal density, $P_X(x)$, follow a GGD, the mutual information of (4) can be computed in a closed form. This follows from the equalities

$$I(X;Y) = \sum_c P_Y(c) KL\left[P_{X|Y}(x|c)||P_X(x)\right], \quad (6)$$

and

$$KL[P_X(x;\alpha_1,\beta_1)||P_X(x;\alpha_2,\beta_2)] = \quad (7)$$
$$\log\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right) + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1},$$

where $KL[p||q] = \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$ is the Kullback-Leibler (K-L) divergence between $p(x)$ and $q(x)$. In this case, the computation of discriminant saliency at an image location only requires the estimation of the α and β parameters, for the center and surround widows, at that location. These parameters can be estimated by the method of moments, through the following equalities,

$$\sigma^2 = \frac{\alpha^2\Gamma(\frac{3}{\beta})}{\Gamma(\frac{1}{\beta})} \quad \text{and} \quad \kappa = \frac{\Gamma(\frac{1}{\beta})\Gamma(\frac{5}{\beta})}{\Gamma^2(\frac{3}{\beta})} \quad (8)$$

where σ^2 and κ are variance and kurtosis, defined as,

$$\sigma^2 = E_X[(X - E_X[X])^2] \;\text{ and }\; \kappa = \frac{E_X[(X - E_X[X])^4]}{\sigma^4}.$$

The estimation only requires computing sample moments from the center and surround windows, and is very efficient. It has also been shown to produce good fits to natural images [8].

3.4. Implementation details

The combination of all observations discussed in the previous sections leads to a discriminant saliency detector whose implementation is illustrated in Figure 2. The first stage, feature decomposition, follows the proposal of [11], which closely mimics the earliest stages of biological visual processing. The image to process is first subject to a feature decomposition into an intensity map and four broadly-tuned color channels, i.e. $I = (r+g+b)/3$, $R = \lfloor\tilde{r} - (\tilde{g}+\tilde{b})/2\rfloor_+$, $G = \lfloor\tilde{g} - (\tilde{r}+\tilde{b})/2\rfloor_+$, $B = \lfloor\tilde{b} - (\tilde{r}+\tilde{g})/2\rfloor_+$, and $Y = \lfloor(\tilde{r}+\tilde{g})/2 - |\tilde{r}-\tilde{g}|/2\rfloor_+$, where $\tilde{r} = r/I$, $\tilde{g} = g/I$, $\tilde{b} = b/I$, and $\lfloor x\rfloor_+ = \max(x, 0)$. The four color channels are, in turn, combined into two color opponent channels, $R - G$ for red/green and $B - Y$ for blue/yellow opponency. These

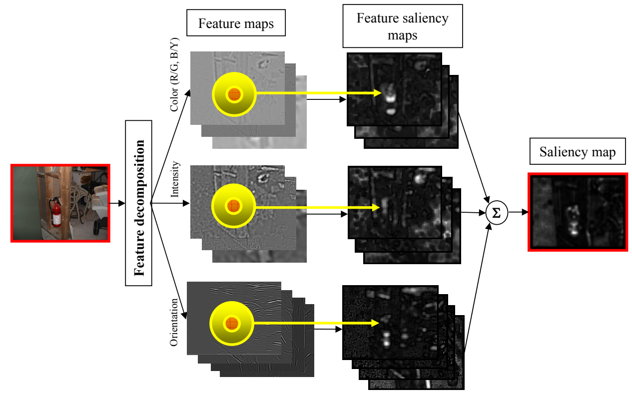


Figure 2. Bottom-up discriminant saliency detector.

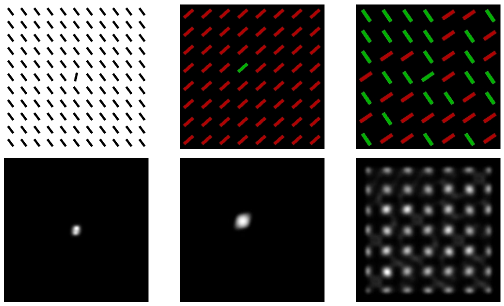
and the intensity map are convolved with three Mexican hat wavelet filters, centered at spatial frequencies 0.02, 0.04 and 0.08 cycle/pixel, to generate nine feature channels. The feature space $\mathcal{X}$ consists of these channels, plus a Gabor decomposition of the intensity map, implemented with a dictionary of zero-mean Gabor filters at 3 spatial scales (centered at frequencies of 0.08, 0.16, and 0.32 cycle/pixel) and 4 directions (evenly spread from 0 to π) [15].

For the computation of the feature saliency maps in the second stage, all window sizes are guided by studies from psychophysics and neurophysiology [18, 4]. For the psychophysics experiments of Section 4, we followed the common practice [21, 9] of setting the size of the center window to a value *comparable* to that of the display items. For natural images we used a window of 25×25 pixels. In all cases, the side of the surround window is 6 times larger than that of the center, at all image locations. Informal experimentation with these parameters has shown that the saliency results are not substantively affected by variations around the values adopted.

At each location, the parameters of the class conditional, and marginal densities are estimated with (8), and the mutual information between each feature X_i and class label Y is computed with (7) and (6). These mutual informations are finally added, according to (4), to generate the overall saliency map. To improve their intelligibility, the saliency maps shown in this paper were subject to smoothing, contrast enhancement (by squaring), and a normalization that maps the saliency value to the interval $[0, 1]$.

4. Experimental evaluation

The performance of discriminant saliency was evaluated by measuring its ability to 1) replicate classic psychophysics results in visual search, and 2) predict human eye fixations on natural images.



(a) (b) (c)

Figure 3. Saliency output for single basic features (orientation (a) and color (b)), and conjunctive features (c). Brightest regions are most salient.

4.1. Consistency with psychophysics

We start with an evaluation on a series classical displays used in the studies of visual attention [21, 22]. Discriminant saliency is compared with both human data and the model of [11], which is arguably the most popular biologically inspired model in the literature. All comparisons are based on the code available in [25].

4.1.1 Pop-out and conjunctive feature search

One classical observation from visual attention is that for basic features, such as color and orientation, the search for a target which differs from a set of distractors by a single feature is efficient. While, in this case, the target "pops-out", the same does not occur when the difference is defined by a conjunction of two basic features. Without top-down guidance, searching for conjunctions can be very difficult.

Some examples of this behaviour are shown in the top row of Figure 3, where a target differs from a field of distractors in terms of (a) orientation, (b) color, and (c) a conjunction of orientation and color (green right-tilted bar among green left-tilted and red right-tilted bars). The saliency maps produced by discriminant saliency are shown below each display. Note that, like human subjects, the detector produces a very unambiguous judgement of saliency for single feature search ((a) and (b)), but is unable to assign a high saliency to the conjunctive target in (c) (bar in the 4^{th} line and 4^{th} column).

The difference between single and conjunctive search has long been known, and probably best explained by Treisman's feature-integration theory (FIT) [21]. This theory predicts that the visual stimulus is projected into basic feature maps, which are then combined into a *saliency* map that drives attention. The saliency map is scalar and only registers the degree of relevance of each location to the search, but not which features are responsible for it. Hence, a target defined by a basic feature "pops-out", but a conjunctive target does not.

While the difficulty of searches for conjunctive targets is widely acknowledged in the literature, we are aware of no previous *computational* explanation of why the preattentive vision would choose to disregard conjunctions. Discriminant saliency justifies this behaviour, by explaining it as optimal, in a decision-theoretic sense, under sensible approximations that exploit the regularities of natural stimuli to achieve computational parsimony. To the degree that (2) holds for natural scenes, i.e. that feature dependencies are not informative for discrimination of image classes, restricting search to the analysis of individual feature maps has no loss of optimality.

4.1.2 Visual search asymmetries

Another classical result in psychophysics is the existence of saliency asymmetries [22]. While, in general, the presence in the target of some feature absent from the distractors produces pop-out, the reverse (pop-out due to the absence, in the target, of a distractor feature) does not hold. We applied the discriminant saliency detector to a set of classic displays and observed the same asymmetric responses. An example is shown in Fig. 4. On the left display, the target (a "Q") differs from the distractors ("O") by the addition of a vertical line, and is highly salient. However, as shown on the right, when the target ("O") differs from the distractors ("Q") by the absence of the same line, it is not salient.

In a quantitative study, Treisman [22] showed that 1) asymmetries also exist for weaker and stronger responses (i.e. when target differs from distractors only *quantitatively*, along one feature dimension), and 2) they follow Weber's law. For this, she designed a set of experiments involving displays where the target (a vertical bar) differs from distractors (a set of identical vertical bars) only in terms of length [22]. One example of such displays is shown in Figure 5 (a), while (b) presents a scatter plot of measurements of discriminant saliency across the set of displays.

Each point in (b) corresponds to the target saliency in one display, and the dashed line shows that, like human perception, discriminant saliency follows Weber's law: target saliency is approximately linear in the ratio between the difference of target/distractor length (Δx) and distractor length (x). For comparison, Figure 5 (c) presents the corresponding scatter plot for the model of [11], which is shown not to replicate human perception.

4.2. Predicting human eye movement data

In addition to classical psychophysics, saliency was evaluated by measuring how well saliency detectors can predict human eye fixations. In this case, the performance of discriminant saliency was compared to those of the methods of [11] and [2]. The experimental set-up followed [2], using the evaluation metric of [20], namely the area under ROC

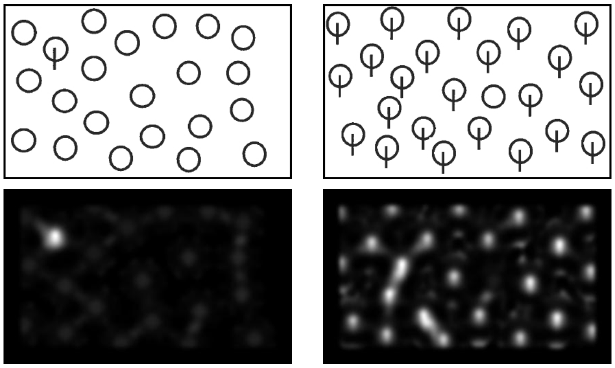
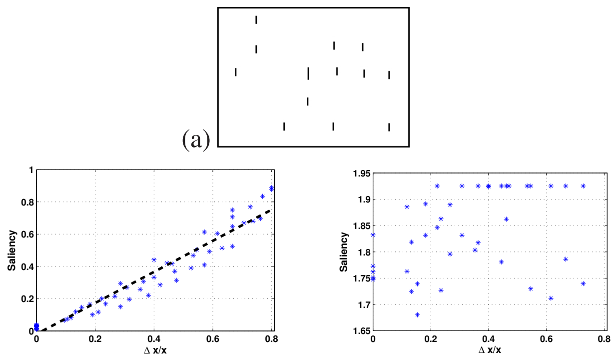


Figure 4. A pop-out asymmetry.



(a) (b) (c)

Figure 5. An example display (a) and performance of saliency detectors (discriminant saliency (b) and [11] (c)) on Treisman's Weber law experiment.

Saliency model	Discriminant	[11]	[2]
ROC area	0.7694	0.7287	0.7547

Table 1. ROC areas for different saliency models with respect to all human fixations.

curve for a binary prediction of eye fixations.

Table 1 presents average ROC areas for all detectors, across the entire image set. It shows that discriminant saliency achieves the best performance among the three saliency detectors.

Acknowledgments

The authors thank Neil Bruce for kindly sharing the eye fixation data, and saliency predictions of [2]. This research was supported by NSF award IIS-0448609.

References

[1] H. B. Barlow. Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, page 217. MIT Press, Cambridge, MA, 1961.

[2] N. D. Bruce and J. K. Tsotsos. Saliency based on information maximization. In *Proc. NIPS*, 2005.

[3] R. Buccigrossi and E. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, 8:1688–1701, 1999.

[4] J. Cavanaugh, W. Bair, and J. Movshon. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.*, 88:2530–2546, 2002.

[5] D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In *Proc. NIPS*, pages 481–488, 2004.

[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.

[7] A. B. Hillel, T. Hertz, and D. Weinshall. Efficient learning of relational object class models. In *Proc. IEEE ICCV*, 2005.

[8] J. Huang and D. Mumford. Statistics of Natural Images and Models. In *Proc. IEEE Conf. CVPR*, 1999.

[9] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.*, 28:229–289, 1965.

[10] L. Itti and P. Baldi. A principled approach to detecting surprising events in video. In *Proc. IEEE Conf. CVPR*, San Siego, CA, Jun 2005.

[11] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506, 2000.

[12] T. Kadir and M. Brady. Scale, saliency and image description. *Int'l. J. Comp. Vis.*, 45:83–105, Nov. 2001.

[13] Z. Li. A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1):9–16, 2002.

[14] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. PAMI*, 11(7):674–693, 1989.

[15] B. S. Manjunath and W. Y. Ma. Texture feature for browsing and retrieval of image data. *IEEE Trans. PAMI*, 18(8):837–842, 1996.

[16] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[17] V. Navalpakkam and L. Itti. An integrated model of top-down and bottom-up attention for optimal object detection. In *Proc. IEEE CVPR*, pages 2049–2056, 2006.

[18] H. C. Nothdurft. Salience from feature contrast: variations with texture density. *Vis. Res.*, 40:3181–3200, 2000.

[19] A. Sha'ashua and S. Ullman. Structural saliency: the detection of globally salient structures using a locally connected network. In *Proc. ICCV*, pages 321–327, 1988.

[20] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist. Visual correlates of fixation selection: effects of scale and time. *Vision Research*, 45:643–659, 2005.

[21] A. Treisman and G. Gelade. A feature-integratrion theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[22] A. Treisman and S. Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95:14–58, 1988.

[23] N. Vasconcelos. Feature selection by maximum marginal diversity: optimality and implications for visual recognition. In *Proc. CVPR*, 2003.

[24] N. Vasconcelos. Scalable discriminant feature selection for image retrieval and recgnition. In *Proc. CVPR*, 2004.

[25] D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19:1395–1407, 2006.

[26] L. Williams and D. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. In *Proc. ICCV*, pages 408–415, 1995.

[27] J. M. Wolfe. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238, 1994.

[28] A. Yarbus. *Eye movements and vision*. Plenum, New York, 1967.