SALIENCY DETECTION VIA DIVERGENCE ANALYSIS: A UNIFIED PERSPECTIVE

BY

JIA-BIN HUANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Professor Narendra Ahuja

ABSTRACT

Computational modeling of visual attention has been a very active area over the past few decades. Numerous models and algorithms have been proposed to detect salient regions in images and videos. We present a unified view of various bottom-up saliency detection algorithms. As these methods were proposed from intuition and principles inspired from psychophysical studies of human vision, the theoretical relations among them are unclear. In this thesis, we provide such a bridge. The saliency is defined in terms of divergence between feature distributions estimated using samples from center and surround, respectively. We explicitly show that these seemingly different algorithms are in fact closely related and derive conditions under which the methods are equivalent. We also discuss some commonly-used center-surround selection strategies. Comparative experiments on two benchmark datasets are presented to provide further insights on relative advantages of these algorithms.

To my parents, for their love and support.

ACKNOWLEDGMENTS

This thesis would not be possible without the support of many people. I am thankful to my advisor, Prof. Narendra Ahuja, for providing helpful guidance with great patience. Also, thanks my labmates in the Computer Vision and Robotics Laboratory for inspirational suggestions and discussions. Finally, I thank Crystal for always being supportive, and my brother and parents for their unconditional love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

LGN	Lateral geniculate nucleus
SVM	Support Vector Machine
KL	Kullback-Leibler
CS	Cauchy-Schwarz

CHAPTER 1

INTRODUCTION

Visual saliency is the perceptual quality which makes some items in the scene pop out from their surroundings and immediately attract our attention. It is well-known that humans can detect salient regions effortlessly even in highly clutter and complex scenes.

The process of filtering out irrelevant and redundant visual information and detecting the most relevant parts of an image can significantly reduce the complexity of visual processing. An effective computational model for automatically generating saliency maps from images is thus of great interest to the computer vision community because it can facilitate many important computer vision and graphics applications, including adaptive image compression [7], object detection and recognition [8], thumbnail generation [9], content-aware image re-targeting [10], photo collage [11] and non-photorealistic rendering [12], among many others. The study of computational saliency modeling can also provide insights on the computational aspects of the underlying neurological attention mechanisms.

Visual saliency detection has received a lot of attention from both psychology and computer vision communities. Inspired by several principles of human visual attention supported by psychological studies, many saliency detection algorithms have been proposed over the past 25 years. Here we list several general principles that have been extensively exploited in the literature:

- Rarity: Less frequently-occurring visual features in the image are considered salient because they carry more information content [3, 4, 13, 14, 15].
- Local complexity: Unpredictability and complexity of a local image region indicate high saliency values [16].
- Contrast: High center-surround contrast draws visual attention [1, 17, 2, 18, 19, 20, 21].
- Learning: High-level factors learning directly from data, e.g., faces [22, 23,

24, 25, 26].

- Center-bias: Exploiting the priors that salient objects are usually placed near the center of the photograph taken [27, 28, 29].

Many of these methods claim performance improvement over others; however, due to the difference in the principles used and the implementation details, the underlying relations among these methods remain hard to understand and their fundamental capabilities are unclear. To better understand the progress of this field, some recent survey papers have been published to qualitatively discuss the state-of-the-art methods on bottom-up saliency detection [30] and visual attention modeling [31, 32]. On the other hand, several benchmark datasets and evaluation methodologies have also been proposed to quantitatively evaluate the performance of existing saliency detection algorithms [33, 34, 35]. These works usually provide a taxonomy using the above principles and cover a large body of existing methods. While the categorization of these methods suggests some characteristics of the methods within the class (and dissimilarity between classes) to a certain extent, differences in feature selection, saliency measure definition, problem assumptions and tuning of the various components make it difficult to understand the underlying computational mechanism and to conduct comparative evaluation.

In this thesis, we propose a unified perspective on *bottom-up* saliency detection algorithms [36]. The saliency of a specific location in an image is defined as the divergence between the probability distributions estimated using samples from center and surround regions centered at that location, respectively. We explicitly show that most of the existing bottom-up saliency models are in fact special cases within our formulation. We derive different assumptions and approximation approaches under which these methods are equivalent (Chapter 3). Therefore, our computational formulation provides a standardized interpretation of the quantities involved in them. Moreover, as divergence has well-known fundamental connections with various fields, e.g., information theory and statistical decision theory [37], we can understand these saliency detection algorithms in a principled way. In addition, we also discuss commonly-used center-surround selection strategies in designing saliency detection algorithms (Chapter 4). Together with the divergence measure selection, these two factors can span a wide spectrum of algorithms covering many existing works. The experiments on benchmark saliency detection datasets can thus provide insights of which aspects of each method are particularly important for good performance (Chapter 5).

The contribution of this thesis is threefold:

1. We propose a unified perspective for bottom-up saliency detection. We explicitly show that many of the existing models are in fact special cases of the proposed model.

2. Our formulation using divergence measure and center-surround selection spans a wide spectrum of saliency detection algorithms.

3. Through our unified formulation and evaluation, we are able to reveal the important aspects which lead to improved performance, shedding new light on how to improve saliency models in the future.

The thesis is organized as follows: We first review relevant works in Chapter 2 based on several categorization factors. We present in Chapter 3 various state-of-the-art bottom-up saliency detection algorithms and explicitly show the theoretic relations among them in a coherent formulation. In Chapter 4, we discuss the selection of center-surround support and the implications. We present comparative experimental results in Chapter 5. Chapter 6 concludes the thesis.

CHAPTER 2

RELATED WORKS

2.1 Computational Modeling of Saliency

There have been numerous computational and psychophysical studies on saliency detection over the last few decades. An exhaustive review of existing saliency detection algorithms is beyond the scope of this thesis. We refer readers to [30, 31, 32] for further details.

Here we briefly introduce existing saliency detection models based on several important factors. All models take an image as an input and transform it into a two-dimensional intensity map (i.e., saliency map [38]). A saliency map is an image with high intensity indicating the high salience regions, i.e., regions that attract our attention most.

2.1.1 Bottom-up vs. top-down

A main factor for categorizing saliency models is whether the top-down influences are incorporated. In terms of attention-guided visual search, the bottom-up approaches correspond to the task-free viewing case while the top-down process account for task-specific search. Research along this direction has explored sources of top-down influences coming from (1) objects [26, 39, 4, 40, 41], (2) scene context [25, 24], or (3) task [42]. In contrast to the top-down driven saliency measure, bottom-up saliency is purely image stimulus driven and task-independent. In this thesis, we consider only the computational formulation of bottom-up saliency measure.

2.1.2 Biologically-inspired vs. pure computational

Itti et al. [1] in one of the early papers on the subject detect saliency in terms of a biologically plausible architecture proposed by Koch and Ullman [38]. The bottom-up saliency is derived using the center-surround difference of image features across multi-scale decomposition. Based on the architecture of Koch and Ullman [38], Le Meur et al. [43] proposed a model leveraging the understanding of Human Visual System behavior, including contrast sensitivity function, perceptual decomposition, visual masking, and center-surround interactions. In contrast to biologically-motivated models, some models derive saliency based on well-known information-theoretic or decision-theoretic quantities. While not directly inspired by the biologically mechanism, a wide range of statistical computations have biologically plausible implementations in simple and complex cells [44].

2.1.3 Rarity vs. contrast

Many of the bottom-up approaches can be viewed as defining saliency either in terms of *rarity* or *contrast* of the local image features. These two terms, rarity and contrast, however, are often misunderstood and used interchangeably in some works.

Rarity-based methods Rarity-based methods measure saliency in terms of how rare the local image content is relative to its surroundings. The notion of rarity can be mathematically represented using probability density distributions. Usually, the probability distribution of local features is first empirically estimated and the Shannon self-information is employed to quantify saliency for task-free visual search [3, 45] and the bottom-up component in task-specific visual search [4, 25, 24] using the Bayesian framework. The rarity principle has also been applied in the frequency domain to detect globally distinct regions, e.g., [6, 46, 47]. Due to the use of the probability for characterizing saliency measures, the rarity-based methods can find meaningful interpretation from information theory and coding theory [3, 45, 48, 49, 50, 51].

Contrast-based methods Contrast-based methods, on the other hand, measure saliency in terms of how an image feature differs from the features in its surroundings. From the nature of direct comparisons of image features, contrast-

based methods have connections with decision-theoretic methods [44] and discriminative approaches [52]. Due to its simplicity and effectiveness, the principle of contrast as saliency measure has been extensively exploited in the literature [53, 54, 1, 55, 19, 56].

2.1.4 Pixel-based vs. region-based

There have been many design choices for the element scale in saliency analysis. For example, several works use pixels as its basic elements [14]. The resultant saliency map thus could preserve the sharp boundaries of salient objects. However, the saliency measures computed from pixel values features may not be reliable especially when the image is noisy. To address this issue, patch-based analysis has been proposed [5, 19]. It is, however, difficult to choose the appropriate scale for the patch size. Region-based methods alleviate this problem by oversegmenting the images, providing regional hypothesis (superpixels) for saliency analysis while maintaining the potential object boundaries [20, 57].

2.1.5 Local vs. global

The notion of local and global saliency is related to the scale of the surround regions. For example, some saliency measures are computed with respect to a localized region [1, 55, 19]. In contrast, some detect globally salient regions by considering the whole image as the background [6, 14, 20]. Recently, the local and global surround are combined in [58].

2.1.6 Learning-based

Inspired by the success in learning-based approaches for object detection, learning methods had been applied to improve saliency detection algorithms. In Liu et al. [59], saliency detection was formulated as a conditional random field trained using a large set of example images with annotations. Similarly, in [60, 23], off-the-shelf classifiers such as SVMs were used to classify input image features into salient and non-salient regions.

CHAPTER 3

A UNIFYING FRAMEWORK

3.1 Center-Surround Divergence

The notion of center-surround operation is ubiquitous in the early stage of human vision processing, e.g., the receptive fields of V1 cells, lateral geniculate nucleus (LGN) cells. It is thus natural to model the saliency of a region through center-surround process. In this chapter, we show that most of the saliency detection algorithms can be viewed as computing center-surround divergence.

Denote $x_i \in \mathcal{M}, 1 \leq i \leq N$ as the i_{th} pixel location in an image with N pixels and spatial support $\mathcal{M}$ and $f_{x_i} \in \mathbb{R}^d$ as the features extracted at position x_i, e.g., luminance, color, orientation, texture, or motion. For a pixel located at x_i, we first define two disjoint spatial supports, namely center $\mathcal{C}_i$ and surround $\mathcal{S}_i$. We also denote their union as $\mathcal{A}_i = \mathcal{C}_i \cup \mathcal{S}_i$ and the patch centered at x_i as $\mathcal{N}_i$ (i.e., a common choice for surround).

The saliency of x_i can thus be defined as the divergence between the two feature distributions estimated using samples from center and surround:

$$s_{x_i} = D(P_{\mathcal{C}_i} || P_{\mathcal{S}_i}), \tag{3.1}$$

where $D(\cdot||\cdot)$ is a function which establishes the *dissimilarity* of one probability distribution to the other on a statistical manifold. The most frequently used class of divergences is the so-called f-divergence, which includes the well-known Kullback-Leibler divergence (KL divergence) as a special case.

In the following sections, we show that most of the saliency detection algorithms in the literature share the same form as Eqn. 3.1. That is, with certain assumptions and approximations, these methods are equivalent. Table 3.1 presents a summary of various saliency detection algorithms categorized based on the underlying quantity involved.

Table 3.1: A summary of saliency detection algorithms using divergence analysis

Basic form	Ref	Notes (assumptions)	Center-surround selection
$D_{KL}(P_{\mathcal{C}_i}\|\|P_{\mathcal{S}_i})$	[61]	Independence among feature	Multi-scale, patch-based
	[3]	Self-information	$\mathcal{C}_i$: $\{x_i\}$, $\mathcal{S}_i$: $\mathcal{N}_i \setminus x_i$
	[4]	Self-information	$\mathcal{C}_i$: $\{x_i\}$, $\mathcal{S}_i$: $\mathcal{N}_i \setminus x_i$
	[5]	Difference of self-information	Single-scale patch-based
	[16]	Surround distribution $P_{\mathcal{S}_i} \sim P_U$	$\mathcal{C}_i$: adaptive, $\mathcal{S}_i$: $\mathcal{M} \setminus x_i$
$D_{KL}(P_{\mathcal{S}_i}\|\|P_{\mathcal{C}_i})$	[53]	Downsample image for speedup	$\mathcal{C}_i$: $\{x_i\}$, $\mathcal{S}_i$: $\mathcal{N}_i \setminus x_i$
	[54]	Luminance feature, look-up table	$\mathcal{C}_i$: $\{x_i\}$, $\mathcal{S}_i$: $\mathcal{M} \setminus x_i$
	[1]	Contrast as center-surround difference	$\mathcal{C}_i$: fine, $\mathcal{S}_i$: coarse
	[2]	Contrast as center-surround difference	$\mathcal{C}_i$: fine, $\mathcal{S}_i$: coarse
	[14]	Replace all samples with its mean	$\mathcal{C}_i$: $\{x_i\}$, $\mathcal{S}_i$: $\mathcal{M} \setminus x_i$
	[62]	Maximum symmetric surround	$\mathcal{C}_i$: $\{x_i\}$, $\mathcal{S}_i$: adaptive
	[19]	K nearest neighbor for approximation	$\mathcal{C}_i$: $\{x_i\}$, $\mathcal{S}_i$: center-weighted
$D_{\lambda}(P_{\mathcal{C}_i}\|\|P_{\mathcal{S}_i})$	[18]	Discriminant center-surround	Single-scale, patch-based
$D_{CS}(P_{\mathcal{C}_i}\|\|P_{\mathcal{S}_i})$	[20]	Sparse histogram comparison	$\mathcal{C}_i$: regions

3.2 From Center to Surround

We first consider saliency s_{x_i} as the KL divergence from center to surround, i.e., from $P_{\mathcal{C}_i}$ to $P_{\mathcal{S}_i}$:

$$s_{x_i} = D_{KL}(P_{\mathcal{C}_i}||P_{\mathcal{S}_i}) = \sum_f P_{\mathcal{C}_i}(f) \log \frac{P_{\mathcal{C}_i}(f)}{P_{\mathcal{S}_i}(f)}. \tag{3.2}$$

Feature independence assumption By assuming independence among the dimensions in f_{x_i}, one can compute the KL divergence in each feature channel and fuse these channels to form the final saliency map [61].

Single pixel center support By shrinking the center support to a single pixel x_i, i.e., $C_i = \{x_i\}$, we have $P_{\mathcal{C}_i}(f_{x_i}) = 1$. The equation in Eqn. 3.2 is then simplified into

$$s_{x_i} = I(f_{x_i}) = -\log P_{\mathcal{S}_i}(f_{x_i}), \tag{3.3}$$

which yields the Shannon self-information as used in AIM [3] and SUN [4] models. Here, we connect the rarity-based methods using the proposed divergence measure.

Self-information Difference The difference between the self-information of observing f_{x_i}evaluated using $P_{\mathcal{A}_i}$ and $P_{\mathcal{C}_i}$ has the form

$$-\log P_{\mathcal{A}_i}(f_{x_i}) - (-\log P_{\mathcal{C}_i}(f_{x_i})) = \log \frac{P_{\mathcal{C}_i}(f_{x_i})}{P_{\mathcal{A}_i}(f_{x_i})}, \tag{3.4}$$

which gives rise to the saliency measure defined in [5].

Complexity-based methods By assuming the surround distribution $P_{\mathcal{S}_i}$ to be uniform P_U, we can build connection with the local complexity-based methods [16], which uses entropy of a local region as a saliency measure:

$$H(P_{\mathcal{C}_i}) = \log |\mathcal{F}| - D_{KL}(P_{\mathcal{C}_i}||P_U), \tag{3.5}$$

where $\mathcal{F}$ is the set of the feature values.

3.3 From Surround to Center

As the KL divergence is not symmetric, one can compute the saliency as the KL divergence from the opposite direction [1]:

$$s_{x_i} = D_{KL}(P_{\mathcal{S}_i}||P_{\mathcal{C}_i}) = \sum_f P_{\mathcal{S}_i}(f) \log \frac{P_{\mathcal{S}_i}(f)}{P_{\mathcal{C}_i}(f)}. \tag{3.6}$$

The meaning of Eqn. 3.6 and 3.2 can be better understood via the fundamental connection between the KL divergence and the likelihood theory [63].

$$D_{KL}(P_{\mathcal{S}_i}||P_{\mathcal{C}_i}) = -H(P_{\mathcal{S}_i}) - \sum_f P_{\mathcal{S}_i}(f) \log P_{\mathcal{C}_i}(f), \tag{3.7}$$

where the second term of the right-hand side can be rewritten as the minus log-likelihood function:

$$-\sum_f P_{\mathcal{S}_i}(f) \log P_{\mathcal{C}_i}(f) = \frac{-1}{|\mathcal{S}_i|} \sum_{j:x_j \in \mathcal{S}_i} \log P_{\mathcal{C}_i}(f_{x_j}). \tag{3.8}$$

[1] From the neuroscience literature, we know that there are two types of retinal ganglion cells: "on-center/off-surround" and "off-center/on-surround". The divergence measure computing from center to surround and from surround to center may have tight connections with these neuroscience findings.

We can then interpret the quantity $D_{KL}(P_{\mathcal{S}_i}||P_{\mathcal{C}_i})$ as how well the probabilistic model of center $P_{\mathcal{C}_i}$ can explain the samples from surround. If the probability density function estimated using samples from the center support $P_{\mathcal{C}_i}$ can provide a good fit (i.e., high likelihood) of the surrounding samples, then the saliency s_{x_i} is small, and vice versa. On the other hand, Eqn. 3.2 measures saliency as how well the model of surround $P_{\mathcal{S}_i}$can explain samples from center.

We can view the likelihood model in Eqn. 3.7 as a generalization of many contrast-based methods [54, 14, 53, 62, 19], which makes different assumptions and approximations.

Single pixel center support For example, by shrinking the center support to a single pixel x_i and assuming the form of $P_{\mathcal{C}_i}$ as Gaussian distribution with mean f_{x_i} and variance σ^2, the minus log-likelihood in Eqn. 3.8 becomes

$$\frac{1}{|\mathcal{S}_i|}\sum_{j:x_j\in\mathcal{S}_i}\frac{(f_{x_i}-f_{x_j})^2}{\sigma^2}+\text{const} \tag{3.9}$$

Many of the contrast-based methods measure saliency by approximately evaluating Eqn. 3.9 or its variants. For example, Achanata et al. [14] replaced all f_{x_j} with its mean ($\frac{1}{|\mathcal{S}_i|}\sum_{j:x_j\in\mathcal{S}_i} f_{x_j}$) for approximating Eqn 3.9. In biologically-inspired saliency detection algorithms [1, 2], the difference between fine and coarse scales in Gaussian pyramids is used.

Other contrast-based methods Other variants include adaptive surround $\mathcal{S}_i$ [62], Laplacian distribution on $P_{\mathcal{C}_i}$ [54] and k nearest neighbor for efficient computation [19].

3.4 Symmetrised Divergence

In contrast to the non-symmetric KL divergence, some symmetrised divergences have also been proposed. One example is the λ divergence:

$$D_\lambda(P||Q)=\lambda D_{KL}(P||A)+(1-\lambda)D_{KL}(Q||A), \tag{3.10}$$

where P, Q, A are probability distributions and $A=\lambda P+(1-\lambda)Q$. By appropriately choosing λ as the prior probability of the center $\lambda=|\mathcal{C}_i|/|\mathcal{A}_i|$, the λ

divergence between center and surround is

$$D_\lambda(P_{\mathcal{C}_i}||P_{\mathcal{S}_i}) = \lambda D_{KL}(P_{\mathcal{C}_i}||P_{\mathcal{A}_i}) + (1-\lambda)D_{KL}(P_{\mathcal{S}_i}||P_{\mathcal{A}_i}), \tag{3.11}$$

which is the mutual information of feature distribution and center-surround label used in [18].

Another alternative is the Cauchy-Schwarz divergence [64], which is given by

$$D_{CS}(P||Q) = -\log \frac{\int P(x)Q(x)\mathrm{d}x}{\sqrt{\int P(x)^2\mathrm{d}x \int Q(x)^2\mathrm{d}x}}. \tag{3.12}$$

When estimating the probabilistic density P, Q using non-parametric density estimation techniques (known as Parzen windowing), the Cauchy-Schwarz divergence can be easily evaluated in closed form. Specifically, we estimate the pdf of $P_{\mathcal{C}_i}$ using

$$\hat{P}_{\mathcal{C}_i}(f_x) = \frac{1}{|\mathcal{C}_i|} \sum_{j:x_j \in \mathcal{C}_i} W_{\sigma^2}(f_x, f_{x_j}), \tag{3.13}$$

where $W_{\sigma^2}(\cdot,\cdot)$ is a Gaussian kernel with parameter σ^2. Then the Cauchy-Schwarz divergence $D_{CS}(P_{\mathcal{C}_i}||P_{\mathcal{S}_i})$ has the form

$$-\log \frac{\sum_{l:x_l \in \mathcal{C}_i} \sum_{j:x_j \in \mathcal{S}_i} K_{l,j}}{\sqrt{\sum_{l,l':x_l,x'_l \in \mathcal{C}_i} K_{l,l'} \sum_{j,j':x_j,x'_j \in \mathcal{S}_i} K_{j,j'}}}, \tag{3.14}$$

where $K_{l,j}$ denotes $W_{2\sigma^2}(f_{x_l}, f_{x_j})$. This gives rise to the histogram contrast saliency measure in [20].

CHAPTER 4

CENTER-SURROUND SUPPORT SELECTION

The center-surround hypothesis for saliency detection is inspired by the center-surround mechanisms in the early stages of biological vision [65, 66], e.g., in V1 cells and LGN cells. However, the selection of the center and surround in an image is a not a trivial task. Here, we investigate various strategies for selecting support of center and surround and discuss their implications.

4.1 Selection of Center Support

The center support represents the basic elements for saliency analysis. The simplest choice for center support is to use a single pixel [3, 4, 19, 62]. The main advantage of using a single pixel as center support is that the potential object boundaries could be preserved in the resultant saliency map, facilitating further processing such as object segmentation. However, estimating the distribution of center with one single observation clearly introduces high variance. As ways of increasing the sample size for estimating the distribution, patch-based or window-based approaches have been proposed [5, 61]. Yet, without knowing image discontinuities in the vicinity, the optimal patch/window size of the center cannot be estimated, which can be only partly mitigated by added complexity of multi-scale computation. Region-based approaches emerge as a good choice for spatial support estimation of center. While not biologically-inspired, region-based methods provide appropriate spatial scales and directly involve potential object boundaries in the saliency analysis. Therefore, region-based analysis has become a promising way for measuring saliency [20, 21]. Note that region-based saliency is different from region-enhanced saliency, which serves as a post-processing of pixel/patch-based saliency by averaging them over segments [67].

4.2 Selection of Surround Support

The selection of the surround is closely related the notion of local and global saliency. For example, by choosing surround as the whole image, the algorithm predicts globally salient regions. For local saliency, finite support or center-weighted kernels can be used.

CHAPTER 5

EXPERIMENTAL RESULTS

In this chapter, we quantitatively evaluate these bottom-up saliency detection algorithms to provide a comparative study.

Dataset We show performance comparison on two publicly available datasets: the MSRA dataset [22] and the McGill dataset [68]. For the MSRA dataset, we use a subset of 1,000 images where groundtruth segmentation are available [14]. The McGill dataset contains 235 natural images with rough categorization based on difficulty.

Evaluation metric To quantitatively evaluate the performance of these saliency detection algorithms, we use binary (thresholded) saliency masks derived from the saliency map and compare with human segmentation to compute the precision and recall curve. A perfect agreement with the human segmentation mask will result in the curve passing though the top-right corner in the precision and recall curve.

State-of-the-art saliency detection methods: We conduct comparative study using the following state-of-the-art methods:

- Rarity-based: AIM [3], SUN [4], SW [5] from Section 3.2
- Contrast-based: IT [1], GB [2], CA [19], AC [62], FT [14], LC [54] from Section 3.3 and HC [20], RC [20] from Section 3.4.
- Spectrum-based: SR [6]

We choose these methods based on the principle used (rarity vs. contrast), the selection of center and surround support (pixels, patches, or regions), and features in different domain (e.g., spectrum-based method [6]).

Quantitative results: In Figures 5.1 and 5.2, we show the mean precision-recall curves on the MSRA dataset. Similar results are shown in Figure 5.3 and 5.4 for the McGill dataset. We use two separate plots to avoid the line clutter.

Several observations can be seen in the comparative experiments. First, for rarity-based methods, the self-information methods AIM [3] and SUN [4] have similar performance regardless of using image-specific or generic statistics of natural images. The SW [5] improves the self-information based method by the use of patch-based analysis. Second, for contrast-based methods, we can see that from LC [54] to FT [14], the performance improves as the feature sets are richer (from intensity to color features). AC [62] further improves upon FT [14] with the adaptive surround support. Third, we can see the effect of using regions (as opposed to patches) as basic elements for saliency detection by comparing RC [20] and HC [20].

Note that similar results can be observed in Figures 5.3 and Figure 5.4. Yet, we can see that the overall performance from all the methods tested on the McGill dataset is lower, suggesting that the images in MSRA 1000 dataset may be too idealized. For example, in MSRA 1000 dataset, it is usually assumed that there is only one salient object (which almost never occurs in real world scenarios).

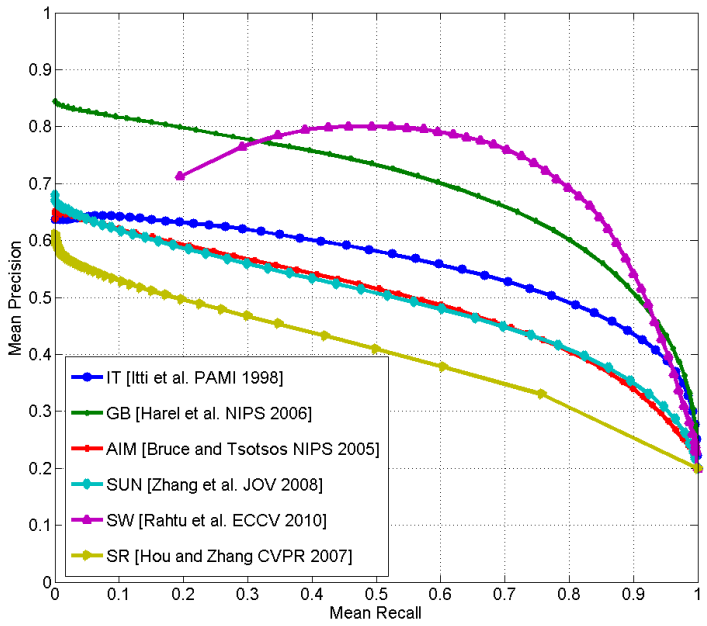

Figure 5.1: Quantitative comparison of on MSRA dataset. IT [1], GB [2] are biologically inspired methods, AIM [3], SUN [4], and SW [5] are representative methods exploiting rarity principle, and SR [6] is spectrum-based approach.

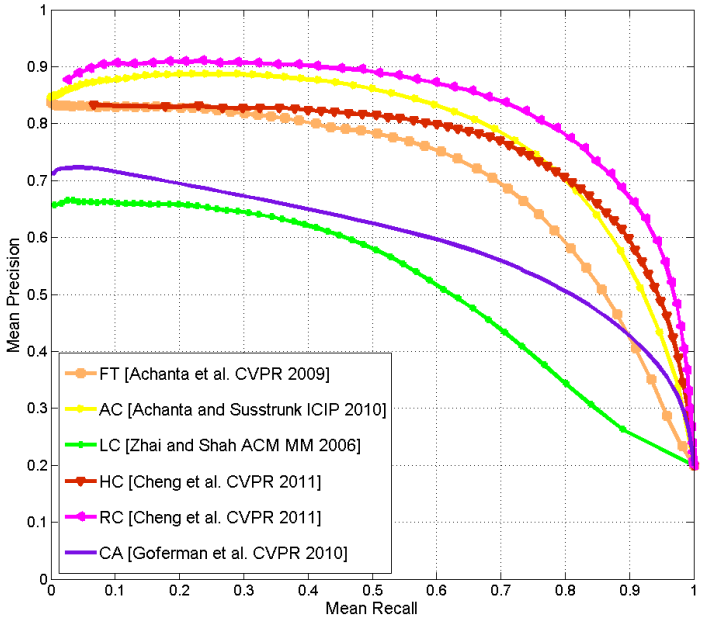

Figure 5.2: Quantitative comparison of on MSRA dataset. All of these methods belong to the contrast-based approaches with different features, assumptions and support selection.

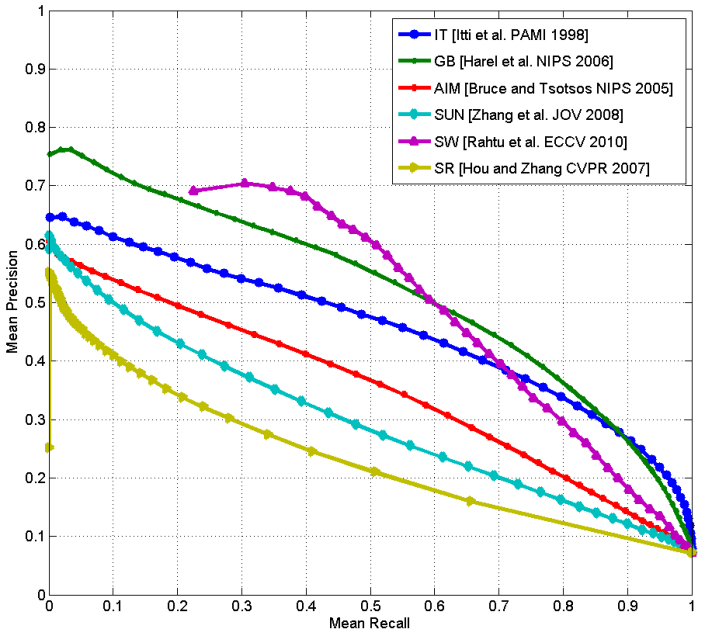

Figure 5.3: Quantitative comparison on McGill dataset. IT [1], GB [2] are biologically inspired methods, AIM [3], SUN [4], and SW [5] are representative methods exploiting rarity principle, and SR [6] is spectrum-based approach.

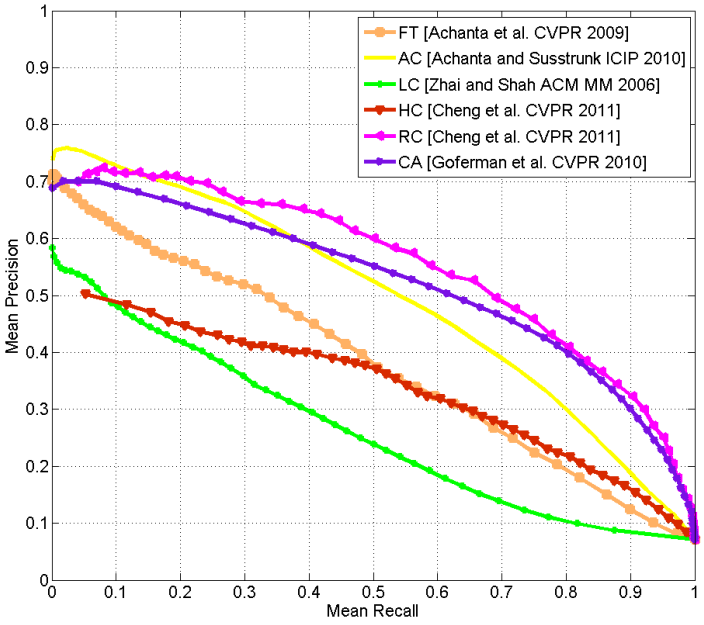

Figure 5.4: Quantitative comparison on McGill dataset.

CHAPTER 6

CONCLUSION

We have shown theoretical connections among various bottom-up saliency detection algorithms. By providing a standardized interpretation, the unified perspective sheds new light on current methods in the literature. From the divergence measure, we can now directly compare the quantity involved in different methods and show conditions under which they are equivalent. We also discuss the several center-surround selection strategies, argue the motivations of the method design and empirically show the effects. Comparative evaluation on two publicly available datasets provides further insights on relative strengths of each method. The list of saliency detection methods in this thesis is not exhaustive. For example, spectrum-based approaches are missing from the list and methods using background and boundary priors are hard to interpret using the proposed scheme. In the future, we plan to extend our framework to include those methods as well as build a common ground for detailed comparison, for example, by controlling all conditions when comparing different divergence measure, feature, and center-surround selection schemes. The initial attempt we made is illustrated in Figure 6.1, where we created an interface for selecting method design choices, center-surround support scale, and divergence measures.

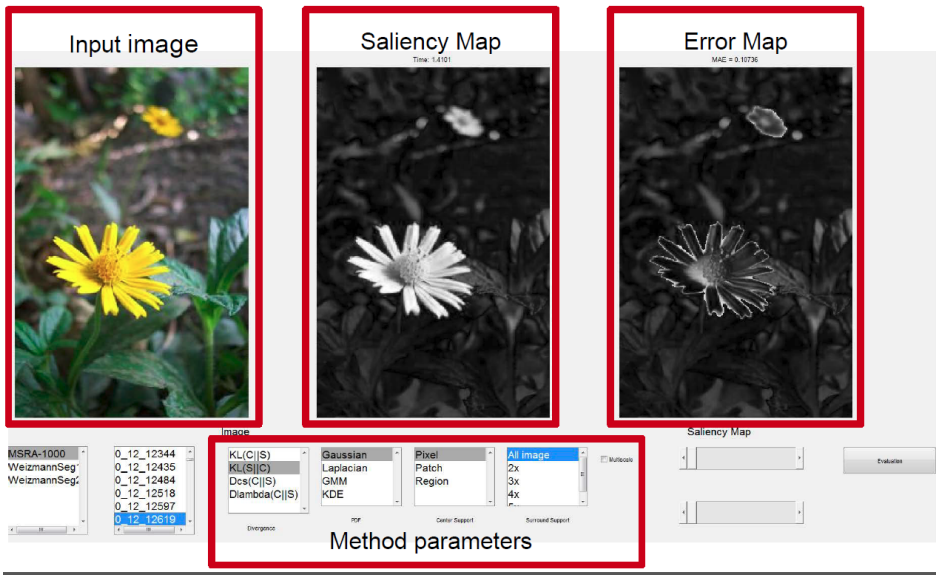

Figure 6.1: Interface for detailed analysis of saliency detection algorithms using controlled experiments.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *PAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.

[2] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *NIPS*, 2006.

[3] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *NIPS*, 2005.

[4] L. Zhang, M. Tong, T. Marks, H. Shan, and G. Cottrell, "Sun: A Bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8, no. 7, pp. 1–20, 2008.

[5] E. Rahtu, J. Kannala, M. Salo, and J. Heikkil, "Segmenting salient objects from images and videos," in *ECCV*, 2010.

[6] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *CVPR*, 2007.

[7] C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in low level vision," in *ICPR*, 2002.

[8] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *CVPR*, 2004.

[9] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs, "Automatic thumbnail cropping and its effectiveness," in *Proceedings of the 16th annual ACM symposium on User interface software and technology*. ACM, 2003, pp. 95–104.

[10] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," in *ICCV*, 2009.

[11] J. Wang, L. Quan, J. Sun, X. Tang, and H. Shum, "Picture collage," in *CVPR*, vol. 1. IEEE, 2006, pp. 347–354.

[12] D. DeCarlo and A. Santella, "Stylization and abstraction of photographs," in *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, 2002, pp. 769–776.

[13] C. Kanan, M. Tong, L. Zhang, and G. Cottrell, "Sun: Top-down saliency using natural statistics," *Visual Cognition*, vol. 17, no. 6-7, pp. 979–1003, 2009.

[14] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009.

[15] H. J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *Journal of Vision*, vol. 9, no. 12, 2009.

[16] T. Kadir and M. Brady, "Saliency, scale and image description," *IJCV*, vol. 45, no. 2, pp. 83–105, 2001.

[17] R. Rosenholtz, "A simple saliency model predicts a number of motion popout phenomena," *Vision research*, vol. 39, no. 19, pp. 3157–3163, 1999.

[18] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," in *NIPS*, 2007.

[19] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *CVPR*, 2010.

[20] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, "Global contrast based salient region detection," in *CVPR*, 2011.

[21] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*. IEEE, 2012, pp. 733–740.

[22] T. Liu, J. Sun, N. N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," in *CVPR*, 2007.

[23] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *ICCV*. IEEE, 2009, pp. 2106–2113.

[24] A. Oliva, A. Torralba, M. Castelhano, and J. Henderson, "Top-down control of visual attention in object detection," in *ICIP*, 2003.

[25] A. Torralba, A. Oliva, M. Castelhano, and J. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search," *Psychological review*, vol. 113, no. 4, p. 766, 2006.

[26] M. Cerf, J. Harel, W. Einhäuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection," in *NIPS*, vol. 20. MIT Press, 2008.

[27] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *ECCV*. Springer, 2012, pp. 29–42.

[28] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *CVPR*, 2013.

[29] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *ICCV*, 2013.

[30] A. Toet, "Computational versus psychophysical image saliency: A comparative evaluation study," *PAMI*, vol. 99, no. 1, 2011.

[31] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *PAMI*, 2012.

[32] S. Frintrop, E. Rome, and H. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Transactions on Applied Perception (TAP)*, vol. 7, no. 1, p. 6, 2010.

[33] A. Borji, D. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," *TIP*, 2012.

[34] A. Borji, D. Sihite, and L. Itti, "Salient object detection: A benchmark," in *ECCV*, 2012.

[35] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," *Massachusetts Institute of Technology Computer Science and Artificial Intelligence Lab Technical Report 2012-001*, 2012.

[36] J.-B. Huang and N. Ahuja, "Saliency detection via divergence analysis: A unified perspective," in *International Conference on Pattern Recognition*, November 2012.

[37] M. D. Reid and R. C. Williamson, "Information, divergence and risk for binary experiments," *JMLR*, vol. 12, pp. 731–817, 2011.

[38] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry." *Human Neurobiology*, vol. 4, no. 4, pp. 219–27, 1985.

[39] M. Cerf, E. Frady, and C. Koch, "Faces and text attract gaze independent of the task: Experimental data and computer model," *Journal of Vision*, vol. 9, no. 12, 2009.

[40] D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," in *NIPS*, 2004, pp. 481–488.

[41] W. Einhäuser, M. Spain, and P. Perona, "Objects predict fixations better than early saliency," *Journal of Vision*, vol. 8, no. 14, 2008.

[42] R. J. Peters and L. Itti, "Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention," in *CVPR*. IEEE, 2007, pp. 1–8.

[43] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *PAMI*, vol. 28, no. 5, pp. 802–817, 2006.

[44] D. Gao and N. Vasconcelos, "Decision-theoretic saliency: computational principles, biological plausibility, and implications for neurophysiology and psychophysics," *Neural Computation*, vol. 21, no. 1, pp. 239–271, 2009.

[45] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, 2009.

[46] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *CVPR*. IEEE, 2008, pp. 1–8.

[47] J. Li, M. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *PAMI*, 2012.

[48] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," in *NIPS*, 2006.

[49] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," in *NIPS*, 2008.

[50] Y. Li, Y. Zhou, J. Yan, Z. Niu, and J. Yang, "Visual saliency based on conditional entropy," in *Asian Conference on Computer Vision*, 2009.

[51] Y. Cao and L. Zhang, "A novel hierarchical model of attention: maximizing information acquisition," in *Asian Conference on Computer Vision*, 2009.

[52] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," in *ICCV*. IEEE, 2007, pp. 1–6.

[53] Y. Ma and H. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *ACM MM*, 2003.

[54] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *ACM MM*, 2006.

[55] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *NIPS*, vol. 19. MIT; 1998, 2006, p. 545.

[56] Y. Pritch, P. Krahenbuhl, F. Perazzi, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*. IEEE, 2012, pp. 733–740.

[57] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior." in *BMVC*, vol. 3, no. 4, 2011, p. 7.

[58] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *CVPR*. IEEE, 2012, pp. 478–485.

[59] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," *PAMI*, vol. 33, no. 2, pp. 353–367, 2011.

[60] W. Kienzle, F. Wichmann, B. Schölkopf, and M. Franz, "A nonparametric approach to bottom-up visual saliency," in *NIPS*, 2007.

[61] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *ICCV*, 2011.

[62] R. Achanta and S. Susstrunk, "Saliency detection using maximum symmetric surround," in *ICIP*, 2010.

[63] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons., 2006.

[64] R. Jenssen, J. Principe, D. Erdogmus, and T. Eltoft, "The Cauchy-Schwarz divergence and Parzen windowing: Connections to graph theory and mercer kernels," *Journal of the Franklin Institute*, vol. 343, no. 6, pp. 614–629, 2006.

[65] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[66] R. Wurtz et al., "Visual receptive fields of striate cortex neurons in awake monkeys," *J Neurophysiol*, vol. 32, no. 5, pp. 727–742, 1969.

[67] F. Liu and M. Gleicher, "Region enhanced scale-invariant saliency detection," in *ICME*, 2006.

[68] J. Li, M. Levine, X. An, and H. He, "Saliency detection based on frequency and spatial domain analyses," in *BMVC*, 2011.