

SALIENCY MAPPING ENHANCED BY SYMMETRY FROM LOCAL PHASE

Shane Grant, Laurent Itti

Department of Computer Science
University of Southern California, Los Angeles, California
wgrant@usc.edu, itti@pollux.usc.edu

ABSTRACT

We describe a method of generating saliency maps that combines symmetry with traditional contrast information. Log Gabor filtering is used to obtain local frequency information that can be used to both calculate local symmetry and edge responses. The symmetry information is combined with center-surround responses from color (either in RGB or LAB), orientation, and intensity features. The algorithm is evaluated on the Kootstra dataset and shown to significantly outperform three state-of-the-art models that utilize either contrast only or symmetry only based responses. A subjective evaluation by 13 human observers of image regions selected by 3 variants of our algorithm shows the significant benefit of including symmetry information in addition to contrast-based features.

Index Terms— log gabor, saliency, symmetry

1. INTRODUCTION

When observing a crowded scene, the human eye can quickly filter out noise and focus on interesting (i.e., salient) regions. Modeling this natural behavior is quite difficult because human vision is not well understood. Numerous attempts have been made to build models of this natural behavior using a wide variety of methodologies including cognitive, bayesian, graphical, and spectral approaches [1]. Saliency maps have been used for numerous applications, including predicting eye movements [2] and aiding object recognition [3].

We present a model to compute a saliency map using a stimulus-driven (bottom-up) methodology based upon feature contrast and symmetry. At the low level, we differ from many models in our choice of color space and the use of log Gabor filters to produce edge responses. In addition, we utilize the response of the log Gabor filters to compute local symmetries as an additional channel to enhance the saliency map.

2. THE MODEL

Input is provided in the form of a color image in any resolution. This image is split into several visual feature maps based upon different aspects of human vision. A linear combination

of these maps forms the final saliency map, which highlights the most conspicuous locations in the image.

2.1. Color Space Feature Maps

A common way of representing a color image on a computer involves using the RGB color space. Our approach uses the CIE LAB color space, which was designed to closely mimic how human vision is believed to perceive color [4]. It describes color in terms of luminosity, red-green opponency, and blue-yellow opponency.

Each color channel is subjected to a center-surround operation to emphasize regions of contrast. This is implemented by filtering with difference of gaussian (DoG) filters at several scales. DoG filters closely resemble the receptive fields of neurons in early visual processing; the center is excited while the surrounding area is inhibited [5]. The filters are generated by:

$$DoG(x, y) = \frac{1}{\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}} - \frac{1}{(1.6\sigma)^2} e^{-\frac{x^2+y^2}{(1.6\sigma)^2}} \quad (1)$$

over four scales ($\sigma = 4, 8, 16, 32$ pixels), giving the model some degree of scale invariance. The response for each channel is then:

$$\left| \sum_{\sigma} DoG_{\sigma} * channel \right| \quad (2)$$

We refer to the results of this operation as L , A , and B , for each respective color channel.

2.2. Orientation Feature Maps

The orientations of edges provide another cue to finding the most salient objects within an image [6]. Traditionally, these edge responses are computed by filtering the image with Gabor filters over multiple scales and orientations. We differ here in that we opt to use log Gabor filters, which have a Gaussian transfer function when viewed on a log scale and allow arbitrarily large bandwidth filters while maintaining a zero DC component in the even symmetric filter. It has been suggested that these filters better model the responses of cortical cells [7].

The log Gabor filter bank is constructed using several parameters: the number of scales $S_{\#}$, the number of orientations $O_{\#}$, the minimum filter wavelength min_wl , the multiplier between successive scales $mult$, and σ_f , the ratio of the standard deviation of the Gaussian describing the log Gabor filter’s transfer function in the frequency domain to the filter center frequency. The filters are built in the frequency domain from the product of a radial component, which controls the frequency response, and an angular component, which controls the orientation response.

The radial component is defined by:

$$G(f) = \exp \frac{-(\log \frac{f}{f_0})^2}{2(\log \frac{\sigma_f}{f_0})^2} \quad (3)$$

where f_0 is the filter’s center frequency and is related to our current scale s by $f_0 = (min_wl \times mult^s)$.

The angular component is then constructed:

$$G(\theta) = \frac{\cos \Delta\theta(\theta) + 1}{2} \quad (4)$$

where $\Delta\theta(\theta)$ is constructed from the angular distance from the filter’s orientation θ_0 :

$$ds(\theta) = \cos(\theta) \cos(\theta_0) - \sin(\theta) \sin(\theta_0)$$

$$dc(\theta) = \cos(\theta) \cos(\theta_0) + \sin(\theta) \sin(\theta_0)$$

$$\Delta\theta(\theta) = \min(|\text{atan2}(ds(\theta), dc(\theta))|, \pi)$$

In our experiments, we utilize a log Gabor bank with $S_{\#} = 7$, $O_{\#} = 6$, $min_wl = 3.0$, $mult = 1.5$, and $\sigma_f = 0.55$. The parameters were chosen such that the log Gabor bank spanned roughly two octaves with some degree of overlap between successive filters. The primary effect of adjusting these parameters is to vary the scale of regions which respond strongly to symmetry processing - thus they were chosen to compromise between small and large objects.

Only the first three scales are used to compute the orientation responses; the rest are only utilized in the symmetry calculation. This is because the larger scales are not beneficial in identifying most edges.

The magnitude response of each log Gabor filter is run through the same DoG filtering pipeline as each color channel was to achieve a total of 18 (from 3 spatial scales and 6 orientations) orientation center-surround responses. These 18 responses are then summed together to create one overall orientation response, known as O , of an image.

2.3. Symmetry Feature Maps

The last cue we consider for saliency comes from local symmetry. Objects likely to be salient, such as man-made objects, plants, and animals, often have pronounced symmetry; hence it is believed to be an important aspect of bottom-up detection [8].

To identify symmetry within an image, the local frequencies are analyzed to determine local symmetries and asymmetries. The aforementioned log Gabor filters are used to accomplish this: the even component for local symmetry, and the odd component for local asymmetry [9]. The differences of these even and odd filters are taken over several orientations to yield the symmetry map:

$$S(x) = \frac{\sum_n (|e_n(x)| - |o_n(x)| - T)}{\sum_n A_n(x) + \epsilon} \quad (5)$$

where e_n is the even cosine function, o_n is the odd sine function, A_n is the magnitude of the filter response, ϵ is a near zero term to prevent division by zero, and T is a noise compensation term based upon the median filter response at the smallest scale of processing, as described in [10].

Unlike the previously generated maps, the symmetry response is not passed through the DoG filters since we are interested in regions with a strong symmetric response - not necessarily in regions where there is a marked spatial change in symmetric response.

2.4. Final Saliency Map

The final saliency map is a weighted linear combination of the individually generated feature maps:

$$\begin{aligned} \text{Saliency} &= w_L N(L) + w_A N(A) + w_B N(B) \\ &+ w_O N(O) + w_S N(S) \end{aligned}$$

$$\text{Saliency}_{final} = N(\text{Saliency}) \quad (6)$$

We set all weights equal with the exception of w_A and w_B , which receive half the value of the other weights, to give the net contribution of color the same weight as any other feature.

The normalization operator $N(\cdot)$ is an iteratively applied DoG filter, which serves to emphasize responses sufficiently greater than their neighbors and inhibit similarly valued regions [11]:

$$M \leftarrow [M + M * DoG_N - C_{inh}]_{\geq 0} \quad (7)$$

where C_{inh} is a constant inhibitory term and all negative values are replaced with zero. On the initial iteration M is initialized to zero and the input map is used for the filtering. Parameterization of the DoG is identical to [11].

3. RESULTS AND EVALUATION

We evaluated our model under two different metrics: a standard scoring scheme (shuffled AUC, designed to minimize center-bias [5]) using a publicly available dataset derived from eye tracking experiments, and a subjective analysis of the importance of regions identified by the algorithm as salient by human participants.

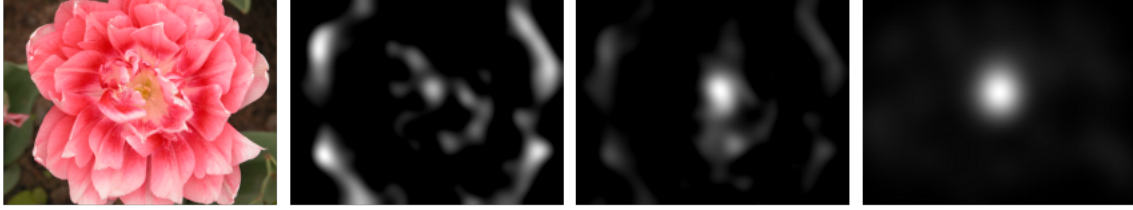


Fig. 1. A sample image from the Kootstra dataset that benefits from a symmetry response. Ordering from left to right is: input, LAB, SYM, human eye fixation data. LAB and SYM are explained in section 3.2.

3.1. Eye Movement Prediction

We evaluated our model on the Kootstra eye tracking dataset [12], which features a total of 101 images in the categories of animals, street scenes, buildings, flowers, and nature. The images in the dataset often do not contain explicit objects or salient regions and are not guaranteed to contain local symmetries. Shuffled AUC (area under the ROC curve) scores, computed from an ROC curve where for each image, human fixations form positive samples and fixations from all other images the negative samples [1], are shown in figure 2. Also included for comparison is the original Itti model [11], the bayesian approach of [5], Kootstra’s symmetry based approach [12], as well as human inter-observer and constant gaussian blob performance. Scores were calculated using the raw saliency maps output by each model with no post processing.

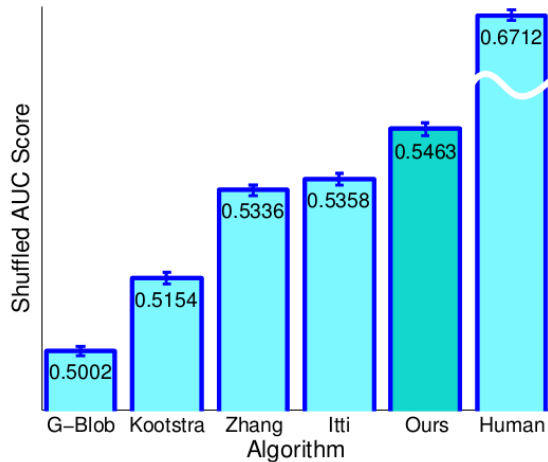


Fig. 2. Shuffled AUC scores on the Kootstra dataset. Our results are significantly ($p < 10^{-4}$) better than the closest performer, Itti, but also significantly below human performance.

3.2. Subjective Performance

A survey of 13 naïve individuals evaluated saliency maps computed by three variants of our model (no symmetry with

an RGB color scheme (RGB), no symmetry with LAB color (LAB), and symmetry with LAB color (SYM)) on a set of 24 images that each contained one or more prominent objects. All non salient regions according to each model variant were masked out, and participants ordered the masked images from best to worst. Best was defined as the image that best captured whatever the individual found most important in the original image.

To ensure validity of the evaluation, none of the images were used in the development of the models, and images had no pre-determined biases towards any of the models. The order in which the three separate models were displayed was random, and each model was simply labeled “A”, “B”, or “C”. For each image, the model with the highest ranking was given a value of 3, followed by a value of 2, and a value of 1 for the lowest ranking. Ties were allowed and the point values were averaged accordingly. Overall, the SYM model masking had the highest average value of 2.37, followed by 1.94 for LAB, and 1.70 for RGB.

A two sample t-test was run to test whether the scores were significantly different from one another. The corresponding t and p values were:

Models	t-score	p-value
LAB > RGB	1.5214	0.0675
SYM > LAB	3.4031	$6.9468 * 10^{-4}$
SYM > RGB	4.5370	$2.0399 * 10^{-5}$

It can be concluded that the inclusion of symmetry clearly helps the salient regions, but the null hypothesis cannot be outright rejected for LABs improvement over RGB.

To test whether LAB was superior to RGB, the SYM results were discarded, and instead we ran a test to see whether LAB is superior to RGB over 50% of the time. A z-value of 2.6042 and a p-value of 0.004604 were reported. Thus there is some further evidence that LAB is superior to RGB for significance levels above 0.01.

4. DISCUSSION

Our results show that our model with the inclusion of symmetry features significantly outperforms three popular state-of-the-art models on the Kootstra dataset. In addition, subjective

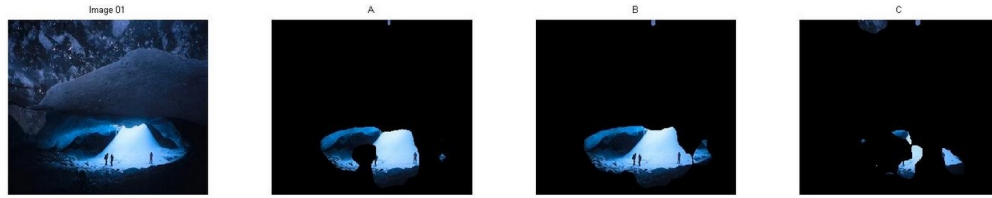


Fig. 3. An example of an image shown to survey participants. The original image followed by three randomly ordered masked variants. In this case, the ordering is LAB, SYM, RGB.

testing favors the model variant that uses LAB and symmetry. Our results provide evidence that symmetry is a relevant feature in attentional selectivity.

With the eye tracking data, we found that symmetry computation is particularly beneficial when indeed there is natural symmetry in the image (as in many of the flower and nature images). However, it can be less useful in scenes with little relevant symmetry, such as many of the street scenes, where our model would detect, e.g., the street as being symmetric although that symmetry tends to not strongly attract human gaze. Because we do not use symmetry in isolation but combine it with the other feature maps through iterative normalization, spurious symmetry responses are inhibited in the final saliency map.

The masking experiment gives a different domain to analyze the performance of the saliency maps: here we see that if the salient regions are observed in isolation, areas with strong symmetric responses seem to be highly rated as interesting and relevant by human observers.

Our approach does have some limitations when it comes to issues of occlusion and scale. Depending on the scale at which symmetry is computed, objects which were previously not important can suddenly become salient. In addition, a symmetrical image that is covered or distorted may lose the qualities which made it symmetric and the filtering will produce little response. Though we attempt to address the scaling issue by filtering across multiple scales, there is little that can be done to address issues such as occlusion without higher-level knowledge of the scene.

Acknowledgments: Supported by the National Science Foundation (grant number BCS-0827764), the Army Research Office (W911NF-11-1-0046), the Defense Advanced Research Projects Agency (HR0011-10-C-0034), and the U.S. Army (W81XWH-10-2-0076). The authors affirm that the views expressed herein are solely their own, and do not represent the views of the United States government or any agency thereof.

5. REFERENCES

- [1] Ali Borji and Laurent Itti, “State-of-the-art in visual attention modeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press. 1, 3
- [2] D. Parkhurst, K. Law, and E. Niebur, “Modelling the role of salience in the allocation of overt visual attention,” *Vision Research*, vol. 42, pp. 107–123, 2002. 1
- [3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, “Is bottom-up attention useful for object recognition?,” in *International conference on computer vision and pattern recognition*, 2004, vol. 2, pp. 37–44. 1
- [4] G. Hoffmann, “Cielab color space,” Tech. Rep., University of Applied Sciences in Emden, 2003. 1
- [5] Lingyun Zhang, Matthew H. Tong, Tim K. Marks, Honghao Shan, and Garrison W. Cottrell, “Sun: A bayesian framework for saliency using natural statistics,” *Journal of Vision*, vol. 8, no. 7, 2008. 1, 2, 3
- [6] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1254–1259, 1998. 1
- [7] David J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *J. Opt. Soc. Am. A*, vol. 4, pp. 2379–2394, 1987. 1
- [8] S.-J. Park, J.-K. Shin, and M. Lee, “Biologically inspired saliency map model for bottom-up visual attention,” *2nd Workshop on Biologically Motivated Computer Vision*, pp. 418–426, 2002. 2
- [9] P. Kovesei, “Symmetry and asymmetry from local phase,” *Tenth Australian Joint Conference on Artificial Intelligence*, pp. 185–190, 1997. 2
- [10] P. D. Kovesei, “MATLAB and Octave functions for computer vision and image processing,” Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, Available from: <<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>>. 2
- [11] L. Itti and C. Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention,” *Vision Research*, vol. 40, pp. 1489–1506, 2000. 2, 3
- [12] A. Nederveen G. Kootstra and B. De Boer, “Paying attention to symmetry,” in *Proceedings of the British Machine Vision Conference*, 2008, pp. 1115–1125. 3