

Detecting and Ranking Saliency for Scene Description

William D. Ferreira and D bio L. Borges

Pontifical Catholic University of Paran  (PUCPR)
Laboratory for Vision and Image Science
Rua Imaculada Conceicao, 1155, CEP 80215-901 Curitiba, Pr Brazil
{wdferreira,dibio}@ppgia.pucpr.br

Abstract. There is a long tradition in Computational Vision research regarding Vision as an information processing task which builds up from low level image features to high level reasoning functions. As far as low level image detectors are concerned there is a plethora of techniques found in the literature, although many of them especially designed for particular applications. For natural scenes, where objects and backgrounds change frequently, finding regions of interest which were triggered by a concentration of non-accidental properties can provide a more informative and stable intermediate mechanism for scene description than just low level features. In this paper we propose such a mechanism to detect and rank salient regions in natural and cluttered images. First, a bank of Gabor filters is applied to the image in a variety of directions. The most prominent directions found are then selected as primitive features. Starting from the selected directions with largest magnitudes a resultant is computed by including directional features in the image neighborhood. The process stops when inclusion of other points in the region makes the resultant direction change significantly from the initial one. This resultant is the axis of symmetry of that salient region. A rank is built showing in order the salient regions found in a scene. We report results on natural images showing a promising line of research for scene description and visual attention.

1 Introduction

Deciding upon what it is really important in a scene and how to extract and reason about that information can be regarded as a summary of all Computational Vision research. The idea behind saliency works is that it would be possible to find regions of interest (ROI) which are more important than others, and possibly to work towards understanding a scene without much trouble with segmentation or getting stuck with low level features uncertainty.

More recently many works have appeared in the literature proposing saliency detectors and visual attention mechanisms. In [4] Kadir and Brady show a saliency detector that starts by finding corners in the image and then analyzes the optimal scale by maximizing entropy surrounding that corner area. Many applications for the detector are given in their paper. Other interest points could

be used in the first step of a similar saliency detector, however one has to consider their intrinsic properties especially if the final task would be retrieval or visual attention. In [9] one can find an evaluation of many known interest point detectors. Reissfeld et. al. have proposed in [8] a symmetry transform which works as a saliency detector by looking for centers of symmetrical patterns in an image. Another approach to saliency has been to use more than one low level feature such as color, texture, or an edge map, and propose a way to weight and integrate those measures in order to reach a decision about the ROI. [2] and [6] are examples of that type of approach. Ranking ROI can be considered a way to compress an image, and special purpose compress algorithms based on visual saliency have been proposed. An example of such work is [10] which finds ROI in the wavelet space and use the saliency to order and divide blocks to be decomposed and reconstructed regarding their importance, or saliency rank.

Directional features are important cues in visual perception. In [1] a Self-Organizing Map (SOM) is proposed as a classification tool to detect salient structures in infrared images. The features considered as inputs to the SOM are directional and they are extracted using Gabor filters. The saliency detector proposed by us here considers directional features extracted by Gabor filtering as first primitives, and then build a novel algorithm to detect and rank a salient region by finding the more significant directions and their resultants, and finally using them as an axis of symmetry for scaling the region to be selected.

The remainder of this paper is organized as follows. Next section gives details of the proposed saliency detector dividing it into four main steps. Section 3 shows results on natural images of the detector and compares to another saliency detector proposed in [4]. Section 4 summarizes the main conclusions.

2 A Saliency Detector for Directional Visual Attention

Saliency frequently serves the purpose of compressing data by prioritizing non-redundancy. In order to understand a scene with unexpected and cluttered data as it is common in a natural image, the Human Visual System (HVS) seems to fixate around, or prioritize local information on an image. We offer here no general solution to the scene description problem, since it is known to be a kind of a local-global cycle depending on the visual task being realized [5, 11]. Our proposal is to have a saliency detector constructed around a property, believed to be present in many mammals visual system [3], that is to respond strongly to oriented patterns. By detecting and ranking salient regions in a scene, further tasks such as recognition and retrieval, or others in Active Vision which rely on visual attention may be achieved more efficiently.

We propose a saliency detector that finds and ranks regions upon which there is a strong directional feature. For this we begin by applying a bank of Gabor filters in the image, then we select the directions with highest probability. A region growing process considers then the peaks of the magnitude as seeds and compute a resultant direction until it finds an end point, or a consistently different direction. This resultant will be the axis of symmetry of the salient region found, and both its magnitude and final diameter grown will act as saliency importance, making it possible to rank the regions for visual attention or recognition tasks.

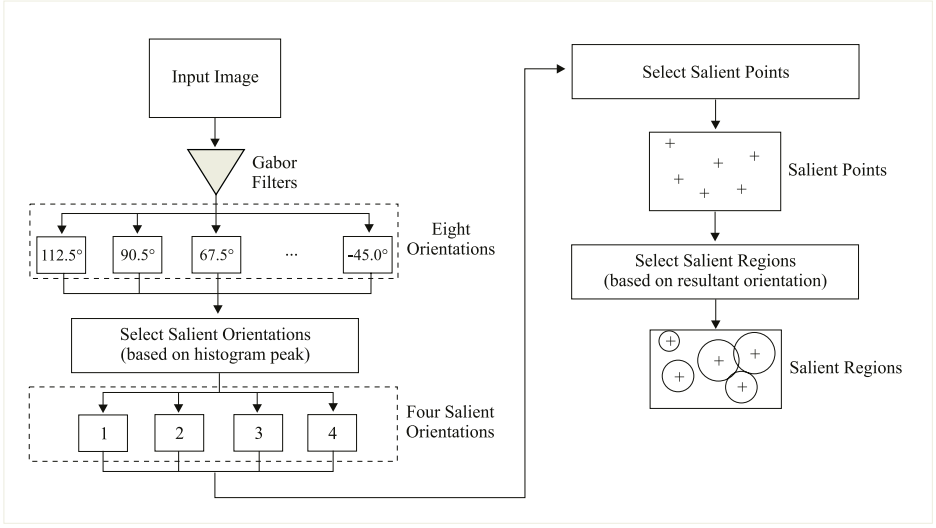


Fig. 1. General architecture of the proposed model.

Figure 1 shows a functional diagram for the proposed saliency detector. We give more details of it in the next sections dividing it into four steps: 1) Image filtering; 2) Preferred orientations; 3) Saliient points; and 4) Saliient regions.

2.1 Image Filtering

Directionality is one of the most important features to human visual perception [3], and some studies suggest that primitives of image representation in human vision have a wavelet form as Gabor filters or wavelets.

Gabor filters are a group of wavelets which can produce magnitude outputs favoring specific resolutions and orientations. A set of filtered images can be obtained by convolving an image with Gabor filters, and each of the images will represent the information at certain scale and orientation. The first step of our approach is to run Gabor filters using 1 scale, and 8 equally distant orientations (from 112.5° to -45.0°). A two dimensional Gabor function $g(x, y)$ and its Fourier transform $G(u, v)$ can be written as

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W x \right] \quad (1)$$

$$G(u, v) = \exp \left\{ -\frac{1}{2} \left[\frac{(u - W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right\}, \quad (2)$$

where $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$, and W is the frequency of the filter. A mother Gabor wavelet can be obtained by dilations and rotations of $g(x, y)$ through the function

$$g_{mn}(x, y) = a^{-m} G(x', y'), \quad a > 1, m, n = \text{integer} \\ x' = a^{-m}(x \cos \theta + y \sin \theta), \quad \text{and} \quad y' = a^{-m}(-x \sin \theta + y \cos \theta), \quad (3)$$

where $\sigma = n\pi/K$ and K is the total number of orientations. In our approach, for each input image $I(i,j)$ with dimension $m \times n$, a set of filtered images ($O = \{O_1, O_2, O_3, \dots, O_8\}$) is obtained by convolving I with Gabor filters in 8 orientations (equally spaced from 112.5° to -45.0°), 1 scale and filters dimension 60. The set of filtered images O is passed to the next step for selecting the orientations.

2.2 Preferred Orientations

Starting from the group of eight filtered images (O), obtained in previous step, a full histogram with the eight filtered images is computed. The four (4) highest peaks of this histogram are selected as the preferred orientations. Beginning by the filtered images from the previous step $O_1, O_2, O_3, \dots, O_8$, the histogram (H) is an array defined by the Equation 4.

$$H(k) = \sum_{i=1}^m \sum_{j=1}^n O(i, j), k = 1, 2, 3, \dots, 8 \quad (4)$$

The four (4) highest values in H are the chosen histogram peaks, with corresponding orientations.

2.3 Selecting Salient Points

There are $m \times n$ points in each chosen orientation given by the previous step. These points are joined in a new image named P . In the case where points will have responses in more than one orientation, the orientation with the largest magnitude value will be the one selected. In this way the image P has dimension $m \times n$, i.e. the same dimension of the input image I .

2.4 Selecting Salient Regions

The regions in our model are based on the resultant orientation. For selecting the salient regions the points obtained in the previous step are classified on order of importance, i.e. depending on the magnitude of its directional feature, and around each point a seed salient region is selected with ray r equals one, i.e. the selected region corresponds to a square matrix with dimension $m' \times n'$ which is then called candidate region (ρ). For the experiments shown here the initial salient region ($r = 1$) has $m' = 6$ and $n' = 6$. Each point $\rho(i,j)$ in candidate region has its corresponding orientation called $\theta(i,j)$.

The resultant orientation RO is computed in candidate region using Equation 5. If the RO has a value which diverts more than 22.5° of the central point orientation, the growing process stops. However, if the mentioned condition is not fulfilled, the candidate region ray is increased of one unit ($r = r + 1$), and the process is repeated.

$$RO = \arctan \left(\frac{\sum_{i=1}^{m'} \sum_{j=1}^{n'} \rho(i, j) \times \sin(\theta(i, j))}{\sum_{i=1}^{m'} \sum_{j=1}^{n'} \rho(i, j) \times \cos(\theta(i, j))} \right) \quad (5)$$

In the experiments shown here candidate regions grow until size 64x64, however an automatic stop criterion for this could be further explored looking at redundancy, either for recognition or retrieval tasks. Figure 2 shows a complete example, highlighting the outputs of the four steps of the method.

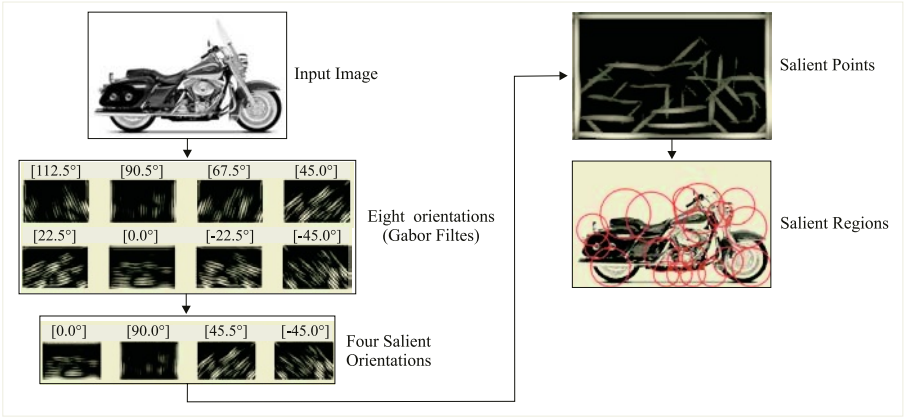


Fig. 2. An example of the steps of the proposed saliency detector.

3 Results on Natural Images

Natural images are usually cluttered because of sunlight and textured objects appearing at different distances. We tested the proposed saliency detector in a variety of natural images and show here some typical results. For these experiments the maximum size of the region would have diameter of 64 pixels. For the scale of the images shown this diameter finds meaningful regions, although this variable can be set differently depending on the image size, scale, and task it is aimed at describing.

Other saliency detectors have been proposed in the literature, and as we commented before each prioritizes either an specific feature to be searched for, or an integration model. Visually comparing results would serve more to enhance what each considers more salient and stable. In [4] Kadir and Brady propose an entropy based saliency detector, which starts at finding corners in an image and then enlarge its window until entropy is maximized. In Figures 3, 4, and 5 it is shown results with both saliency detectors for comparison. Numbers appearing on the regions borders indicate their importance rank in our detector.

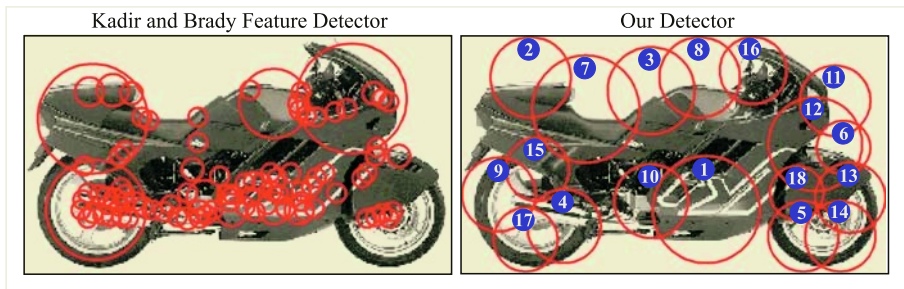


Fig. 3. Results of both saliency detectors (Kadir and ours) on images with object and clear background.

The Figure 3 shows an image with clear background, and it is observed that Kadir and Brady feature detector selects important regions to characterize the image. However, important regions were detected by the proposed model also, and the number of regions selected by it is less than Kadir and Brady feature detector. The reduced number of regions is an important characteristic to diminish redundancy.

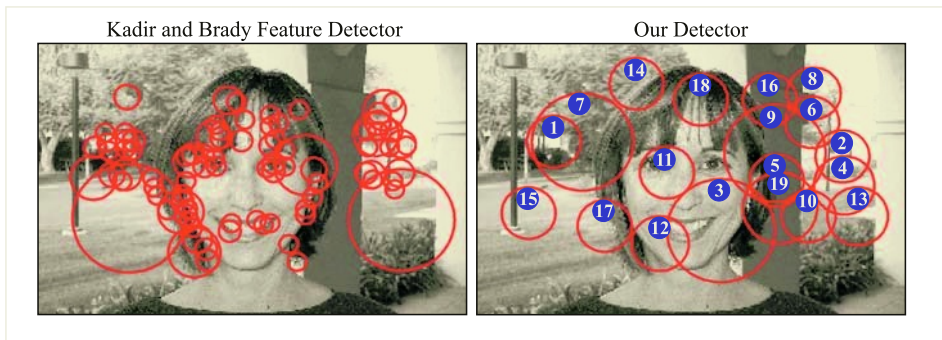


Fig. 4. Results of both saliency detectors (Kadir and ours) on images with object and cluttered background.

It is possible to verify in Figure 4 that in an image with textured background the feature detector proposed by Kadir and Brady selects large regions on the background, and it leaves out important regions in the face area, like mouth, nose and hair. Those regions are regarded as salient in the detector here proposed as shown in Figure 4. In the car sequence images of Figure 5 stability of the regions selected can be observed in our detector by looking at regions 1 and 4. Their relative importance and size remain closely similar (from one frame to the other), while the other detector changes its consideration on the object car.

As it can be verified by the examples shown the proposed saliency detector finds very interesting and informative regions in natural and cluttered scenes. For the purpose of demonstration we have also included results of another known

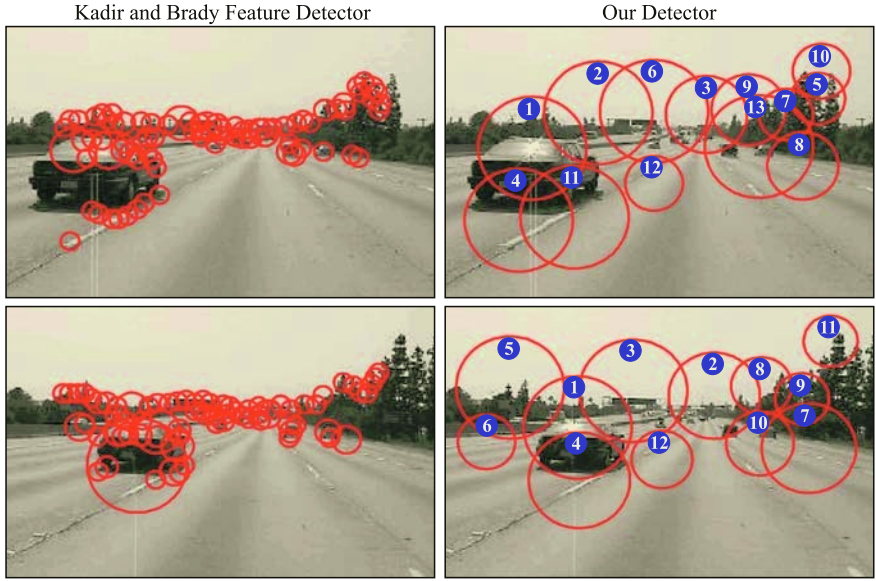


Fig. 5. Results of both saliency detectors (Kadir and ours) on a sequence of images with objects in different positions and cluttered background.

saliency detector [4] from the literature. Our proposed saliency detector prioritizes direction and symmetry notions, and serves the purpose of a reliable tool for scene description, especially if meant to visual attention tasks in Active Vision, recognition, and retrieval. Next section summarizes our main conclusions and point to future work.

4 Conclusions and Future Work

In this paper we have presented a region saliency detector based on the idea that regions of interest in a natural scene can be found around oriented patterns. A complete approach using Gabor filtering initially to locate the main responses to orientations, and then a selection and growing process using a resultant direction as an axis of symmetry of the region of interest is built and tested. Further, a recognition or retrieval process can make efficient use of such a scheme, since no segmentation is necessary for finding and ordering those regions as it was shown here.

Visual attention mechanisms and saliency provide a new paradigm for scene description in Computational Vision research, and there is an increase of interesting solutions appearing in the literature more recently. Unveiling the integration of attention models and saliency in the Human Visual System (HVS) is yet an out of reach target, and saliency detectors such as the one here proposed are meant to be just one of the useful mechanisms to be further explored. Future

work is planned exploring the output regions of the detector for recognition and retrieval.

Acknowledgements

This research is in part supported by a grant from PUCPR (PRPPG Research Fund), and a grant from CNPq - MCT - Brazil to DLB.

References

1. M. Carreira, M. Mirmehdi, B. Thomas, and J. Haddon. Grouping of Directional Features using an Extended Hough Transform. In *Proc. of Int. Conf. on Pattern Recognition (ICPR)*, pages 990-993, 2000.
2. A. Dimai. Invariant Scene Description based on Salient Regions for Preattentive Similarity Assessment. In *Proc. of Int. Conf. on Image Analysis and Processing (ICIAP)*, pages 957-962, 1999.
3. D. Hubel. *Eye, Brain, and Vision*. W. H. Freeman and Company, USA, 1995.
4. T. Kadir and M. Brady. Saliency, Scale and Image Description. *Int. Journal of Computer Vision*, vol.45, n.2, pages 83-105, 2001.
5. S. Kosslyn. *Image and Brain*. MIT Press, USA, 1994.
6. J. Luo and A. Singhal. On Measuring Low-Level Saliency in Photographic Images. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1084-1089, 2000.
7. D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, USA, 1982.
8. D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context Free Attentional Operators: the Generalized Symmetry Transform. *Int. Journal of Computer Vision*, vol. 14, pages 114-130, 1995.
9. C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of Interest Point Detectors. *Int. Journal of Computer Vision*, vol. 37, No. 2, pages 151-172, 2000.
10. T. Slowe and I. Marsic. Saliency-based Visual Representation for Compression. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 554-557, 1997.
11. S. Ullman. *High-level Vision*. MIT Press, USA, 1996.