Saliency Detection Based on Heuristic Rules

Diana E. Martinez-Rodriguez, Victor Ayala-Ramirez[(✉)],
and Uriel H. Hernandez-Belmonte

División de Ingenierías, Campus Irapuato- Salamanca DICIS,
Universidad de Guanajuato, Carr. Salamanca-Valle Km. 3.5+1.8, Palo Blanco,
36700 Salamanca, Mexico
{demr,hailehb}@laviria.org, ayalav@ugto.mx

Abstract. Detecting salient regions in images aims at finding regions
which contains relevant information, where a more detailed process can
be applied. Saliency detection is useful in many computer vision tasks
such as object segmentation, object detection, image retrieval, place
recognition, among others. In this paper, we propose a method based
on heuristic rules that uses color and spatial features which allows to get
a good approximation to the salient region in a very low time compared
with other methods in the state of the art. The tests were performed
over the images of a benchmark dataset.

1 Introduction

Humans can easily identify the relevant objects in a scene because their inher-
ent mechanism of visual attention. Human visual system models suggest that
humans selectively process perceived information instead of taking all in mind
[18]. The visual attention model attributed to Neisser states that there are two
stages in the visual saliency task: a pre-attentive stage and an attentive stage.
In the first stage, the features are detected and in the second stage, the visual
system finds relationships between them.

Visual saliency computation methods try to emulate the human visual atten-
tion methods. The first computational approach for saliency detection was pro-
posed by Itti *et al.* [8]. The main contribution of this work is the proposal of
the saliency map. This map is an image-like representation where the intensity
is proportional to the relevance of the corresponding pixels.

Saliency detection has been widely used in computer applications. It has
been used in object segmentation [11], object recognition [17] adaptive image
compression [4] and place recognition [19]. In particular, saliency detection is
useful because it reduces the computational cost of these tasks by focusing the
process in reduced regions instead of processing the entire image.

Nothdurft [16] and Ma and Zhang [15], coincide in the statement that there
is not a feature *per se* that captures human attention. For example, the saliency
of an object does not depends on a particular color but on its contrast with
respect to its neighborhood.

In recent works, the color contrast has been received too much attention. Luo *et al.* [14] consider the global color contrast as the key property for saliency and they also assign more importance to the center of the image. Ma and Zhang [15] consider local contrast in the CIELuv space, it computes the pixel saliency by using a local window. Liu and Gleicher [12] create a Gaussian pyramid of the image in order to be invariant to scale, the distance is computed by using a L_2 norm in CIELuv color space.

Saliency detection methods can be classified as biologically based [8], purely computational based [5,14], or a combination of both [3]. We propose a pure computational method that uses color and spatial features. The proposed approach reduces the computational cost by using a grid representation of the image. Each cell is composed of a rectangular region of a given size. This operation does not significantly reduce the quality of the saliency detection because the humans are attracted by objects and not by individual pixels. In a second step, cells that are connected and that exhibit similar color properties are grouped into non regular regions. Heuristic rules are applied to determine which regions are the seeds of the background and foreground regions. Another set of heuristic rules is used to assign the rest of the regions either into the salient foreground or the background. The final foreground is a meaningful region, in the sense proposed by [17]. The main advantage of our method is that obtains comparable results to the methods in the state of the art but in considerably less time.

We consider to find salient regions as a binary classification labeling: salient or not salient. This is to avoid the problem of thresholding the image, something that is necessary for most of the applications, as in the object segmentation problem. We have found that the thresholding method affects significantly the results.

The content in this paper is organized as follows: In Sect. 2 the proposed approach is described. In Sect. 3 we present the tests and the results. A comparison with previous approaches is also presented. Finally conclusions and perspectives are presented in Sect. 4.

2   Heuristic Rules Based Saliency Detection

In this section a Heuristic Rules based Saliency Detection method is presented (HRSD for short). The proposed approach looks for a meaningful region that satisfies a set of constraints both in color contrast and in spatial arrangement to consider it as a salient region. The objective of the method is to partition the image into two regions: the foreground region (R_F) and the background region (R_B). Each of these regions are constructed through a number of steps described below.

1. An input image is partitioned into a grid of rectangular regions, each of them containing $m \times n$ pixels. Each of these regions will be assigned either to R_F or to R_B. Let us name the cells of the grid as C_i, $i \in \{1, 2, \ldots, u \times v\}$ with u and v being respectively the number of rows and columns resulting from the grid partition of the input image.

2. Each cell is characterized by the mean value $\bar{C}_i$ of the color coordinates of all the $m \times n$ pixels belonging to it. We use the YUV color space to represent pixel color in this work. That is:

$$\bar{C}_i = [\bar{C}_{iY}\ \bar{C}_{iU}\ \bar{C}_{iV}] \tag{1}$$

Each of the components of $\bar{C}_i$ is computed as follows:

$$\bar{C}_{iY} = \frac{1}{m \times n} \sum_{C_i} c_{iY}(i,j) \tag{2}$$

where $C_{iY}(i,j)$ is the color coordinate of the pixels in the region R_i. Similarly for U and V coordinates, we have:

$$\bar{C}_{iU} = \frac{1}{m \times n} \sum_{C_i} c_{iU}(i,j) \tag{3}$$

$$\bar{C}_{iV} = \frac{1}{m \times n} \sum_{C_i} c_{iV}(i,j) \tag{4}$$

3. Cells with similar color features are then grouped using a connected component labeling-like procedure explained later in this paper. As a result, we obtain a list of component regions R_i, $i \in \{1, 2, \ldots, r\}$, with r the total number of connected components found in the image. Each of these regions groups cells that are similar among them with respect to its mean color $\bar{C}_i$ and that are spatially connected.
4. The cells of each region R_i that are located in the boundary cells of the image are counted and the sum is recorded as B_i.
5. The initial selection of R_B and R_F is done by choosing the pair of more contrastive regions in $R = \{R_1, R_2, \ldots, R_r\}$. For doing this, we compute a table of distances D where the element d_{ij} is the distance between the color mean values of the regions R_i for each pair of regions R_i and R_j, $i \neq j$. The color distance is computed using an Euclidean distance in the YUV color space, weighted by a factor depending on the size of both regions as shown in Eq. 5. Where the variables $size_i$ and $size_j$ are the sizes of the regions i and j respectively. The variable $size$ is the number of pixels of the whole image.

$$dist_{ij} = d_{ij}\left(\frac{size_i + size_j}{size}\right) \tag{5}$$

$$dcolor = \sqrt{(Y_1 - Y_2)^2 + (U_1 - U_2)^2 + (V_1 - V_2)^2}$$

For the more contrastive pair, R_B is chosen as the region covering the larger number of cells.

We have chosen to use the YUV color space because it has shown better performance when compared to CIELab, CIELuv and HSI color spaces for image

representation in our experiments. The parameters were tuned experimentally to optimize the performance evaluation measure of the system.

There exist some heuristic rules that R_F and R_B should satisfy. If that is not the case, the next distance in the rank of region color distances is used to choose the R_F and R_B and the verification of the heuristic rules is repeated.

The heuristic rules are as follows:
(a) The salient object is in the center of the image. The R_F (foreground) is limited to have at most 5 cells in the boundary of the image. This is to avoid selecting a region of the background as the foreground.
(b) The size of the representative objects must be above a 3 cell area threshold. The R_F and R_B selected have to be initially composed by at least 3 cells. This is to try to avoid choosing an artifact of the image as a salient region.
6. In the following step, the rest of the regions are grouped either to R_F or to R_B. This procedure is guided by another set of heuristic rules that includes spatial relationships. The rules to determine if a region R_i is salient are the following:

(a) $dist_{R_i R_F} < dist_{R_i R_B}$,
(b) R_i does not contain cells in the contour of the image.

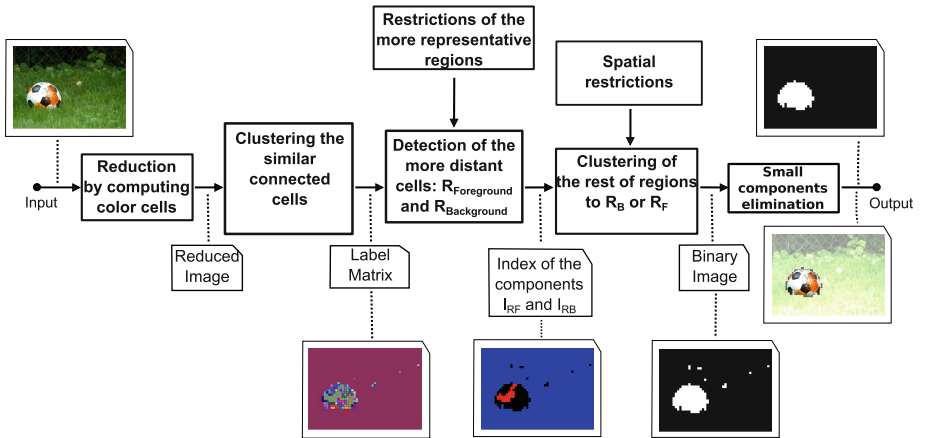In the Fig. 1, a graphical block diagram of our method is presented.

Fig. 1. Heuristic Rules based Saliency Detection Method

2.1   Color Connected Component Labeling (CCL) Procedure

As it was mentioned before, we generate regions from cells of a grid. The CCL task is performed by extending the work by Hernadez-Belmonte *et al.*
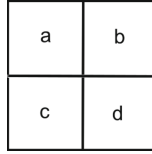
Fig. 2. Reduced Connectivity Mask [7].

[7]. The key concept of this work is the use of a Reduced Connectivity Mask (RCM) and the use of a lookup table that determines if the regions need to be connected as a component or not. We consider a neighborhood of these cells in the grid, as it is shown in the Fig. 2. Let as assume that d is the cell under analysis. The scanning of the image using the RCM, is performed in a left to right, top to bottom sequence.

- If the cells (b, c) are similar, join their labels.
- If cell (d) is similar to one of the neighbor labels (a, b, c). In the Table 1 the operations in each case are presented.
- If d is not similar to the other cells, create a new label.

The criteria for defining the similarity of two cells is the Euclidean distance in YUV coordinates. YUV coordinates include one luminance (brightness) and two chrominance (color) components [10]. If the distance between two colors is lower than a threshold we consider that colors are similar. We selected color because is a very important feature, but another features (e.g. texture-related features) could also be used by using an appropriate threshold.

3   Tests and Results

For the evaluation of our method we used the standard images from the MRSA, a widely used image dataset for saliency evaluation. We compare our approach with other state of the art methods using the standard F metrics. We also compare our system in execution time with the best approach in or knowledge [5]. Our system was implemented in Language C and the tests were executed using an Intel Core i7-4700MQ machine with 8 GB RAM.

Table 1. Operations which have to be computed in each case for the color CCL procedure. (1) The two grids are similar, (0) The two grids are not similar (-) It is not necessary to verify.

d, b	d, a	d, c	Operation
1	-	-	Join d to b.
0	1	-	Join d to a.
0	0	1	Join d to c.
0	0	0	Create a new label

3.1   Test Protocol

In order to evaluate our method, we use the MSRA image dataset images. The images in this dataset present a variety of situations: there are images for indoor scenes and outdoor scenes; there are also images including natural and artificial objects. The salient regions in these images represent humans, animals, plants, and objects. Achanta *et al.* provide the ground truth for a subset of 1000 images of the original dataset MSRA [1]. We use same subset of 1000 images to compare the proposed method to the other methods in the same conditions. The metrics used to evaluate the results are the well known *precision* and *recall* metrics combined into the *F-measure* presented in the Eq. 6. The *precision* and *recall* metrics are computed using the Eqs. 7 and 8. Where B are the salient pixels detected by the HRSD method method and G are the salient pixels in the ground truth; x and y are the coordinates of the pixel under analysis. We use $\beta = 0.3$ to weight precision more than recall. That is more convenient in objects segmentation, and that is used by most of the automatic saliency methods

$$F = \frac{(1 + \beta)(P \cdot R)}{\beta(P) + R} \tag{6}$$

$$P = \frac{\sum_{(x,y)} B(x, y)G(x, y)}{\sum_{(x,y)} B(x, y)} \tag{7}$$

$$R = \frac{\sum_{(x,y)} B(x, y)G(x, y)}{\sum_{(x,y)} G(x, y)} \tag{8}$$

3.2   Results

The best results of the Heuristic Rules based Saliency Detection (HRSD) method were obtained by using the YUV color space for image representation, and a cell size of 8×8 pixels.

In the Fig. 3, we present some qualitative results obtained with the proposed method. In the Fig. 4 a histogram of the F-measure results for the entire dataset is presented, we can see that more of 400 images obtain a very high F-measure value between 0.9 and 1.0.

3.3   Comparisons

At first, we present the results of the F-measure of some methods in the state of the art. This is in order to establish how well perform our system against the other methods. The results presented are the reported results in the works [9,14] with the codes provided by original authors. Finally, we show the comparison with [5]. We use the last version of the implementation provided by the authors. This is in order to compare the executing time with our system. The both methods were tested in the same machine.

Most of the saliency methods obtain a continuous saliency values. For comparison purpose we need a binary image. To binarize the image we need to choose
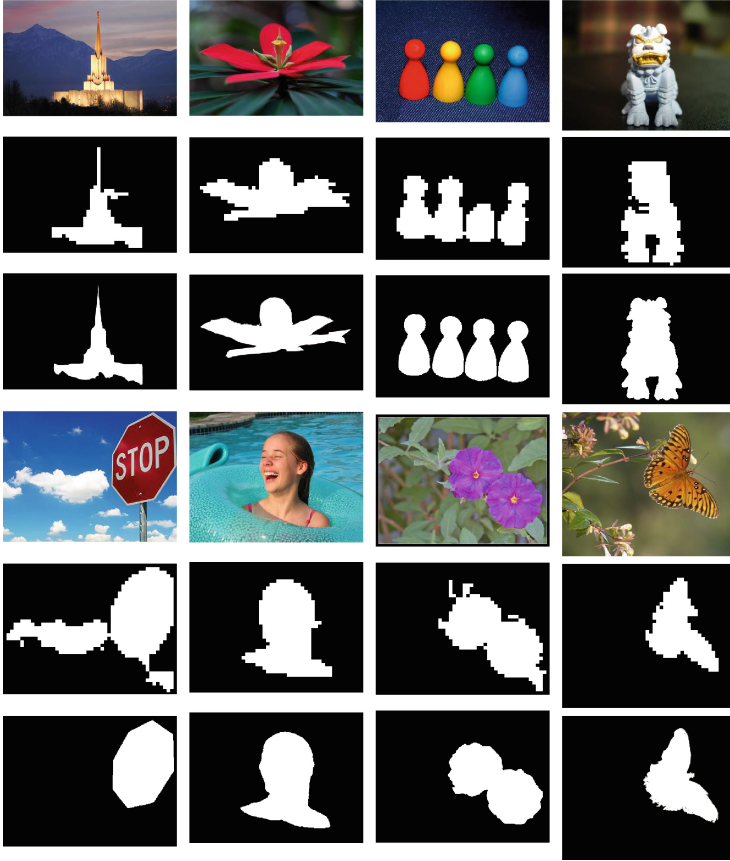
Fig. 3. Qualitative examples of the output of the HRSD method compared to the ground truth. In the first and fourth rows the original images are shown; in the second and fifth rows the HRSD output is presented; in the third and sixth rows the available ground truth is shown.

a threshold. For this reason, the results reported in works by Luo et al. [14] and Kannan et al. [9] may vary from the results reported by the original authors when a non optimal threshold is used.

In the Fig. 5(a), we present the results reported by Luo [14] including the methods labeled CA [6], KD [13], RC [5], OS [21], using Otsu as the method to choose the threshold. In the Fig. 5(b), we present the results reported by Kannan [9] for the methods labeled CA [6], FT [1], RC [5], UL [20], HC [5] using Eq. 9 to compute the threshold. At the end of both graphs, we have added a column with our results. In the comparisons presented, they report different results for RC and CA, this is caused by the use of a different method to define the threshold.
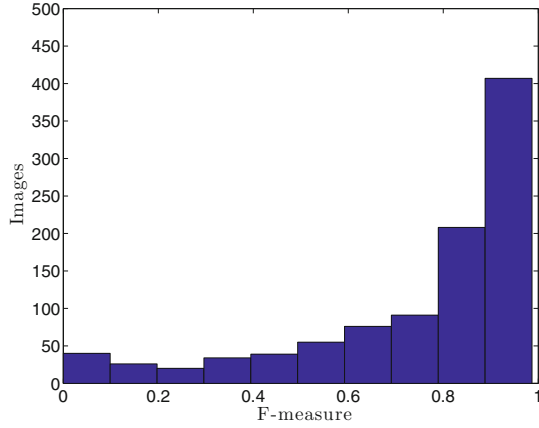
$$Th = 2 \sum S_{(x,y)} \tag{9}$$

Fig. 4. Histogram of the F-measure of the 1000 images resulting from the application of the HRSD method.
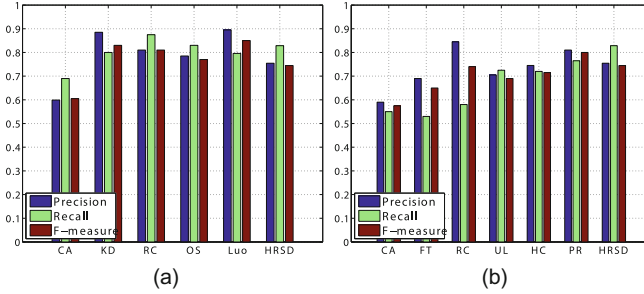

Fig. 5. (a) Results comparison by Luo [14]. (b) Results comparisons by Kannan [9]. We add to both graphics the average results of HRSD method (Color figure online)

In Table 2 we present the results obtained by our approach when compared with the RC method, using the code available from https://github.com/MingMingCheng/CmCode.git. We compare our results with the RC method [5] because it is, in our best knowledge, the faster method in saliency detection with relatively good results. The time of most of the other methods which performs in several seconds. As we can see, the F-measure value is practically the same,

Table 2. Comparison results in F-measure and Time.

Method	F-measure	Time
RC	0.77	41.2 ms
HRSD	0.74	20.8 ms

but the time needed by the proposed approach to perform the computation is about the half of the time spent by the RC method.

4   Conclusions and Perspectives

In this paper, we presented a method to find the salient regions in images. This method considers in addition to the use of color contrast among regions, the use of spatial information to determine the salient regions.

We process the image by using a grid of regularly spaced cells in both horizontal and vertical directions and we group similar image cells by using a color connected component labeling algorithm in the YUV space. The regions formed in such way are then classified into foreground or background regions according to a heuristic set of rules.

The results obtained by our approach are comparable to the results obtained by the state of the art method proposed by Cheng. However, the method computes the saliency output in half the time in the average than that method.

Future work will be directed towards implementing a computational intelligence algorithm for the automatic setting of the parameters of the approach. This could improve the efficiency of heuristic rules for the correct foreground association of the grouped regions. Another line of research will be to make the heuristic rules adaptive to take advantage of specific features of each image.

References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision andPattern Recognition (CVPR), pp. 1597–1604 (2009)
2. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: a benchmark. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 414–429. Springer, Heidelberg (2012)
3. Bruce, N., Tsotsos, J.: Attention based on information maximization. J. Vis. **7**(9), 950 (2007)
4. Chen, T., Cheng, M.M., Tan, P., Shamir, A., Hu, S.M.: Sketch2photo: internet image montage. ACM Trans. Graph. **28**(5), 124 (2009)
5. Cheng, M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **37**(3), 569–582 (2015)
6. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **34**(10), 1915–1926 (2012)
7. Hernandez-Belmonte, U.H., Ayala-Ramirez, V., Sanchez-Yanez, R.E.: Enhancing CCL algorithms by using a reduced connectivity mask. In: Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Rodríguez, J.S., di Baja, G.S. (eds.) MCPR 2012. LNCS, vol. 7914, pp. 195–203. Springer, Heidelberg (2013)

8. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 1254–1259 (1998)
9. Kannan, R., Ghinea, G., Swaminathan, S.: Salient region detection using patch level and region level image abstractions. IEEE Signal Process. Lett. **22**(6), 686–690 (2015)
10. Kekre, H.B., Thepade, S.D., Athawale, A., Parkar, A.: Using assorted color spaces and pixel window sizes for colorization of grayscale images. In: International Conference and Workshop on Emerging Trends in Technology, pp. 481–486 (2010)
11. Ko, B.C., Nam, J.Y.: Object-of-interest image segmentation based on human attention and semantic region clustering. J. Opt. Soc. Am. A **23**(10), 2462–2470 (2006)
12. Liu, F., Gleicher, M.: Region enhanced scale-invariant saliency detection. In: IEEE International Conference on Multimedia and Expo (ICME), pp. 1477–1480 (2006)
13. Liu, Z., Xue, Y., Shen, L., Zhang, Z.: Nonparametric saliency detection using kernel density estimation. In: IEEE International Conference on Image Processing (ICIP), pp. 253–256 (2010)
14. Luo, S., Liu, Z., Li, L., Zou, X., Le Meur, O.: Efficient saliency detection using regional color and spatial information. In: European Workshop on Visual Information Processing (EUVIP), pp. 184–189 (2013)
15. Ma, Y.F., Zhang, H.J.: Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of the Eleventh ACM International Conference on Multimedia, pp. 374–381 (2003)
16. Nothdurft, H.C.: Salience from feature contrast: additivity across dimensions. Vis. Res. **40**(10), 1183–1201 (2000)
17. Ren, Z., Gao, S., Chia, L.T., Tsang, I.W.H.: Region-based saliency detection and its application in object recognition. IEEE Trans. Circ. Syst. Video Technol. **24**(5), 769–779 (2014)
18. Sha, C., Li, X., Shao, Q., Wu, J., Bian, S.: Saliency detection via boundary and center priors. In: 6th International Congress on Image and Signal Processing (CISP), vol. 2, pp. 1066–1071 (2013)
19. Siagian, C., Itti, L.: Rapid biologically-inspired scene classification using features shared with visual attention. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **29**(2), 300–312 (2007)
20. Siva, P., Russell, C., Xiang, T., Agapito, L.: Looking beyond the image: unsupervised learning for object saliency and detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3238–3245 (2013)
21. Zhang, X., Ren, Z., Rajan, D., Hu, Y.: Salient object detection through over-segmentation. In: IEEE International Conference on Multimedia and Expo (ICME), pp. 1033–1038 (2012)