# Salient Region Detection Using Contrast-Based Saliency and Watershed Segmentation

**Christopher Wing Hong Ngau, Li-Minn Ang, Kah Phooi Seng**

*School of Electrical and Electronic Engineering*
*The University of Nottingham Malaysia Campus*
*Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia*
*Tel: +603 8924 8000, Fax: +603 8924 8002*
*E-mail: {keyx8nwh, kezklma, kezkps}@nottingham.edu.my*

**ABSTRACT**

*Salient region detection is useful for many applications such as image segmentation, compression, image retrieval, object tracking, and machine vision systems. In this paper, an approach to detect salient regions in a visual scene using contrast-based saliency and watershed segmentation is presented. The approach allows salient objects to be detected and extracted for analysis while preserving the actual boundaries of the salient objects. The approach can be executed in parallel making it efficient for real time applications.*

**Keywords**

*Region Detection, Saliency, Watershed segmentation.*

## 1.0 INTRODUCTION

With the increase in application concerning visual images in machine vision systems, considerable amount of research have been done in recent years to provide robust and intelligent visual interpretations on data from visual sensors or cameras (Aziz et al., 2008). One important role in visual interpretation is to locate important objects which are most likely salient to the human visual system. However, to locate these objects using a machine as a human does will require deep understanding of the human visual and psychology which is beyond the current available research. One alternative is to provide an estimate of the area of the important objects through visual saliency which simulates the low-level stimuli of the human vision. Visual saliency has been applied to areas of visual search in complex scenes (Rajashekar et al., 2002), traffic signs detection (Won et al., 2007), object tracking (Ouerhani et al., 2003), image retrieval (Bamidele et al., 2004), image watermarking (Park et al., 2002), image compression (Bradley et al., 2003), and video compression (Itti, 2004).

Locating objects in visual scenes where high level information such as faces, text, and colours are available can be very straight-forward. However, in certain visual scenes where this information is unavailable, detection of salient objects could be difficult. Nevertheless, in almost any visual scene be it grayscale or coloured images, saliency in the form of contrast exists. The contrast could be in the form of differences in colour, intensity, size or any feature which has a gradual change from one area to another area in the same image. Contrast-based saliency operates on the areas of feature change and highlights parts where the changes occur. The regions where the change occurs are usually the salient parts in the image.

Many pixel-based saliency models (Itti, 2000), (Ma et al., 2003), (Walther et al., 2006), (Liu et al., 2006) use several scales to compute the contrast in the image. In this method, images are first computed into different pyramid levels. Images in each level are then resized to a certain size to provide fine and coarse representations of the contrast features. The resized images are summed to obtain the pre-saliency map or the feature map. Although this method is effective in emphasizing the feature contrast, there is a possibility that some salient objects appearing in certain scales will be lost in the final saliency map. Furthermore, the resolution of the saliency map in this method causes the boundaries of salient objects to be roughly highlighted. When this saliency map is used together with the segmented regions from the watershed transform, inaccurate boundary detection will occur as the saliency values will overflow into non-salient parts of the segmented image.

In this paper, an approach on salient region detection using the contrast-based saliency model in (Achanta et al., 2008) and watershed segmentation is presented to increase salient detection performance and to provide a more accurate border representation of salient objects. Achanta et al., 2008 proposed a method to maintain the resolution of the saliency map by using resizable block filters. The time taken in computing the saliency map is reduced using integral images. The resulting saliency map is then used to calculate the average saliency per region on the watershed transformed image. Regions with average saliency values above a predetermined threshold value will be extracted. Extraction based on region segmentation and a high resolution saliency map will ensure a more accurate border representation of the salient objects.

## 2.0 SALIENT REGION DETECTION

In this section, details on the determination of the salient regions and extraction will be discussed. The algorithm consists of two separate modules which can be executed in parallel for real-time applications. Figure 1 shows the overview of the salient region detection algorithm. The first module computes the saliency maps and the second module computes the segmented regions using the watershed transform. The modules are discussed in the following sections.

### 2.1 Contrast-based Saliency

Contrast can be found in many different forms depending on the type of visual input. Among the commonly found contrasts in images are the colour contrast and the luminance contrast. Normally in visual scenes, colour contrast is more preferable due to the abundance of colour information. There are several efficient methods available (Ma et al., 2003), (Liu et al., 2006), and (Achanta et al., 2008) to determine contrast saliency for colour images. The method proposed by (Achanta et al., 2008) is preferred due to its ability to maintain the original image resolution.

For the contrast-based saliency map, the saliency is evaluated as the contrast between an image sub-region and its surrounding neighbourhood at various scales. The image sub-region is denoted as $R_1$ and the surrounding neighbourhood pixels enclosed as a region is denoted as $R_2$, shown in Figure 2. The region $R_1$ is taken to be a single pixel although a small region of $N \times N$ can be used to allow detection in a noisy image. The size of region $R_2$ is varied to allow filtering at different scales while maintaining the image resolution.

The generic contrast is given as the distance between the average feature vector of the sub-region and the average feature vector of the pixels in the neighbourhood. In this detection approach, since the sub-region consists of a single pixel, the contrast-based saliency value of a pixel at any point on the image is given as:

$$c_{i,j} = D\left[\mathbf{v}_p, \left(\frac{1}{N}\sum_{q=1}^{N}\mathbf{v}_q\right)\right] \quad (1)$$

where $N$ is the number of pixels in region $R_2$, $\mathbf{v}$ is the feature vector elements corresponding to a pixel, and $D$ is the Euclidean distance between the feature vector $\mathbf{v}_1$ and the average feature vector of $\mathbf{v}_2$ (region $R_2$).

The input image colour space is first converted from RGB to CIELab to provide a more natural colour range. Then, the vector corresponding to the pixel location $(i, j)$ is given as

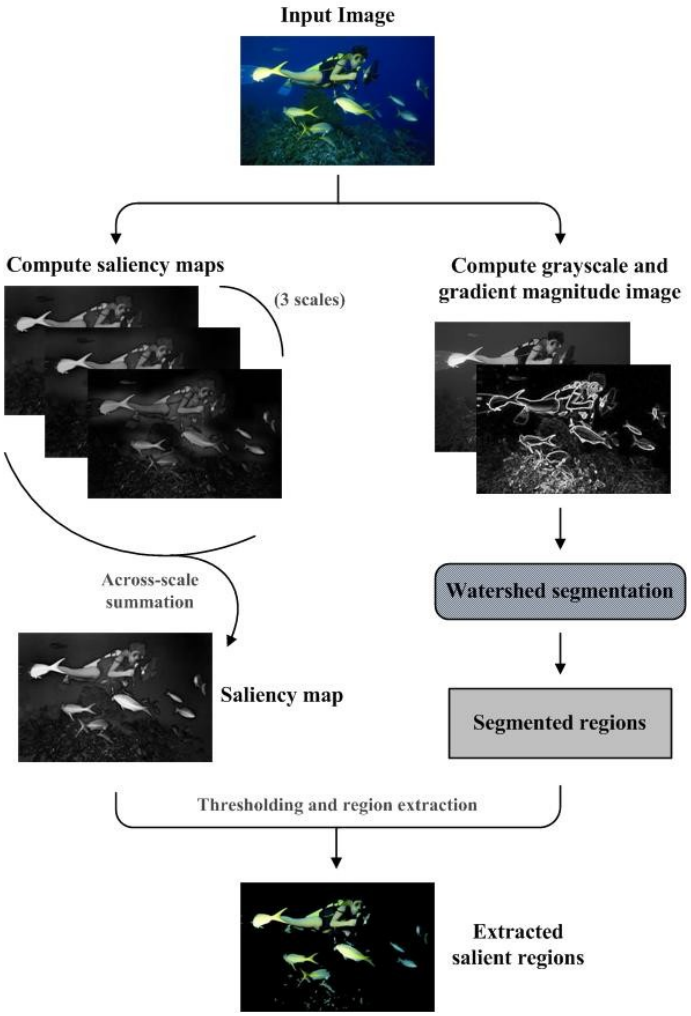


*Figure 1:* An overview of the salient detection algorithm.

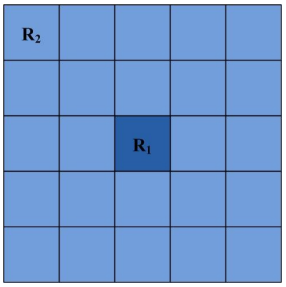


*Figure 2:* Filter for detecting contrast with sub-region $R_1$ and the surrounding neighbourhood pixels (region $R_2$).

$\mathbf{v}_1 = [L_1, a_1, b_1]^T$ and the average feature vector of region $R_2$ is given as $\mathbf{v}_2 = [L_2, a_2, b_2]^T$. Since the contrast at location $(i, j)$ is simply the Euclidean distance between $\mathbf{v}_1$ and $\mathbf{v}_2$, the contrast saliency can be written as:

$$c_{i,j} = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \qquad (2)$$

*Figure 3:* Contrast-based saliency map at three different scales

The final saliency map is determined by across-scale summation of the saliency maps at different scales shown in Equation (3).

$$m_{i,j} = \sum_{S} c_{i,j} \qquad (3)$$

where $S$ is the number of scales.

### 2.2 Watershed Segmentation Based On Rainfall Simulation and Region Extraction

Watershed segmentation based on rainfall simulation is a method which simulates the rain fall over the surface associated with the image. The rain that falls on a point will flow along a path with the steepest descent until it reaches a minimum point. The point where the rain fall is labelled belongs to the catchment basin associated with this minimum. The rain fall process is repeated throughout the whole image until all the points are assigned to a minimum. The surface is then divided into catchment basins. Each point then can be said to belong to a catchment basin and there will be no watershed line as compared to the watershed segmentation based on flooding.

In the region extraction procedure, the watershed transformation algorithm based on rainfall simulation in (Osma-Ruiz et al., 2007) is applied to the gradient magnitude image to provide segmentation of sensible regions. The watershed algorithm of Osma-Ruiz et al. is used to allow efficient computations of catchment basins and region labelling. Compared to the immersion based watershed algorithm by Vincent et al., 1991, this algorithm not only is computationally efficient, but at the same time overcomes the problem of watershed lines having the width of more than one pixel. In this watershed segmentation, the regions are labelled and separated without a watershed line to give accurate representations of borders of the neighbourhood regions. A comparison of the immersion and rainfall simulation watershed segmentations is shown in Figure 4.

The Sobel operator is used to generate the gradient magnitude image by convolving the grayscale of the original image with a small 3-by-3 separable filter in both the horizontal and vertical directions. Given the grayscale image $I(x, y)$; $H_x$ and $H_y$ are the two filters which detect changes in the vertical and horizontal directions respectively:

$$H_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, H_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad (4)$$

Vertical $G_x$ and horizontal $G_y$ gradients can be found by convolving the grayscale image with the two filters respectively:

$$G_x = I(x, y) * H_x \text{ and } G_y = I(x, y) * H_y \qquad (5)$$

where the symbol $*$ denotes convolution.

The gradients in the vertical and horizontal direction can be combined to give the gradient magnitude using the Euclidean distance equation:

$$G(x, y) = \sqrt{G_x^2 + G_y^2} \qquad (6)$$

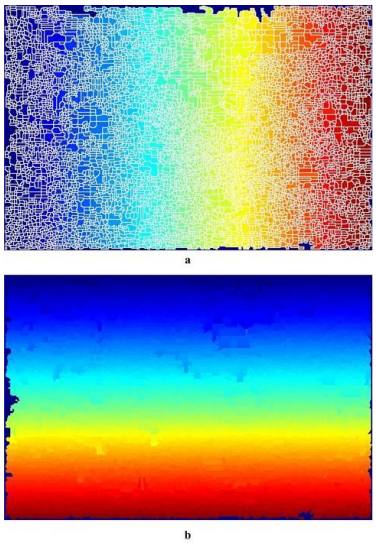


*Figure 4:* Watershed segmentation performed on the airplane image. a) Immersion based watershed b) watershed by rainfall simulation.

The watershed transform is then applied to the gradient magnitude image. The detailed algorithm of the watershed transform can be found in (Osma-Ruiz et al., 2007).

After the segmented regions $r_k$ for $k = 1,2,\ldots K$ are obtained, the average saliency of each region is calculated by summing the contrast-based saliency values of each pixels in the particular region using the final saliency map and dividing the sum with the number of pixels in that region as shown in Equation (7).

$$Rs_k = \frac{1}{Nr_k} \sum_{i,j \in r_k} m_{i,j} \tag{7}$$

where $Nr_k$ is the number of pixels in region $r_k$ . The segmented regions whose average saliency exceeds a threshold $T$ is segmented out and the rest are discarded.

## 3.0 SIMULATION RESULTS AND DISCUSSION

Simulations were performed to compare the detection and extraction of salient regions using the contrast-based saliency map and Itti's saliency map. Both contrast-based saliency map and Itti's saliency map are first computed using colour test images from the Berkeley image database (http://www.eecs.-berkeley.edu/). Itti's version of the saliency map is generated using the Saliency Toolbox (Walther et al., 2006). Then, segmented regions which have an average saliency value above a pre-determined threshold are extracted for comparison. The threshold value used in this simulation is set to 0.15 for most images except for the deer image in Figure 5 where the threshold value is set to 0.25.

Figure 5 show the comparison between the results generated using the contrast-based saliency map and Itti's saliency map. Since there are no established methods for salient region comparison, the simulation results are compared visually. It can be seen that by using the contrast-based saliency map and the watershed segmentation, more or less correct salient regions are managed to be extracted. In the case of using Itti's visual saliency model, only certain salient locations are extracted. Furthermore, the number of salient regions extracted using Itti's model is far less compared to the contrast-based saliency model.

In our approach, watershed segmented regions infused with saliency values from the saliency map computed using contrast provide a more accurate detection and extraction of salient objects. Contrast computation using distance between pixels and their surrounding neighbourhoods while only varying the neighborhood size will result in a high resolution saliency map. Since the resolution of these maps is considerably good, regions of high saliency value tend to smoothly spread throughout a certain region shape instead of clustering like a

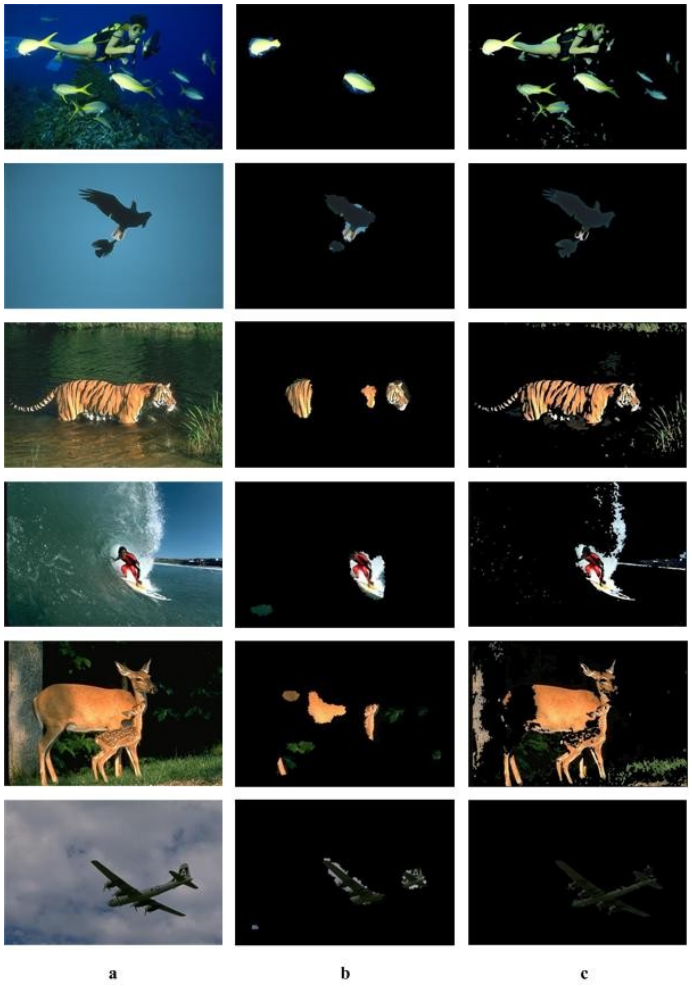


*Figure 5:* Salient detection results. a) Original image, b) extracted salient regions using Itti's model and watershed segmentation, c) extracted salient region using contrast-based saliency and watershed segmentation.

crude patch of pixel blocks. Because of this, it is very likely that regions forming a salient object will survive the thresholding step. These regions tend to preserve the boundary of the salient object.

In Itti's model, the saliency map is computed by resizing the salient points (usually of a single pixel or a small pixel block) in the summed conspicuity maps. When these points are resized, they lose resolution causing them to look like large blobs with random sizes. Due to this factor, not all the regions forming an object will have a saliency value above the threshold value. This causes the irregular salient regions shapes to be formed as seen in Figure 5b.

In our approach, there are two noticeable limitations. The first limitation is that small non-salient regions are extracted as seen in some of the result images. This is due to the nature of over-segmentation caused by the watershed transform. These non-

salient regions usually have low saliency values corresponding to the saliency map. However, due to their small region area, the average saliency values of these regions are sufficient to survive the thresholding step. The threshold value is raised particularly for the deer image due to many of these regions appearing in the image.

The second limitation is that certain large regions do not have the sufficient saliency value to overcome the thresholding step while their neighbouring regions have far much higher value than the threshold value. Should the threshold value be lowered to make providence for the large regions, many non-salient regions will be extracted along with the salient ones. The effect of insufficient salient value in large regions can be seen in the deer image in Figure 4. Both of the mentioned limitations can be corrected with pre and post-processing to reduce the number of regions before the extraction part takes place.

## 4.0 CONCLUSION

A salient region detection approach using contrast-based saliency and watershed segmentation is proposed. It provides a means of accurate salient region representation and extraction. The approach can be performed in parallel to allow efficient computations when applied to real time applications. This approach has promising and significant usefulness in particular applications especially in the emerging research area of unsupervised video segmentation and attention-based image retrieval.

## REFERENCES


A. Bamidele, F. Stentiford, and J. Morphett (2004). An Attention-based Approach to Content-Based Image Retrieval. *BT Technology Journal, 2004, 22(7), pp. 151-160.*

A. P. Bradley and W. M. Stentiford (2003). Visual Attention for Region of Interest Coding in JPEG 2000. *J. Vis. Commun. Image Represent., pp. 232-250.*

Berkeley Database from http://www.eecs.berkeley.edu/

D. Walther and C. Koch (2006). Modelling Attention to Salient Proto-objects. *Neural Networks 19, pp. 1395-1407.*

F. Liu and M. Gleicher, Region Enhanced Scale-Invariant Saliency Detection. *Multimedia and Expo, 2006 IEEE Conference on Volume, Issue, July 9-12, 2006, pp. 1477-1480.*

L. Itti (2004). Automatic Foveation for Video Compression using a Neurobiological Model of Visual Attention. *IEEE Trans. Image Processing, Oct 2004, Vol 13. No. 10, pp. 1304-1318.*

L. Itti (2000). Models of Top-Down and Bottom-Up Visual Attention. *PhD Thesis, California Institute of Technology, Pasadena, California.*

L. Vincent and P. Soille (1991). Watershed in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Trans. Pattern Analysis Machine Intelligence 13(6), pp. 583-598.*

M. C. Park and K. J. Cheoi (2002). A Digital Image Watermarking Using A Feature-Driven Attention Module. *Proceeding (364) Visualization, Imaging, and Image Processing.*

M. Z. Aziz and B. Mertsching (2008). Fast and Robust Generation of Feature Maps for Region-Based Visual Attention. *IEEE Transaction on image Processing, Vol.17 No. 5, May.*

N.Ouerhani and H.Hugli (2003). A Model of Dynamic Visual Attention for Object Tracking in Natural Image Sequences. *Proceedings of IWAN, pp. 702-709.*

R.Achanta, F. Estrada, P. Wils, and S. Susstrunk (2008). Salient Region Detection and Segmentation. *Computer Vision Systems, Springer Berlin / Heidelberg, pp. 66-75.*

U. Rajashekar, L. Cormack, and A. Boyik (2002). Visual search: Structure from Noise. *Proceedings of Eye Tracking Research and Applications Symposium, New Orleans, LA, USA, March 25-27, pp.119-123.*

V. Osma-Ruiz, J. I. Godino-Llorente, N. Saenz-Lechon, and P. Gomez-Vilda (2007). An Improved Watershed Algorithm Based on Efficient Computation of Shortest Paths. *Pattern Recognition 40, pp. 1078-1090.*

W.-J. Won, S. Jeong, and M. Lee (2007). Road Traffic Sign Saliency Map Model. *Proceedings of Image and Vision Computing New Zealand 2007, Hamilton, New Zealand, December 2007, pp. 91-96.*

Y.-F. Ma and H.-J. Zhang, Contrast-based Image Attention Analysis by Using Fuzzy Growing. *International Multimedia Conference, Proceedings of The 11th ACM International Conference on Multimedia, 2003, pp. 374-381.*