# An Improved Image Segmentation Algorithm for Salient Object Detection

Yuee Liu [1,2], Jinglan Zhang [1], Dian Tjondronegoro [1], Shlomo Geva [1], Zhengrong Li [1,2]

[1] Faculty of Information Technology, Queensland University of Technology, Australia
[2] College of Information Engineering, Northwest A & F University, China

Email: y53.liu@student.qut.edu.au

## Abstract

*Semantic object detection is one of the most important and challenging problems in image analysis. Segmentation is an optimal approach to detect salient objects, but often fails to generate meaningful regions due to over-segmentation. This paper presents an improved semantic segmentation approach which is based on JSEG algorithm and utilizes multiple region merging criteria. The experimental results demonstrate that the proposed algorithm is encouraging and effective in salient object detection.*

**Keywords**: semantic segmentation, salient object, JSEG, region merging

## 1 Introduction

With the advancement of communication networks and Internet services, there has been rapid increase in the use of multimedia contents. A huge amount of images are generated and stored every day. As a result, techniques for content based image retrieval (CBIR) are becoming increasingly important. In most cases, low-level features (e.g. colour, shape, texture) are extracted for representing the image content. However, it is discovered that the desired content of an image is localized rather than a global view of the image [1]. Investigation also shows that humans naturally try to capture semantic objects first when they watch a visual scene because they are perceptually dominant [2]. This is called "visual saliency" which refers to that certain parts of a scene are pre-attentively distinctive and create some forms of immediate significant stimulation within the early stages of Human Vision Systems (HVS).

Although object extraction from images is critical, it is still not fully achieved using current technologies. S*alient object*s are firstly proposed as a more effective middle-level representation of image content [3]. They do not exactly correspond to the real objects, but they could capture most common visual properties of object classes. It is defined as visually distinguishable image compounds that can characterize visual properties of corresponding object classes. For example, a salient object "sky" could be described as a set of connected image regions with dominant image components that can be detected semantically by human as "sky". Accurate segmentation can be used to improve: local feature extraction, shape detection, and multi label image annotation, etc.

An important approach for salient object detection is segmentation [4]. Image segmentation is the process of partitioning an image into non-overlapping regions which allows users to identify meaningful concepts or objects which would be perceptual to human. Therefore, developing a suitable image segmentation technique which effectively partitions image into salient regions is an important step toward salient object detection.

The JSEG (J-measure based segmentation) algorithm provides colour-texture homogeneous regions which are useful for salient region detection [5]. According to a recent survey, JSEG seems to be the most widely used method in natural image segmentation so far [6]. Although it has been shown to be robust and computationally effective on a variety of natural images, it still often fails to produce accurately segmented salient objects. This is caused by its weakness in handling the spatially variation of illumination, which usually results in over-segmentation. In this paper, several extensions to the well-known JSEG algorithm are proposed among which the most important is a novel merging criteria. Firstly, JSEG segmentation is applied to over-segment the image. Salient objects are then obtained by merging the similar regions based on a homogenous criterion. It is re-defined based on colour and syntactic features. These extensions have improved JSEG in producing more complete salient regions.

## 2 The Proposed Method

In all segmentation methods, an image is usually partitioned into either too coarse or too fine components (i.e. under-segmentation and over-segmentation. Compared to under-segmentation,

978-1-4244-2582-2/08/$25.00 ©2008 IEEE

over-segmentation increases the chance to extract the important boundaries [7].

The good segmentation is usually evaluated only using a colour criterion. For example, J-measure based on the dominant colours is used for "goodness" of segmentation in JSEG. However, a single measure can not capture the rich content of the image. Image segmentation depends on many factors (e.g. homogeneity, compactness). In this paper, we combine *colour* and *geometric* criterion to extract the salient objects.

## 2.1 Initial Segmentation

JSEG segmentation relies on the J criteria to calculate the ratio of distance between different classes and distance between members in the same class. A small J indicates the good segmentation.

Firstly, an image is quantized using the peer group filtering method, and then regions are grown from seeds. After that, regions are merged based on their colour histogram in the perceptually uniform CIE LUV colour space.

Experiment shows that although results are satisfying, most boundaries of objects can be kept, over-segmentation often occurs. Therefore, a single criterion (i.e. colour property) cannot describe whether good regions are found.

## 2.2 Merging Strategy

Syntactic features, which are firstly proposed by Ferran and Casas, represent geometric properties of regions and their spatial constraints like homogeneity, compactness, regularity, inclusion or symmetry [8]. It has been shown that these features can partition an image into more meaningful regions without considering any application dependent knowledge. Therefore, besides commonly used colour criteria, syntactic features are also taken into account.

### 2.2.1 Colour Homogeneity Criteria

All region growing segmentation methods (e.g. JSEG ) start from seeds and extends to other regions by calculating the homogeneity between two adjacent regions. Dominant colours of regions are most commonly used. They are represented using

$$f_c = \{(c_i, p_i), i = 1,...,M, p_i \in [0,1]\} \quad (1)$$

where $c_i$ is the dominant colour that is a 3-D vector in a perceptual colour space (e.g. HSV, Luv, Lab); $p_i$ is the corresponding percentages of that dominant colour; $M$ is the total number of all dominant colours of a region.

However, that commonly used dominant colour descriptor $f_c$ assumes that the colours of images are not sensitive to the variations of the illumination and perspective. This is true for the fabric images, but does not stand in the natural images. Spatially adaptive dominant colours is proposed to make up this inefficiency [9]. We adapt this dominant colour descriptor to the perceptual uniform Luv colour space, because Luv is proven to be more robust than the widely used HSV colour space. This dominant colour descriptor comprises a set of locally adapted dominant colours and their corresponding percentage of occurrence of each colour within a certain neighbourhood. Mathematically, it is described

$$f_c(x, y, N_{x,y}) = \{(c_i, p_i), i = 1,...,M, p_i \in [0,1]\} \quad (2)$$

where $c_i$ is a 3D vector that contains the dominant colour; $p_i$ is the corresponding percentages of that colour; $N_{x,y}$ denotes the neighbourhood of the pixel $(x, y)$ and $M$ is the total number of colours in the neighbourhood. $M$ depends on different image regions, and typically it is $M = 4$.

Optimal colour composition distance (OCCD) is used to measure the perceptual similarity between colour regions. It has been proven better than other colour similarity metrics in that it takes both colour and area differences [9].

### 2.2.2 Syntactic Feature Criteria

Two syntactic features (i.e. adjacency and regularity) are used to evaluate the goodness of the salient object detection method [8, 10].

- **Adjacency**

It is well known that "real world" objects tend to be compact. So they exhibit adjacency characteristics of their constituent parts. This characteristics is represented as following

$$C_{adj}(i, j) = 1.0 - \frac{l_{ij}}{\min\{l_i, l_j\}} \quad (3)$$

where $l_{ij}$ is the length of the common boundary between two neighbouring regions $r_i$ and $r_j$; $l_i$ and $l_j$ are their corresponding perimeter lengths.

This measure provides quite rich information about geometrical relations between both regions. If values of $C_{adj}(i, j)$ approximates to zero, it indicates that almost total inclusion whereas if values are close to 1, it indicates weak adjacency. Therefore, this measure provides strong evidence against merging.

- **Regularity (low complexity)**

Intuitively, more complex shapes have a longer perimeter compared to simple shapes. Therefore, global shape complexity $x_i$ of a region $r_i$ can be measured as the ratio between its perimeter length $l_i$ and the square root of its area $a_i$ :

$$x_i = l_i / \sqrt{a_i} \tag{4}$$

Average changes of global shape complexity $x_i$ for two neighbouring regions $r_i$ and $r_j$ are measured as:

$$C_{cpx}(i,j) = \frac{x_{ij}}{\left( \frac{a_i x_i + a_j x_{ji}}{a_i + a_j} \right)} \tag{5}$$

where $x_{ij}$ is the shape complexity of a hypothetical region by merging $r_i$ and $r_j$.

Low values of $C_{cpx}(i,j)$ results in a merging of the two regions. If values are below 1.0, it indicates that the complexity of the joint region is lower than the average of the complexities of two original regions.

### 2.2.3 Region Merging Using Multiple Criteria

After colour homogeneity and syntactic features are calculated, they are combined together to determine whether two adjacent regions should be merged or not. This region merging process continues until a certain stop criteria are reached. The operation sequence is as follows:

(1) For each region $r_i$, calculate the adjacency and regularity of all its adjacent regions.

(2) If the regularity of a region $r_j$ is below 1, and its adjacency is larger than a threshold, merge $r_i$ and $r_j$.

# 3 Experiment and Discussion

The effectiveness of segmentations is evaluated on a collection of images that are collected from LabelMe[11]. It is a web-based image annotation tools that allows researchers to label images and share the annotations with the rest of the community. The images in LabelMe is not only annotated but also segmented collaboratively. 500 images and their corresponding manual segmentations are downloaded for testing. Figure 1 demonstrates several sample images and their ground-truth segments. We intentionally collect images with various objects, such as animals, buildings, mountains, etc.

The segmentation efficacy is usually characterized by accuracy given the ground truth. The degree to which the segmented results agree with the manual segmentations could be obtained by visual evaluation (matching the segmentation results with human segmentations) and quantitative evaluation.

The quantitative evaluation is based on Jaccard coefficient $P$ of the region coincidence between the segmentation result and the ground truth [12]. It is proven that this method is insensitive to small variations in the ground-truth construction and incorporates the precision and recall measurement into a unified function. Let the whole image $\Omega$ be partitioned into a set of disjoint segments $\{r_1,...,r_n\}$, where $n$ is the number of final segments, $r_i \cap r_j = \Phi$ and $\cup_{i=1}^{n} r_i = \Omega$, precision rate $P$ is mathematically shown as following [12]

$$P(r_1, r_2, ..., r_n; A) = \frac{|A \cap R|}{|A \cup R|} \tag{6}$$

where segmentation result $R = \bigcup_{i:\rho(r_i,A)>0.5} r_i$ ; $A$ is ground truth, and $\rho(r_i, A) = \max\{\frac{|r_i \cap A|}{|r_i|}, \frac{|r_i \cap A|}{|A|}\}$.

**Figure 1:** Sample images and their manually-created ground-truth

Region number is a very important criterion to determine the "goodness" of segmentations. If the region number of final segments in images is equal to the exact number of objects, the segmentation results are much higher acceptable. Based on this reason, region number is incorporated as the penalty factor. The revised $P$ is shown below.

$$P = \frac{1}{\sqrt{TR}} \times \frac{|A \cap R|}{|A \cup R|} \tag{7}$$

where $TR$ is the number of the final segments.

Figure 2 demonstrates some segmentation results generated by hand, JSEG and our algorithm. In Figure 3, JSEG segmentation results with different region

merging threshold are visually compared with our algorithm. Figure 4 shows the objective comparisons between JSEG and our proposed. In Figure 3, the horizontal axis represents each objects of ground truths that are manually extracted, and the vertical axis represents the detection precision of each object.

It is discovered that the segmentations produced by the proposed algorithm matches the manual ground truths much more than JSEG in Figure 2. Regions that belong to a salient object are merged together. Compared to the ground truths, some objects are missed not only by JSEG but also by our algorithm. For example, in row 2 of Figure 2, two people can not be detected. In this case, it is not important whether they can be well segmented because they are too small, very similar to the contextual environment, and not perceptually dominant. Additionally, whether they can be well extracted does not influence the recognition of the beach scene because Concept Sea and Beach are determinate. Therefore, integration of colour and geometric features can significantly improve the segmentation results. Figure 4 provides more visual evaluation results between ground truth, JSEG algorithm, and our algorithm.

From the objective evaluation in Figure 4, as JSEG algorithm partitions the image into too many fine regions, the precision of object detection is smaller than our algorithm.

In the experiment, the parameter region merging threshold of JSEG is to be 0.3. If the value of this parameter is higher, more regions will be merged, as shown in Figure 3. However, images are still over-segmented, and small regions can not be integrated to form a complete salient object. Additionally, some regions are merged wrongly. For example, jauger is merged with the background. Our algorithm extracts more complete salient objects.

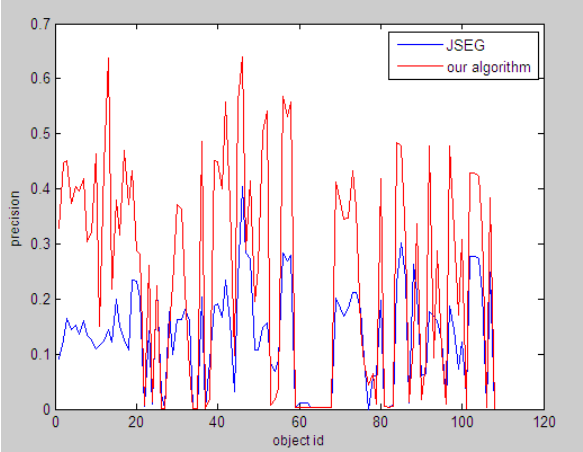


**Figure 4:** Object evaluation result between JSEG and our algorithm

**Figure 2:** Selected segmentation results from top to bottom: Original images, Segmentations by hand, JSEG algorithm results and proposed segmentations. The region merging threshold for JSEG algorithm is set to $l = 0.3$

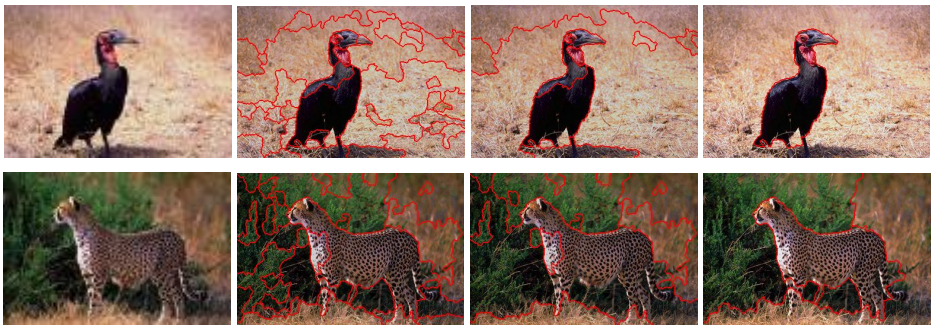

**Figure 3:** Segmentation result comparison between JSEG algorithm (with different region merging parameter) and our algorithm. The images from left and right: original image, JSEG result with different merging threshold $l = 0.3$ , $l = 0.7$ and our algorithm respectively

# 4 Conclusion

In this paper, an improved JSEG is proposed by combining multiple criteria including spatially adaptive dominant colours and syntactic features. Experiments show that the segmentation results satisfied the purpose of salient object detection, and provide an effective solution to the over-segmentation problem in the JSEG algorithm.

# References


[1] R. Rahmani, S. A. Goldman, H. Zhang, J. Krettek, and J. E. Fritts, "Localized content based image retrieval," in ACM Internatonal Workshop on Multimedia Information Retrieval Singapore, 2005.

[2] N. Zlatoff, G. Tyder, B. Tellez, and A. Baskurt, "Content-based image retrieval: On the way to object features," in The 18th International Conference on Pattern Recognition Hong Kong, 2006.

[3] J. Fan, Y. Gao, H. Luo, and G. Xu, "Salient Objects: Semantic Building Blocks for Image Concept Interpretation," in 3rd ACM International Conference on Image and Video Retrieval, Dublin, Ireland, 2004.

[4] Y.-H. Kuan, C.-M. Kuo, and N.-C. Yang, "Color-based Image Salient Region Segmentation Using Novel Region Merging Strategy," IEEE Transactions on Multimedia, vol. 10, pp. 832-845, 2008.

[5] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 800-810, 2001.

[6] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," Pattern Recognition, vol. 40, pp. 262-282, 2007.

[7] "Superpixel: Empirical Studies and Applications." vol. 2008, 2004.

[8] C. F. Bennstrom and J. R. Casas, "Binary-partition-tree creation using a quasi-inclusion criterion," in Proceedings of 8th International Conference on Information Visualisation, London, UK, 2004.

[9] J. Chen, T. N. Pappas, A. Mojsilovie, and B. E. Rogowitz, "Adaptive perceptual color-texture image segmentation," IEEE Transactions on Image Processing, vol. 14, pp. 1524-1536, 2005.

[10] T. Adamek and N. E. O'Connor, "Using Dempster-Shafer Theory to Fuse Multiple Information Sources in Region-based Segmentation," in IEEE International Conference on Image Processing (ICIP2007), San Antonio, Texas, USA, 2007.

[11] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: a Database and Web-based Tool for Image Annotation," International Journal of Computer Vision, vol. 77, pp. 157-173, 2008.

[12] F. Ge, S. Wang, and T. Liu, "Image-Segmentation Evaluation From the Perspective of Salient Object Extraction," in IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006.

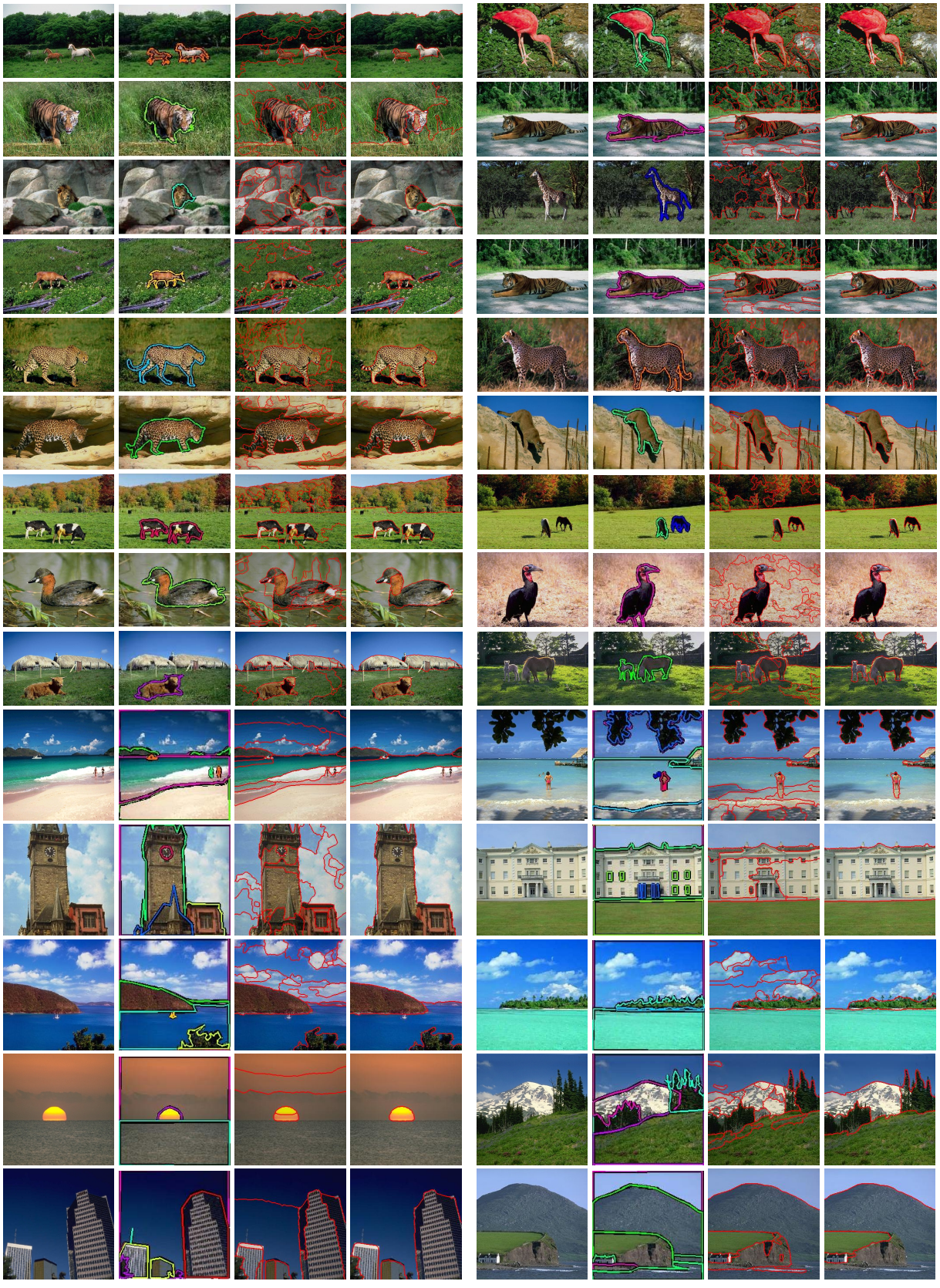


**Figure 5:** Selected segmentation results. Parameters for JSEG algorithm is the region merging threshold $l = 0.3$ for JSEG algorithm. (a) and (e) are original images; (b) and (f) are manual segmentations; (c) and (g) are JSEG results; (d) and (h) are our results.