

# Saliency Detection in Aerial Imagery Using Multi-Scale SLIC Segmentation

Samir Sahli<sup>1</sup>, Daniel. A. Lavigne<sup>2</sup>, and Yunlong Sheng<sup>1</sup>

<sup>1</sup>Center of Optics, Photonics, and Lasers, Image Science group, Laval University, Quebec, QC, Canada

<sup>2</sup>Defence Research and Development Canada – Valcartier, Quebec, QC, Canada

**Abstract** - Object detection in a huge volume of aerial imagery requires first detecting the salient regions. When an image is over-segmented by the superpixels, the latter will adhere to object boundaries, resulting in their shape deformation and size variation, which can be used as the saliency measure. The normalized Hausdorff distances from the inner pixels to boundary of the superpixels are then transformed to the posterior probability of saliency useful to build the saliency map and the salient region map. The method implemented by the multi-scale Simple Linear Iterative Clustering (SLIC) is simple without a priori knowledge on the object, computational efficient, robust to low image contrast, free of parameter tuning and is therefore suitable for aero-surveillance applications.

**Keywords:** segmentation, salient region, aerial imagery, superpixel, SLIC, multi-scale

## 1 Introduction

With advent of new sensing hardware and fast growing data transmission capability, the gap between the amount of data and the means at disposal for processing them rapidly increases. In aero-surveillance applications, large-scale high-resolution aerial images are acquired in huge quantity in order to preserve awareness of the global scene. However, the objects of interest to be detected are usually small and localized in the regions much smaller than the size of the aerial imagery. Thus the problem to be addressed is “what object in the image attracts the most the visual attention?” In the context of aerial imagery, resources should be prior allocated to salient regions detection. This change of paradigm can be seen as a divide and conquer strategy. Once the salient regions are highlighted, a further analysis would be dedicated for the objects detection. This mechanism is similar to that of the brain and the Human Visual System (HVS).

Many computer vision methods have been developed for saliency measure and salient region detection [1-9]. Koch and Ullman were among the first to introduce a set of topographical feature maps that, once combined, provide a saliency map measuring the degree of “conspicuity” of each pixel over the entire image [1]. Rarity and surprise can be also used as saliency measure [2-7]. However, their detection

requires feature extraction, selection and combinations steps. The center-surround filter has been proposed [7-9] to mimic the human visual system that tends to focus its attention on the center of the visual field. This approach may be counter-productive in aerial imagery where the targets may appear anywhere in the image. The adaptive Gaussian Mixture Model was suggested [8] for partitioning images into background and attention regions. This approach requires a supervised learning and the tuning of a set of parameters. The complexity detected as the maximum of Shannon’s entropy in the histogram has been used as a saliency measure [10]. In aerial imagery, self-similar or fractal-like structures such as forest or trees may appear in the image. They are known to have a high complexity but also as non salient structures. Hence, the use of complexity as criterion for salient region detection, as proposed in [10], is irrelevant. Saliency detection may be also treated as a segmentation task using statistical method [7] or combining salient features classified with a mixture of linear Support Vector Machines (SVMs) [11].

Most of the existing methods for saliency detection are not specifically designed for aerial imagery surveillance applications, where one needs detecting the salient regions with the methods, which are computational efficient to process large size imagery, without training and parameter tuning nor a priori knowledge on the objects. Rosin’s method [12] only requires the calculation of Distance Transform (DT) on a set of thresholded binary images, and the application of an adaptive threshold to their summation. Hou and Zhang [13] propose a method independent of any features. The saliency map is created by inverting the Fourier Transform of the spectral residual firstly obtained by subtracting redundant spectral information from the original log-magnitude spectrum.

In this paper, we propose a method for capturing salient regions in aerial imagery using the superpixel image segmentation with the multi-scale Simple Linear Iterative Clustering (SLIC) algorithm. Over-segmenting an image with SLIC produces superpixels which the contours would deform to tightly adhere with the boundaries objects present in the image. Hence, the size and shape irregularity of the superpixels can be used as a new saliency measure. They are then quantified using the Hausdorff distance and the sigmoid posterior probability function in order to create the saliency map and the binary salient region map. The method is fast as an  $O(N)$  operator, where  $N$  is the number of pixels in the images, and is robust to image low contrast. The SLICs with

multiple scales of the superpixels are applied to the image, but without modifying its original resolution.

The paper is organized as follows. Section 2 describes the proposed method; section 3 depicts experiments on the aerial imagery with a comparison with the state-of-the-art methods; section 4 concludes with a discussion on further improvements.

## 2 Method

### 2.1 Superpixel segmentation

Superpixels are coherent local clusters of pixels, which preserve the homogeneity and most image information of the image. The use of superpixels instead of pixels reduces the complexity of subsequent pixel-based image processing tasks. The superpixels may be used for image segmentation, combining the region and extracting edge information. We selected the Simple Linear Iterative Clustering (SLIC) method [14] because it outperforms the state-of-the-art methods, such as Normalized Cut [15], Turbopixel [16], Mean-shift [17], and Quick-shift [18]. SLIC requires a low computational cost while it achieves a high degree of segmentation quality typically measured by boundary recall and under-segmentation error [14].

In the SLIC, the only one parameter to input is the desired number of superpixels  $K$ , similarly to that in the  $K$ -mean clustering. For an image of  $N$  pixels, the average size of a superpixel is  $S \times S$  pixels with  $S = \sqrt{N/K}$ . For superpixel clustering, a grid of  $K$  seeds equally spaced by the distance  $S$  is first set in a color image. Then, the seeds are moved to the lowest gradient position in the color space away from edges in the image. After this step, each image pixel is associated to the nearest cluster center according to the Euclidian distance in the 5-dimensional space, combining the *CIELAB* color space and the  $X$ - $Y$  space. Competition between the neighbouring seeds delimits the spatial extent and forms boundaries of the respective superpixels. After all the pixels are associated to the cluster centers, the clusters centers are moved to the centroids of the newly formed superpixels, and each image pixel is again reassigned to the new centroids to form new set of superpixels. The clustering procedure iterates until the cluster centers converge to stable positions.

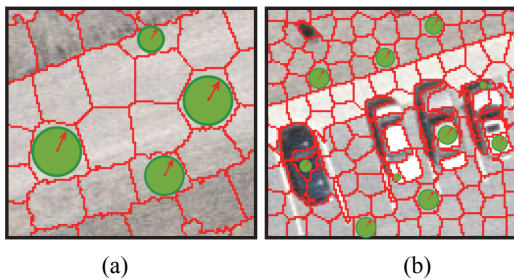


Fig. 1 (a) Superpixels of the average size and regular shape in the background, except those close to boundary of route; (b) superpixels become irregular in size and shape with presence of the vehicles, for instance.

In the image segmented by the superpixels, the shapes and sizes of the superpixels depend on their scale  $S$  with respect to the size of the regions. In the large and mostly uniform background regions, such as land, sea and snow field, the segmentation by small scale superpixels results in the superpixels of average size and regular shape, as shown in Fig.1a. However, with the presence of the objects of interest, the superpixels will tightly adhere to the object structural boundaries, resulting in strongly deformed shapes with varying sizes, as shown in Fig.1b. Observing the over-segmentation results, we noticed that when the scale of the superpixels is much larger than the size of the objects of interest, the objects and their structural details may be ignored in the segmentation process. On the contrary, when the superpixel scale is smaller than the sizes of the objects of interest, the irregularity in shape and size of the superpixels may serve as a measure of saliency.

**Input :** Color image  $I$ ; Superpixel Scales  $S_i$  ( $i = 1..p$ )

**Output :** Saliency map  $Smap$  ; Mask of salient regions  
 $Mask\_Salient\_Reg$ .

**ForEach** scale  $S_i$  **do**

(1) Segment  $I$  using SLIC at scale  $S_i$  :

$SLIC(I, S_i) \rightarrow Seg$

(2) Get a superpixel boundary map from  $Seg$  :

$BinaryMap = getBoundaries(Seg)$

(3) Calculate the distance transform on the binary map :

$DTmap = DT(BinaryMap)$

**ForEach** superpixel in  $Seg$  **do**

(4) Find the maximum in  $DT$  map as Hausdorff distance :

$H = \max\{DTmap\}$ ;

(5) Compute the shape factor from the normalized  $H$  :

$ShF = PlNorm(H)$ ;

(6) Update the saliency map at scale  $S_i$  :

$Smap\_i(pixel) = ShF$ ;

**End**

(7) Accumulate saliency map from each scale  $S_i$  :

$Smap += Smap\_i$ ;

(8) Capture salient region from scale  $S_i$  using pixel-based probabilistic decision rule

$Smap\_i(pixel) \rightarrow Mask\_Salient\_Reg_i$

**End**

(9) Create final mask with salient regions

$$Mask\_Salient\_Reg = \bigcup_{i \in 1..p} Mask\_Salient\_Reg_i$$

Fig. 2 Multi-scale SLIC for capturing salient regions

In the case when the sizes of the objects of interest are unknown, or the objects of interest have various sizes over the scene, the multi-scale SLIC with a set of preset scale values  $S$  may be applied. Hence, the final saliency map may be a summation of the saliency maps resultant of contribution from multiple scale levels. As when the superpixel scale is much larger or much smaller than the size of the object of interest, most superpixels will be of average size and regular

shape, which do not contribute much to the saliency. The main contributions to the final saliency map come only from the superpixels, which over-segment the objects of interest. Note also that we apply the multi-scale SLIC to the original image without scaling the image.

The overall process for computing the saliency map is described in Fig.2. We first segment a color image  $I$  using the SLIC at multiple scales  $S_i = \sqrt{N / K_i}$  with  $i=1\dots p$ ,  $K_i$  the desired number of superpixels at scale  $S_i$  (step 1). At some scales  $S_i$  smaller than the size of the objects of interest, the superpixels nearby the object boundaries are deformed. The Distance Transform (DT) is applied to the binary map of superpixel boundaries (step 2, 3) and the Hausdorff distance from the superpixel inner pixels to the boundary pixels is computed (step 4) and normalized (step 5) in order to compute the shape factor, which describes superpixels shape and size changes as a saliency measure (step 6). The final saliency map is an addition of the salient maps computed at all the scales  $S_i$  with  $i = 1\dots p$  (step 7). The salient regions,  $Mask\_Salient\_Reg_i$ , are extracted at each scale using a pixel based probabilistic decision rule (step 8), defined in section 2.2. The final salient regions is a logic OR of the salient regions at all the scales (step 9).

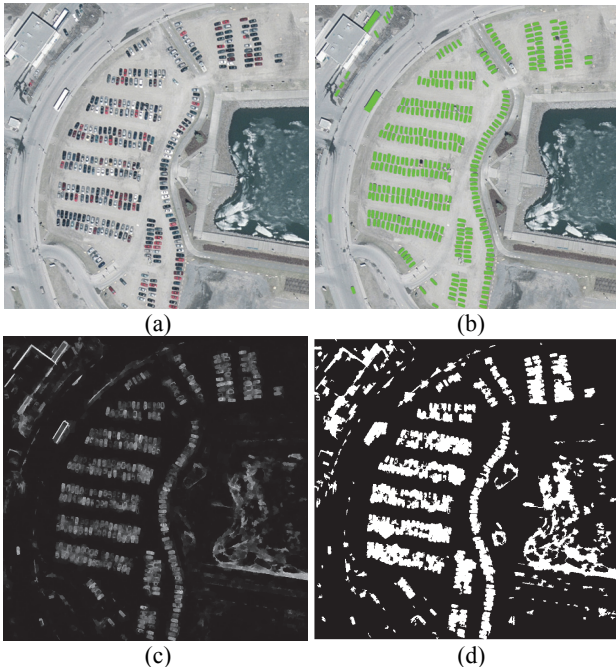


Fig. 3 Example resulting from the proposed method illustrated in Fig.2, (a) aerial imagery of an urban scenario; (b) its associated ground truth; (c) its saliency map and (d) the salient regions detected with multi-scale SLIC.

To describe the variation in size and shape of a superpixel, associated to the presence of the object of interest, the Hausdorff distance is defined in the superpixel as a *string-to-string* presentation. Considering the inner pixels in a superpixel and the pixels on its boundary as two strings of points  $A=a_1a_2\dots a_n$  and  $B=b_1b_2\dots b_m$ , respectively, as shown in Fig.4a.

The Hausdorff distance [19] between the two strings of points  $A$  and  $B$ , is defined as  $H(A,B) = \max\{h(A,B), h(B,A)\}$ . For each point  $a$  in  $A$ , a minimum distance to all the points in  $B$  denoted as  $dB(a)$  exists, as illustrated in the Fig.4b, with

$$dB(a) = \min_{b \in B} (d(a,b)), \quad a \in A \quad (1)$$

Among all the points in  $A$ , at least one point exist for which  $dB$  is maximum, as illustrated with the red arrow in Fig.4c. Thus, the direct Hausdorff distance from  $A$  to  $B$ ,  $h(A,B)$ , is defined as

$$h(A,B) = \max_{a \in A} \min_{b \in B} (d(a,b)) \quad (2)$$

As the direct Hausdorff distance from the boundary point set  $B$  to the inner point set  $A$ ,  $h(B,A)$ , never exceeds one pixel when  $B$  is a closed curve. The  $H(A,B)$  is always given by the direct Hausdorff distance  $h(A,B)$  from the inner point set  $A$  to the boundary point set  $B$ . In fact  $H(A,B)$  represents the radius of a circle inscribed in the superpixel that covers a maximum area, as illustrated by the green disc in Fig.4c. By this definition, the Hausdorff distance is not sensitive to spikes and wiggles that may appear in the superpixel boundary.

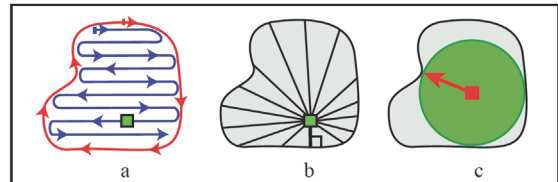


Fig. 4 (a) Strings of points of inner pixels (blue) and boundary pixels (red); (b) minimum distance from an inner pixel to the boundary; (c) Hausdorff distance  $H$  represented by the red arrow as the radius of the green disc inscribed in the superpixel.

A significant advantage for dealing with large scale aerial image is the estimating of the  $H(A,B)$  requires a low computational cost. In fact, the distance transform needs to be applied only once for all the superpixels in the image. Hence, both the SLIC and DT are computational efficient. The overall process described in Fig.2, with 7 scales of the multi-scale SLIC, took 4.36s for a color image of 500x500 pixels size with Matlab<sup>®</sup> on a Duo Core 2 PC at 2.5 GHz. Furthermore, the proposed method is also robust, as even in the low contrast images, since the SLIC can still delimit superpixels with closed boundaries and does not need a threshold, comparing favourably to most edge detectors.

The Hausdorff distance normalized by the average scale of the superpixels,  $f = H(A,B)/(S_i/2)$ , is independent of the scale  $S_i$  and the orientation of the superpixel. The value of  $f$  close to unity means that the superpixel has average size and regular shape. On the contrary, a small value of  $f$  corresponds to the superpixel of a size smaller than the average or of a deformed shape. Furthermore, the value of  $f$  is, by definition, independent of the extreme points in the superpixel boundary. Thus, the normalized Hausdorff distance can be considered as a measure of compactness for the superpixels which also reflects size and shape changes.

From the normalized Hausdorff distance  $f$  we define a shape factor,  $ShF$ , as a posterior probability function, using the monotone sigmoid, as

$$ShF = P(f) = \frac{1}{1 + \exp(\alpha f + \beta)}$$

with  $f = 2H(A, B) / S_i$

which decreases with the increase of  $f$ . The sigmoid parameters  $\alpha=12$  and  $\beta=-6$  have been optimized for the overall performance in the detection of salient map. In fact, the slope of the sigmoid is  $\alpha/4$ . The abscissa of the inflection point is  $-\beta/\alpha$  and  $P(1)$  tends to 1,  $P(0)$  tends to 0 and  $P(0.5) = 0.5$ . The sigmoid transformation permits a more pronounced distinction between salient or no salient classes for a given pixel, especially in the range of the value  $f$  close to 0.5.

The  $ShF$  value computed for a superpixel is propagated to all the inner pixels in this superpixel in order present the probability for a pixel to belong to a salient region. In this way, the saliency map remains the same resolution than the input image at each scale. The saliency maps at the scale  $S_i$  is denoted by  $Sm_{ap\_i}$ , where  $i = 1 \dots p$ .

## 2.2 Salient regions

The salient regions are extracted by applying a pixel-based probabilistic decision rule to the saliency map at each scale  $Sm_{ap\_i}$  to form a binary salient region map. Then, the final salient regions  $\mathfrak{R}$  are obtained by a logic OR of the binary salient maps at all the scales. A pixel  $t$  is set to belong to  $\mathfrak{R}$  in the final binary map, if once it is set to 1 in a binary salient map at any scale. Formally, the pixel-based probabilistic decision rule is applied to every pixel  $t$  in the image  $I$  as the following:

$$\forall t \in I, t \in \mathfrak{R} \Leftrightarrow \exists e \in S / P(X_t=1 | \varepsilon_t=e, f) \geq 0.5 \quad (4)$$

where  $\{S\}$  is ensemble of the scales with  $\{S_1, S_2, \dots, S_p\}$ ,  $X_t$  is a binary decision with  $X_t=1$  if  $t$  belongs to  $\mathfrak{R}$  and  $X_t=0$ , otherwise. The “hidden” local scale random variable  $\varepsilon_t$  [20] depicts the fact that information from all the scales that contribute to the final salient region map is hidden and imperceptible in the overall process.

The  $ShF$  value defined in Eq.3 as the sigmoid posterior probability can also be expressed as the conditional probability  $P(X_t=1 | \varepsilon_t, f)$ . In this case, Eq.4 is interpreted as a decision rule based on estimation of  $ShF$  through all the scales.

The multi-scale superpixel framework guarantees that no isolated pixels will appear in the composition of  $\mathfrak{R}$  since the conditional probability  $P(X_t | \varepsilon_t, f)$  is propagated to all inner pixels in a given superpixel. Therefore, the individual salient region map at every scale  $S_i$  consists of an aggregate of superpixels with the associated  $ShF$  value exceeds or is equal

to 0.5. The spatial coherence induced by the superpixel use prevents the apparition of isolated pixels in the final salient region  $\mathfrak{R}$ . The decision rule introduced in Eq.4 operates without any consideration about the rest of pixels in the image. The process to extract salient regions does not depend on the content of the entire image but only on a local decision taken for a given pixel  $t$  in regard of information content coming from possible scales  $S_i$ . In practice, the probability is normalized and scaled in the range of  $[0, 255]$  for illustration in Fig.3c and comparison in Fig.6.

## 3 Experiments

Experiments were performed on a collection of large-scale high-resolution aerial images taken in nadir view. The image size is 1500x1500 pixels side with a resolution of 11.2 cm/pixel. Such images may correspond to images that are usually taken by Wide Field of View (*WFOV*) cameras deployed on *UAVs*. Their high-resolution and the large-scale make possible the identification of detected targets while preserving situation awareness of the global scene. Two of these images are shown in Fig.3a and Fig.5.

Define “*what is salient in an image*” is mostly subjective. In the Fig.3a, we manually select all the objects of interest, e.g. vehicles in the image as the ground truth, as shown in Fig.3b. Figure3c and Fig.3d shows the salient map and the salient regions resulting from the use of our proposed method.

A quantitative comparison is performed between the ground truth and the salient regions resulting from our proposed method. In this goal, three evaluation criteria [7] are employed. The Precision and Recall defined as ratios:

$$Precision = \frac{\sum_x g_x dx}{\sum_x dx} \quad Recall = \frac{\sum_x g_x dx}{\sum_x g_x}$$

where  $g_x=1$  for ground-truth pixels and  $g_x=0$  otherwise,  $dx=1$  for detected salient pixels and  $dx=0$  otherwise, and the overall salient region detection performance, F-measure defined as the weighted harmonic mean of precision and recall :

$$F - measure = \frac{(1 + \theta) \times Precision \times Recall}{\theta \times Precision + Recall}$$

where  $\theta$  is set to 0.5 as in [3] and [7]. For instance, for the Fig.3, the Recall, the Precision and the F-measure are equal to 87.47%, 41.21% and 50.03%, respectively.

Most of the existing methods for saliency detection are not suitable for aerial imagery surveillance applications. An adequate method should reach some requirements such as, be computational efficient to process large size imagery, ideally works without parameters tuning nor *a priori* knowledge about the objects.

Assuming that the location of salient regions coincides with regions in the image with high density of edges, Rosin [12] proposed a simple non-parametric method for the saliency detection, based on estimation of the spatial density of edges in the image. The gradient image computed from an input

gray-scale image is sliced to a stack of binary images using a set of equally spaced thresholds. Then, the DT is applied to the set of edge maps newly formed, resulting in a set of DT map, which are finally summed up to produced the final salient map. Then, the salient regions are captured by applying an adaptive threshold.

Rosin’s method is based on the edges extracted from local gradients in gray-scale images, while the proposed method is based on the shape and size of the superpixel boundaries, extracted by combining the region and edge information in color images. As a consequence, the edge-based method [12] may produce salient regions in the background, such as the sea area, while the proposed method will not, as shown in Fig.5. This is a significant advantage, especially when dealing with aerial images for which large and inconsistent background may exist. In this case “nothing” could be also considered as a valuable response to the question: “what is salient in an image?”.

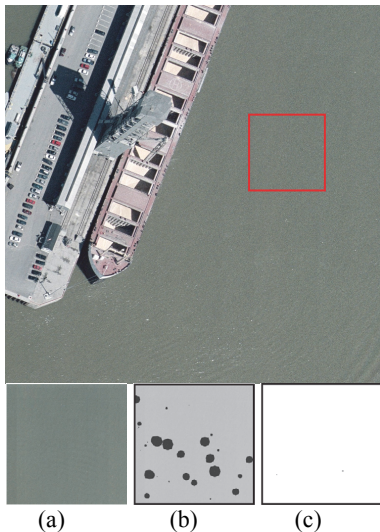


Fig.5 (a) A patch in sea area; (b) salient regions detected by Rosin method [12]; (c) by multi-scale SLIC segmentation, with the color removed.

Another advantage is that in Rosin’s method edge information is propagated to whole image by the DT without spatial restriction. As a result, edges of the object of interest are lost in Fig.6c even if the method is based on edges. In the proposed method, the decision rule in Eq.4 is applied to every pixel  $t$  without consideration on its neighborhood. In addition, the superpixel boundaries act as a *cut-off* for limiting the spatial extent of the DT that allows us to describe more precisely the object boundaries that compose the salient regions. The decision is entirely local, so that the sharp structure details and a high contrast to the background are conserved as shown in Fig.6d.

The Spectral Residual (SR) [13] is the second method used for comparison. As to Rosin’s method, the SR method is 1) simple to implement, free of any parameters; and 2) the saliency map is computed independently of any feature.

In the SR method, it is considered that information content in an image can be decomposed in two parts: *redundant information* related to a prior and common knowledge and *innovation information* which appear as a statistical singularity in the input image.

Hou and Zhang [13] propose to suppress redundant information to stand out innovative part of the image which is associated to salient region. The analysis is done in the spectral domain. The saliency map is created in two steps. Firstly, the redundant spectral information is subtracted from the image to conserve the innovative spectral part of the image, noticed as the *spectral residual*. Secondly, the saliency map is created by inverting the Fourier Transform of the spectral residual. By applying a simple threshold empirically chosen equal to 3 times the average intensity on the saliency map, the salient regions are captured.



Fig.6 Saliency map computed for the three methods; (a) the input color image (top left); (b) Spectral Residual (top right); (c) Rosin’s method (bottom left) and (d) multi-scale SLIC (bottom right).

Observing the result in Fig.6, we noticed that the Rosin’s method produces quite good results even if halos of brightness spread beyond the limits structure dimensions. For instance, vehicles in the parking lot are fused to continuous regions, instead of appearing disjoint as in the input image. For SR, the majority of salient structure locations are stand out but it remains difficult to discern shape of structures.



Fig. 7 (a) Multi-scale SLIC is applied on entire image (top left) and (b) on a sub-image (top right). Their salient regions are respectively given by (c) and (d), the bottom-left and the bottom right images. Cropping the image has no effect since a pixel-based decision rule is used, Eq.4.

The pixel based probabilistic decision rule introduced in Eq.4 permits to classify every pixel  $t$  of the image with no consideration about neighbor pixels and none local or global threshold have to be set. In this sense, the decision is done locally, as mentioned before. The inherent advantage is that the input image can be divided in sub-images as illustrated in Fig.7, treated in parallel using the multi-scale SLIC and recombine to create salient regions, without altering the quality result.

In this way, it is possible to create a mosaic of salient regions from a large-scale area under surveillance. The capacity of awareness of the entire scene is thus preserved while the inherent constraints in the use of very large images coming from *WFOV* cameras are overcome.

The methods that produce salient regions by applying a threshold on the saliency map do not present this advantage. In fact, cropping or subdividing the input image leads to different salient regions since the threshold value is estimated to the entire image and thus may change according to the information content of the image.

A quantitative evaluation of the three methods was performed on the entire database of 18 large-scale high-resolution images similar with that presented in Fig.3a and Fig.5. The Table 1 gives the evaluation criteria averaged over the 18 images.

Table.1: evaluation of criteria averaged over 18 large-scale high-resolution aerial imagery.

	Precision	Recall	F-measure
SpectralResidual[13]	3.51%	79.82%	5.13%
Rosin[12]	8.86%	<b>93.96%</b>	12.56%
Multi-scale SLIC	<b>22.20%</b>	80.84%	<b>27.57%</b>

The F-measure represents the overall performance measurement of the algorithms. As noticed in [7], for attention detection, recall rate is not as important as precision. For instance, a recall rate of *100%* can be easily achieved by simply select the whole image as salient region.

The result of the experiment illustrated in Table.1 shows that the multi-scale SLIC has a F-measure at least 2 times higher than the compared methods while its recall measure belongs to the same order of magnitude.

The precision measure reflects the number false alarms and this measure is directly related to the number of pixels classified as salient while they are not.

In the context of aero-surveillance, reducing the false alarms by a factor 2.5, as with the proposed method, is significant. In practice, reducing the number of false alarm decreases the number of regions that would be retained for a further analysis, such identification of objects. This gain directly affects the computational cost and permits to reallocated computational resources to address other tasks in the chain of image processing.

## 4 Conclusions and further work

We have proposed a new method for detecting saliency and extracting salient regions in the aerial images, using the irregularity in the size and the compactness of the superpixels in image segmentation as a saliency measure. The implementation with the multi-scale SLIC is simple, fast, robust and requires no a priori knowledge about the objects.

To evaluate the proposed method, we have proceed to quantitative comparison with two other algorithms which have in common properties of the multi-scale SLIC. Using an annotated database composed of 18 large-scale high-resolution aerial imagery, we evaluate Precision, Recall and F-measure for each method. The result of the experiment shows that the multi-scale SLIC has a F-measure at least 2 times higher than the other methods while its recall measure remains within the same order of magnitude than compared methods.

In this paper, we have used a collection of static images. However, with the GPU implementation of SLIC recently proposed by [21]; it may be possible to deploy the multi-scale SLIC for real-time application.

## Acknowledgement

This work has been supported by grants from Natural Sciences and Engineering Research Council of Canada (NSERC); Fonds Quebecois de Recherche sur la Nature et les Technologies (FQRNT); and Defence Research and Development Canada– Valcartier.

## References

- [1] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4:219-227, 1985.
- [2] Y-H. Kuan, C-M. Kuo and N-C. Yang. Color-based image salient region segmentation using novel region merging strategy. *IEEE Trans. on Multimedia*, 10(5):831-845, 2008.
- [3] R. Valenti, N. Sebe and T. Gevers. Image saliency by isocentric curvedness and color. In *ICCV*, 2185-2192, 2009.
- [4] M. Z. Aziz and B. Mertsching. Fast and robust generation of feature maps for region-based visual attention, *IEEE Trans. on Image Processing*, 17(5):633-644, 2008.
- [5] V. Gopalakrishnan, Y. Hu and D. Rajan. Salient region detection by modeling distributions of color and orientation. *IEEE Trans. on Multimedia*, 11(5):892-905, 2009.
- [6] Z. Li and L. Itti. Saliency and gist features for target detection in satellite images. *IEEE Trans. on Image Processing*, 20(7):2017-2029, 2011.
- [7] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang and H-Y. Shum. Learning to Detect a Salient Object. *IEEE Trans. on PAMI*, 33(2):353-367, 2011.
- [8] W. Zhang, Q.M.J Wu, G. Wang and H. Yin. An adaptive computational model for salient object detection. *IEEE Trans. on Multimedia*, 12(4):300-316, 2010.
- [9] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on PAMI*, 20(11):1254–1259, 1998.
- [10] T. Kadir and M. Brady. Saliency, scale and image description. *IJCV*, 45(2):83-105, 2001.

- [11] P. Khuwuthyakorn, A. Robles-Kelly and J. Zhou. Object of interest detection by saliency learning. In ECCV, 636-649, 2010.
- [12] P. L. Rosin. A simple method for detecting salient regions. *Pattern Recognition*, 42(11):2363-2371, 2009.
- [13] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In CVPR, 1-8, 2007.
- [14] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk. SLIC Superpixels, EPFL Technical Report 149300, June 2010.
- [15] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. on PAMI*, 22(8):888-905, 2000.
- [16] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson and K. Siddiqi. Turbopixels: fast superpixels using geometric flows. *IEEE Trans. on PAMI*, 31(12):2290-2297, 2009.
- [17] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. on PAMI*, 24(5):603-619, 2002.
- [18] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In ECCV, 705-718, 2008.
- [19] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Trans. on PAMI*, 15(9):850-863, 1993.
- [20] B. Chalmond, B. Francesconi and S. Herbin. Using hidden scale for salient object detection. *IEEE Trans. on Image Processing*, 15(9):2644-2656, 2006.
- [21] C.Y. Ren and I. Reid. gSLIC: a real-time implementation of SLIC superpixel segmentation. University of Oxford, Department of Engineering, Technical Report, 2011.