

IRREGULARITY-BASED IMAGE REGIONS SALIENCY IDENTIFICATION AND EVALUATION

Mohammad Al-Azawi^(1,2), Yingjie Yang⁽¹⁾, Howell Istance⁽¹⁾

¹Center for Computational Intelligence, De Montfort University, Leicester, UK

²Department of Computer Science, Oman College of Management and Technology, Oman
(m.a.al-azawi@ieee.org, yyang@dmu.ac.uk, hoi@dmu.ac.uk)

Abstract— Saliency or Salient regions extraction from images is still a challenging field since it needs some understanding for the image and the nature of the image. The technique that is suitable in some application is not necessarily useful in other application, thus, saliency enhancement is application oriented. In this paper, a new technique of extracting the salient regions from an image is proposed which utilizes the local features of the surrounding region of the pixels. The level of saliency is then decided based on the global comparison of the saliency-enhanced image. To make the process fully automatic a new Fuzzy-Based thresholding technique has been proposed also. The paper contains a survey of the state-of-the-art methods of saliency evaluation and a new saliency evaluation technique was proposed.

Keywords: Irregularity, Saliency, Image Processing, Thresholding, Fuzzy.

1 Introduction

Salient points, objects, or regions extraction is still a challenging topic of research and its definition is still very general. Different literature have defined the saliency in different ways. In general and based on human vision system HVS, the salient object is the object that may capture the attention or attract the

HVS. HVS uses two stages to identify the objects, pre-attentive and attentive stages (Kadir & Brady, 2001). In pre-attentive stage, the local regions of image that present spatial discontinuity are detected. In the attentive stage, relationships between these regions are found, and grouped together.

Many efforts have been done to develop new saliency extraction techniques based on different image techniques such as Fourier transform, wavelet, corners, etc. Different techniques were proposed based on the definition of saliency according to the authors' points of view.

The rest of the paper will be organized as follows: previous work is presented in section 2. In Section 3, we will explain the proposed saliency extraction technique based on the irregularity of the regions in an image. In section 4, a new automatic thresholding technique will be proposed. Points clustering and region merging will be discussed in section 5. Section 6 will review the saliency evaluation technique and introduce a new method that is suitable to the application at hand. In section 7 we will discuss the experimental results and the optimum window size selection. Finally, conclusions will be derived in section 8.

2 Previous Work

Wavelet was used by many authors such as in (Loupias, et al., 2000), (Tian, et al., 2001), (Song, et al., 2006), (Lin & Yang, 2007), and (Arivazhagan & Shebiah, 2009) in which they have applied the principles of wavelet transform to extract the salient points. Loupias *et al.* have used orthogonal Haar wavelet in extracting the salient points (Loupias, et al., 2000). Song et al. (Song, et al., 2006) have proposed the use of wavelet to identify the salient points in colour image using the three bands RGB. The use of wavelet may give good results in a non-homogeneous image, like the cameraman example that was used in (Loupias, et al., 2000), but it will not give that good results when applied on image with high texture.

Location of the object was considered as a measure of object saliency by Kim, et al. in 2003 (Kim, et al., 2003), they have used the central position and colour contrast as the features that give importance to the object, which is not always true, since objects are not necessarily placed in the centre of the image.

Background suppression, in which the background unimportant details of the image are removed and the salient details are kept only was used by Davis & Sharma (Davis & Sharma, 2004) and Zhang & Goldman (Zhang & Goldman, 2006). This method incorporates a pre-knowledge of the objects to be searched for in the image.

Geometric features, such as corners, were used to identify the points' saliency. Corners were considered as a measure of saliency firstly by Schmid and Mohr in 1997 as part of their effort to identify interest points

locally (Schmid & Mohr, 1997) (Kapsalas, et al., 2008). In their paper published in 2000, Loupias et al. (Loupias, et al., 2000) have criticized the use of corner-based techniques and showed the limitations of using this method. The limitations they presented are; first, important visual features are not necessarily to be corners, and second, corners may gather in small regions, like in texture images.

Feature maps that can be defined as the maps of some features extracted from the image was used to show the regions saliency also. The saliency of a point or a region is measured based on the combination of these maps. Low-level features such as colour, intensity, texture, orientation, etc. are commonly used. Early work in this field was done by Koch and Ullman (Koch & Ullman, 1985) and Itti et al (Itti, et al., 1998). Itti *et al.* have developed a model to extract the regions of attention; in their model, they used image Gaussian pyramid. Intensity, colour, and orientation features were used to distinguish the salient regions. 42 feature maps were generated from the features that were calculated by a set of centre surround operations. The saliency map is extracted by combining the above maps. The inhibition of return was used to prohibit the algorithm from considering the same salient object more than once (Ivanoff & Klein, 2008). The main drawback of this technique is the large number of feature maps; in addition to the use of image pyramid, these two drawbacks may affect the speed of the algorithm. In addition, it extracts the interest regions sequentially, which means that one region at each iteration using winners-takes-all paradigm, which increases the computation time.

Frequency domain was used to extract the saliency of the objects by many literature such as (Bruce, et al., 2007), (Hou & Zhang, 2007), (Li, et al., 2007), (Achanta, et al., 2009), (Zhou, et al., 2010), and (Fang, et al., 2012). The idea behind using the frequency domain is that, they considered that salient points usually of higher change in frequency domain both in magnitude and in orientation. The main limitation of such techniques is that they are considering regions with high frequency as the salient regions. Region of high frequency are regions with high local changes such as edges, and edges not necessarily are salient regions.

From the above discussion, and in order to develop a robust salient regions extraction algorithm one needs to identify the problems incorporated with the available techniques. Several issues have been identified from analysing the above techniques, which shall be considered in the developed technique. The first issue is that in most of the available techniques, they need to have pre-knowledge about the nature of the image like in (Zhang & Goldman, 2006). Specifying some parameters manually, like specifying the value of the threshold, is another issue like in (Davis & Sharma, 2004). In techniques like frequency, corners, and wavelet based techniques, the problem of falsely detected salient points FDSP due to the noise and the nature of the image and the extractor, should be taken care of.

In this paper, we shall introduce a robust and general saliency extraction method that is fully automated and does not need any intervention from the user. In

addition, it does need any pre-knowledge about the image.

3 Irregularity as a Measure of Saliency

In this Section, we shall discuss the proposed technique. First, we shall define the saliency as any irregular region in the image with nature differs from the nature of the image as a whole; for example, a region with texture in a uniform region, or vice versa, thus we shall define irregularity as the main measure of saliency.

Let $\mathcal{J}(x, y)$ be the image intensity at position (x, y) , where $x = 1, 2, \dots, W, y = 1, 2, \dots, H$, with W and H are the image width and height respectively. The image is represented as a set of pixels and is defined as:

$$\begin{aligned} \mathbb{I} = \{ & \mathbf{p}_{xy} \mid \mathbf{x}, \mathbf{y} \in \mathbb{N} \wedge \mathbf{x} = \mathbf{1}, \mathbf{2}, \dots, \mathbf{W}, \\ & \wedge \mathbf{y} = \mathbf{1}, \mathbf{2}, \dots, \mathbf{H} \} \\ & \mathbf{p}_{xy} = \mathcal{J}(\mathbf{x}, \mathbf{y}) \end{aligned} \quad (1)$$

Where p_{xy} is the pixel's value at location (x, y) and \mathbb{N} is the set of natural numbers. Furthermore, we shall define the set $\mathbb{P}_{\mathbb{I}} = \{s \mid s \subseteq \mathbb{I}\}$ as the power set of \mathbb{I} . According to the application at hands, the main constrain is that; all the elements of a given subset should belong to a connected region which means they should be in a given neighbouring area.

The sub-image, which we shall refer to as a region, \mathbb{I}_{ij} is a subset of the image with elements that belong to the same region. The size of this sub-image is $w \times h$. Considering that the regions are disjoint, they can be defined as:

$$\begin{aligned} \mathbb{I}_{ij} = \{ & \mathbf{p}_{kl} \mid \mathbf{p}_{kl} \in \mathbb{I} \wedge \mathbf{j} \times \mathbf{w} \leq \mathbf{k} \\ & \leq (\mathbf{j} + \mathbf{1})\mathbf{w} \wedge \\ & \mathbf{i} \times \mathbf{h} \leq \mathbf{l} \leq (\mathbf{i} + \mathbf{1})\mathbf{h} \} \end{aligned} \quad (2)$$

$$\mathbb{R} = \left\{ r_{ij} \mid \mathbf{0} \leq j \leq \frac{W}{w} \wedge \mathbf{0} \leq i \leq \frac{H}{h} \right\} \quad (3)$$

$$\mathbb{R} \subseteq \mathbb{P}_{\mathbb{I}}$$

Assuming that there is no overlapping between the adjacent regions, the total number of regions is equal to $\frac{W}{w} \times \frac{H}{h}$. In most cases there should be overlapping between the moving windows, if we denote the overlapping value by δ , then the total numbers of windows will be $\frac{W}{w-\delta} \times \frac{H}{h-\delta}$.

In addition, we shall define the description space, in which each region r_{ij} will go through features extractors to convert them to a set of measures that can be used in describing the region. We shall define mapping ϕ from the regions space to the description space as given below:

$$\Phi: \mathbb{R} \rightarrow \mathbb{D} \quad (4)$$

Where \mathbb{D} is the set of descriptors of the regions in \mathbb{R} and it is given by:

$$\mathbb{D}_{ij} = \Phi(r_{ij})$$

$$\mathbb{D} = \left\{ \mathbb{D}_{ij} \mid \mathbf{0} \leq j \leq \frac{W}{w} \wedge \mathbf{0} \leq i \leq \frac{H}{h} \right\} \quad (5)$$

\mathbb{D}_{ij} might be a single value or a set of values.

The selection of the mapping ϕ is important and it should give high values for regions with high irregularity and low for other regions. Since pixel's intensity can be treated as a random variable, then probability and statistical measures can be used to measure the features of a region. Two important statistical measures can be used to identify the regularity and irregularity of an image, which are the expectation and the variation. The expectation value is very close to the pixels values in the regular regions and the variation such as standard deviation value is

small. In contrast the and vice versa in the case of irregular regions, then the measure of variation (irregularity) can be derived from these two measures. The block diagram of the proposed model is given in Figure 1, in this diagram; the image should go first through a saliency enhancement process, irregularity measuring in our case. The resultant saliently enhanced image is then goes through thresholding process to remove points with low irregularity. The extracted salient points is then clustered to form groups that can be merged together to form the salient regions.

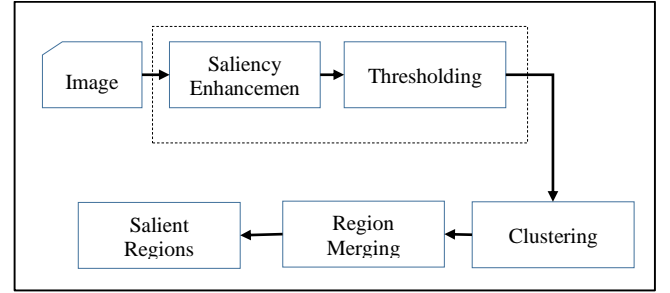


Figure 1 the block diagram for the proposed model. Figure 2 shows the original and intensity images. It also shows the mean image and the difference between the intensity image and the mean image. The space representation shows that the values corresponding to the salient objects is higher than other values.



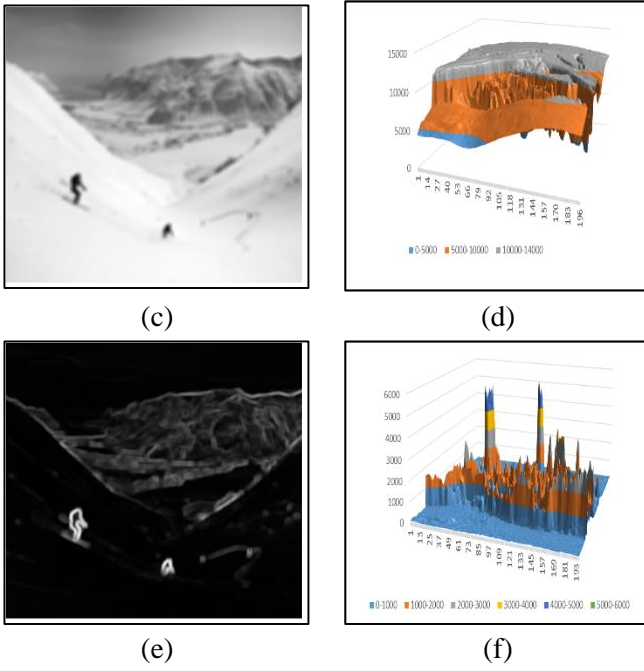


Figure 2 Testing image and space representation, (a) original image, (b) intensity image, (c) mean image, (d) space representation of the mean image, (e) difference between intensity and mean images, (f) space representation of the difference image.

The expectation value (mean) μ_{ij} and the variation value (standard deviation) σ_{ij} are calculated using pixel surrounding region, thus the irregularity can be extracted as a function of the expectation value and the variation value:

$$v_{ij} = \phi(p_{xy}, \mu_{ij}, \sigma_{ij})$$

$$v_{ij} = \sum_{y=1}^h \sum_{x=1}^w |p_{xy} - \mu_{ij}| \sigma_{ij} \quad (6)$$

Where v_{ij} is the measure of irregularity in region \mathbb{R}_{ij} . In this case \mathbb{d}_{ij} shall be represented in one value which is v_{ij} . In order to make the values acceptable and comparable, we shall normalize the value of \mathbb{d}_{ij} by dividing them by the maximum value, i.e. $\mathbb{d}_{ij}^* = \frac{\mathbb{d}_{ij}}{\text{Max}(\mathbb{D})}$.

Figure 3 shows the results of applying the above described method on an image, it is clear from the

figure that the values corresponding to the salient objects have become even higher and more distinguishable.

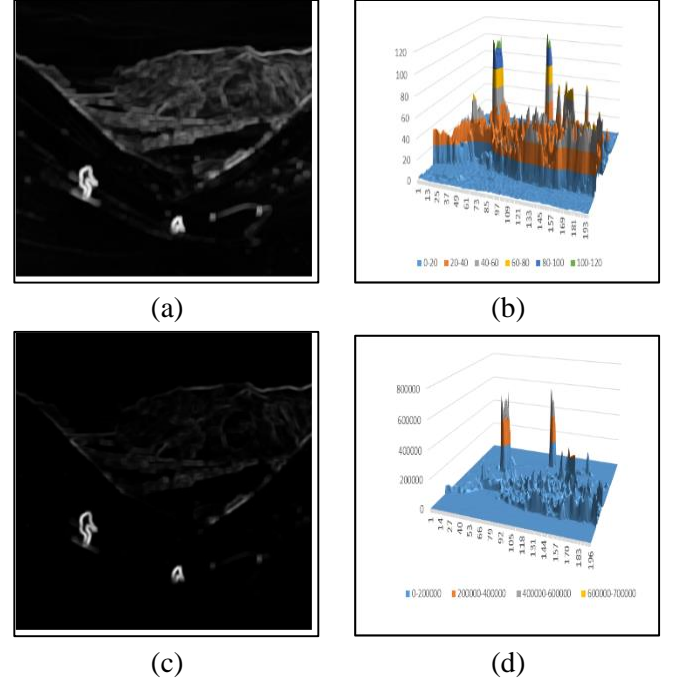


Figure 3 The results of applying the irregularity measures on an image, (a) standard deviation image, (b) space representation of the standard deviation image, (c) the variation values image, (d) the space representation for the variation image.

4 Unimportant Details Suppression and Thresholding

Extracting the salient object means removing unimportant details. Based on the local descriptor the regions are divided into different classes, $\mathbb{R} = \mathbb{R}_I \cup \mathbb{R}_U$ with \mathbb{R}_I represents the set of important region and \mathbb{R}_U is the set of unimportant regions.

There should be some way to separate these two regions; thresholding can be used for this purpose, in which a suitable threshold value, which is extracted from the available data, is select to isolate important from unimportant regions. The value should be selected carefully since it plays important role in discriminating important from unimportant regions. Small threshold may lead to consider some of the

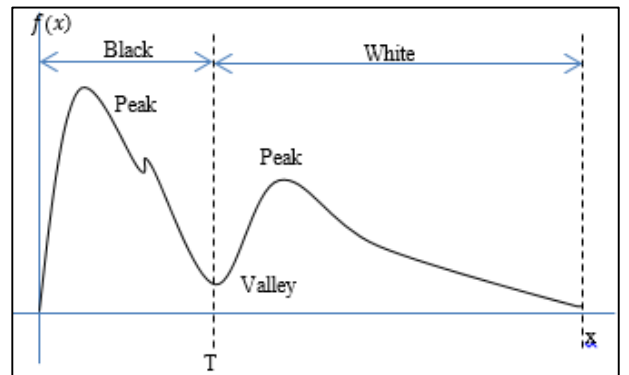
unimportant regions as important and vice versa. Regions with descriptor values larger than or equal the threshold value T are considered as important regions while others are considered as unimportant as given in the following equation:

$$\begin{aligned} \mathbb{R}_I &= \{r_{ij} \mid d_{ij}^* \geq T\} \\ \mathbb{R}_U &= \{r_{ij} \mid d_{ij}^* < T\} \end{aligned} \quad (7)$$

The value of T can be derived from the available data in \mathbb{D} . Many techniques can be used to extract the value of T . One possible technique is by using histogram thresholding. Histogram gives good information about the distribution of the intensity in the image. Since most of the algorithms assume the input image as a dark object in a light background, or vice versa, thus the histogram is expected to have two peaks and one valley. The grey level corresponding to the valley, which represents the minimum point between the two peaks, is considered as the threshold value. This thresholding technique is known as bimodal thresholding. Another solution is by considering the average of the pixels values as the threshold value. These techniques consider that the distribution of the pixels intensity is uniform and the edges are well-defined, which is not the real case.

Since the term salient is not clearly defined and fuzzy, then Fuzzy logic can give reasonable results. In this paper, we shall introduce a new thresholding technique utilizing the principle of membership function in the fuzzy logic in bimodal histogram thresholding technique (Al-Azawi & Ibrahim, 2012) (Al-Azawi, 2013). The method is a modification to the bimodal thresholding technique. Figure 4 (a) shows the traditional bimodal thresholding technique, in

which the value of T was selected as the gray level value corresponding to the valley of the histogram. The main problem with the method given in (a) is that it assumes that the border between the object and its background is well defined and crisp, which is not the real case in real images. (b) in the same figure shows the use of Fuzzy membership function in extracting the value of T . The histogram is approximated by two membership functions; $\mu_b(x)$ which measure how much the pixel is black, and $\mu_w(x)$ which measure the white membership value. Thus, the luminance space will be divided into two subspaces or subsets B and W. According to this figure, the histogram is divided into 4 region rather than 2 regions as in crisp thresholding. One of the regions is the black region and the second one is the white region. The third and fourth regions, which are the overlapping between the black and white regions, these regions have been generated due to the uncertainty in image luminance value and they may represent the fading or the shadow. The value of T can be found in different ways, such as, the intersection of the two lines of the membership functions, by calculating the entropy of the regions, and using AI techniques like Genetic Algorithms and Neural Networks.



(a)

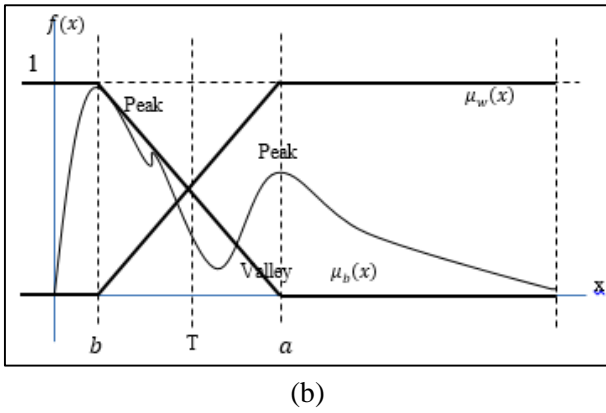


Figure 4. Bimodal thresholding technique, (a) normal technique, (b) the application of fuzzy membership function on bimodal thresholding technique.

From Figure 4, it is clear that the histogram has been divided into four regions; each region has different saliency based on the intensity of the region. We shall define a Fuzzy Linguistic set $FL = \{NS, LS, S, VS\}$. The properties of each Fuzzy Linguistic Variable (FLV) is given as follows:

Table 1 Fuzzy linguistic variables meanings and ranges

FLV	Meaning	Range
NS	Not Salient Region	$0 \leq x < b$
LS	Less Salient Region	$b \leq x < T$
S	Salient Region	$T \leq x < a$
VS	Very Salient Region	$a \leq x \leq 100$

Based on the table above, it is clear that user can select the level of saliency. The best selection might be the S and VS together as it will be shown in the experimental results section.

5 Salient Points Clustering and Regions Merging

Clustering the extracted salient points to form a salient region is important part of the proposed algorithm. As per Estivill-Castro (Estivill-Castro, 2002), the notion of cluster cannot be precisely defined; that is why there are many different clustering algorithms. Most

of the algorithms are application-oriented algorithms, i.e. they differ from each other's based on the application at hand. The algorithm that is suitable for a particular problem may not work effectively in other problem, thus, the selection of appropriate algorithm for a particular problem is a crucial issue and it depends on the eye of the beholder (Estivill-Castro, 2002). For this reason, dozens if not above a hundred of algorithms were proposed (Anon., 2012). Many reviews have been published about this topic such as (Milligan & Cooper, 1987), (Tryfos, 1998) (Jain, et al., Sept. 1999), (Xu & Wunsch, May, 2005), (Tan, et al., 2006), (Cook & Swayne, 2007), (Burns & Burns, 2008), and (Rai & Singh, 2010). In ref (Jain, et al., Sept. 1999) most of the clustering algorithms have been discussed including the algorithms that utilize artificial intelligence such as neural nets, uncertainty, and evolutionary techniques.

Based on the nature of our application we shall adopt the Blobs-Based Clustering (BBC) technique (Al-Azawi, et al., 2014). In this clustering technique, each point is replaced by the small surrounding region (blob). If any adjacent regions are intersecting, they are merged to form a single region. The process is repeated until no more merging is needed. The size of the blobs is increased with every iteration and the merging process is performed until the increase in the blob size will not affect the merging process anymore.

The following figure shows the BBC process:



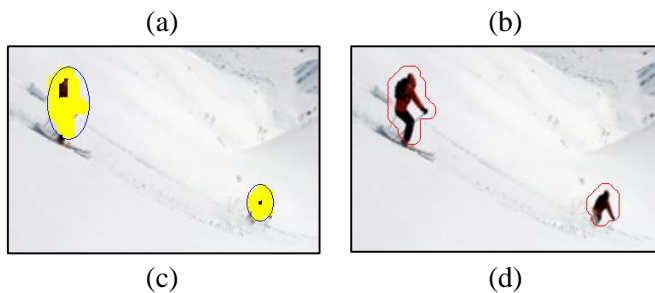


Figure 5 Points clustering and regions merging using BBC, (a) points to be clustered, (b) obtained regions after first iterations, (c) obtained regions after the iterations stopped (3 iterations), (d) the obtained regions.

Figure 5 (a) shows the points obtained from the salient points extraction technique. (b) in the same figure shows the result of the merging process after one iteration, in which one can see that only some points have been merged to form regions. In the next iteration, the size of the blob is increased and the new regions are merged. The process is repeated until the increase in the size of the blob will not affect the number of regions for few successive iterations. In the example given in Figure 5, the iterations stopped after three iterations as shown in (c). By taking the borders of the obtained regions one can identify the salient region and hence the salient object.

6 Saliency Evaluation

For any developed algorithm, there should be a measure to evaluate how good this algorithm is, and how suitable is it for the application at hand. First, we need to setup some rules that will be used to evaluate the saliency extraction algorithms, these rules should satisfy the idea of saliency and attention; also, they should consider the application at hands requirements. Many Surveys that tried to investigate the most state-of-the-art evaluation methods such as (Sebe , et al., 2002), (Toet, 2011), (Gide & Karam, 2012), (Borji, et al., October 7-13, 2012), (Judd, et al., January 13,

2012), (Zhao & Koch, 2011) and others are available but no general method can be used with all algorithms. Some literature judged the results visually from the observers' points of views, which is the easiest method but it depends totally on the feedback from the observers and there is no fixed measure in this method (Hu , et al., 2005), (Einhauser, et al., 2008). Area Under the Curve was used for this purpose in which the saliency map is converted to binary image and then the AUC is calculated and compared to the AUC extracted from the ground truth data (Lin, et al., 2013) (Gide & Karam, 2012), (Zhao & Koch, 2011), (Erdem & Erdem, 2013), and (Kim & Milanfar, 2013). Tatler supported the idea of empirical evaluation of the saliency since they were interested more in mimicking the natural behaviour of the humans' eyes (Tatler, et al., 2011). Precision and Recall, is another method was used by many publications such as (Fang, et al., 2012), (Stöttinger , et al., 2009), (Liu, et al., 2011), (Achanta, et al., 2009), (Liu, et al., 2007) and (Anon., 2013) to evaluate the saliency algorithms. Correlation-based measures also were widely used, in which, the correlation between the saliency map generated by the saliency extraction algorithms and those that generated by using fixation data (Zhao & Koch, 2011), (Gide & Karam, 2012), (Mancas, 2008), (Parkhurst , et al., 2002), (Masciocchi, et al., 2009), (Rosin, 2009) and (Le Meur, et al., May 2006). Some other techniques were proposed, such as, Least Square Index (Zhao & Koch, 2011), Earth Mover's Distance (Zhao & Koch, 2011), (Lin, et al., 2013), (Judd, et al., January 13, 2012), (Pele & Werman, 2008), (Lin, et al., 2010) and (Rubner, et al., 2000), Receiver

Operating Characteristics (Zhao & Koch, 2011), (Erdem & Erdem, 2013), (Zhang, et al., 2008), Normalized Scanpath (Zhao & Koch, 2011) (Peters, et al., 2005), String-Edit Distance (Zhao & Koch, 2011), (Choi, et al., 1995), and Info Contents (Schmid, et al., 2000).

In the method above, the order of the extracted points is important and they work well with eye tracking data since the proposed algorithms were to develop a method that can simulate the fixations of humans. Although there are some other limitations discussed in the references given above and out of the scope of our research. In this paper, we shall develop a method that can compare two images based on the salient regions contained in each image. The method we will introduce is suitable for application at which the order is not of big importance.

To evaluate the obtained results, we shall compare them with the ground truth data obtained both from manual labelling and from eye-tracking data. In general, let $P = \{p_i | p_i = (x_i, y_i) \wedge i = 1, 2, \dots, N\}$ be the extracted points set and $H = \{h_i | h_i = (x_i, y_i) \wedge i = 1, 2, \dots, M\}$ as the ground truth data. It is not possible to compare the points individually since it is not possible to have two points on the same x, y coordinate, thus, we may use the centroid of the data instead, which is the simplest form of comparison. We shall refer to this method as single centroid distance (SCD); the centroid is calculated as follows:

$$\begin{aligned} C_{Px} &= \frac{\sum_{i=1}^{|P|} x_i}{|P|} \\ C_{Py} &= \frac{\sum_{i=1}^{|P|} y_i}{|P|} \\ C_P &= (C_{Px}, C_{Py}) \end{aligned} \quad (8)$$

In the same way, one can find the centroid of the ground truth data $C_H = (C_{Hx}, C_{Hy})$, the geometrical distance between the two centroids can be used as a measure for the distances between the centroids.

$$D(P, H) = \sqrt{(C_{Px} - C_{Hx})^2 + (C_{Py} - C_{Hy})^2} \quad (9)$$

Figure 6 shows the result of comparing two sets of points the first set was extracted using the local irregularity salient extraction method and the second set was obtained using manual labelling.

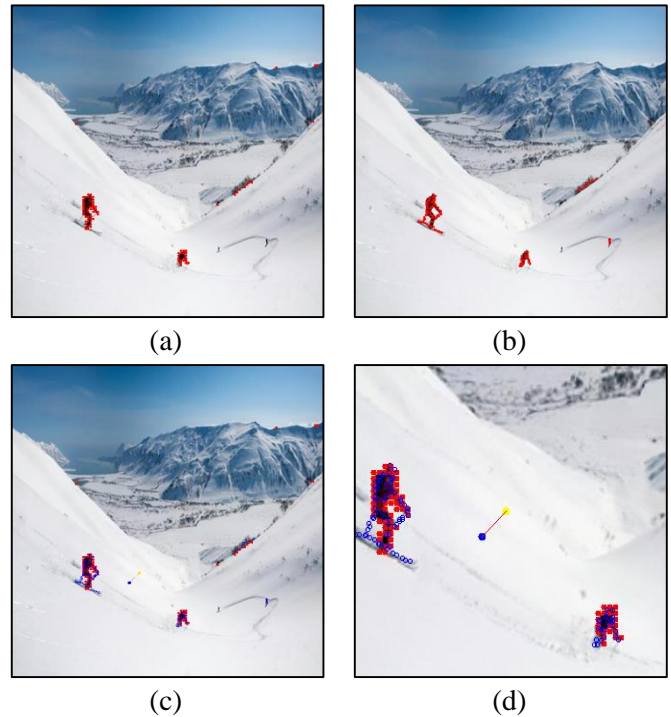


Figure 6 Comparing two points sets, (a) automatically extracted points, (b) Manually extracted points, (c) the distance between centroids, (d) the same distance in (c) scaled up for clarification

In the above example, the distance is very small (30) which means that the maps are close enough to be

considered as similar, in spite of the difference in the point distribution of the two sets.

Another possible comparison to evaluate the salient point extraction technique is by comparing the extracted data with the one obtained from the eye-tracking device. The following figure shows the distance between the extracted data and the eye-tracking data. The distance between the two centroids was (30.1).

From the above discussion, it was obvious that the comparison is not very precise, since it compares all the points at the same time, which is strongly affected by the falsely detected salient points FDSPs and the density of the points.

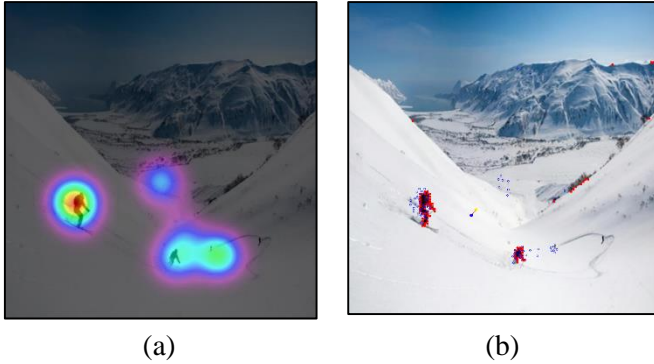


Figure 7 The distance between the centroid of the computationally extracted data and the eye tracking data, (a) fixations map obtained from the eye tracking data, (b) the distance between the centroids.

In this part, we shall consider the distance between the regions after clustering, i.e. after forming the regions of interest from the points of interest. One possible measure is by taking the exclusive OR between the saliency regions, in this technique, the pixels in the salient regions are set to one, while others are set to zero. Pixel-wise exclusive or is applied on the two maps, similar pixels will produce one and different pixels will produce zero. The total number of zeroes

represents the difference between the maps. We shall refer to this method as Exclusive Or Distance (EOD).

$$EOD = \sum_{y=0}^H \sum_{x=0}^W (1 - [h(x,y) \oplus m(x,y)]) \quad (10)$$

Where $h(x,y)$ and $m(x,y)$ are the value of the human map and the machine map respectively.

The following figure shows an example on the application of the EOD. In this figure, (a) and (b) show the maps we want to compare, (c) and (d) the maps after binarisation, and (e) is the result of applying the exclusive or operation on the maps in (c) and (d).

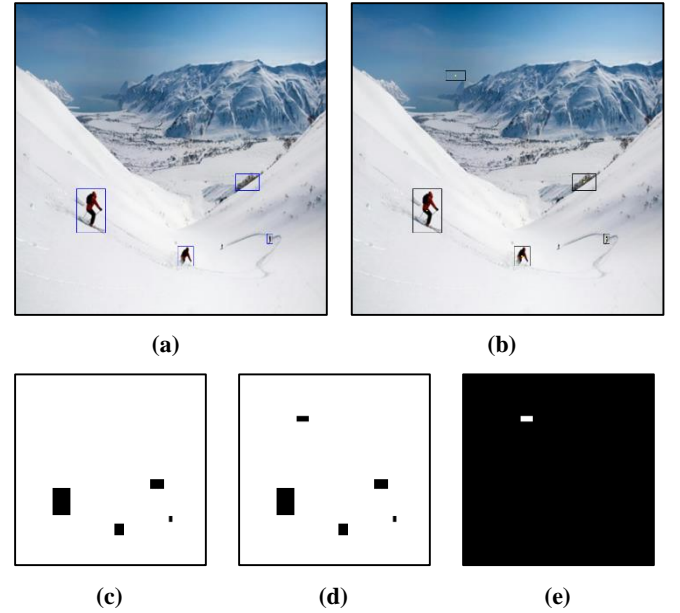


Figure 8 Comparing two sets of salient regions using EOD, (a) first attention map, (b) second attention map, (c) binary attention map of (a), (d) binary attention map of (b), (e) the result of applying EOD on (c) and (d).

The effect of falsely detected salient points FDSP on the SCD and EOBD has been studied on different image, to test the effect of falsely detected region with different sizes, and the effect of the distribution of the FDSPs on the accuracy of the distance measure. The following figure shows the curves of the distance with the increase of the number of FDSPs.

From Figure 9, one can notice that In SCD the curve is changing with respect to the change of points, which gives very good impression about the effect of FDSPs. EOD curve also gives excellent impression about the effect of the FDSPs.

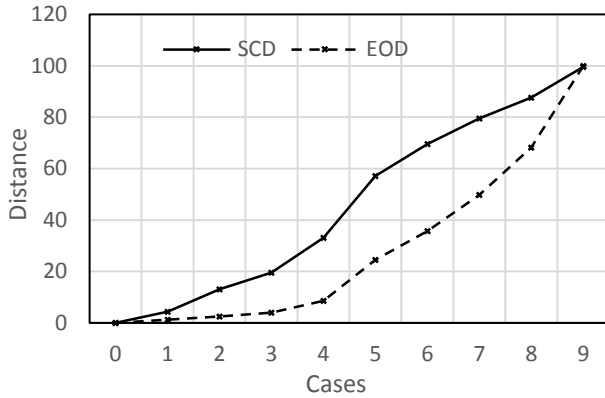


Figure 9 The effect of the number (quantity) of the falsely detected salient points FDSP

7 Experimental Results

The above described technique was applied on a set of images of size of 400×400 and windows size of 7×7 with overlapping of 3 between the windows. The following is a discussion for the results obtained. Suitable size of the sub image is important because it may affect the salient details, selecting a very large size may cover more than one object, while selecting a small size may consider the fine details such as texture as salient regions. In most literature the size usually selected form the common windows sizes such as 3×3 , 5×5 , or 7×7 (Kim & Milanfar, 2013), (Banerjee & Kundu, 2003), and (Rahmani, et al., 2008).

Different sub-image sizes were tested, in each test the distance between the extracted and the ground truth data is measured as shown in Figure 10, which represents the relation of window size vs. the SCD distance. Form this figure it is shown that the

minimum distance is at 7 and at 13, with window size of 13×13 the calculation will be very much higher than 7×7 , thus the latter size will be selected.

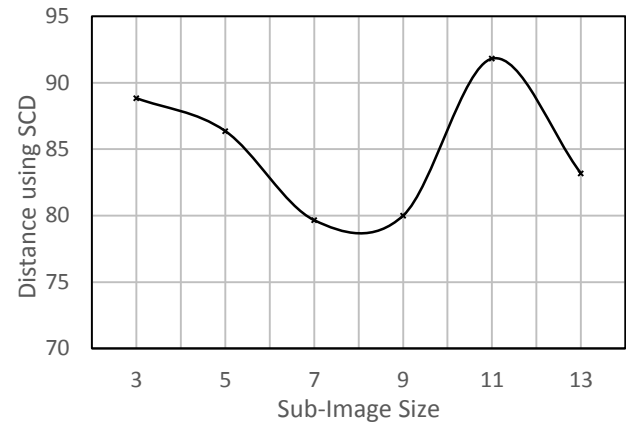


Figure 10 Average distance measured using SCD, vs. sub-image size.

In the same way we have selected the value of the overlapping between the adjacent windows, the optimum value for overlapping was 3.

The values of the histogram of the image is given in the following table which shows the bins, the frequency of occurrence (F), the log scaled frequency (LF), the normalized log scaled frequency (NLF) and the smoother NLF (SNLF).

Table 2 the distribution of the normalized variation measures for the image given in Figure 3

Bin	F	LF	NLF	SNLF
0	100	2	0.479501	0.545931
5	678	2.83123	0.678789	0.570505
10	203	2.307496	0.553224	0.551599
15	58	1.763428	0.422783	0.447727
20	34	1.531479	0.367173	0.39135
25	40	1.60206	0.384095	0.35074
30	18	1.255273	0.300952	0.304609
35	9	0.954243	0.22878	0.224692
40	4	0.60206	0.144344	0.162505
45	3	0.477121	0.11439	0.124375
50	3	0.477121	0.11439	0.11439
55	3	0.477121	0.11439	0.148432
60	8	0.90309	0.216516	0.186562
65	9	0.954243	0.22878	0.196547
70	4	0.60206	0.144344	0.17249
75	4	0.60206	0.144344	0.13436
80	3	0.477121	0.11439	0.13436
85	4	0.60206	0.144344	0.110302
90	2	0.30103	0.072172	0.072172
95	1	0	0	0.024057
100	1	0	0	0

The following figure shows the graphical representation for the histogram data given in the above table. The range of the data is very large and since we are interested more in the size (which frequency is larger), we shall reduce the difference in range by drawing the histogram in Log scale.

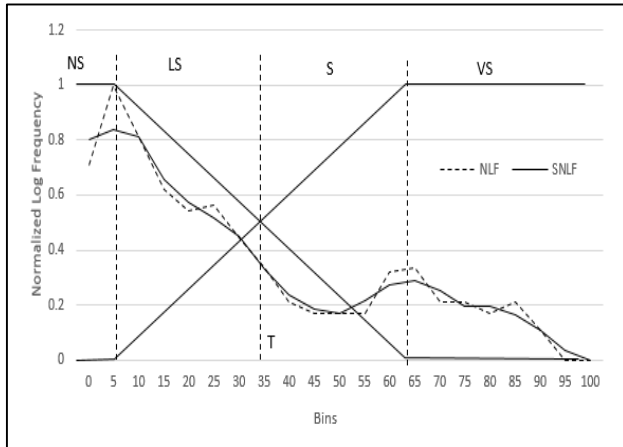


Figure 11 Fuzzy membership function representation of the histogram

The following table shows the threshold ranges. for salient regions extraction we shall consider the salient (S) and very salient (VS) and suppress the not salient (NS) and less salient (LS). The values in the table were extracted from the histogram approximation given in Figure 11.

Table 3 the ranges of the salient regions

FLV	Meaning	Range
NS	Not Salient Region	$0 \leq x < 5$
LS	Less Salient Region	$5 \leq x < 35$
S	Salient Region	$35 \leq x < 65$
VS	Very Salient Region	$65 \leq x \leq 100$

By using the threshold values in Table 3, which were extracted from the histogram given in Figure 11, the results in Figure 12 were obtained. From the figure,

(a) shows the not salient regions (white) which is in the range of (0 to 5). It is clear that most of the redundant information has been suppressed such as the sky and the snow on the mountain. (b) in the same figure shows the less salient regions, (c) shows the salient regions and (d) shows the very salient regions. The best threshold choice is by considering salient and very salient regions.

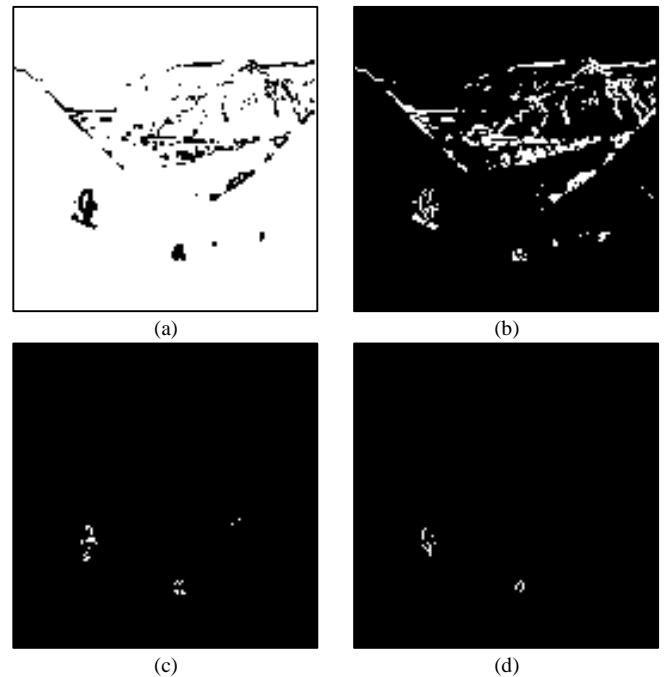


Figure 12 applying Fuzzy bimodal histogram thresholding, (a) not important region, (b) less important regions, (c) important regions, and (d) very important regions.

With the threshold obtained above the results of applying the algorithm are as shown in the following figure

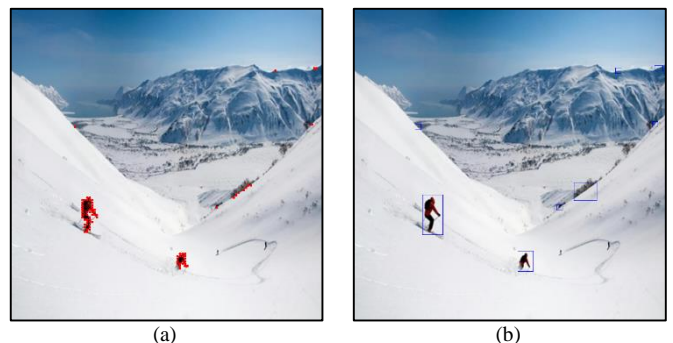


Figure 13 Applying the thresholding technique on the saliency enhanced image, (a) important regions (blue squares), (b) important object

After specifying suitable thresholding value and use it in unimportant details suppression, clustering process should be applied to cluster the important regions in clusters to form the object. Points growing technique was used for merging the important regions to form larger regions that contain the object.

The algorithm has been applied on images from different datasets that are available online for saliency tests, such as the datasets in (Liu, 2007), (Judd, 2012), (Achanta, 2010), (Li, 2010), in addition to the dataset that was created by the authors. A comparison with different state-of-the-art saliency extraction techniques was carried out. The evaluation has been done by comparing the obtained saliency map with two kind of maps, the first one was obtained by manual labelling, and the second one was obtained from eye tracking experiment. The distance between the ground truth data was calculated to each image in the used dataset and then the average is extracted for each algorithm and compared to other methods. Figure 14 shows the curves obtained from comparing the proposed method with FT (Achanta, et al., 2009), HFT (Li, et al., 2007), Itti (Itti, et al., 1998), and MSSS (Achanta & Susstrunk, 2010).

From the obtained curves it is clear that the distance is minimum at the point corresponding to Irregularity Saliency Detector ISD, which means that the result that we have obtained is better than the rest of the methods as shown in **Table 5**, which gives the result of the methods mentioned above for a sample of images. The maximum distance was at Itti method,

this is because Itti concerned about the sequence much.

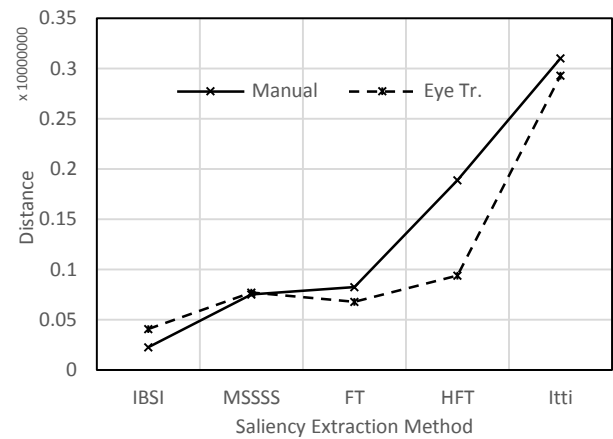


Figure 14 the distance measured between the saliency maps and ground truth data using EOBDD.

8 Conclusions

In this paper, a new saliency extractor has been developed based on the irregularity in some regions of the image; the new method is fully automatic since it needs no interference from the user. In order to achieve this a new thresholding technique was developed based on the available data. The proposed method has been compared with some state-of-the-art techniques and evaluated using two developed saliency evaluation methods. The obtained results showed that the proposed method has worked well on different kinds of images. Images with high texture, and with multiple salient regions have been tested as well.

Table 4 More Results of the proposed algorithm.


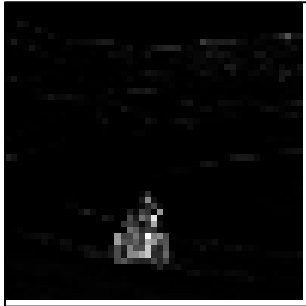
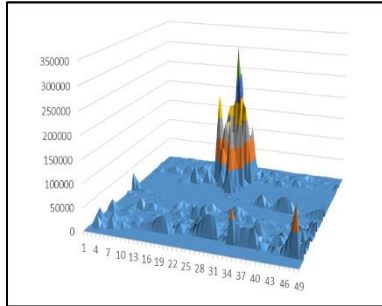



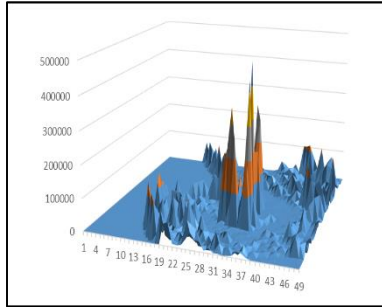


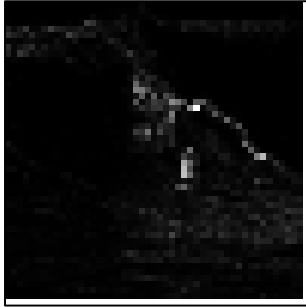
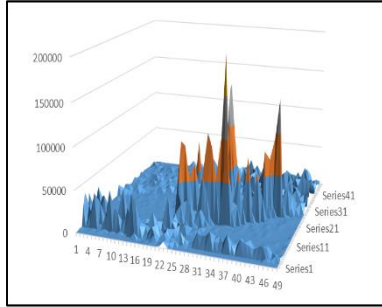








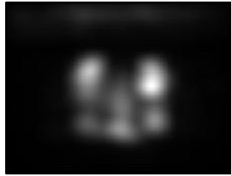






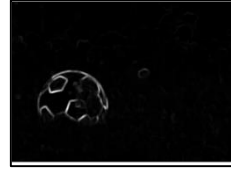
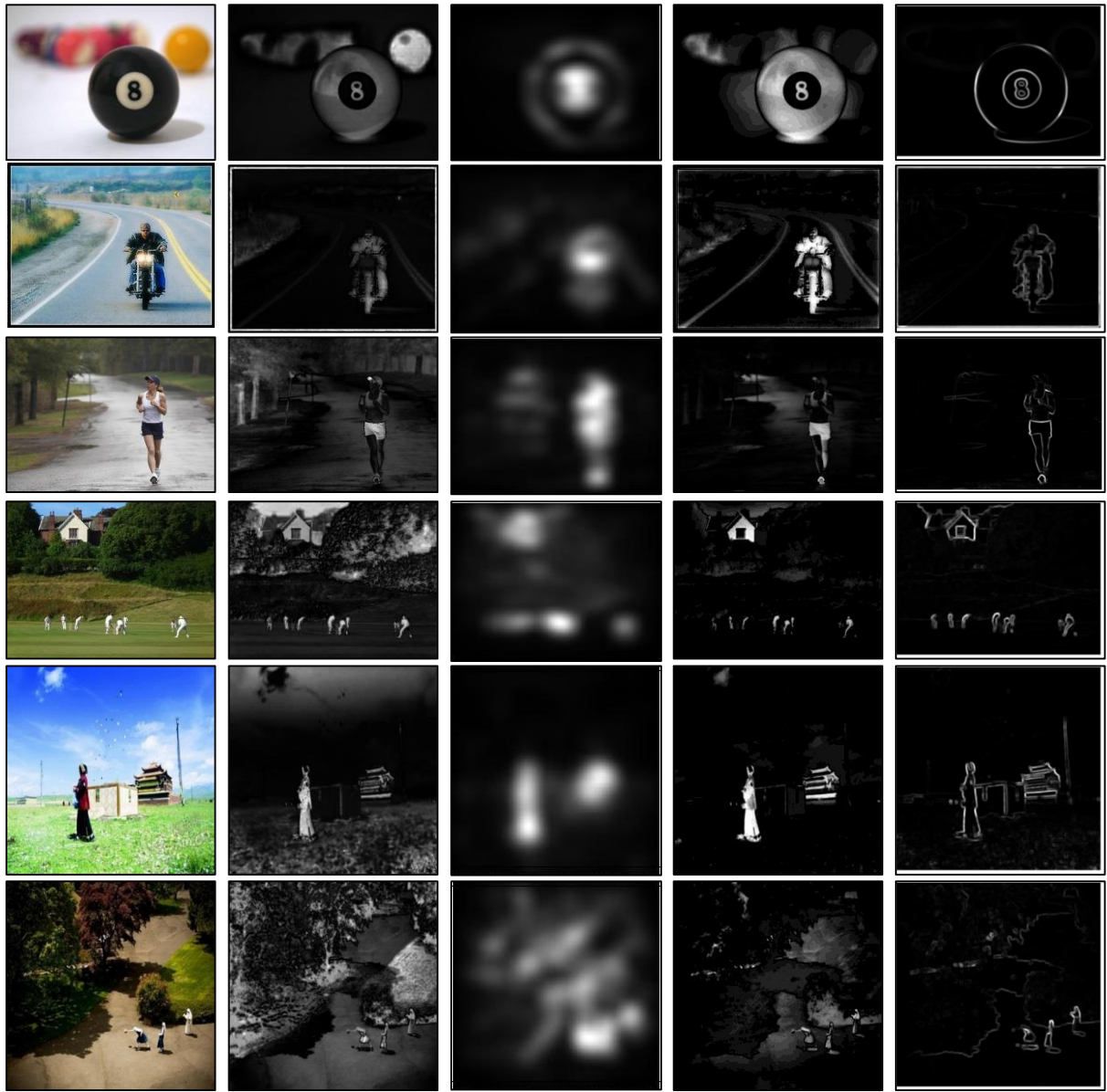
Original Image	Applying Saliency	3D space Representation	Saleint object
			
			
			

Table 5 The Saliency Maps obtained from different Saliency Extractors

Original	FT (Achanta, et al., 2009)	HFT (Li, et al., 2007),	MSSS (Achanta & Susstrunk, 2010)	ISD
				
				
				



9 References

- Achanta, R., 2010. *Image and Visual Representation Group IVRG Dataset*. [Online] Available at: <http://ivrg.epfl.ch/research/saliency/MSSS.html> [Accessed 12 2012].
- Achanta, R., Hemami, S., Estrada, F. & Susstrunk, S., 2009. *Frequency-tuned Salient Region Detection*. s.l., IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2009.
- Achanta, R. & Susstrunk, S., 2010. *Saliency detection using maximum symmetric surround*. Hong Kong, s.n.
- Al-Azawi, M., 2013. Image Thresholding using Histogram Fuzzy Approximation. *International Journal of Computer Applications*, 83(9), pp. 36-40.
- Al-Azawi, M. & Ibrahim, N. K., 2012. *Bimodal Histogram Based Image Segmentation Using Fuzzy-Logic*. Massana, Mussna College of Technology.
- Al-Azawi, M., Yang, Y. & Istance, H., 2014. *A new gaze points agglomerative clustering algorithm and its application in regions of interest extraction*. Gurgaon, IEEE International Advance Computing Conference (IACC).
- Amini, A., Wah, T. Y., Saybani, M. & Yazdi, S., 2011. *A Study of Density-Grid based Clustering Algorithms on Data Streams*. s.l., s.n., pp. 1652-1656.
- Anon., 2012. *Cluster analysis*. [Online] Available at: en.wikipedia.org/wiki/Cluster_analysis [Accessed 13 2013].
- Anon., 2013. *Precision and recall*. [Online] Available at: https://en.wikipedia.org/wiki/Precision_and_recall [Accessed 15 8 2013].
- Arivazhagan, S. & Shebiah, R. N., 2009. Object Recognition Using Wavelet Based Salient Points. *The Open Signal Processing Journal*, Volume 2, pp. 14-20.
- Banerjee, M. & Kundu, M. K., 2003. Edge based features for content based image retrieval. *Pattern Recognition*, Issue 36, pp. 2649-2661.
- Borji, A., Sihite, D. N. & Itti, L., October 7-13, 2012. *Salient Object Detection: A Benchmark*. Florence, Italy, 12th European Conference on Computer Vision Proceedings, Part II, pp. 414-429.
- Bruce, N., Loach, D. & Tsotsos, J., 2007. *Visual Correlates of Fixation Selection: A Look at the Spatial Frequency Domain*. San Antonio, TX, s.n.
- Burns, R. B. & Burns, R. A., 2008. Chapter 23: Cluster Analysis. In: *Business Research Methods and Statistics Using SPSS*. s.l.: Sage Publications Ltd., pp. 552-567.
- Choi, Y. S., Mosley, A. D. & Stark, L. W., 1995. String editing analysis of human visual search.. *Optometry and Vision Science*, Volume 72, pp. 439-45.
- Chowdhury, Nirmalya, Murthy, C. A. & Pal, S. K., 1995. Cluster detection using neural networks. *IEEE International Conference on Neural Networks, Proceedings.*, Volume 4, pp. 2166-2170.
- Cook, D. & Swayne, D. F., 2007. chapter 5: Cluster Analysis. In: *Interactive and Dynamic Graphics for Data Analysis*. s.l.:Springer, pp. 103-128.
- Davis, J. & Sharma, V., 2004. *Robust Background Subtraction for Person Detection in Thermal Imagery*. s.l., Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04), pp. 128-132.
- Einhauser, W., Spain, M. & Perona, P., 2008. Objects predict fixations better than early saliency. *Journal of Vision*, 8(14)(18), pp. 1-26.

- Erdem, E. & Erdem, A., 2013. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, 13(4)(11), pp. 1-20.
- Estivill-Castro, V., 2002. Why so many clustering algorithms: a position paper. *ACM SIGKDD Explorations Newsletter*, 4(1), pp. 65-75.
- Fang, Y. et al., 2012. Bottom-Up Saliency Detection Model Based on Human Visual Sensitivity and Amplitude Spectrum. *IEEE TRANSACTIONS ON MULTIMEDIA*, 14(1), pp. 187-198.
- Gide, M. & Karam, L., 2012. *Improved foveation- and saliency-based visual attention prediction under a quality assessment task*. s.l., s.n., pp. 200-205 .
- Gour, B., Bandopadhyaya, T. K. & Sharma, S., 2008. *ART Neural Network Based Clustering Method Produces Best Quality Clusters of Fingerprints in Comparison to Self Organizing Map and K-Means Clustering Algorithms*. s.l., s.n., pp. 282-286.
- Hou, X. & Zhang, . L., 2007. *Saliency Detection: A Spectral Residual Approach*. Minneapolis, MN, s.n.
- Hu , Y. et al., 2005. *Salient Region Detection Using Weighted Feature Maps Based on the Human Visual Attention Model*. s.l., Proc. Fifth Pacific Rim Conf. Multimedia, pp. 993-1000.
- Huang, G.-Y., Liang, D.-P., Hu, C.-Z. & Ren, J.-D., 2010. *An Algorithm for Clustering Hetrogeneous Data Stream with Uncertainty*. s.l., s.n.
- Huang, Y.-P., Chang, T.-W. & Huang, C.-Z., 2003. *A fuzzy feature clustering with relevance feedback approach to content-based image retrieval*. s.l., s.n., pp. 57-62.
- Itti, L., Koch, C. & Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), pp. 1254-1259.
- Ivanoff, R. M. & Klein, J., 2008. Inhibition of return. *Scholarpedia*, 3(10), p. 3650.
- Jain, A. K., Murty, M. N. & Flynn, P. J., Sept. 1999. Data Clustering: A Review. *ACM Computing Surveys (CSUR) Surveys*, 31(3), pp. 264-323.
- Judd, T., 2012. *Achanta*. [Online] Available at: <http://people.csail.mit.edu/tjudd/SaliencyBenchmark/>
- Judd, T., Durand, . F. & Torralba, A., January 13, 2012. *A Benchmark of Computational Models of Saliency to Predict Human Fixations*. s.l., s.n.
- Kadir, T. & Brady, M., 2001. Scale, Saliency and Image Description. *International Journal of Computer Vision*, 45(2), pp. 83-105.
- Kapsalas, P., Rapantzikos, K., Sofou, A. & Avrithis, Y., 2008. *Regions of interest for accurate object detection*. London, s.n., pp. 147-154 .
- Kim, C. & Milanfar, P., 2013. Visual saliency in noisy images. *Journal of Vision*, 13(4)(5), pp. 1-14.
- Kim, S., Park, S. & Kim, M., 2003. *Central object extraction for object-based image retrieval, CIVR'03*. Heidelberg , s.n., pp. 39-49.
- Koch, C. & Ullman, S., 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, Volume 4, p. 219-227.
- Krishnapuram, R. & Keller, J. M., 1993. A Possibilistic Approach to Clustering. *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, 1(2), pp. 98-110.
- Le Meur, O., Callet, P. L., Barba, D. & Thoreau, D., May 2006. A Coherent Computational Approach to Model Bottom-Up Visual Attention. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(5), pp. 802-817.
- Li, J., 2010. *LI, Jian Dataset*. [Online] Available at: <http://www.cim.mcgill.ca/~lijian/>

[database.htm](#)

[Accessed 1 12 2012].

Li, J. et al., 2007. Visual Saliency Based on Scale-Space Analysis in the Frequency Domain. *JOURNAL OF LATEX CLASS FILES*, 6(1), pp. 1-16.

Lin, . D.-W. & Yang, S.-H., 2007. Wavelet-Based Salient Region Extraction. In: *Advances in Multimedia Information Processing – PCM 2007. Vol. 4810*. Honk Kong: Springer, pp. 389-392.

Lin, Y., Fang, B. & Tang, Y., 2010. *A Computational Model for Saliency Maps by Using Local Entropy*. Atlanta, Georgia, USA, Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10).

Lin, Y. et al., 2013. A Visual-Attention Model Using Earth Mover's Distance-Based Saliency Measurement and Nonlinear Feature Combination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2), pp. 314-328.

Liu, T., 2007. *MSRA Salient Object Dataset*. [Online]

Available at: http://research.microsoft.com/en-us/um/people/jiansun/SalientObject/salient_object.htm

[Accessed 1 12 2012].

Liu, T. et al., 2007. *Learning to Detect a Salient Object*. s.l., Proc. IEEE CS Conf. Computer and Vision Pattern Recognition, pp. 1-8.

Liu, Z., Xue, Y., Yan, H. & Zhang, Z., 2011. Efficient saliency detection based on Gaussian models. *IET Image Processing*, 5(2), pp. 122-131.

Loupias, E., Sebe, N., Bres, S. & Jolion, J.-M., 2000. *Wavelet-based salient points for image retrieval*. Vancouver, BC, Canada, s.n., pp. 518-521.

Mancas, M., 2008. *Image Perception: Relative Influence of Bottom-Up and Top-Down Attention*. s.l., Proc. Int'l Workshop Attention and Performance in Computer Vision in

conjunction with the Int'l Conf Computer Vision Systems, pp. 94-106.

Masciocchi, C. M., Mihalas, S., Parkhurst, D. & Niebur, E., 2009. Everyone knows what is interesting: Salient locations which should be fixated. *Journal of Vision*, 9(11)(25), pp. 1-22.

Milligan, G. & Cooper, M., 1987. Methodology Review: Clustering Methods. *Applied Psychological Measurement*, 11(4), pp. 329-354.

Parkhurst, D., Law, K. & Nieb, E., 2002. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, Issue 42, pp. 107-123.

Pele, O. & Werman, M., 2008. *A Linear Time Histogram Metric for Improved SIFT Matching*. Marseille, France, Proceeding ECCV '08 Proceedings of the 10th European Conference on Computer Vision: Part III, pp. 495-508.

Peters, R., Iyer, A., Itti, L. & Koch, C., 2005. Components of bottom-up gaze allocation in natural images. *Vision Research*, Volume 45, pp. 2397-2416..

Rahmani, R. et al., 2008. Localized Content Based Image Retrieval. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, SPECIAL ISSUE*, pp. 1-10.

Rai, P. & Singh, S., 2010. A Survey of Clustering Techniques. *International Journal of Computer Applications*, 7(12), pp. 1-5.

Rosin, P. L., 2009. A simple method for detecting salient regions. *Pattern Recognition*, Volume 42, pp. 2363-2371.

Rubner, Y., Tomasi, C. & Guibas, L., 2000. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2), pp. 99-121.

Schmid, C. & Mohr, R., 1997. Local greyvalue invariants for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(5), pp. 530-535.

- Schmid, C., Mohr, R. & Bauckhage, C., 2000. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2), pp. 151-172 .
- Sebe , N. et al., 2002. Evaluation of Saliient Point Techniques. *Image and Vision Computing*, Volume 21, pp. 367-377.
- Song, H., Li, B. & Zhang, L., 2006. *Color Salient Points Detection Using Wavelet*. Dalian, China, Proceedings of the 6th World Congress on Intelligent Control and Automation.
- Stöttinger , J., Hanbury , A. , & Gevers, T., 2009. *Lonely but attractive: Sparse color salient points for object retrieval and categorization*. Miami, FL, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009., pp. 1-8 .
- Tan, P.-N., Steinbach, M. & Vipin , K., 2006. Cluster Analysis: Basic Concepts and Algorithms. In: *Introduction to Data Mining*. s.l.:Addison-Wesley Companion, pp. 488-568.
- Tatler, B. W., Hayhoe, M. M., Land, M. F. & Ballard, D. H., 2011. Eye guidance in natural vision: Reinterpreting saliency. *Journal of Vision*, 11(11)(5), pp. 1-23.
- Tian, Q. et al., 2001. Image retrieval using wavelet-based salient points. *J. Electron. Imaging*, 10(4), pp. 835-849.
- Toet, A., 2011. Computational versus Psychophysical Bottom-Up Image Saliency: A Comparative Evaluation Study. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 33(11), pp. 2131-2146.
- Torn, A. A., 1977. Cluster Analysis Using Seed Points and Density-Determined Hyperspheres as an Aid to Global Optimization. *IEEE Transactions on Systems, Man and Cybernetics*, , 7 (8), pp. 610-616.
- Tryfos, P., 1998. Chapter 15: Cluster analysis. . In: *Methods for Business Analysis and Forecasting: Text & Cases*, . s.l.:Wiley.
- Tseng, V. S. & Kao, C.-P., 2007. A Novel Similarity-Based Fuzzy Clustering Algorithm by Integrating PCM and Mountain Method. *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, , 15(6), pp. 1188-1196.
- Xu, R. & Wunsch, D., May, 2005. Survey of Clustering Algorithms. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 16(3), pp. 645-678.
- Zhang, H. & Goldman, S. A., 2006. *Image Segmentation Using Salient Points-Based Object Templates..* Atlanta, GA , 2006 IEEE International Conference on Image Processing,, pp. 765-768.
- Zhang, L. et al., 2008. SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision* , 8(7)(32), pp. 1-20.
- Zhao, Q. & Koch, C., 2011. Learning a saliency map using fixated locations in natural scenes. *Journal of Vision*, 11(3)(9), pp. 1-15.
- Zhou, B., Hou, X. & Zhang, L., 2010. *A phase discrepancy analysis of object motion*. Queenstown, New Zealand, Proceedings of the 10th Asian conference on Computer vision ACCV'10, pp. 225-238.