Pertanika J. Sci. & Technol. 27 (4): 1735 - 1758 (2019)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Improved Salient Object Detection via Boundary Components Affinity

Nur Zulaikhah Nadzri[1,2]*, Mohammad Hamiruce Marhaban[1], Siti Anom Ahmad[1], Asnor Juraiza Ishak[1] and Zalhan Mohd Zin[2]

[1]Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[2]Industrial Automation Section, Universiti Kuala Lumpur Malaysia France Institute, Section 14, 43560, Bandar Baru Bangi, Selangor, Malaysia

ABSTRACT

Referring to the existing model that considers the image boundary as the image background, the model is still not able to produce an optimum detection. This paper is introducing the combination features at the boundary known as boundary components affinity that is capable to produce an optimum measure on the image background. It consists of contrast, spatial location, force interaction and boundary ratio that contribute to a novel boundary connectivity measure. The integrated features are capable to produce clearer background with minimum unwanted foreground patches compared to the ground truth. The extracted boundary features are integrated as the boundary components affinity. These features were used for measuring the image background through its boundary connectivity to obtain the final salient object detection. Using the verified datasets, the performance of the proposed model was measured and compared with the 4 state-of-art models. In addition, the model performance was tested on the close contrast images. The detection performance was compared and analysed based on the precision, recall, true positive rate, false positive rate, F Measure and Mean Absolute Error (MAE). The model had successfully reduced the MAE by maximum of 9.4%.

Keyword: Boundary connectivity, boundary ratio, force interaction

ARTICLE INFO

Article history:
Received: 18 January 2019
Accepted: 13 May 2019
Published: 21 October 2019

E-mail addresses:
eka.nadzri@gmail.com (Nur Zulaikhah Nadzri)
mhm@upm.edu.my (Mohammad Hamiruce Marhaban)
sanom@upm.edu.my (Siti Anom Ahmad)
asnorji@upm.edu.my (Asnor Juraiza Ishak)
zalhan@unikl.edu.my (Zalhan Mohd Zin)
* Corresponding author

INTRODUCTION

Salient detection is the ability to detect the most prominent object on a particular scene or region. This is adapted from the ability of human eyes to distinguish distinctive

ISSN: 0128-7680
e-ISSN: 2231-8526

(salient) objects on the visual field. Out of large inputs entering our eyes, most of it are filtered and the ones left are useful for cognitive purpose. The factor that influences visual saliency lies in two mechanisms, which are bottom-up attention and top-down attention. The bottom-up mechanism is fast and stimulus-driven attention in which the visual saliency is based on the low-level features including color, intensity, orientation, texture, and motion. On the other hand, the top-down mechanism is slower as it is goal-driven which referring to internal guidance of attention based on prior knowledge (Katsuki & Constantinidis, 2014) . It is a voluntary allocation of attention to certain features, objects, or regions in space (Pinto et al., 2013)

There have been a large number of studies in salient object detection in recent years. This area has attracted researchers in which hundreds of computational models were proposed to obtain the most real-time and optimum result on the detection.

The salient object detection model has been applied in many applications such as object detection and segmentation (Liu et al., 2008; Xiuli et al., 2017; Zhang et al., 2006), image retargeting (Chou & Su, 2016; Pritch et al., 2013), image compression (Guo & Zhang, 2010; Itti, 2004; Srivastava et al., 2016) and image quality assessment (Xiao & Yeh, 2017). The models have been successfully applied in many multi-discipline areas including multimedia (Li et al., 2017; Luz et al., 2017), medical (Woodbridge et al., 2011; Ahn et al., 2017), remote sensing (Li et al., 2016; Zhang et al., 2016) and robotics (Jiang et al., 2015; Liu et al., 2013).

The earliest salient detection model was proposed by Itti et al. (2001) which was based on the selective attention mapping by Koch and Ullman (1985) together with the basis of other models of Baluja (1997) and Milanese et al. (1995). It is also related to the physiological theories of visual attention called the Feature Integration Theory by Treisman and Galade (1980) that explains the human visual strategies. The theory suggests that the attention onto an object involved in separate process where the early perceived stimulus will put all the features in parallel as a pre-attentive stage and the individual features are combined to select the focus location. The model has become hit in a way that it was used as a basis for multi cross discipline in cognitive psychology, neuroscience, and computer vision.

In recent years, multiple approaches have been used as the basis of hypothesis on the salient detection model in which the base theorem is always being referred to the term 'prior' of the presented models. To list a few, which are contrast prior to either local (Achanta et al., 2008; Itti et al., 2001) or global (Cheng et al., 2011; Perazzi & Kr, 2012; Zhang et al., 2016), edge/shape prior (Jiang et al., 2011; Yang et al., 2017), texture prior (Hu et al., 2016; Zhang et al., 2017), background prior (Ahn et al., 2017; Jing et al., 2014; Wei et al., 2012), or foreground prior (Wang et al., 2015; Zhou et al., 2017). The contrast prior has been the most applied assumption in which the salient object is highlighted whenever

the contrast between the object and the surroundings are high. Besides, boundary prior models (Luo et al., 2016; Manke & Jalal, 2016; Niu, 2018; Tang et al., 2018; Wu et al., 2013) have proven good results in salient detection where the image boundary is being assumed as the image background.

The earlier model that has exploited the boundary prior is by Wei et al. (2012) . The model is related to the assumption that associates image boundary and their connectivity within the image patches. The saliency detection is improved with the integration of geodesic saliency that measures the shortest path of the image patches to the background patch. Motivated by the model, Zhu et al. (2014) had come out with saliency detection that relied on the measure of the image patches as background only when the region it belonged to was strongly connected to the image boundary. The model by Wu et al (2013) uses the *Lo* smoothing filter and the Principal Component Analysis to make categorization of the salient object, and at the same time produces the boundary information for the background merging and boundary scoring stages. On adapting the boundary prior to salient detection model, Manke et al. (2016) used the Poisson distribution to highlight the salient object by computing the difference of pixel intensity and mean of boundary pixel of an image. Considering other perspectives within the boundary prior to the assumption, the models by Luo et al., (2016); and Tang et al., (2018) are focusing on the probability that the salient object could be located at the image boundary.

Among those models, the model by Zhu et al (2014) has been listed as the top 6 models for the salient detection supporting along a detailed study comparison by Borji et al. (2015). With proven computation for the saliency detection, the detection of the model still falsely shows some image patches that should be assigned as the image background. It is due to their boundary connectivity computation that measures the contrast difference towards the whole image boundary and not considering the weight or ratio of the patches towards the individual side of the boundary. Thus, the large contrast difference of patches will be considered as the image foreground even though they are highly connected to the image boundary. As a result, any high connectivity of patches at the boundary with higher contrast difference will be highlighted as the salient object. Figure 1 shows the saliency map by Zhu et al. (2014) named as RBD where other high contrast patches were pop-out as the salient object.

This paper presents a new method for the saliency detection based on the image boundary prior. The main contribution is the combination features extracted that consist of contrast, spatial location, force interaction and boundary ratio that is assigned as the Boundary Components Affinity (BCA). Apart from only taking into account the patches' contrast and location difference using the Euclidean distance, considering the patch force interactions is able to strengthen the patch differences in grouping them according to the object they belong (foreground or background). On the other hand, the boundary ratio is

Figure 1. (From left to right) Original image, ground truth and saliency map of RBD (Zhu et al., 2014)

used to identify the ratio of a patch connected to which image boundary. These features are integrated as the BCA and act as a measure of the boundary connectivity. This measure is able to assign patches as foreground or background of the image. Higher boundary connectivity means that the patch is highly connected to the image boundary and hence it would be assigned as the image background. With these combination features, the proposed saliency measure is computed from novel integrated boundary components in which this method is able to highlight the salient object by producing the optimum background measure.

THE PROPOSED FRAMEWORK OF BCA

The following section discusses the detail of the proposed salient detection algorithm. The core of this saliency detection arises from the background measurement. The measurement is made by integrating the contrast, spatial location, force interaction and boundary ratio of the patches to be put together as the boundary components and hence measure the connectivity of the patches towards the image boundary. The spatial force measurement is able to highlight the salient object by obtaining the force of each patch related to its spatial contrast and distance. Thus, the high force of the patches is grouped based on the object it belongs to. The boundary ratio is used to strengthen the computation of patches-boundary connectivity. The detail about the algorithm will be discussed in the next section. The

summary of the stated algorithm is depicted as in Figure 2. It is inspired by the RBD model (Zhu et al., 2014) where the extracted BCA are integrated for computing the boundary connectivity as they are highlighted in red dotted square which illustrates the additional steps that have been added in the proposed BCA in comparison with the RBD model. The details of the proposed algorithm are discussed in the following subsections.

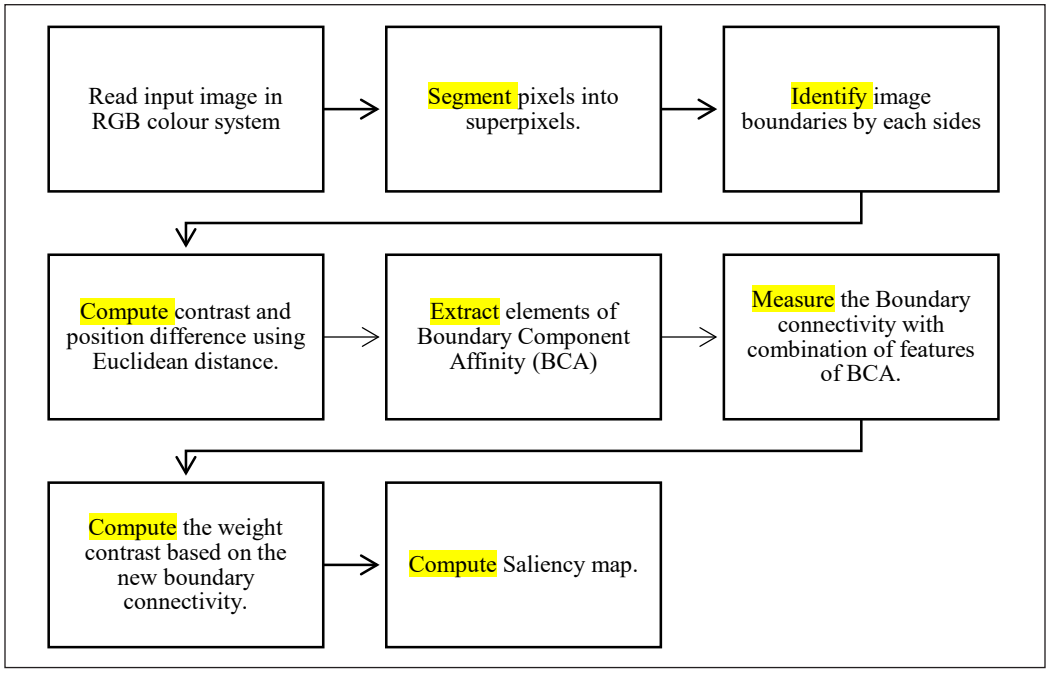


Figure 2. The proposed model block diagram

Pre-processing

Having the fastest detection time is the vital criteria for any salient detection algorithm as processing a large number of pixels, i.e. 300 x 400 pixels, may result in a higher computational time. Thus, the image is segmented using the Simple Linear Iterative Clustering (SLIC) technique (Achanta et al., 2012) where clustering the pixels is 5-dimensional including color and plane space to obtain a compact superpixel. The unique distance measure produces compactness and regularity in the superpixel shapes and thus resulting an efficient superpixel segmentation. With a good resulting segmentation, the images will be in accurate group of patches with reduce number of component to be processed as the saliency measurement rather that individual pixels. Figure 3 shows the result of SLIC superpixel segmentation.

Figure 3. (From left to right) Original image and superpixel segmentation using SLIC

Pixel Force Feature

The concept of spatial force is taken from the concept of pixel-force field (Hurley et al., 1999). The mathematical modelling of object force interaction in physics is adopted for the pixel force transformation. This feature has been demonstrated to benefit numbers of application such as segmentation (Bucha et al., 2007), extraction and recognition (Bucha et al., 2006; Hurley et al., 1999), map vectorization (Bucha et al., 2007), and image registration (Ghayoor, 2010). Each pixel in an image is considered as a single particle that contains a specific scalar and vector value towards other pixels(Hurley et al., 1999) and it can be visualized as in Figure 4. In transforming image into the force field, each pixel is presumed to produce symmetrical force field, $\mathbf{F}_i(\mathbf{r})$ towards other pixel of $P(\mathbf{r}_i)$ where $\mathbf{r}$ is the location vector. It can be defined mathematically as Eqation [1],

$$\mathbf{F}_i(\mathbf{r}) = P(\mathbf{r}_i)\frac{\mathbf{r}_i - \mathbf{r}}{|\mathbf{r}_i - \mathbf{r}|^3} \qquad [1]$$

Generally, a single pixel p_i in an image can be presented as their two dimensional spatial location vectors and the color appearance q in n dimensional vector, $\mathbf{Q_p} = (q1, \ldots, qn)$. The definition of the pixel interaction force, $\mathbf{F}_{i,j}$ can be referred to Bucha et al (2007). The force, $\mathbf{F}_{i,j}$ between two pixels p_i and p_j is defined by the pixel spatial relation $D(\mathbf{X_p}(i), \mathbf{X_p}(j))$

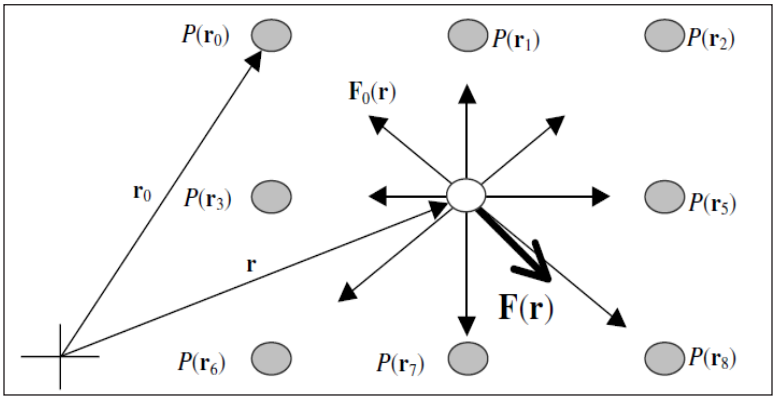


Figure 4. Illustration of force field geometry (Hurley et al., 1999)

and their color dissimilarities $C(\mathbf{Q_p}(i), \mathbf{Q_p}(j))$. Thus, the pixel interaction force between p_i and p_j can be presented as Equation2.

$$\mathbf{F}(p_i, p_j) = D(\mathbf{X_p}(i), \mathbf{X_p}(j)) \,.\, C(\mathbf{Q_p}(i), \mathbf{Q_p}(j)) \,.\, \mathbf{r}_{i,j} \quad [2]$$

The two dimensional vector, $\mathbf{r}_{i,j}$, is the direction of the pixel interaction force from the coordinate location of pixel, p_i, to the coordinate location of pixel, p_j.

Image Patch Force Interaction

In applying this pixel force field interaction into the model, the pixels are replaced with the superpixel patches which are obtained from the superpixel segmentation earlier. Note that the patch force interaction is presented as $\mathbf{F}(P_i, P_j)$ where P_i is a particular patch of the superpixel. As a single patch consists of a number of pixels, the location can be presented as the mean spatial location $\overline{\mathbf{X}}_\mathbf{P}(i) = (\bar{x}_a, \bar{y}_a)$ where $a \in i$. The component *a* is the set of pixels that are bounded in the patch *i*.

To imitate the human vision, the CIE Lab color system is used for the color vector where the pixel appearance, $\mathbf{Q}_p$ is presented such that,

$$\mathbf{Q_p} = (q_{L,} q_a, q_b) \quad [3]$$

Where q_L, q_a and q_b are pixel appearances on the L, a and b channel.

However, the vector needs to be presented as a vector for the superpixel patch $\overline{\mathbf{Q}}_\mathbf{P} = (\bar{q}_{L,} \bar{q}_a, \bar{q}_b)$ as the mean of CIE Lab color of a pixel set *a* of the patch *i* originates from the RGB input pixels.

The scalar force interaction between the patches P_i and P_j is obtained by substituting the relevant values in equation (2) and presented as in Equation [4].

$$F_{ij} = D(\bar{X}_P(i), \bar{X}_P(j)) \cdot K(\bar{Q}_P(i), \bar{Q}_P(j)) \quad [4]$$

Both elements $D(\bar{X}_P(i), \bar{X}_P(j))$ and $K(\bar{Q}_P(i), \bar{Q}_P(j))$ are the Euclidean distance of the spatial coordinate and CIE Lab difference between patches respectively which are based on the equation introduced by Bucha et al. (2007). They can be written as Equations [5], [6] and [7].

$$D(\bar{X}_P(i), \bar{X}_P(j)) = \frac{1}{d(\bar{X}_P(i), \bar{X}_P(j))} = \frac{1}{\sqrt{\sum_{i=1}^{n} \left(\left|\bar{x}_i - \bar{x}_j\right|\right)^2 + \left(\left|\bar{y}_i - \bar{y}_j\right|\right)^2}} \quad [5]$$

$$K\left(\bar{Q}_P(i), \bar{Q}_P(j)\right) d\left(\bar{Q}_P(i), \bar{Q}_P(j)\right) = \sqrt{\sum_{i=1}^{n} \left(\bar{q}_{L_i} - \bar{q}_{L_j}\right)^2 + \left(\bar{q}_{a_i} - \bar{q}_{a_j}\right)^2 + \left(\bar{q}_{b_i} - \bar{q}_{b_j}\right)^2} \quad [6]$$

Combining both equations [5] and [6] into [4], the final force interaction of each patch, $\mathbf{F}_i(p)$ can be presented by,

$$\mathbf{F}_i(P) = \sum_{i=1}^{n} \frac{\sqrt{\sum_{i=1}^{n}\left(\bar{q}_{L_i} - \bar{q}_{L_j}\right)^2 + \left(\bar{q}_{a_i} - \bar{q}_{a_j}\right)^2 + \left(\bar{q}_{b_i} - \bar{q}_{b_j}\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(\left|\bar{x}_i - \bar{x}_j\right|\right)^2 + \left(\left|\bar{y}_i - \bar{y}_j\right|\right)^2}} \cdot \mathbf{r}_{ij} \qquad [7]$$

The calculated force interaction is drastically able to group the similar patch and hence it is able to contribute to the coarse object detection. It is because the force interaction is combining the color feature and spatial feature in one single value. This feature is not integrated in the previous models whereby those models only use the Euclidian distance on both color feature and spatial coordinate separately. The equation in [7] is used in the proposed model by ignoring the vector force direction, $\mathbf{r}_{i,j}$ since this is not an important element to be considered for computation.

Boundary Ratio

In computing saliency, this model considers how near the location of a patch to a specific boundary. The concept is illustrated as in Figure 3. The boundary ratio is the ratio of the patch center (red dot of each patch) to the nearest boundary either at the top, left, right or down boundary. This boundary ratio is used to improve the measure of the boundary connectivity.

The term boundary connectivity has been introduced by Zhu et al. (2014). The boundary connectivity, *b* of a patch, *P* can be measured by taking the ratio of pixels, *p* of a patch , *P* at the image boundary, *m* to the square root of the pixels at patch region, *R* (Equation [8]).

$$b(P) = \frac{|\{p|P \in m\}|}{\sqrt{|\{p|P \in R\}|}} \qquad [8]$$

From the defined equation, we can see that the measure of boundary connectivity is taking all the four side boundaries into account and this will result in average measure on the boundary connectivity.

To have a more precise computation, the measurement on boundary connectivity is integrated with the boundary ratio to strengthen the patches that are highly connected to the boundary. Here, the model is taking the center of each patch and obtain its nearest distance to a specific boundary. The boundary ratio, b_r of a patch, *P* is defined as the ratio of pixels, *p* of patch *P* towards a specific boundary, *m* to its center position with respect to a boundary side, *m*, whether it referred to top, down, left or right side.

Figure 5 is the illustration on boundary ratio computation. Consider that each color represents a single patch. By locating the mean location of each patch, the distance from the patch center to the nearest boundary is measured. As the patch has the lowest contrast towards the boundary and it is near the patch to any boundary side, the larger the ratio. Therefore, any low contrast patch that is close to the boundary has higher probability to be accounted as the image background. This value will be used to strengthen the boundary connectivity of each patch from the equation [8]. This is the missing consideration in the previous model (Zhu et al., 2014) where the boundary connectivity is measured towards all boundaries. As a result, any small high contrast patches will appear as salient region as in Figure 1.

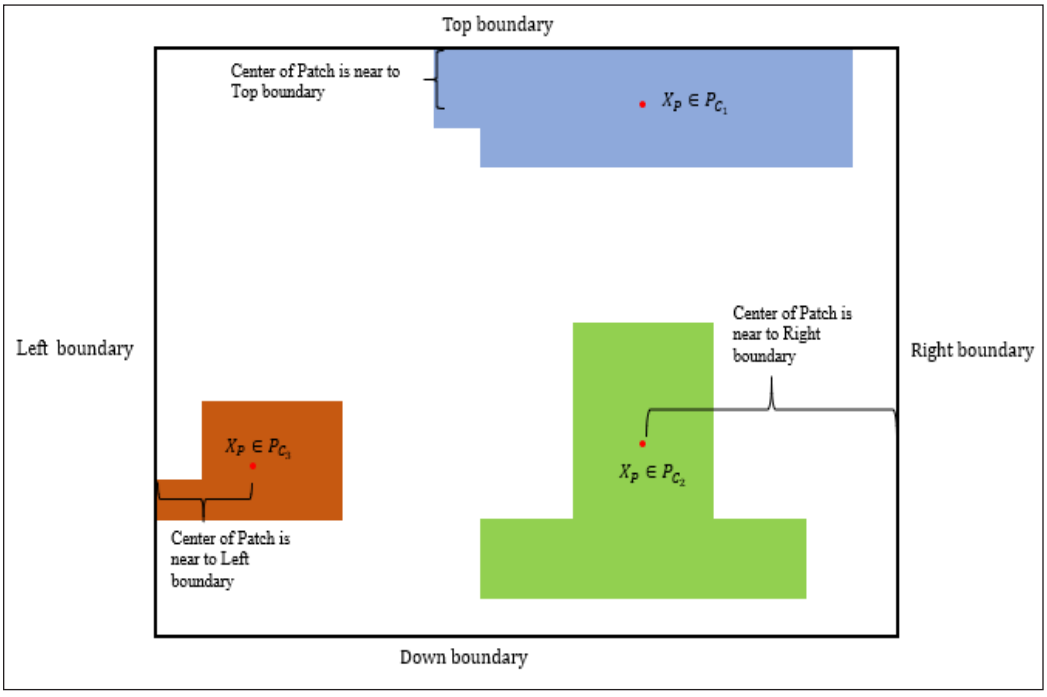


Figure 5. The boundary ratio of each patch towards its nearest boundary

Instead of taking all boundaries in the computation, the ratio of patches towards each boundary, b_r is computed at different boundary side, m which refers to top, down, left and right boundaries, respectively. The boundary ratio is calculated by taking the summation of mean appearance difference of patches, P and P_i towards a specific boundary m, over the distance of the patch mean location towards the boundary which is given by,

$$b_r(P) = max \sum_{i=1}^{N} \frac{\bar{q}(P, P_i) \in m}{(X_P | P_c \in m)} \qquad [9]$$

Then, the maximum of all ratio is obtained to know that the patch is near to which boundary. By referring to the concept of boundary connectivity from (Zhu et al., 2014), the patch on boundary length and the patch area are computed from the geodesic distance between two superpixels $g(p_i, p_j)$, where q is the appearance difference between the neighboring patches. It can be defined as,

$$g(p_i, p_j), = \min_{pi=p,p2,\ldots pn=q} \sum_{i=1}^{n-1} q(p_i, p_{i+1}) \quad [10]$$

The geodesic distance is a method that is used to compute the shortest path between the patch on the graph (Wei et al., 2012). Thus the patch area, a and patch length at the boundary, l are computed by,

$$a(P) = \sum_{i=1}^{n} e^{\left(\frac{-g^2(p,p_i)}{2\tau^2}\right)} = H(P, P_i \Bigg) \quad [11]$$

$$l(P) = \sum_{i=1}^{N} H(P, P_i).\delta(P_i \in m) \quad [12]$$

The value of τ is set to 10 as being experimented by Zhu et al (2014) model. $H(P,P_i)$ is the spanning area of the superpixel patch. The length at the boundary, $l(P)$ is defined from the spanning area with respect to the image boundary, m. The constant δ is set to 1 for the patch superpixel at the boundary and 0 otherwise. The new boundary connectivity, $b(P)$ are the multiplication of the patch length, $l_m(P)$ with the patch ratio, $b_r(P)$, the patch force interaction, $F_m(P)$ and the inverse square root of patch area, $a(P)$, where m is the measurement towards the image boundary. The patch force interaction, $F_m(P)$ *is* taken from the equation in [7]. It is given by,

$$b(P) = \frac{l_m(P)}{\sqrt{a(P)}} \cdot b_r(P) \cdot F_m(P) \quad [13]$$

This is the major contribution of this model where the terms $l_m(P)$, $b_r(P)$, $F_m(P)$, and $a(P)$ are the elements of BCA where the component contrast and spatial location are considered explicitly in the geodesic distance computation in equation [10]. The combination in BCA produces a new definition for boundary connectivity that is able to diminish the high contrast non-salient patch that is highly connected to the boundary since the measurement towards the boundary has been strengthened by combination of the features in BCA.

Background Measurement

The core of the background measure relies on the contrast computation where a reliable measurement will contribute to precise salient detection. Many previous models use the region contrast with respect to its neighbor that acts as saliency cue (Cheng et al., 2011; Perazzi & Kr, 2012; Yan et al., 2013). Generally, the weight contrast, *u(P)* from (Zhu et al., 2014) is the summation of contrast difference, *q* between patches *p* and p_i of the superpixel, multiplied with the distance weight, *w* on the superpixel patch. It can be written as,

$$u(P) = \sum_{i=1}^{N} q(P, P_i) \cdot w(P, P_i) \qquad [14]$$

With $\beta = 0.25$ and $z(P,P_i)$ as the spatial distance between patch *P* and P_i, the distance weight, *w* is obtained by,

$$w(P, P_i) = \exp\left(\frac{-z(P, P_i)}{2\beta^2}\right).z(P, P_i\right) \qquad [15]$$

Zhu et al. (2014) has extended the background weight contrast, d_i which includes the boundary connectivity, *b(P)* as in equation [9] and μ is set to 1. Thus it can be defined as,

$$d_i(P) = 1 - \exp\left(\frac{-b^2(P)}{2\mu^2}\right) \qquad [16]$$

With this new definition, the re-defined weight contrast of each patch *u(P)* based on equation [16] is,

$$u(P) = \sum_{i=1}^{N} q(P, P_i) w(P, P_i) d_i(P) \qquad [17]$$

To have an optimum measure on the background contrast, the new boundary connectivity measure, *b(P)* in equation [16] is used. Thus, the new optimum background contrast can be defined as,

$$d_i(P) = 1 - \exp\left(\frac{-b^2(P)}{2\sigma_b^2}\right).z(P, P_i\right) \qquad [18]$$

Saliency Map Computation

The optimum saliency map computation, *M* measure is adopted from Zhu et al., (2014) model where they combined the terms background, d_i foreground, v_i and smoothness, t_{ij} in a single computation as,

$$M = \sum_{i=1}^{N} d_i(P)s_i^2 + \sum_{i=1}^{N} v_i(P)(s_i - 1)^2 + \sum_{i=1}^{N} t_{ij}(s_i - s_j)^2 \quad [19]$$

The term foreground, v_i is computed by subtracting 1 from the background, d_i. The combination of the three terms are able to put the superpixel of large background probability to the small value of s_i (close to 0), and the patch with high foreground probability to the large value of s_i (close to 1). In addition, the smoothness term promotes the enhancement of the saliency values. The smoothness term t_{ij} can be obtained by $w_{ij} = \exp\left(\frac{-d_c(p_i,p_j)}{2\sigma_{clr}^2}\right) + \mu$. The constant μ is set to 0.1 that is used to regularize the cluttered image regions.

EXPERIMENTS

The evaluation of the proposed model was tested on the ASD (Achanta et al., 2009) and ECSSD (Li et al., 2013) datasets. Both datasets consisted of 1000 images with the manually segmented ground truth images for the salient object detection model evaluation. The images in ASD datasets were considered simple containing single salient object with clean background while the ECSSD datasets contained semantically meaningful images and were quite complex.

Using these verified datasets, the performance of the proposed model are measured and compared with 4 state-of-the-art salient detection models which are, *Robust Background Detection*, RBD (Zhu et al., 2014), *Saliency Filter*, SF (Perazzi & Kr, 2012), *Geodesic Saliency*, GS (Wei et al., 2012) and *Manifold Ranking*, MR (Yang et al., 2013). Apart from that, the model performance was also tested on the close contrast images that were manually selected from the ASD dataset. This type of image was tested to evaluate the model detection performance as this type of image had become part of unsolved issues in the study of salient detection (Borji et al., 2014).

Out of all 4 models, the RBD, is listed (Borji et al., 2015) to be among the top models for salient object detection based on exhaustive comparison study done. In evaluating the proposed model, the precision-recall (PR), Receiver-Operating Characteristic (ROC), F-Measure and Mean Absolute Error (MAE) were used, where these measurements are universally-agreed and the standard measurement that can be used for evaluating the salient object detection model (Borji et al., 2015).

Models Comparison on Detection

The precision, γ is the ratio of salient pixels accurately detected and allocated as,

$$\gamma = \frac{B \cap G}{B} \quad [20]$$

On the other hand, recall is the ratio, η of detected salient pixel over its ground truth as,

$$\eta = \frac{B \cap G}{G} \qquad [21]$$

The detail of the measurement can be defined as in the equations [20] and [21], where B is the saliency detected converted in binary mask, based on the saliency image map M, and G is the image ground truth. On the other hand, the Receiver Operating Characteristic (ROC) curve reports the relationship of True Positive Rate (TPR) , ρ and False Positive Rate (FPR), φ of the saliency map S_b and the ground truth, G. They can be finely written as,

$$\rho = \frac{|S_b \cap G|}{|G|} \qquad [22]$$

$$\varphi = \frac{|S_b \cap G|}{|S_b \cap G| + |\overline{S_b} \cap \bar{G}|} \qquad [23]$$

For simplification, comparison is made based on the saliency map S_b that is normalized into [0,255] range. The curve is plotted by taking the average of computed precision and recall on the dataset. The F-measure, F_a and mean Absolute Error (MAE), ε are defined as,

$$F_a = \frac{(1+\alpha)\rho \times \eta}{\alpha \times \rho + \eta} \qquad [24]$$

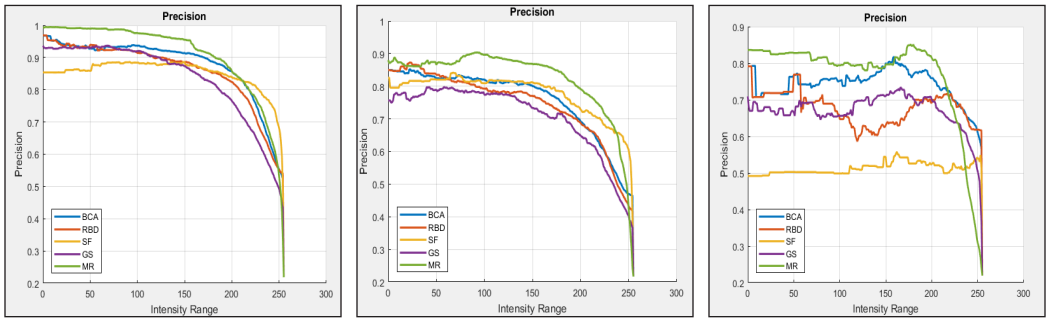
$$\varepsilon = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left\| S_b(x,y) - G(x-y) \right\| \qquad [25]$$

Where W and H is the width and height of the image respectively. The constant , α for F_a measurement is set to 0.3 to raise the importance of precision as suggested by Achanta et al. (2009).

Detection Evaluation on Datasets

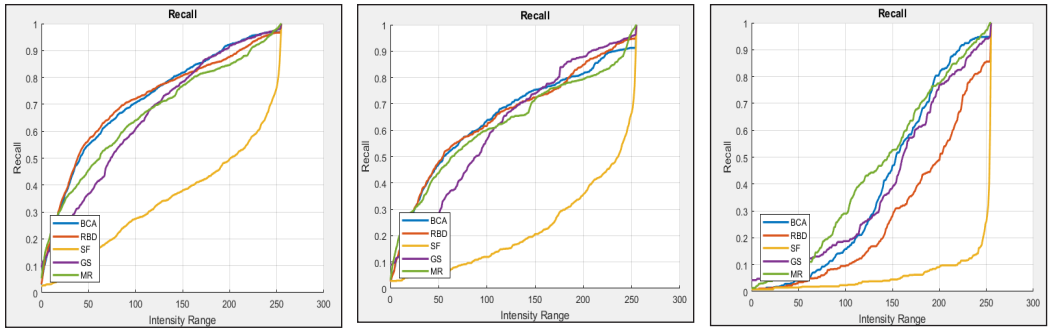
All performance comparisons were put on 3 tests using the ASD and ECSSD datasets as well as the close contrast images collected from the ASD dataset. The performances were compared side by side based on precision and recall with intensity range, precision over recall, and ROC curves. Histograms were used to compare the performances on MAE, precision, recall and F-measure. Samples of tested images, ground truth and saliency were compared as well. In each result comparison, the proposed model is labeled as BCA as the other 4 state-of-the-art models are labeled as RBD, SF, GS and MR. The details of the results analysis are discussed in the next section.

Precision and Recall over the Intensity Range. In these comparisons, the obtained saliency maps are normalized to [0,255] to generate the binary masks. From the curves in Figure 6, it can be observed that the proposed model has high precision compared to other models with MR as the highest. However, the model's precision is low on ECSSD dataset due to its difficult images. The high precision indicates that the saliency map obtained had lower fall-out rate. The curves obtained for close contrast images were quite uneven since there were small number of those types of images in the ASD data set. Referring to the recall curve over the intensity range in Figure 7, the proposed model has the highest and consistent recall value compared to MR on ASD and ECSSD datasets. It indicates that the saliency map obtained from BCA has less miss rate. It can also be observed that the model SF produces an abnormality curve compared to others. It is because its global contrast assumption would result in noisy segmentation as the image has low contrast due to lighting or similar foreground background appearance. Thus, the resulted saliency map would be in high false negative pixels. The min-cut segmentation could be the solution of the issue as discussed in their research paper (Perazzi & Kr, 2012).



(a) (b) (c)

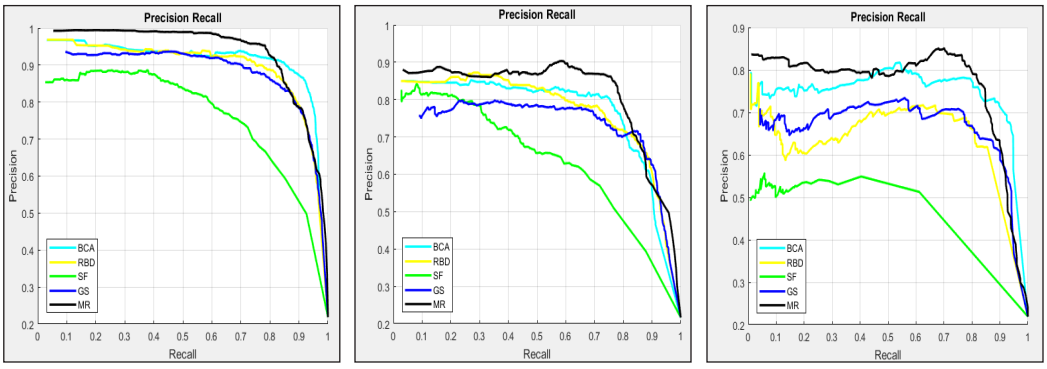
Figure 6. Precision over intensity range on the BCA comparison with other state-of-the-art model using: (a) ASD dataset; (b) ECSSD dataset; and (c) close contrast images collected from ASD dataset.



(a) (b) (c)

Figure 7. Recall over intensity range on the BCA comparison with other state-of-the-art model using: (a) ASD dataset; (b) ECSSD dataset; and (c) close contrast images collected from ASD dataset.

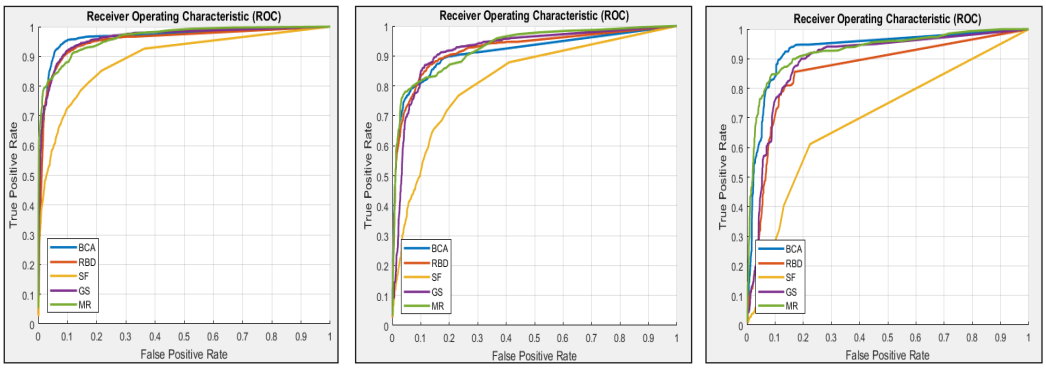
Precision over Recall Curve. Comparing both precision and recall of BCA with other state-of-the-art models in Figure 8, BCA had the second highest precision after MR on the 3 testing datasets. However, BCA had the highest recall when being tested on ASD and close contrast images. The higher and closer the curve to the right, the lesser the miss rate and false alarm of the obtained saliency map. It indicates that the model is able to produce correct salient detection region with lesser false salient region.



(a) (b) (c)

Figure 8. PR curve on the BCA comparison with other 4 state-of-the-art models using: (a) ASD dataset; and (b) ECSSD dataset; and (c) close contrast images collected from ASD dataset.

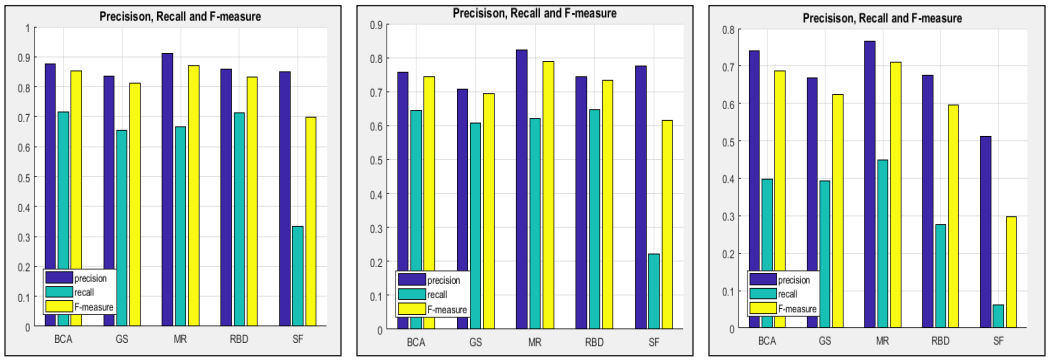
ROC Curve. In Figure 9, the proposed model BCA had the highest True Positive Rate when being tested with the ASD and close contrast images. It indicates that the salient detection obtained from the model had the highest rate when being compared to its ground truth. The lower curve indicates that the saliency map obtained consisted of many fall-out rates where many patches were falsely detected as the salient region.



(a) (b) (c)

Figure 9. PR curve on the BCA comparison with other state-of-the-art models using: (a) ASD dataset; (b) ECSSD dataset; and (c) close contrast images collected from ASD dataset.

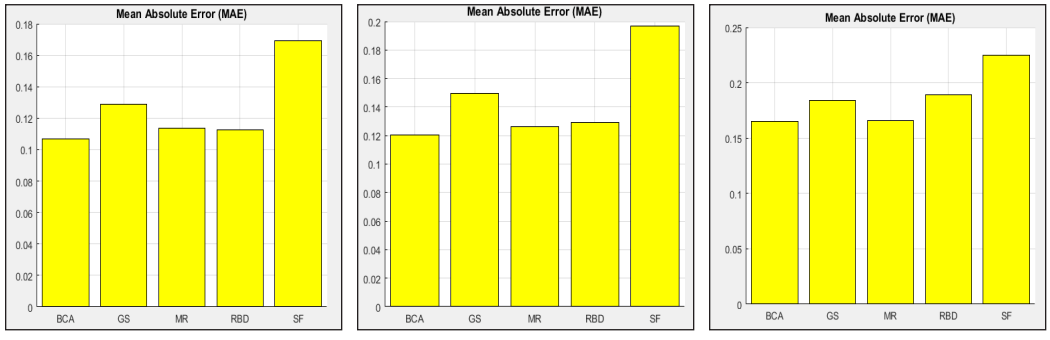
Precision, Recall and F-Measure. In the comparison made in Figure 10, the average of precision, recall and F-measure were obtained from the 3 testing images. From this histogram, the proposed model BCA had the second highest value of all after MR for all tests. Higher F-measure indicates that saliency map obtained has high value for both precision and recall.



(a) (b) (c)

Figure 10. Precision, Recall and F Measure comparison with the FBR as proposed model, using (a) ASD dataset,(b) ECSSD dataset and (c) close contrast images collected from ASD dataset.

Mean Absolute Error (MAE). Table 1 is the value of computed MAE for different test images. When compared with the RBD as a similar base algorithm assumption, BCA has been able to reduce the MAE value in the range of 3.0% to 9.4%. When comparing the MAE of BCA against other models as in Figure 11, the BCA gives the lowest MAE for all 3 test images. This indicates that the saliency map obtained is close to the ground truth. The high value of MAE shows that there are many patches on the saliency map when compared to the ground truth. Thus, it can be said that the BCA is able to produce clean background on its detection.



(a) (b) (c)

Figure 11. MAE value comparison with the proposed model, FBR using the (a) MSRA10K dataset, (b) ECSSD dataset and (c) close contrast images collected from ASD dataset.

Table 1
MAE comparison for the three testing datasets

No	Test Images Set	BCA	GS	MR	RBD	SF
1	ASD	0.10486	0.12905	0.11361	0.10807	0.16910
2	ECSSD	0.11351	0.14979	0.12648	0.12124	0.19656
3	Close Contrast Images	0.18131	0.18421	0.16610	0.20013	0.22519

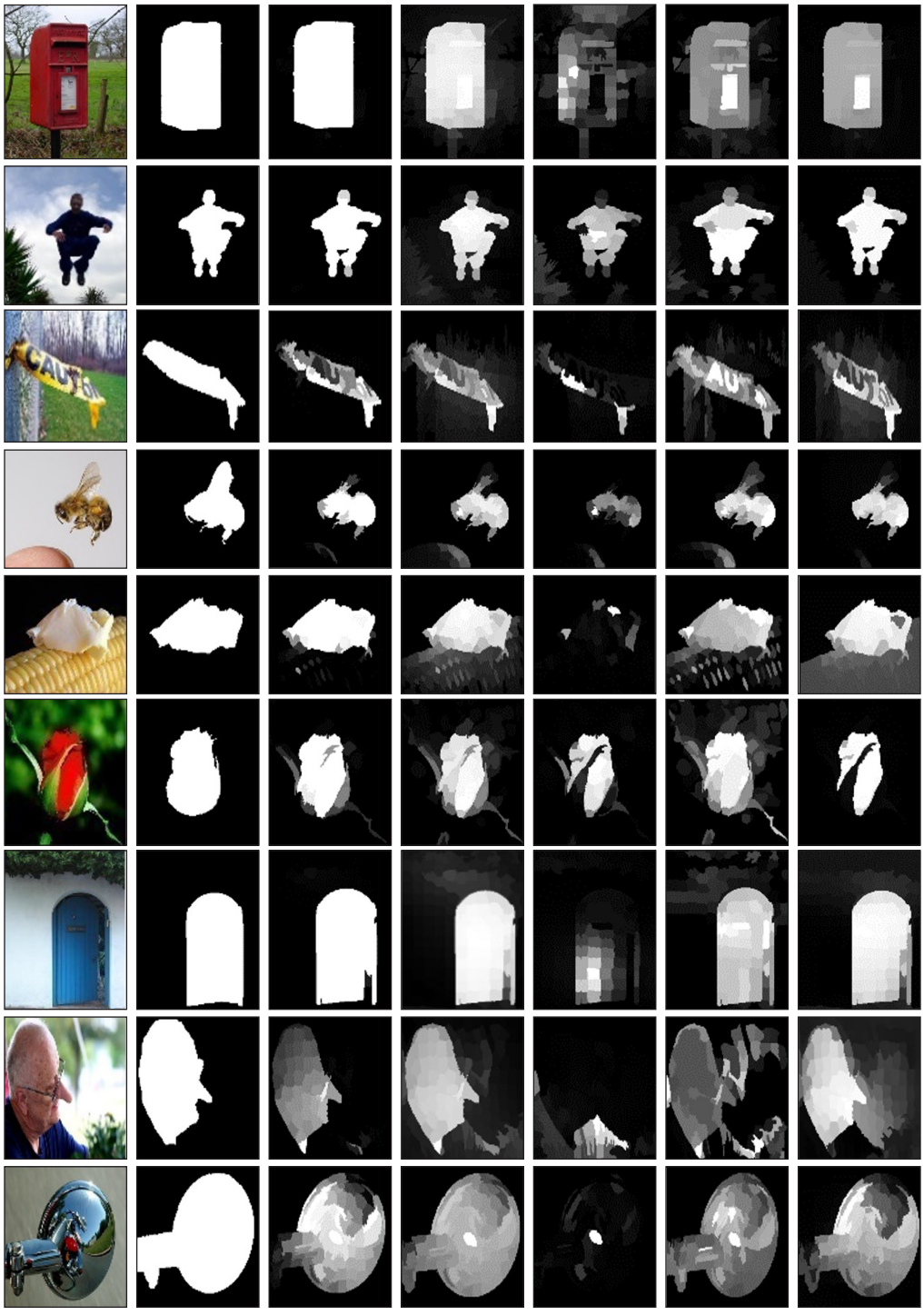
Execution Time of Salient Object Detection Comparison. The intension of salient detection is to reduce the processing time of some semantic processing purpose as the output of the detection is able to mask out the image background and only left with the most prominent object in the image. Therefore, the evaluation of detection execution time is very important and needs to be put as the result comparison. The codes of the other 4 models were accessible online as in MATLAB file (m file). They were run and compared using the MATLAB R2017b version software on a machine with the Intel i7-7500U CPU 2.70GHz, 2901 as the processor and 4G of RAM. The execution time comparison can be referred to Table 2.

Table 2
Execution time (s) for 3 different test images

No	Test Images Set	BCA	GS	MR	RBD	SF
1	ASD	0.1555	0.1315	0.1360	0.1503	0.1401
2	ECSSD	0.1466	0.1220	0.1295	0.1435	0.1299
3	Close Contrast Images	0.1579	0.1437	0.1444	0.1568	0.1535

The execution time of BCA and RBD were quite consistent on all type of images. However, BCA had a slightly longer execution time in the range of 3.45% to 0.7% higher compared to RBD. This is due to additional computation that were considered in the model.

Input Images, Ground Truth and Salient Detection Comparison. The visual comparison of the saliency map obtained from BCA with other models, the BCA was able to produce clean background, resulting in minimum white patches as shown in Figures 12, 13 and 14. The results are consistent on the 3 types of test images. These are the valid reasons the BCA is able to produce the lowest MAE in previous comparison. Even though the saliency maps obtained for close contrast images do not really match the ground truth, they still produce better detections compared to the other models.



(a) (b) (c) (d) (e) (f) (g)

Figure 12. Comparison results for salient object detection images from the ASD dataset. (a) Original image, (b) Ground truth, (c) BCA (proposed), (d) RBD, (e) SF, (f) GS, (g) MR

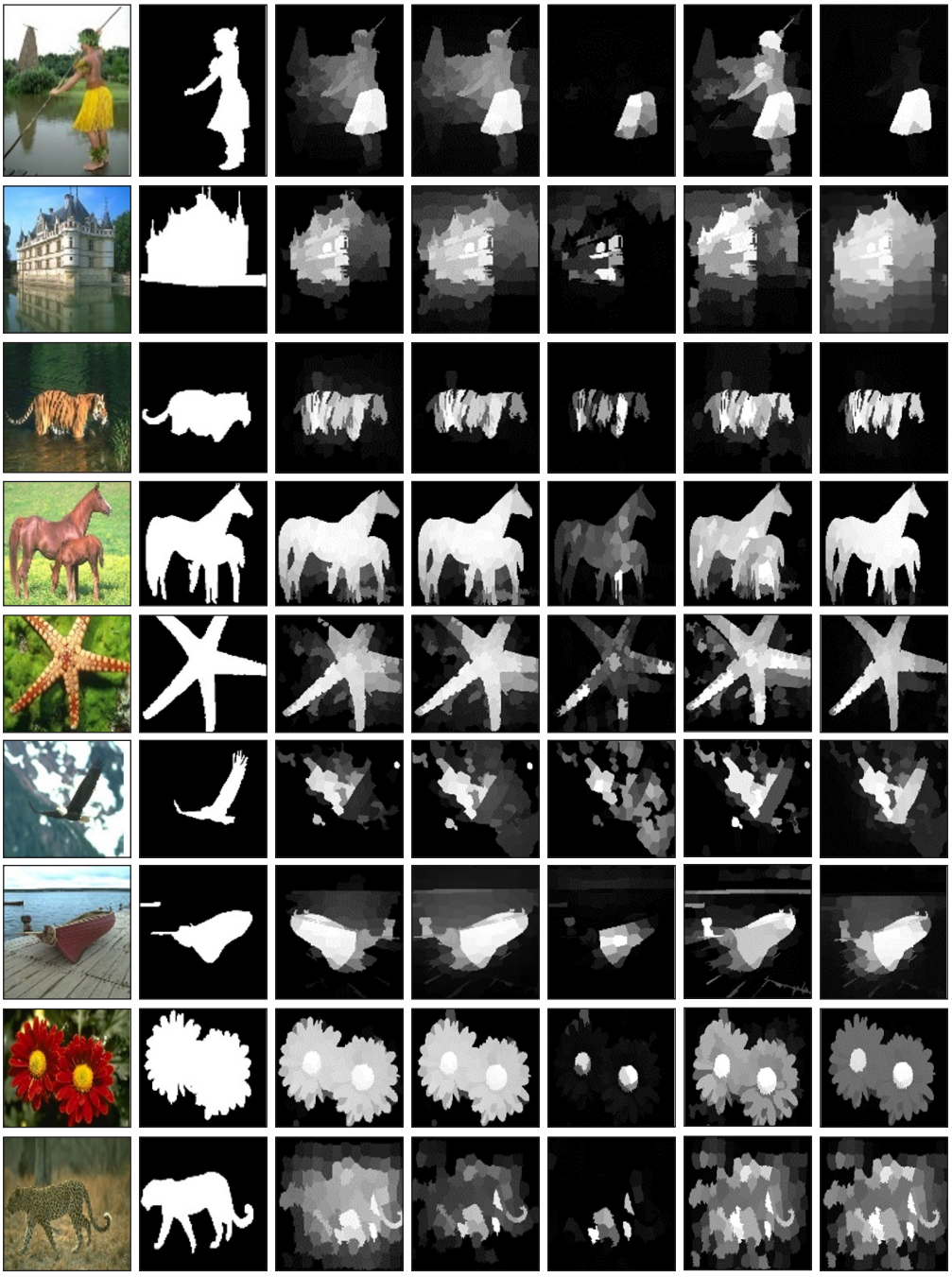


Figure 13. Comparison results for salient object detection images from the ASD dataset. (a) Original image, (b) Ground truth, (c) BCA (proposed), (d) RBD, (e) SF, (f) GS, (g) MR

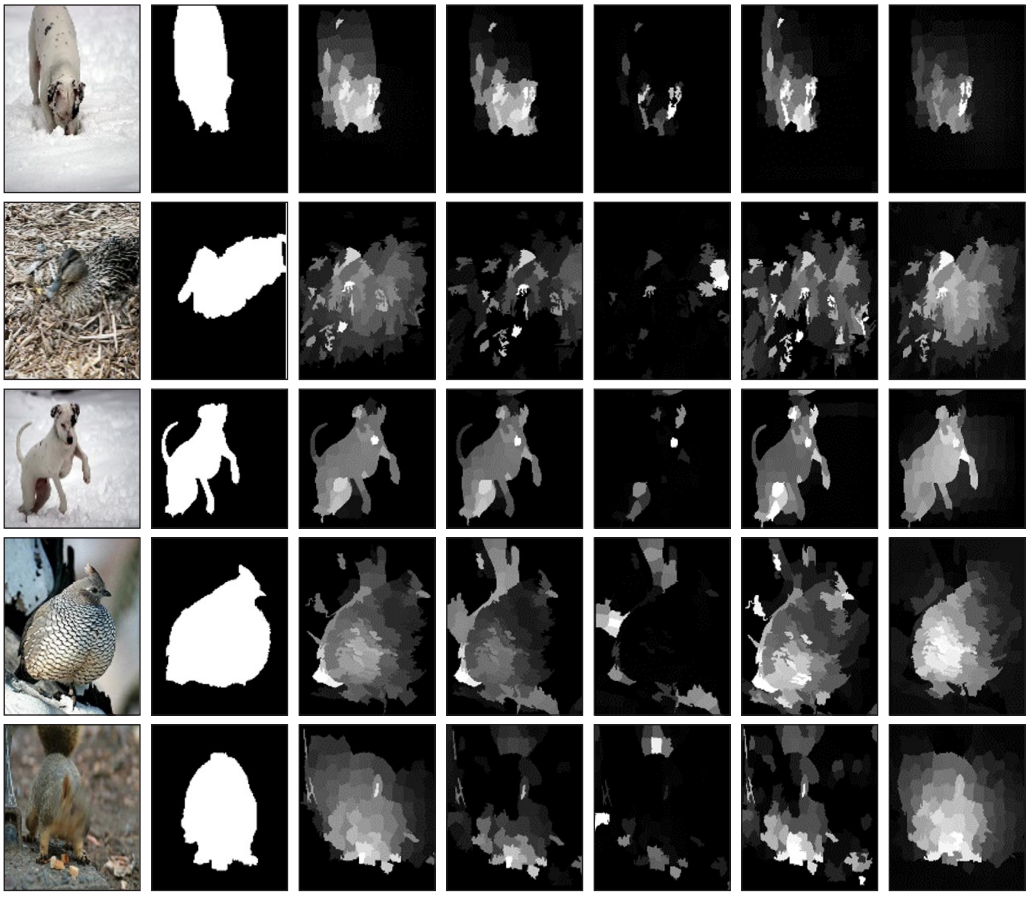


Figure 14. Comparison results for salient object detection images on close contrast image. (a) Original image, (b) Ground truth, (c) BCA (proposed), (d) RBD, (e) SF, (f) GS, (g) MR

CONCLUSION

This paper has presented an improved salient object detection model using combination features known as BCA on its saliency computation. The model has overcome the problem where high contrast background patches are falsely assigned as the foreground. On the other hand, the detection on close contrast image has been improved with lesser false positive rate. With the lowest MAE obtained, it indicates clear background with less white patches successfully obtained. As the lowest MAE is obtained for the close contrast image, it indicates that the proposed model is suitable to be applied for the salient object detection on that kind of image. The integrated features consisting of contrast, spatial location, force interaction and boundary ratio have been able to produce a novel equation of boundary connectivity. This new consideration has resulted in improved saliency map and has outperformed the referred model, RBD in most of the performance comparisons. Despite

that, the model has embedded the force interaction feature that has not previously been applied in any existing salient detection models. Therefore, the combination features in BCA has been successful in producing better salient object detection by its background measure computation. Despite of the improvement on the result of detections, the BCA execution time requires a longer duration due to the additional feature extraction process. This weakness could become the component that can be improved in the future. Furthermore, the learning technique could be integrated with the model to improve detection especially for close contrast images.

ACKNOWLEDGEMENT

We would like to thank to the anonymous reviewers for their constructive comments and opinions that helped us in producing significant improvement on this research paper.

REFERENCES

Achanta, R., Estrada, F., Wils, P., & Süsstrunk, S. (2008). Salient region detection and segmentation. In A. Gasteratos, M. Vincze, & J. K. Tsotsos (Eds.), *Computer vision systems* (pp. 66-75). Heidelberg, Germany: Springer.

Achanta, R., Hemami, S., Estrada, F., Sabine, S., & Epfl, D. L. (2009, June 20-25). Frequency-tuned salient region detection. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1597-1604). Miami Beach, Florida.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Susstrunk, S. (2012). SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2274-2282.

Ahn, E., Kim, J., Bi, L., Kumar, A., Li, C., Fulham, M., & Feng, D. D. (2017). Saliency-based lesion segmentation via background detection in dermoscopic images. *IEEE Journal of Biomedical and Health Informatics*, *21*(6), 1685-1693.

Baluja, S. (1997). Expectation-based selective attention for visual monitoring and control of a robot vehicle. *Robotics and Autonomous Systems*, *22*(3-4), 329-344.

Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2014). Salient object detection: a survey. *Computattional Visual Media, 5*(2), 117-150.

Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2015). salient object detection: a benchmark. *Image Processing, IEEE Transactions On*, *24*(12), 5706-5722.

Bucha, V., Uchida, S., & Ablameyko, S. (2006). Interactive road extraction with pixel force fields. *Proceedings-International Conference on Pattern Recognition*, *4*(i), 829-832.

Bucha, V., Uchida, S., & Ablameyko, S. (2007). Image pixel force fields and their application for color map vectorisation. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, *2*, 1228-1232.

Cheng, M., Zhang, G., Mitra, N. J., Huang, X., & Hu, S. (2011). cn-global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 2011*, *37*(3), 409-416.

Chou, Y., & Su, P. C. (2016, May 27-29). Efficient content-based cropping using visual saliency and SLIC for image retargeting. In *2016 IEEE International Conference on Consumer Electronics* (pp. 1-2). Nantou, Taiwan.

Ghayoor, A., Sadri, A., & Shirazi, A. A. B. (2010, May 26-28). Image registration method based on physical forces for color images. In *2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation* (pp. 350-353). Bornea, Malaysia.

Guo, C., & Zhang, L. (2010). A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Transactions on Image Processing*, *19*(1), 185-198.

Hu, P., Wang, W., Zhang, C., & Lu, K. (2016). Detecting salient objects via color and texture compactness hypotheses. *IEEE Transactions on Image Processing*, *25*(10), 4653-4664.

Hurley, D. J., Nixon, M. S., & Carter, J. N. (1999). Force field energy functionals for image feature extraction. *Image and Vision Computing*, *20*(5-6), 311-317.

Itti, L. (2004). Automatic foveation for video compression using a neurobiological modelof visual attention. *IEEE Transactions on Image Processing, 13*(10), 1304-1318.

Itti, L., Koch, C., & Niebur, E. (2001). A Model of Saliency-Based Viusla Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(6), 674-680.

Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., & Li, S. (2011, August 29-September 2). Automatic salient object segmentation based on context and shape prior. In *Proceeding of British Machine Vision Conference* (pp. 1-12). University of Dundee, Scotland.

Jiang, L., Koch, A., & Zell, A. (2015). Salient regions detection for indoor robots using RGB-D data. *Proceedings-IEEE International Conference on Robotics and Automation*, *2015–June*(June), 1323-1328.

Jing, H., He, X., Han, Q., & Niu, X. (2014). Neurocomputing. *Neurocomputing*, *124*, 57-62.

Katsuki, F., & Constantinidis, C. (2014). Bottom-Up and Top-Down Attention : Different Processes and Overlapping Neural Systems. *The Neuroscientist, 20*(5), 509-521.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*(4), 219-227.

Li, C., Huo, H., & Fang, T. (2016, July 11-12). Oil depots detection from high resolution remote sensing images. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)* (pp. 285-288). Shanghai, China.

Li, J., Member, S., & Levine, M. D. (2013). Visual saliency based on scale-space analysis in the frequency domain. *IEEE transactions on pattern analysis and machine intelligence, 35*(4), 996-1010.

Li, S., Xu, M., Ren, Y., & Wang, Z. (2017). Closed-form optimization on saliency-guided image compression for HEVC-MSP. *IEEE Transactions on Multimedia*, *20*(1), 155-170.

Liu, D., Cong, M., Du, Y., & Gao, S. (2013, December 12-14). Robot behavior selection using salient landmarks and object-based attention. In *2013 IEEE International Conference on Robotics and Biomimetics, ROBIO 2013* (pp. 1101-1106). Shenzhen, China.

Liu, Y., Zhang, J., Tjondronegoro, D., Geva, S., & Li, Z. (2008, November 26-28). An improved image segmentation algorithm for salient object detection. In *2008 23rd International Conference Image and Vision Computing New Zealand* (pp. 1-6). Christchurch, New Zealand.

Luo, Y., Wang, P., Li, W., Shang, X., & Qiao, H. (2016, June 12-15). Salient object detection based on boundary contrast with regularized manifold ranking. In *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)* (pp. 2074-2079). Guilin, China.

Luz, G., Ascenso, J., Brites, C., & Pereira, F. (2017, October 16-18). Saliency-driven omnidirectional imaging adaptive coding: Modeling and assessment. In *2017 IEEE 19th International Workshop On Multimedia Signal Processing (MMSP)* (pp. 1-6). Luton, UK.

Manke, R., & Jalal, A. S. (2016, December 3-4). Salient region detection using fusion of image contrast and boundary information. In *2016 11th International Conference on Industrial and Information Systems (ICIIS)* (pp. 199-202). Roorkee, India.

Milanese, R., Gil, S., & Pun, T. (1995). Attentive mechanisms for dynamic and static scene analysis. *Optical Engineering*, *34*(8), 2428-2434.

Niu, Y. (2018). Salient object segmentation based on superpixel and background connectivity prior. *IEEE Access*, *6*, 56170-56183.

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of vision, 13*(3), 16-16.

Perazzi, F., & Kr, P. (2012, June 16-21). Saliency filters Contrast based filtering for salient region detection. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 733-740). Providence, RI, USA.

Pritch, Y., Sorkine-hornung, A., Mangold, S., & Gross, T. (2013, September 15-18). Content-aware compression using saliency-driven image retargeting. In *2013 IEEE International Conference on Image Processing* (pp. 1845-1849). Melbourne, VIC, Australia.

Srivastava, S., Mukherjee, P., & Lall, B. (2016, June 12-15). Adaptive Image Compression Using Saliency and KAZE Features. In *2016 International Conference on Signal Processing and Communications (SPCOM)* (pp. 1-5). Bangalore, India.

Tang, L., Li, H., Wu, Q., & Ngan, K. N. (2018). Boundary guided optimization framework for saliency refinement. *IEEE Signal Processing Letters*, *9908*(c), 1-1.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136.

Wang, J., Zhang, K., Madani, K., Sabourin, C., & Zhang, J. (2015, July 21-23). Salient foreground object detection based on sparse reconstruction for artificial awareness. In *2015 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO)* (Vol. 2, pp. 430-437). Colmar, France.

Wei, Y., Wen, F., Zhu, W., & Sun, J. (2012). Geodesic saliency using background priors. In *European conference on computer vision* (pp. 29-42). Heidelberg, Germany: Springer.

Woodbridge, J., Lan, M., Sarrafzadeh, M., & Bui, A. (2011, July 26-29). Salient segmentation of medical time series signals. In *2011 First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB 2011)* (pp. 1-8). San Jose, CA, USA.

Wu, P. H., Chen, C. C., Ding, J. J., Hsu, C. Y., & Huang, Y. W. (2013). Salient region detection improved by principle component analysis and boundary information. *Ieee Transactions on Image Processing*, *22*(9), 3614-3624.

Xiao, R., & Yeh, M. (2017, June 12-14). A new method with saliency detection for image quality assessment. In *2017 IEEE International Conference on Consumer Electronics* (pp. 75-76). Taipei, Taiwan.

Xiuli, T., Shuhan, C., & Xuelong, H. (2017, October 20-22). Salient object detection by combing eye fixation prediction and semantic segmentation. In *2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI)* (pp. 407-411). Yangzhou, China.

Yan, Q., Xu, L., Shi, J., & Jia, J. (2013, June 23-28). Hierarchical saliency detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1155-1162). Portland, OR, USA.

Yang, B., Zhang, X., Chen, L., Yang, H., & Gao, Z. (2017). Edge guided salient object detection. *Neurocomputing*, *221*(March 2016), 60-71.

Yang, C., Zhang, L., Lu, H., Ruan, X., & Yang, M. H. (2013, June 23-28). Saliency detection via graph-based manifold ranking. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3166-3173). Portland, Oregon.

Zhang, H., Goldman, S. A., & Louis, S. (2006, October 8-11). Image Segmentation Using Salient Points-Based Object Templates. In *2006 International Conference on Image Processing, Atlanta* (pp. 765-768). Atlanta, GA, USA.

Zhang, L., Lv, X., Chen, J., & Zhang, L. (2016, July 10-15). Region of interest detection based on salient feature clustering for remote sensing images. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 88-91). Beijing, China.

Zhang, Q., Lin, J., & Li, X. (2016). Salient object detection via structure extraction and region contrast. *Journal of Information Science and Engineering*, *32*(6), 1435-1454.

Zhang, Q., Lin, J., Tao, Y., Li, W., & Shi, Y. (2017). Neurocomputing salient object detection via color and texture cues. *Neurocomputing*, *243*, 35-48.

Zhou, L., Chen, Y., & Yang, Z. (2017). Foreground and background propagation based salient region detection. *2016 9th International Symposium on Computational Intelligence and Design (ISCID)*, *1*, 110-113.

Zhu, W., Liang, S., Wei, Y., & Sun, J. (2014, June 23-28). Saliency optimization from robust background detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2814-2821). Columbus, Ohio.