Saliency detection is one of the extraordinary abilities of the human visual system (HVS), and also provides a powerful tool for predicting where humans tend to focus in the free-viewing process. In this paper, we propose a novel method for computing image saliency. At first, an image is subject to $L_0$ smoothing to characterize its fundamental constituents while diminishing insignificant details. Distance-transform-based saliency detection is then applied to the smoothed image, to extract the general salient regions and form a rough saliency map. Next, the segmentation information generated by normalized cuts is used to improve the saliency detection performance by averaging the saliency values in each segmented block. Finally, we employ the center-bias mechanism to further improve the saliency model. The proposed method is compared with six existing saliency models, and achieves the best performance in terms of the area under the ROC curve (AUC).

Index Terms—Saliency detection, $L_0$ smoothing filter, distance transform, center bias, image segmentation

1. INTRODUCTION

Visual saliency is a selection mechanism of the human visual system (HVS) that quickly focuses on general salient regions. To detect salient regions, many computational models have attempted to simulate the bottom-up and top-down mechanisms proposed by [1] and [2]. In general, whether a region is salient or not may depend on its uniqueness, unpredictability or even rarity, and is highly related to the color, gradient, edge, and boundaries [3] in the image. Saliency detection is useful in many applications, such as image resizing [4], image retrieval [5], and object segmentation [6].

Over the past decades, many computational models have been proposed for computing visual saliency. Among these models, the most famous one was proposed by Itti et al. [7]. The algorithm is based on the center-surround contrast, which is implemented as the difference between the fine and coarse scales. Bruce and Tsotsos [8] proposed a bottom-up model based on the principle of maximizing information sampled from a scene through a probabilistic framework. Harel et al. [9] extended [7] into a graph theory framework namely Graph-based visual saliency. Hou and Zhang [10] presented the first spectral approach for visual saliency detection based on spectral residual, defined as the difference between the perceived log-spectrum and the characteristic log-spectrum of natural images. Achanta et al. [11] proposed using the difference of Gaussian (DoG) filtering to eliminate redundant information, and then generated full-resolution saliency maps with salient objects' boundaries well defined. Hou et al. [12] proposed the concept of an image signature, which is defined as the sign function of the discrete Cosine transform of the image, to predict human-fixation points.

To enhance the saliency detection performance, several studies have introduced image segmentation result into the algorithm. Cheng et al. [13] proposed a model based on regional contrast, which evaluates global contrast differences and spatial coherence. Wu et al. [3] combined principal component analysis with boundary information. Inspired by these methods, we propose to use segmentation result to enhance the performance of saliency detection.

In addition to the computational techniques developed for more advanced saliency detection models, several studies have also observed that subjects' visual attention is often biased towards the center of an image [14]. Tseng et al. [15] also found that photographers tend to place the objects-of-interest near the center of an image.

The main contribution of this paper is to establish a saliency framework that combines the distance transform with image segmentation, and to improve the model with $L_0$ smoothing [16], and the center-bias mechanism [14]. In our method, an image is first subject to $L_0$ smoothing to retain the important salient information and to eliminate noises. The general salient regions are then detected using the distance-transform-based model. Next, the center-bias mechanism is applied to modulate saliency values according to pixels' position. Finally, the saliency values in each segmented block are averaged to form a single value to preserve the outlines of salient objects.

The rest of the paper is organized as follows. Section 2 describes the proposed method in detail. Experimental results and an evaluation of the proposed method are given in Section 3. Finally, Section 4 presents the conclusion.
2. PROPOSED FRAMEWORK

2.1. L₀ smoothing filtering

Most of the existing saliency models do not pay much attention to the smoothing process. However, noise is universal in all images, resulting in high-contrast regions from a local or a global perspective, e.g. the footprints in image (a) and the paragraphs in image (b) in the first row of Fig. 1. Xu et al. [16] proposed a new image-editing method for sharpening major edges by increasing the steepness of the transition while eliminating a manageable degree of low-amplitude structures, as can be seen in the second row of Fig. 1. The main idea of the L₀ smoothing filter is to confine the number of intensity changes among neighboring pixels, which links mathematically to the L₀ norm for information-sparsity pursuit. For an input image I, and the computed result S, the gradient of each pixel p is calculated as the color difference between neighboring pixels in the x and y directions. The number of pixels whose gradient magnitudes are not zero is defined as:

\[ C(S) = \#\{ p \mid \partial_x S_p + | \partial_y S_p | \neq 0 \}, \]  

where \( \partial_x S_p \) and \( \partial_y S_p \) are the gradients for pixel p along the x and the y direction, respectively. Then, with this definition, S is obtained by solving:

\[ \min_S \left\{ \sum_p (S_p - I_p^t) + \lambda \cdot C(S) \right\}, \]  

where \( \lambda \) is a smoothing parameter to control the significance of \( C(S) \). For color images, the gradient magnitude is practically defined as the sum of gradient magnitudes in the r, g, and b channels.

![Fig. 1](image)

Fig. 1. From top to bottom: sample images from MSRA dataset [17], the smoothed images by L₀ smoothing, the saliency maps from the original images in the first row, the saliency maps from the smoothed images in the second row.

An effective smoothing at the preprocessing stage can benefit the next process by both accentuating the true salient objects – thus enhancing the saliency detection accuracy, and eliminating insignificant false salient objects – thus avoiding over-segmentation. As shown in the second row of Fig. 1, the L₀ smoothing filter can remove trivial information in the background and emphasize the most salient objects. It is clear that this approach of saliency detection can produce stronger outputs for salient objects with smoothed images than with original images, as shown in the third and fourth rows of Fig. 1.

2.2. Distance transform based saliency detection with multiple scales

Rosin [18] attempted to use edge density as a measure of saliency, as an edge map is easy to compute and requires no parameters. However, the performance is greatly affected by the noise points and texture patterns. Since L₀ smoothing can eliminate irrelevant noise points to a large extent, distance transform can be a feasible approach to detecting salient regions. The basic idea of distance transform is calculating the distance to the nearest edge: the value of a particular pixel in the transform domain is its distance to the nearest edge. Hence, the larger distance a pixel is from its nearest edge, the higher its transform value is. Usually, salient regions will have a sharp intensity contrast, or color transition, so a salient region generally exhibits a high edge density. Since distance transform is a measure of edge intensity, this allows distance transform to be correlated with saliency detection. In order to generate a saliency map using distance transform, the first step is to apply edge operators, e.g. the Sobel or Canny operators, to produce an edge map. The edges are then subject to distance transform. However, the standard distance transform is defined on binary images rather than on gray-level or color images, therefore it is necessary to threshold images to form the binary image counterparts. We use threshold values starting from an intermediate value, to eliminate some false alarms. The final saliency map is obtained by simply complementing the saliency map or by reversing the sign in the thresholding, since a smaller value indicates more saliency, while a larger value indicates less saliency.

Although L₀ smoothing attenuates the noise to a great extent, it also reduces the details in salient objects. The consequence is that original salient objects become more homogeneous, and may be treated as non-salient regions. Hence, we propose to expand our algorithm into multiple scales, i.e. Gaussian pyramids, to eliminate the homogeneity tendency. We follow [19] in that the number of pyramid levels is defined as: \( n = \log_2(\min(w, h)/10) \), where \( w \) and \( h \) are the width and height of the image, respectively. Firstly, \( n \) spatial scales are created using the Gaussian pyramids, which are computed by progressively low-passing and down-sampling the images into half-sized counterparts. The images at different scales are then subject to Sobel edge detection.
detection, and are summed together to form the final edge map after image resizing. Finally, the scale-enhanced edge map is subject to distance transform to generate the final saliency map.

2.3. Applying the center-bias mechanism

According to [14], visual attention is often biased towards the center of an image. Therefore, pixels near the center should generally provide more saliency than those pixels far away from the center. In order to simulate the human visual system more accurately, the saliency model should consider this common phenomenon. Here, we propose to use the Gaussian function to simulate the center-bias mechanism due to its smooth transition. The center-bias function is defined as follows:

\[
C(x, y) = \exp\left(-\frac{(x-x_0)^2}{2\sigma^2} - \frac{(y-y_0)^2}{2\sigma^2}\right),
\]

where \(H\) and \(W\) are the height and width of the image, respectively. \((x_0, y_0)\) is image’s center; \(\sigma\) is the standard deviation of the Gaussian kernel. In this expression, we choose \(\sigma\) to be 0.2944 such that the attenuation rate at the four corners is 50%. The different standard deviations in the horizontal and vertical directions are used to guarantee the same attenuation rate in both directions. The saliency map after applying the center-bias mechanism is defined as:

\[
S(x, y) = SDT(x, y) \times C(x, y),
\]

where \(S(x, y)\) is the rough saliency map after applying the center-bias mechanism; \(SDT(x, y)\) is the saliency map after using the multiple-scale distance transform; \(C(x, y)\) is the center-bias function defined in (3).

2.4. Segmentation enhancement on saliency map

Image segmentation is the process of partitioning a digital image into multiple segments. Saliency detection results can benefit greatly from high-quality segmentation algorithms, e.g. normalized cuts [20] and mean-shift segmentation [21], since the segmentation algorithms can provide an accurate shape of the objects in images. As the \(L_0\) filter can produce saliency-focused images, the accuracy of the segmentation algorithms can be greatly improved. The segmentation information is combined with the distance-transform-based saliency model to improve the detection of salient objects’ boundaries. In our method, the rough saliency map is boundary-regulated to obtain the accurate object boundary by normalized-cut segmentation [20] is employed. Assume that there are \(N\) segments in an image after segmentation, and each of the segments is denoted as \(R_i\) \((i = 1, 2,...,N)\). \(N\) segmentation masks \(M_i\) \((i = 1, 2,...,N)\) can thus be generated, with the values within \(R_i\) being 1 and other regions being 0. The saliency value \(s_i\) in segment \(R_i\) can be calculated as follows:

\[
s_i = M_i(x, y) \times S(x, y) / |R_i|, \quad i=1, 2,...,N,
\]

Fig. 2. Saliency maps generated from images in the MSRA dataset [17]: (a) Original images, (b) IT [7], (c) AIM [8], (d) GB [9], (e) SR[10], (f) IG [11], (g) DCT [12], and (h) our proposed model.
where \( S(x, y) \) is the rough saliency map obtained in the last sub-section, and \( |R_i| \) is the number of pixels in the segment \( R_i \). As (5) shows, the average saliency value in each segment is considered, thus producing clear object boundaries. In our experiments, we empirically choose \( N \) to be 5. After this operation, the false-alarm rate outside salient objects is greatly reduced, since the saliency leakage is shared by all the pixels in that segment, which will be negligible after thresholding.

3. EXPERIMENTAL RESULTS

The proposed method is evaluated on the MSRA dataset [17], with 5,000 color images and the ground-truth images. As mentioned in Section 1, the proposed method is compared with six existing saliency detection models: IT [7], AIM [8], GB [9], SR [10], IG [11] and DCT [12].

Fig. 2 shows the saliency maps generated by the different methods. The compared results demonstrate that our method can effectively detect salient objects accurately, with the boundaries being much better defined than with the other methods. We also evaluate the performance of our proposed method quantitatively based on the area under the Receiver Operating Characteristic (ROC) curve [22]. The saliency map can be divided into two parts, namely the salient points and the non-salient points; and the ground-truth map can be divided into target points and background points. The percentage of the target points that fall into the salient points in the saliency map is the True-Positive Rate. The percentage of background points that fall into the salient points in the saliency map is the False-Positive Rate. The overall performance can be reflected by the area under the ROC curve, where a larger ROC area indicates a better performance. Perfect prediction corresponds to an area of 1, while random prediction results in an area of 0.5.

In order to threshold the saliency map to salient and non-salient regions, the adaptive threshold is empirically set as the mean of a saliency map. The ROC curves of the different methods and our proposed method are shown in Fig. 3, and the area under the ROC curves (AUC) of these methods are shown in Fig. 4. The overall performance can be reflected by the AUC score, with larger AUC scores indicating high performance. From Fig. 3 and Fig. 4, it can be seen that our method has the largest AUC score, thus achieves the best performance among the different methods.

4. CONCLUSION AND FUTURE WORKS

In this paper, we have proposed a novel saliency detection method. Firstly, \( L_0 \) smoothing is applied to images to highlight their fundamental image constituents while diminishing the insignificant details. Distance-transform-based saliency detection is then performed on the smoothed image in order to extract the general salient regions. Next, the center-bias mechanism is used to simulate the human visual system more accurately, and to form a rough saliency map. Finally, the rough saliency map is boundary-regulated to obtain the accurate object boundary by normalized-cut method. By computing the average saliency value for each segmented block, the saliency detection results can be greatly improved with clear object boundary.

Our future work will focus on the following aspects. Firstly, the center-bias mechanism can be analyzed using more biological plausibility, to determine the attenuation rate on boundaries. Secondly, more advanced algorithms on image segmentation can be applied to refine the salient objects more accurately. Finally, some spectral saliency methods can be used in the generation of rough saliency maps, in order to extract saliency from both the spatial domain and the spectral domain.

5. ACKNOWLEDGEMENT

This work is supported by an internal grant from the Hong Kong Polytechnic University (Grant No. G-YN21).
6. REFERENCES


