SALIENT REGION DETECTION via HIGH-DIMENSIONAL COLOR TRANSFORM AND LOCAL SPATIAL SUPPORT

VANJA.AKHIL (M.Tech) [1]
L.SRINIVAS (M.TECH, Associate Professor) [2]
[1,2] NALANDA INSTITUTE OF ENGINEERING & TECHNOLOGY, Kantepudi , Sattenapalli, Guntur,A.P , AICTE,New Delhi.JNTU, KAKINADA, India,,522438
akiakhil.aki7@gmail.com[1] srinivaslankireddy72@gmail.com[2]

Abstract

In this paper, we introduce a unique approach to robotically discover salient regions in an image. Our technique includes worldwide and neighborhood capabilities, which supplement every extraordinary to compute a saliency map. The first key concept of our paintings is to create a saliency map of a picture through using a linear combination of colors in a excessive-dimensional shade space. This is primarily based on a declaration that salient areas often have distinctive colorations in comparison with backgrounds in human perception; however, human perception is complex and in particular nonlinear. By mapping the low-dimensional purple, inexperienced, and blue shade to a characteristic vector in a high-dimensional shade location, we display that we're able to composite a correct saliency map through locating the most effective linear combination of coloration coefficients in the excessive-dimensional shade area. To similarly improve the overall performance of our saliency estimation, our second key idea is to make use of relative region and shade assessment amongst terrific pixels as features and to clear up the saliency estimation from a tri-map thru a mastering-based totally definitely set of rules. The greater nearby capabilities and gaining knowledge of-based set of rules complement the global estimation from the excessive-dimensional color remodel-based algorithm. The experimental results on three benchmark datasets show that our technique is effective in contrast with the preceding present day saliency estimation strategies.

Index Terms— Salient area detection, super pixel, tri-map, random forest, color channels, high-dimensional color space.

I. INTRODUCTION

We, as humans, are experts at quickly and accurately identifying the most visually noticeable foreground Object in the scene, known as salient objects, and adaptively focus our attention on such perceived important regions. In contrast, computationally identifying such salient object regions [2], [3], that match the human annotators' behavior when they have been asked to pick a salient object in an image, is very challenging. Being able to automatically, efficiently, and accurately estimate salient object regions, however, is highly desirable given the immediate ability to characterize the spatial support for feature extraction, isolate the object from potentially confusing background, and preferentially allocate finite computational resources for subsequent image processing.

While essentially solving a segmentation problem, salient object detection approaches segment only the salient foreground object from the background, rather than partition an image into regions of coherent properties as in general segmentation algorithms [3]. Salient object detection models also differ from eye fixation prediction models that predict a few fixation points in an image rather than uniformly highlighting the entire salient object region [3]. In practice, salient object detection methods are commonly used as a first step of many graphics/vision applications including object-of interest image segmentation [4], object recognition [5], adaptive compression of images [6], content-aware image editing [7], [8], image retrieval [9]–[11], etc.

We focus on bottom-up data driven salient object detection using image contrast (see Fig. 1) under the assumption that a salient object exists in an image [2]. The proposed method is simple, fast, and produces high quality results on benchmark datasets. Motivated by the popular belief that human cortical

cells may be hard wired to preferentially respond to high contrast stimulus in their receptive fields [14], we propose contrast analysis for extracting high-resolution, full-field saliency maps based on the following considerations:

• A global contrast based method, which separates a large-scale object from its surroundings, is desirable over local contrast based methods producing high saliency values at or near object boundaries. Global considerations enable assignment of comparable saliency values across similar image regions, and can uniformly highlight entire objects.

• Saliency of a region mainly depends on its contrast with respect to its nearby regions, while contrasts to distant regions are less significant (see also [15]).

• In man-made photographs, object is often concentrated towards the inner regions of the images, away from image boundaries (see [13]).

• Saliency maps should be fast, accurate, have low memory footprints, and easy to generate to allow processing of large image collections, and facilitate efficient image classification and retrieval.

II. Related Work

Our work belongs to the active research field of visual attention modeling, for which a comprehensive discussion is beyond the scope of this paper. We refer readers to recent survey papers for a detailed discussion of 65 models [12], as well as quantitative analysis of different methods in the two major research directions: fixation prediction [16], [18] and salient object detection [3].

We focus on relevant literature targeting pre-attentive bottom-up saliency region detection, which are biologically motivated, or purely computational, or involve both aspects. Such methods utilize low-level processing to determine the contrast of image regions to their surroundings, and use feature attributes such as intensity, color, and edges [33]. We broadly classify the algorithms into local and global schemes. Note that the classification is not strict as some of the research efforts can be listed under both categories.

Local contrast based methods investigate the rarity of image regions with respect to (small) local neighborhoods. Based on the highly influential biologically inspired early representation model introduced by Koch and Ullman [21], Itti et al. [17] define image saliency using central-surrounded differences across multi-scale image features. Ma and Zhang [18] propose an alternate local contrast analysis for generating saliency maps, which is then extended using a fuzzy growth model. Harel et al. [19] propose a bottom-up visual saliency model to normalize the feature maps of Itti et al. to highlight conspicuous parts and permit combination with other importance maps.

III. Proposed Method

3.1. Initial Saliency Tri-map Generation

In this section, we describe our method to detect the initial location of salient regions in an image. Our method is a learning-based method and it processes an image in super-pixel level.

A. Super pixel Saliency Features

As demonstrated in recent studies, features from super pixels are effective and efficient for salient object detection. For an input image I, we first perform over-segmentation to form super pixels X = $\{X_1, \dots \dots X_N\}$. We use the SLIC super pixel because of its low computational cost and high performance, and we set the number of super pixels to N = 500. To build feature vectors for saliency detection, we combine multiple information that are commonly used in saliency detection. We first concatenate the super pixels' x- and y-locations into our feature vector. The location feature is used because humans tend to focus more on objects that are located around the center of an image. Then, we concatenate the color features, as this is one of the most important cues in the human visual system and certain colors tend to draw more attention than others. We compute the average pixel color and represent the color features using different color space representations.

Next, we concatenate histogram features as this is one of the most effective measurements for the saliency feature, as demonstrated in [33]. The

histogram features of the i th super pixel D_{H_i} is measured using the chi-square distance between other super pixels' histograms. It is defined as

$$D_{H_i} = \sum_{j=1}^{N} \sum_{k=1}^{b} \frac{(h_{ik} - h_{jk})^2}{(h_{ik} - h_{jk})}, \qquad (1)$$

Where b is the number of histogram bins. In our work, we used eight bins for each histogram.

We have also used the global contrast and local contrast as color features. The global contrast of the ith super pixel D_{G_i} is given by

$$D_{G_i} = \sum_{j=1}^{N} d(c_i, c_j), \qquad (2)$$

Where $d(c_i, c_j)$denotes the Euclidean distance between the i th and the j th super pixels' color values, c_i andc_j, respectively. We use the RGB, CIE Lab, hue, and saturation of eight color channels to calculate the color contrast feature so that it has eight dimensions. The local contrast of the color features D_{L_i} is defined as

$$D_{L_i} = \sum_{j=1}^{N} \omega_{i,j}^{p} d(c_i, c_j) \qquad (3)$$

$$\omega_{i,j}^{p} = \frac{1}{Z_i} exp\left(-\frac{1}{2\sigma_p^2} \|p_i - p_j\|_2^2\right) \qquad (4)$$

Where $p_i \in [0, 1] \times [0, 1]$ denotes the normalized position of the ith super pixel and Z_i is the normalization term. The weight function in Eq. (4) is widely used in many applications including spectral clustering [13]. We adopt this function to give more weight to neighboring super pixels. In our experiments, we set $\sigma_p^2 = 0.25$. In addition to the global and local contrast, we further evaluate the element distribution by measuring the compactness of colors in terms of their spatial color variance. .

The aforementioned features are concatenated and are used to generate our initial saliency tri-map. Table I summarizes the features that we have used. In short, our super pixel feature vectors consist of 71 dimensions that combine multiple evaluation metrics for saliency detection.

B. Initial Saliency Tri-map via Random Forest Classification

After we calculate the feature vectors for every super pixel, we use a classification algorithm to check whether each region is salient. In this study, we use the random forest classification because of its efficiency on large databases and its generalization ability. A random forest is an ensemble method that operates by constructing multiple decision trees at training time and decides the class by examining each tree's leaf response value at test time. This method combines the bootstrap aggregating idea and random feature selection to minimize the generalization error. To train each tree, we sample the data with the replacement and train a decision tree with only a few features that are randomly selected. Typically, a few hundred to several thousand trees are used, as increasing the number of trees tends to decrease the variance of the model.

We used a regression method to estimate the saliency degree for each super pixel and classified it via adaptive thresholding. As our goal is to classify each super pixel as foreground and background, we found that using a classification method is more suitable than the regression for tri-map generation. Table II shows a comparison of the tri-map performance, in which the Fg. Precision (FP), Bg. Precision (BP), error rate (E R) are defined as below:

$$F_p = \frac{|\{F_c\} \cap \{F_{GT}\}|}{|\{F_c\}|}, \qquad (5)$$

$$B_p = \frac{|\{B_c\} \cap \{B_{GT}\}|}{|\{B_c\}|}, \qquad (6)$$

$$E_R = \frac{|(\{F_c\} \cap)\{B_{GT}\} \cup (\{B_c\} \cap \{F_{GT}\})|}{|\{I\}|} \qquad (7)$$

in which $|\cdot|$ denotes the number of pixels, F_c and B_c denote the foreground/background candidates, $F_{GT}$and B_{GT}denote the ground-truth annotations' foreground/background, respectively, and I denotes the whole image. The error rate (E_R) denotes the ratio of the area of misclassified regions to the image size, and the unknown rate is the ratio of the area of the

regions classified as unknown to the image size. We used 2,500 images from the MSRA-B dataset [49], which are selected as a training set from Jiang et al. for training data, and we used the annotated ground truth images for labels. We generated N feature vectors for each image. In total, we have approximately one million vectors for the training data We used the code provided by Becker et al. for random forest classification. In our implementation, we use 200 trees and we set the maximum tree depth to 10.

3.2 Saliency Estimation from Tri-map

In this section, we present our global salient region detection via HDCT and learning-based local salient region detection, and we describe a step-by-step process to obtain our final saliency map starting with the initial saliency map.

In section IV-A, we propose a global saliency estimation method via HDCT [2]. The idea of global saliency estimation implicitly assumes that pixels in the salient region have independent and identical color distribution. With this assumption, we depict the saliency map of a test image as a linear combination of high-dimensional color channels that distinctively separate salient regions and backgrounds. In section IV-B, we propose local saliency estimation via learning-based regression. Local features such as color contrast can reduce the gap between an independent and identical color distribution model implied by HDCT and true distributions of realistic images. In section IV-C, we analyze how to combine these two maps to obtain the best result.

A. Global Saliency Estimation via HDCT

Colors are important cues in the human visual system. Many previous studies have noted that the RGB color space does not fully correspond to the space in which the human brain processes colors. It is also inconvenient to process colors in the RGB space as illumination and colors are nested here. Therefore, many different color spaces have been introduced, including YUV, YIQ, CIE Lab, and HSV. Nevertheless, which color space is the best for processing images remains unknown, especially for applications such as saliency detection, which are strongly correlated to human perception. Instead of picking a particular color space for processing, we introduce a HDCT that unifies the strength of many different color representations. Our goal is to find a linear combination of color coefficients in the HDCT space such that the colors of salient regions and those of backgrounds can be distinctively separated. Fig. 4 illustrates the idea of using the linear combination of color coefficients for saliency detection.

The different magnitudes in the color gradients can also be used to handle cases in which salient regions and backgrounds have different amounts of defocus and different color contrasts. In summary, 11 different color channel representations are used in our HDCT space.

To further enrich the representative power of our HDCT space, we apply power-law transformations to each color coefficient after normalizing the coefficient between [0, 1]. We used three gamma values: {0.5, 1.0, and 2.0}.1 this resulted in a high-dimensional matrix to represent the colors of an image:

$$K = \begin{bmatrix} R_1^{\gamma_1} & R_1^{\gamma_2} & R_1^{\gamma_3} & G_1^{\gamma_1} & \dots \\ R_2^{\gamma_1} & R_2^{\gamma_2} & R_1^{\gamma_1} & R_1^{\gamma_1} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_N^{\gamma_1} & R_N^{\gamma_2} & R_N^{\gamma_3} & G_N^{\gamma_1} & \dots \end{bmatrix} \in R^{N\times 1}, \quad (8)$$

In which R_i and G_i denote the test image's i th superpixel's mean pixel value of the R color channel and G color channel, respectively. By using 11 color channels such as RGB, CIELab, hue, and saturation, we can obtain an HDCT matrix K with l = 11 × 3 = 33.

Algorithm 1 HDCT-Based Saliency Estimation

Input: initial tri-map T, and **K**= (Eq. (8))
1: $f \leftarrow$ number of foreground super pixels in the tri-map
2: $b \leftarrow$ number of background super pixels in the tri-map
3: $M \leftarrow f+b$
4: construct $\widetilde{\mathbf{K}} \in\in \mathbb{R}^{M\times l}$ by Eq. (10)
5: construct $\mathbf{U} = \in \mathbb{R}^{M\times 1}$ by Eq. (11)
6: calculate $\alpha^* = \left(\widetilde{\mathbf{K}}^{\mathbf{T}}\widetilde{\mathbf{K}} + \boldsymbol{\lambda}\mathbf{I}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{U}$ by solving Eq. (9)
7: calculate $S_G(X_i) = \sum_{j=1}^{l} \mathbf{K}_{ij}\,\alpha_j^*$ by Eq. (12)
Output: Saliency map S_G

The nonlinear power-law transformation takes into account the fact that our human perception responds nonlinearly to incoming illumination. It also stretches/compresses the intensity contrast within different ranges of color coefficients. Table III summarizes the color coefficients concatenated in our HDCT space. This process is applied to each super pixel in an input image individually.

To obtain our saliency map, we utilize the foreground candidate and background candidate color samples in our tri-map to estimate an optimal linear combination of color coefficients to separate the salient region color and background color. We formulate this problem as an l2 regularized least squares problem that minimizes

$$\min_{\alpha} \left\| \left(U - \widetilde{K}\alpha\right) \right\|_2^2 + \lambda \|\alpha\|_2^2, \qquad (9)$$

where $\alpha \in R^I$ is the coefficient vector that we want to estimate, λ is a weighting parameter to control the magnitude of α, and $\widetilde{K}$ is a M×l matrix with each row of K corresponding to color samples in the foreground/background candidate regions:

$$\widetilde{K} = \begin{bmatrix} R_{FS_1}^{\gamma_1} & R_{FS_1}^{\gamma_2} & R_{FS_1}^{\gamma_3} & G_{FS_1}^{\gamma_1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{FS_f}^{\gamma_1} & R_{FS_f}^{\gamma_1} & R_{FS_f}^{\gamma_1} & R_{FS_f}^{\gamma_1} & \cdots \\ R_{BS_1}^{\gamma_1} & R_{BS_1}^{\gamma_1} & R_{BS_1}^{\gamma_1} & R_{BS_1}^{\gamma_1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{BS_b}^{\gamma_1} & R_{BS_b}^{\gamma_1} & R_{BS_b}^{\gamma_1} & G_{BS_b}^{\gamma_1} & \cdots \end{bmatrix}, \qquad (10)$$

Where F_{s_i} and BS_j denote the i th foreground candidate super pixel among entire super pixels and the j th background super pixel among entire super pixels that are classified at the tri-map generation step, respectively. M is the number of color samples in the foreground/background candidate regions (M N), and f and b denote the number of foreground and background regions, respectively, such that M = f + b. U is an M dimensional vector with value equal to 1 and 0 if a color sample belongs to the foreground and background candidate respectively.

$$\mathbf{U} = [1 \quad 1 \quad \ldots \quad 1 \quad 0 \quad 0 \quad \ldots \quad 0]^T \quad \in \mathbb{R}^{M\times 1} \qquad (11)$$

Since we have a greater number of color samples than the dimensions of the coefficient vector, the l2 regularized least squares problem is a well-conditioned problem that can be readily minimized with respect to α as $\alpha^* = (\widetilde{K}^T\widetilde{K} + \lambda I)^{-1}\widetilde{K}^T U$. In all experiments, we use λ = 0.05 to produce the best results. After we obtain α∗, the saliency map can be constructed as

$$S_G(X_i) = \sum_{j=1}^{l} \boldsymbol{K}_{ij}\alpha_j^*, \quad i = 1,2,\ldots\ldots,N, \qquad (12)$$

Which denotes the linear combination of the color coefficient of our HDCT space? The l2 regularize in the least square formulation in Eq. (9) restricts the magnitude of the coefficient vector to avoid over-fitting to U. With this l2 regularizer, the constructed saliency map is more reliable for the both foreground and background super pixels that are initially classified in the tri-map. We tested several values of λ, and the regularized l2 least square with nonzero λ produces better saliency maps than the least square method without regularizer (λ = 0). Note that the popular l1 regularizer for sparse solution could also be considered, but the l1 regularizer is not essential in our work, since more accurate representation of both foreground and background super pixels in HDCT space are important. Also, it is not necessary for the coefficient vector to be sparse. The overall process of the HDCT-based saliency detection is described in algorithm 1.

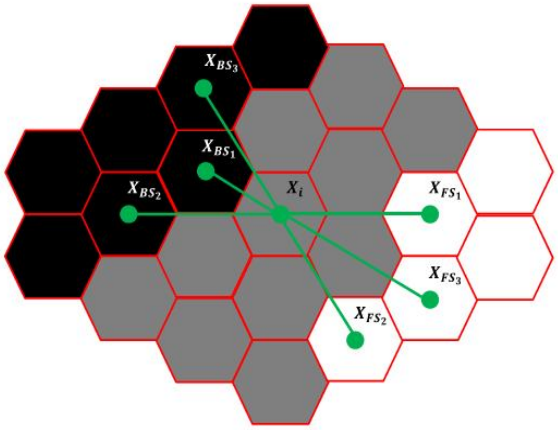

Fig.1. an illustration of local saliency features. Black, white, and gray regions denote background super pixels, foreground super pixels, and unknown super pixels, respectively. We use K -nearest foreground

super pixels and K -nearest background super pixels to calculate a feature vector.

B. Local Saliency Estimation via Regression

Although the HDCT-based salient region detection provides a competitive result with a low false positive rate, this method has a limitation in that it is easily affected by the texture of the salient region, and therefore, it has a relatively high false negative rate. To overcome this limitation, we present a learning-based local salient region detection that is based on the spatial and color distance from neighboring super pixels. Table IV summarizes the features used in this section. First, for each super pixel, we find the K -nearest foreground super pixels and K -nearest background super pixels as described in Fig.1. For each super pixel Xi, we find the K -nearest foreground super pixels X_{FS}= $\{X_{FS_1}$, $\{X_{FS_1}$,. . . , $\{X_{FS_k}\}$ and K -nearest background super pixelsX_{FS}= $\{X_{BS_1}$, $\{X_{BS_1}$,. . . , $\{X_{BS_k}\}$, and we use the Euclidean distance between a super pixel Xi and super pixels XF S or XBS as features. The Euclidean distance to the K-nearest foreground (∈ RK×1) and background (dBSi ∈ RK×1) features of the i th super pixel is defined as follows:

$$\mathbf{d}_{FS_i} = \begin{bmatrix} \| \mathbf{p}_i - \mathbf{p}_{FS_{i_1}} \|_2^2 \\ \| \mathbf{p}_i - \mathbf{p}_{FS_{i_2}} \|_2^2 \\ \cdot \\ \cdot \\ \cdot \\ \| \mathbf{p}_i - \mathbf{p}_{FS_{i_K}} \|_2^2 \end{bmatrix} \mathbf{d}_{BS_i} = \begin{bmatrix} \| \mathbf{p}_i - \mathbf{p}_{BS_{i_1}} \|_2^2 \\ \| \mathbf{p}_i - \mathbf{p}_{BS_{i_2}} \|_2^2 \\ \cdot \\ \cdot \\ \cdot \\ \| \mathbf{p}_i - \mathbf{p}_{BS_{i_K}} \|_2^2 \end{bmatrix}, \quad (13)$$

in which FS_{i_j} denotes the j th nearest foreground super pixel and BS_{i_j} denotes the j th nearest background super pixel from the i th super pixel. As objects tend to be located in a compact region in an image, the spatial distances between a candidate super pixel and the nearby foreground/background super pixels can be a very useful feature for estimating the saliency degree. We also use the color distance features between super pixels. The feature vector of color distances from the i th super pixel to the K-nearest foreground (d_{CF_i}∈ $R^{8k\times1}$) and background (d_{BS_i}∈$R^{8k\times1}$) super pixels is defined as follows:

(d_{CB_i} ∈ R8K×1) super pixels is defined as follows:

$$\mathbf{d}_{CF_i} = \begin{bmatrix} \mathbf{d}\left(\mathbf{c}_i, \mathbf{c}_{FS_{i_1}}\right) \\ \mathbf{d}\left(\mathbf{c}_i, \mathbf{c}_{FS_{i_2}}\right) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{d}\left(\mathbf{c}_i, \mathbf{c}_{FS_{i_K}}\right) \end{bmatrix}, \quad \mathbf{d}_{CB_i} = \begin{bmatrix} \mathbf{d}\left(\mathbf{c}_i, \mathbf{c}_{BS_{i_1}}\right) \\ \mathbf{d}\left(\mathbf{c}_i, \mathbf{c}_{BS_{i_2}}\right) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{d}\left(\mathbf{c}_i, \mathbf{c}_{BS_{i_K}}\right) \end{bmatrix} \quad (14)$$

Although a super pixel located near the foreground super pixels tends to be a foreground, if the color is different, there is a high possibility that it is a background super pixel located near the boundary of an object. We use eight color channels—RGB, CIE Lab, hue, and saturation—to measure the color distance, where c_i,$c_F S_{ij}$, and $c_B S_{ij}$are eight-dimensional color vectors. The distance vector d (c_i,$c_F S_{ij}$) is also an eight-dimensional vector, where each element of d (c_i,$c_F S_{ij}$) is the distance in a single color channel. To decide the optimal number of nearest super pixels K, we calculate the F-measure rate for each parameter. Fig. 8 shows the result, and we set K = 25, which shows the best result.2

For saliency estimation, we used the super pixel-wise random forest regression algorithm, which is effective for large high-dimensional data. We extracted feature vectors using the initial tri-map, and then, we estimated the saliency degree for all super pixels. For this local saliency map, even those classified as foreground/background candidate super pixels in the initial tri-map are reevaluated because they could still be misclassified. It should be noted that the initial tri-map is generated by a random forest classifier and that the next random forest regressor generates a local saliency map. Considering that we have two stages of cascaded random forests, we divided the training data set into two disjoint sets so that the second random forest is trained with more realistic inputs. Toward this end, we trained the first random forest with one data set, and we obtained the training data set for the second random forest from the tri-maps generated for the other data set, which is not used for training the first random forest. This process is repeated in a manner similar to five-fold cross-validation. We used the code provided by Becker et al. [51] for random forest regression using 200 trees and setting the maximum tree depth to 10.

C. Final Saliency Map Generation

After we generated the global and the local saliency maps, we combined them to generate our final saliency map. The examples show that the HDCT-based saliency map tends to catch the object precisely; however, the false negative rate is relatively high owing to textures or noise. In contrast, the learning-based saliency map is less affected by noise, and therefore, it has a low false negative rate but a high false positive rate. Therefore, combining the two maps is a significant step in our algorithm.

Borji et al. [38] proposed two approaches to combine the two saliency maps. The first approach is to perform the pixel wise multiplication of the two maps, as shown below:

$$S_{mult} = \frac{1}{z}\left(p(S_G) \times p(S_G)\right), \qquad (15)$$

in which Z is a normalization factor, p(.) is a pixel-wise combination function, S_G is the global saliency result (Section IV-A), and S_L is the local saliency result (Section IV-B). However, this combination tends to show darker pixels and suppresses bright pixels, and therefore, some false negative pixels from a global saliency map will suppress the local saliency map, and the merit of the local saliency map will decrease.

The second approach is to combine the two maps using a summation:

$$S_{sum} = \frac{1}{z}\left(p(S_G) \times p(S_G)\right), \qquad (16)$$

In our study, we combine the two maps more adaptively to maximize our performance. Based on Eq. (16), we adopt p(x) = exp(x) as a combination function to give greater weight age to the highly salient regions. The weight values are determined by comparing the saliency map with the ground truth. We calculate the optimal weight values for the linear summation by solving the nonlinear least-squares problem, as shown below:

$$\min_{\substack{\omega_1 \geq 0, \omega_2 \geq 0 \\ \omega_3 \geq 0, \omega_4 \geq 0}} \| \omega_1 p(\omega_2 S_G) + \omega_3 P(\omega_4 S_L) - GT \|_2^2 \qquad (17)$$

in which GT is the ground truth of an image in the training data. To find the most effective weights, we iteratively optimize the nonnegative least-squares objective function in Eq. (17) with respect to each variable. As the objective function in Eq. (17) is bi-convex, it must converge after a few optimization steps; however, different local solutions are obtained by the different initializations. To obtain the best solution (i.e., the solution that yields the smallest value of the objective function in Eq. (17) among several local solutions), we repeat the optimization process with randomly initialized variables several times, and the final solution for the objective function in Eq. (17) is obtained as ω_1 = 1.15, ω_2 = 0.74, ω3 = 1.57, and ω_4 = 0.89. We found that our performance further improves with the values of the solution. Finally, we defined the equation of the final saliency map combination as

$$S_{final} = \frac{1}{Z}\left(\omega_1 p(\omega_2 S_G) + \omega_3 P(\omega_4 S_L)\right) \qquad (18)$$

We observe that the performance greatly improves after combining the two maps: highly salient regions that have been caught by the local saliency map are preserved, and the false negative region that is vaguely salient is discarded.

To evaluate the effectiveness of our local saliency estimation, we compare the precision-recall curve with that of the spectral matting algorithm that extracts foregrounds from the user input. We use the tri-map result instead of the user input for automatic matting.. Although the matting algorithm can provide a reasonable result without being influenced by textures, we found that the matting method heavily relies on the input tri-map and is therefore easily affected by misclassified super pixels. On the other hand, the learning-based method can determine the saliency degree by observing the spatial distribution of the nearest foreground and background super pixels, and therefore, our method is more robust to misclassified errors.

IV. RESULTS

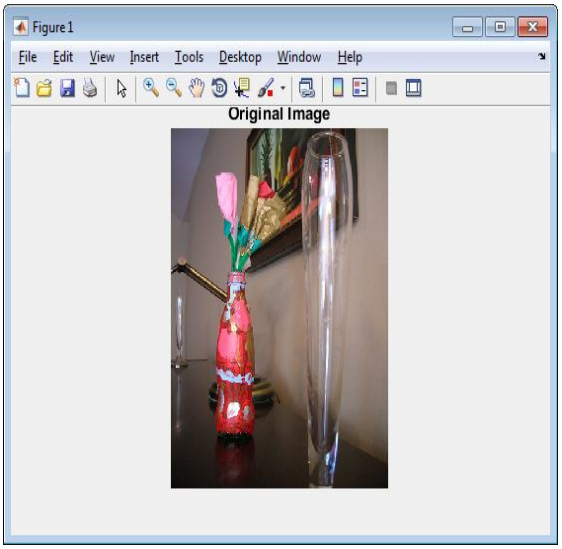

Fig.1 original image

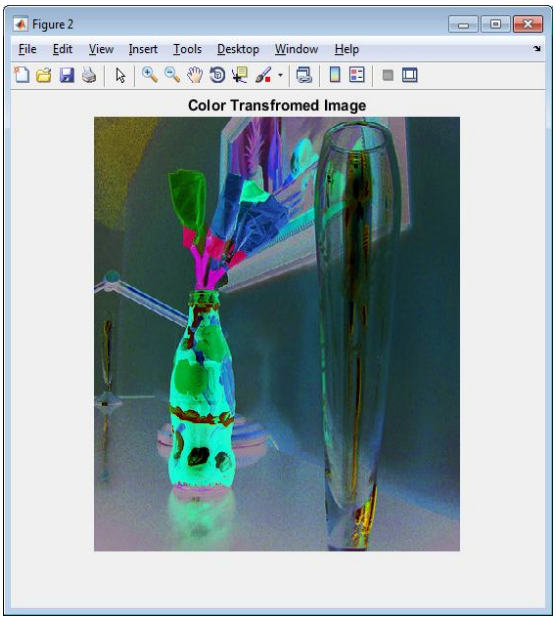

Fig.2.color transformed image.

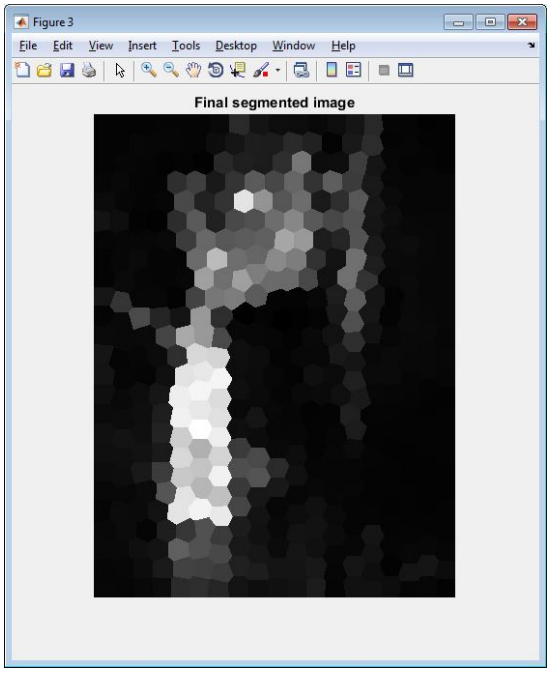

Fig.3.final segmented image.

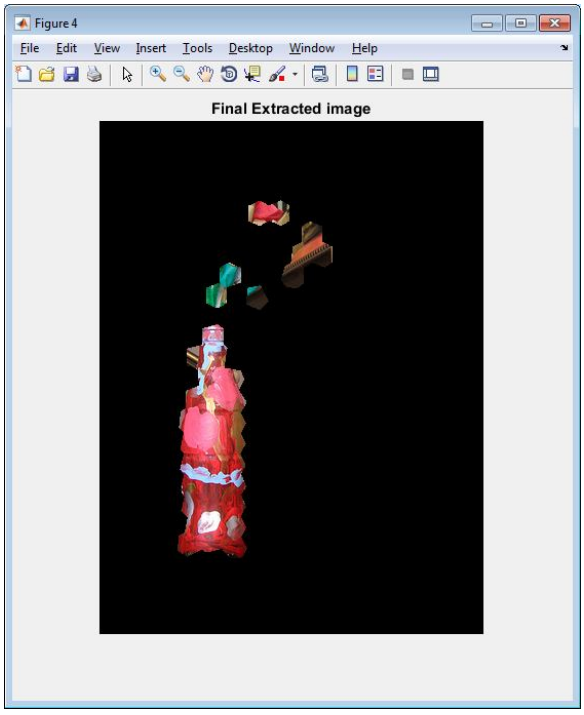

Fig.4.final extracted image

V. CONCLUSION

We have provided a unique robust salient region detection technique that estimates the foreground regions from a tri-map using different methods: global saliency estimation through HDCT and neighborhood saliency estimation through regression. The tri-map-primarily based totally strong estimation overcomes the constraints of inaccurate preliminary saliency kind. As end result, our method achieves proper common overall performance and is computationally green in assessment to the dominion-of-the art work techniques we also confirmed that our proposed technique can further improve DRFI, which is the quality appearing technique for salient location detection. In the destiny, we purpose to boom the features for the preliminary tri-map to further enhance our set of set of rules performance.

REFERENCES

[1] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in IEEE CVPR, 2011, pp. 409–416.

[2] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, T. X., and S. H.Y., "Learning to detect a salient object," IEEE TPAMI, no. 2, 2011.

[3] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in ECCV, 2012.

[4] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, "Saliency driven total variation segmentation," in IEEE ICCV, 2009.

[5] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in IEEE CVPR, 2004.

[6] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," IEEE Trans. Consumer Elec., vol. 46, no. 4, pp. 1103–1127, 2002.

[7] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, "A shapepreserving approach to image resizing," Comput. Graph. Forum, vol. 28, no. 7, pp. 1897–1906, 2009.

[8] M.-M. Cheng, F.-L. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Repfinder: Finding approximately repeated scene elements for image editing," ACM TOG, vol. 29, no. 4, pp. 83:1–8, 2010.

[9] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," ACM TOG, 2009.

[10] S.-M. Hu, T. Chen, K. Xu, M.-M. Cheng, and R. R. Martin, "Internet visual media processing: a survey with graphics and vision applications," The Visual Computer, pp. 1–13, 2013.

[11] Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, and X. Wu, "Visualtextual joint relevance learning for tag-based social image search," IEEE TIP, vol. 22, no. 1, pp. 363–376, 2013.

[12] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," IEEE TPAMI, 2012.

[13] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in BMVC, 2011, pp. 1–12.

[14] J. Reynolds and R. Desimone, "Interacting roles of attention and visual salience in v4," Neuron, vol. 37, no. 5, 2003.

[15] P. Reinagel, A. Zador et al., "Natural scene statistics at the centre of gaze," Network: Computation in Neural Systems, 1999.

[16] A. Borji, D. Sihite, and L. Itti, "Quantitative analysis of humanmodel agreement in visual saliency modeling: A comparative study," IEEE TIP, 2012.

[17] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," MIT tech report, Tech. Rep., 2012.

[18] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in ACM Multimedia, 2003.