

# Salient Object Detection Based on Regional Contrast and Relative Spatial Compactness

**Dan Xu<sup>1,2</sup>, Zhenmin Tang<sup>1</sup>, Wei Xu<sup>1</sup>**

<sup>1</sup>Department of Computer Science & Engineering  
Nanjing University of Science & Technology  
Nanjing, China

<sup>2</sup>Department of Computer Science & Engineering  
Jiangsu University of Science & Technology  
Zhenjiang, China

[e-mail: xudan.zj@gmail.com, tang.zm@mail.njust.edu.cn, njust\_cst\_xuwei@163.com]

\*Corresponding author: Dan Xu

*Received July 8, 2013; revised September 26, 2013; accepted October 23, 2013; published November 29, 2013*

---

## Abstract

In this study, we propose a novel salient object detection strategy based on regional contrast and relative spatial compactness. Our algorithm consists of four basic steps. First, we learn color names offline using the probabilistic latent semantic analysis (PLSA) model to find the mapping between basic color names and pixel values. The color names can be used for image segmentation and region description. Second, image pixels are assigned to special color names according to their values, forming different color clusters. The saliency measure for every cluster is evaluated by its spatial compactness relative to other clusters rather than by the intra variance of the cluster alone. Third, every cluster is divided into local regions that are described with color name descriptors. The regional contrast is evaluated by computing the color distance between different regions in the entire image. Last, the final saliency map is constructed by incorporating the color cluster's spatial compactness measure and the corresponding regional contrast. Experiments show that our algorithm outperforms several existing salient object detection methods with higher precision and better recall rates when evaluated using public datasets.

---

**Keywords:** Color names, salient object detection, region contrast, spatial compactness

## 1. Introduction

Researchers are attempting to find methods to enable algorithms to move towards the ability, natural to human beings, to focus on important parts of an image with seemingly little effort. This is an important and fundamental research problem in neuroscience and psychology. Recently, it has attracted more attention in computer vision because of its application in adaptive image compression [1,2], image segmentation [3,4], object detection and recognition [5,6], and image editing techniques [7,8,9].

Visual saliency refers to the physical, bottom-up distinctness of image details. It is a relative property that depends on the degree to which a detail is visually distinct from its background [10]. This distinctness may vary in image attributes such as color, orientation, intensity, or edges. The methods used to evaluate distinctness can depend on local or global contrast. Global contrast methods are believed to be preferable to local ones in producing more salient values for uniform inside objects; furthermore, the global methods are more efficient when there is a cluttered background [7,11]. Based on this idea, Cheng et al. [7] proposed a region-based contrast method (RC), which achieved good results on a public dataset.

In this study, we also concentrate on regional contrast salient object detection; however, the primary differences are in the approaches employed for image segmentation and region representation. We introduce the color attribute color names, which contains 11 basic colors and every color corresponding to a range of pixel values [12]. We train the color names distribution on pixel values using probabilistic latent semantic analysis (PLSA), a well-known latent aspect model in the text analysis community. The distribution can be used as prior knowledge to assign pixels to special color names and form different color clusters. Each cluster corresponds to a color name in an image. We explore the relative spatial compactness of the clusters to obtain their initial conspicuity values. In this application, relative means that the spatial compactness of a cluster is measured by its distance relative to other clusters rather than the intra spatial variance of the cluster as used by [13]. Then, each cluster is divided into regions, which are represented by color name descriptor (CN). The saliency value of a region depends on its color contrast with respect to other regions in the image and the initial conspicuity value of the cluster to which it belongs. Fig. 1 shows the flowchart of the proposed framework for salient object detection.

**Contributions.** The contributions of our study include: 1) Color names are introduced into the method of salient object detection for image segmentation and region representation. Comparisons show that the color name descriptor is more distinct and compact than the sparse histogram in RGB color space, commonly used in RC [7]. 2) Our method uses the relative spread of the color cluster as the saliency factor. That is, the less the color cluster spreads, the more salient it is.

We have compared the application of our method on publicly available benchmark datasets with some state-of-the-art saliency methods [7,14,15,16,17,18] and with manually produced ground truth annotations. The experiments show improvements over previous methods both in precision and recall rates.

## 2. Related Work

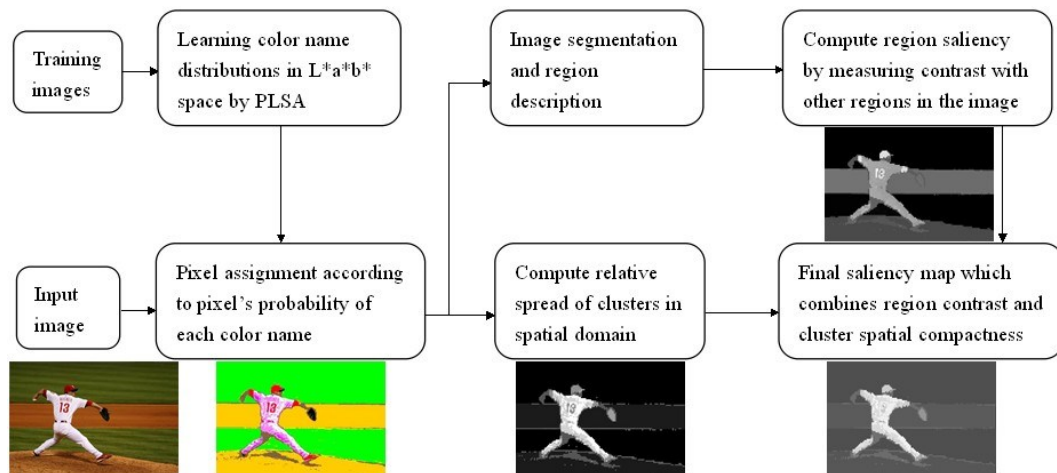
In the past decade, many different algorithms have been proposed to compute visual saliency from digital imagery. According to the theme of our study, we briefly introduce the local and

global contrast methods in the subfield of pre-attentive bottom-up saliency detection.

Local contrast methods consider local structures of image pixels or patches in small neighborhoods. Pixels or patches with high contrast will be assigned a high saliency. Itti et al. [19] used center-surround difference as the filter response during local measurements. Liu et al. [20] proposed a set of novel features, including multiscale contrast, center-surround histograms, and color spatial distribution, to describe a salient object. In [21], the saliency measure is formulated using a statistical framework and local feature contrast in illumination, color, and motion information. The measure is based on applying a sliding window to the image. In each window, the contrast is computed between the distribution of certain features in an inner window and the distribution in the collar of the window. These local contrast methods work well for images with homogenous backgrounds and produce well-defined salient object boundaries, but do not hold for images with cluttered backgrounds and uniform inside objects [11].

Global contrast methods compute saliency in the entire image. Units (which can be pixels, super-pixels or regions) with high contrast or low occurrence frequency are more salient. Achanta et al. [15] proposed a frequency-tuned method that evaluates saliency based on individual pixels. This pixel-level saliency analysis inevitably loses information from the original image, therefore, finding salient objects with it is difficult. Feng et al. [11] detected salient objects by directly measuring the saliency of a window in the original image. Cheng et al. [7] proposed a region contrast saliency extraction algorithm, which computes color contrast and spatial relations at the region level.

There are some previous studies that combined local methods with global ones. Goferman et al. [14] considered local and global clues simultaneously, taking visual organization rules and high-level features into account. To overcome the drawback of [15], i.e., that the saliency maps highlight the background when the salient regions are large, Achanta et al. introduced local scale information and then made assumptions about the scale of the object of detection based on its position in the image [18].



**Fig. 1.** Proposed framework for salient object detection in an image.

### 3. The Proposed Method

#### 3.1 Learning Color Names

Color names are linguistic labels that humans assign to colors in their world [22]. In a linguistic study, Berlin and Kay [23] concluded that the English language contains eleven basic color terms: black, blue, brown, gray, green, orange, pink, purple, red, white and yellow. Color names learning finds the relation between color names and the corresponding pixel values. To learn color names for saliency detection, we have chosen 100 images for each color name from the MSRA public saliency dataset [20] as training data, in which every image is labeled with a corresponding color name. A sample of the training images is shown in Fig. 2.



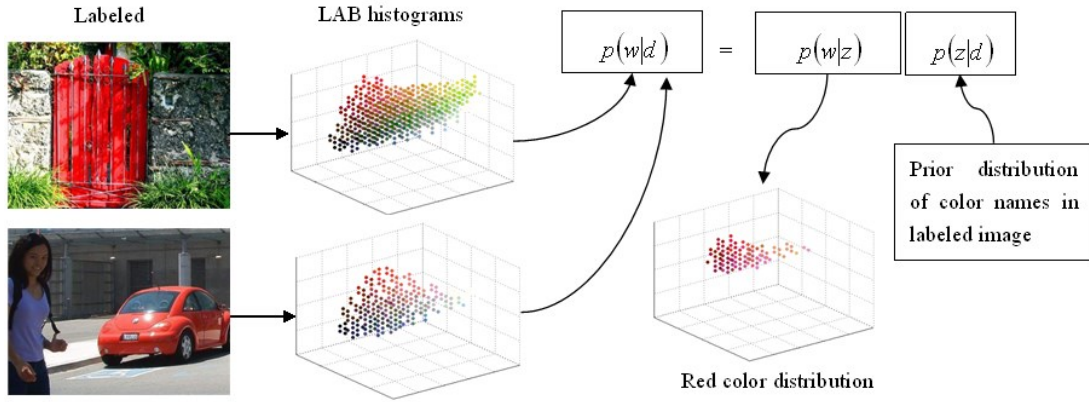
Fig. 2. A sample of training images for color names.

We use probabilistic latent semantic analysis (PLSA) [24], which is a well-known latent aspect model in the text analysis community, to learn the color names. PLSA uses a generative model, to find the latent topics which best explain the observed data. Given a document set  $D = \{d_1, \dots, d_n\}$ , each document described by a vocabulary  $W = \{w_1, \dots, w_m\}$ , the words are generated by latent topics  $Z = \{z_1, \dots, z_k\}$ . PLSA assumes that every pair  $(d, w)$  is generated as follows:

$$p(d, w) = p(d)p(w|d) \quad (1)$$

$$p(w|d) = \sum_{z \in Z} p(w|z) p(z|d) \quad (2)$$

In our work, we treat the image as a document. Pixels in the image are represented by discretizing their values in  $L^*a^*b^*$  space into a histogram by assigning each value using cubic interpolation to a regular  $10 * 20 * 20$  bin. The bin of the histogram is synonymous with the word in PLSA [12]. As shown in Fig. 3, the color histogram corresponds to the component of word-document distribution  $p(w|d)$  in the PLSA model. In addition, every image naturally includes some colors, such as red, green, yellow, etc., just like every document includes a number of topics in PLSA. In short,  $d$  represents an image,  $w$  represents the pixel values and  $z$  denotes the color names appearing in the image. Distributions  $p(z|d)$  and  $p(w|z)$  can be evaluated using the expectation-maximization algorithm. The former represents the conditional probabilities of various color names given an image, which may vary from image to image, the latter represents the word-topic distribution and is shared among all images. The training stage will find the relation between words and topics,  $p(w|z)$ . We will show how it is used for image segmentation and region description in the next stage.



**Fig. 3.** Overview of PLSA used in color names training.

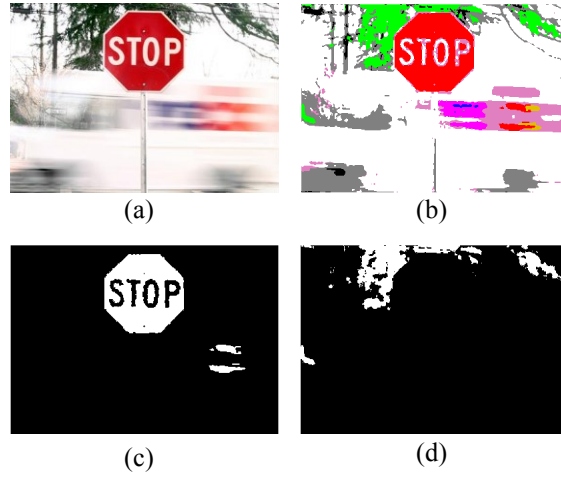
Compared to standard PLSA, in this study, a prior distribution of color names in images,  $p(z|d)$ , is defined according to image labels. The topic corresponding to the label of the image has a higher frequency than the other topics. The prior distribution can be achieved by a parameter vector,  $w_{l_d}$ , where  $l_d$  is the label of the image. The length of the vector  $w_{l_d}$  equals the number of topics. For  $z = l_d$ ,  $w_{l_d}(z) = c \geq 1$ , otherwise  $w_{l_d}(z) = 1$ . By varying  $c$ , the influence of image label  $l_d$  on distribution  $p(z|d)$  can be controlled. In our experiment, we find  $c = 5$  to be optimal.

### 3.2 Image Segmentation and Region Description

After we obtain the topic distribution over the words,  $p(w|z)$ , Bayes theorem is used to evaluate  $p(z|w)$ , which then represents a pixel's probabilities of different color names.

$$p(z|w) \propto p(z)p(w|z) \quad (3)$$

where the prior probability of the color names,  $p(z)$ , is obtained from the training images. The calculation of  $p(z|w)$  is obtained offline, i.e., it is like a lookup table. For pixel  $X$  in an input image, probabilities of different color names can be represented as  $\{p(cn_1|f(X)), p(cn_2|f(X)), \dots, p(cn_{11}|f(X))\}$ , according to  $p(z|w)$ , where  $f(X)$  is the pixel value in  $L^*a^*b^*$  space.  $X$  is assigned to the color name that has the maximum probability for that location. In this formulation,  $p(z|w)$  in Equation (3) is computed based on the individual pixel and will result in too many isolated points. Therefore, we employ a median filter after the pixel assignment. As shown in **Fig. 4**, a spread of color names in spatial space (color clusters) are obtained after pixel assignment and median filtering. Then, each color cluster is segmented into regions. For example, in the red cluster, pixels that belong to red are set to 1, otherwise the pixel is set to 0. Then the cluster is segmented into regions in a binary image. The other clusters in the image are treated in the same way.



**Fig. 4.** Overview of image segmentation. (a) Original image, (b) color clusters after pixel assignment and median filtering, (c) regions of red cluster and (d) regions of green cluster.

If a region is identified as salient, it has at least one feature different from the surroundings [25]. Ideally, the feature descriptor is discriminative, compact and invariant to illumination. In [7], RGB histograms are chosen as the color descriptor to compute contrast between regions. In this study, we use color names to describe region features and compare with RGB histograms. The color name descriptor (CN) is defined as a vector consisting of every color name's probability for given region  $R$ .

$$CN = \{p(cn_1|R), p(cn_2|R), \dots, p(cn_{11}|R)\} \quad (4)$$

with

$$p(cn_i|R) = \frac{1}{N} \sum_{X \in R} p(cn_i|f(X)) \quad (5)$$

where  $X$  stands for the pixel in region  $R$ ,  $N$  is the total number of pixels,  $f(X)$  is the pixel value in  $L^*a^*b^*$  color space and  $p(cn_i|f(X))$  is the conditional probability of color name  $i$  for a given  $f(X)$ , which can be obtained from the learned  $p(z|w)$ .

When comparing CN with RGB histograms, we first examine the illumination invariance. Koen E.A [26] considered RGB histograms as a combination of three 1-D histograms, based on the R, G and B channels, possessing no invariance properties. The color name descriptor displays a certain amount of photometric invariance because colors with small differences that often occur together are more likely to be found in the same topic. For example, the label yellow includes highly saturated yellow and dark yellow that is caused by shadows or shading. In the learning stage, all these colors will be captured by  $p(w|z = \text{yellow})$ . Second, we consider the discrimination and compactness of the two descriptors such as [22] does. The KL-ratio of the descriptor is computed between inter-class KL divergence and intra-class KL divergence in the bounding box of each object category in PASCAL VOC 2007 and the MSRA dataset as follows:

$$KL - ratio = \frac{\sum_{k \in C^m, j \in C^k} \min KL(p_j, p_k)}{\sum_{k \in C^m, i \in C^k, i \neq k} \min KL(p_i, p_k)} \quad (6)$$

where

$$KL(p_i, p_j) = \sum_{x=1}^N p_i(x) \log \frac{p_i(x)}{p_j(x)} \quad (7)$$

and  $p_i$  is the histogram of the bounding box  $i$  over the  $N$  visual words  $x$ . Indices  $i \in C^k$  represent bounding boxes which belong to class  $k$ , while  $j \notin C^k$  are random samples of the bounding boxes that are not the same class as  $k$ . A higher KL-ratio indicates a more discriminative and compact descriptor because the inter-class KL divergence is larger than the intra-class KL divergence. In Fig. 5, we show the average KL-ratio of RGB histograms and the color name descriptor. The RGB histograms used in [7] include 1728 theoretical dimensions; however, to increase the speed of the algorithm, only 85 dimensions are actually used. For comparison, we changed the dimensions of the RGB histograms from 11 (the same as the number of color names) to 85. Fig. 5 shows that CN is superior to the RGB histograms in both compactness and discrimination even when the number of histograms is increased to 85. We also find that both the CN and the RGB histograms result in a better KL-ratio on the MSRA dataset than on PASCAL VOC 2007. The reason is that PASCAL VOC 2007 is shape predominant and the MSRA dataset is color predominant. Based on the analysis and the experimental results, we adopt CN as the color descriptor to compute region contrast.

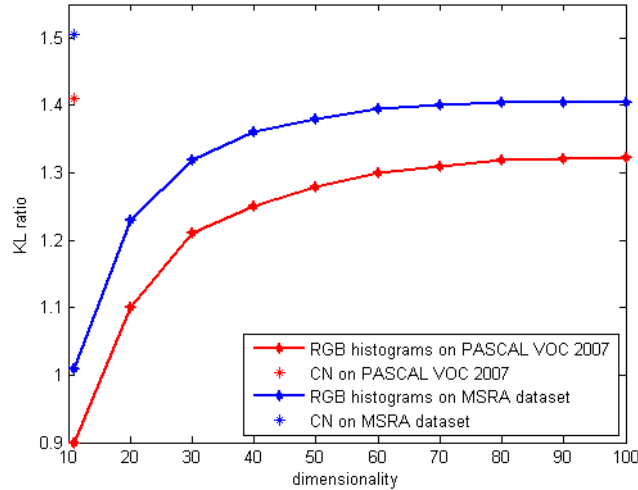


Fig. 5. KL-ratio of CN and RGB histograms on PASCAL VOC 2007 and the MSRA dataset.

### 3.3 Saliency Evaluation

To evaluate the saliency, two features are considered, the spatial compactness of every color name and the color contrast at the region level. Each color name appears in the spatial domain with a certain spread over the image. The less the color name spreads, the more salient it is [27]. The spread of a color cluster is evaluated by the distance in the spatial domain between color clusters rather than by intra spatial variance alone because there can be colors with similar intra spatial variance, but their compactness or relative spread maybe different. The relative spread of a cluster is quantified by the distance from the cluster's pixels to the centroids of the other clusters. For example, the distance  $D_s(i, j)$  between clusters  $i$  and  $j$  quantifies the relative spread of cluster  $i$  with respect to cluster  $j$  in the spatial domain. Every pixel in the image belongs to different color names in a probabilistic manner.

$$D_s(i, j) = \frac{\sum_{X \in C_i} \|X - u_j\|^2 \cdot p_i(X)}{\sum_{X \in C_i} p_i(X)} \quad (8)$$

where  $X$  is a pixel of cluster  $i$  and  $P_i(X)$  is the probability that  $X$  belongs to color name  $i$  in

the original image. Additionally,  $u_j$  is the spatial mean of cluster  $j$ , which is computed as follows:

$$u_j = \begin{bmatrix} M_x(j) \\ M_y(j) \end{bmatrix} \quad M_x(j) = \frac{\sum_{X \in C_j} p_{j(X) \cdot x}}{\sum_{X \in C_j} p_{j(X)}} \quad M_y(j) = \frac{\sum_{X \in C_j} p_{j(X) \cdot y}}{\sum_{X \in C_j} p_{j(X)}} \quad (9)$$

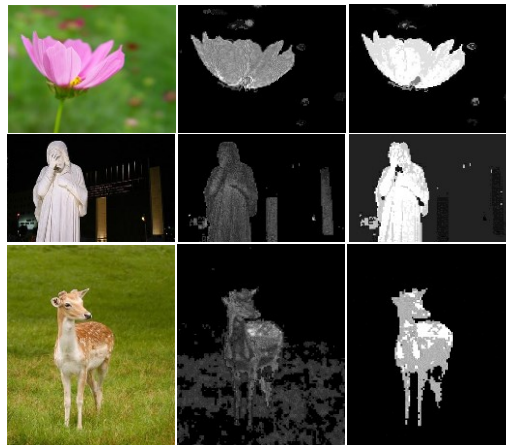
where  $(x, y)$  is the location of pixel  $X$  of cluster  $j$ .

The compactness of the cluster  $i$  in the spatial domain is calculated as follows:

$$S_{isd} = \left( \sum_{j \neq i} D_s(i, j) \right)^{-1} \quad (10)$$

By considering the relative spread, we not only eliminate the background color names that have large spatial variance but also bias the color names that appear in the center of the image with more importance. This happens because the cluster located in the center of an image results in shorter distances to other clusters' centroids compared with the ones located at the image edges. The feature maps created by spatial compactness are shown in **Fig. 6**.

Researchers have previously used color spatial distribution for salient object detection. In **[13]**, the spatial distribution of a component is described by computing the intra spatial variance of the color. In addition, to emphasize the color components that located at the image center, a center weight is introduced. Our method is similar to **[13]** that each pixel is assigned to a color cluster (component) in a probabilistic manner; however, relative spread is introduced in our method to compute the spatial compactness of clusters and bias the colors that occur near the center of the image. **Fig. 6** shows some feature maps produced by our method and the method in **[13]**. Comparison of two methods in the 1000-image dataset is shown in **Fig. 8**.



**Fig. 6.** From left to right are original images, feature maps created by the method in **[13]** and feature maps created by our method.

Now consider saliency caused by regional contrast. In section 3.2, we segmented the image into regions and described each region with a discriminative and compact color descriptor CN. The saliency of a region is based on its contrast with respect to other regions in the image in the color domain. The contrast of region  $k$  belongs in cluster  $i$  can be quantified by the inter-region distance between the color descriptors as follows:



$$S_{ikcd} = \sum_{r_l \neq r_{ik}} D_c(r_l, r_{ik}) \quad (11)$$

with

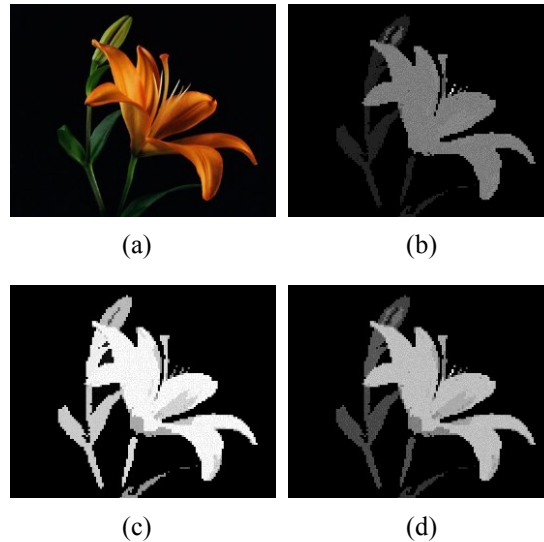
$$D_c(r_1, r_2) = D(CN_1, CN_2) \quad (12)$$

Where  $CN_1$  and  $CN_2$  are the color name descriptors of  $r_1$  and  $r_2$ , respectively, and  $D(CN_1, CN_2)$  is the Euclidean distance between the two color descriptors.

We combine region contrast with cluster compactness by propagating the color cluster's saliency values to the corresponding regions. A more compact and higher contrast region is more salient. Hence, the saliency of region  $k$  in color cluster  $i$  is as follows:

$$S_{ik} = S_{ikcd} \cdot \exp(S_{isd}) \quad (13)$$

In the experiment, we find that the relative spatial compactness measure,  $S_{isd}$ , is more discriminative than the region contrast,  $S_{ikcd}$ . Therefore, we use an exponential function to emphasize spatial compactness [28]. Before converging, both feature maps,  $S_{isd}$  and  $S_{ikcd}$ , are normalized to the range  $[0,1]$  using the formula  $((pixel\_value - Min\_value)) / ((Max\_value - Min\_value))$ . A saliency map is then generated according to Equation 13, and normalized to be a gray-scale image, as shown in Fig. 7.



**Fig. 7.** Illustration of salient object detection using our model. (a) Original image, (b) feature map of cluster spatial compactness, (c) feature map generated by regional contrast and (d) final saliency map.

## 4. Experimental Results and Analysis

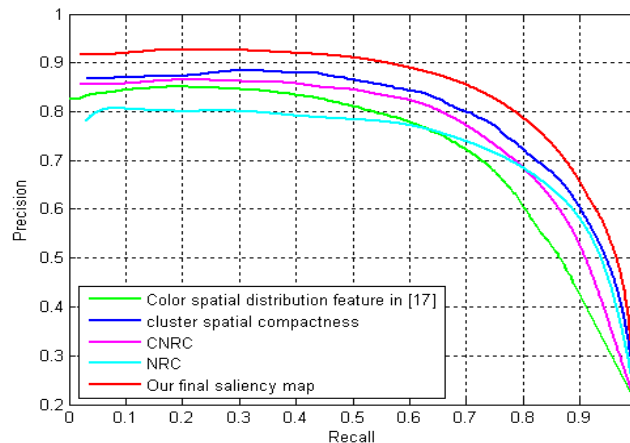
In saliency detection, several datasets are publicly available. One is the MSRA [20] dataset, which includes 20,000 images labeled by three users and 5000 images labeled by nine users. The 1000-image dataset is derived from the MSRA dataset by Achanta et al. [15], creating an object-contour ground truth database. The background structures of images from 1000-image dataset are primarily simple and smooth. Therefore, another complex scene saliency dataset (CSSD) with 200 images is constructed [30]. The images of CSSD are collected from BSD300 [31], VOC dataset and internet. When learning color names, we chose 100 images for every color from the MSRA dataset. For salient object detection, the 1000-image and CSSD datasets

are used.

To evaluate the results of our approach, we employ two evaluation methodologies provided by Achanta [15]. In the first method, a fixed threshold is adopted to obtain a binary segmentation. The segmented image is then compared with the ground truth mask to obtain the precision and recall. To reliably compare how well various saliency detection methods highlight salient objects in images, a threshold,  $T$ , is set from 0 to 255. As  $T$  is changed, different precision-recall pairs are obtained, and a precision-recall curve is the result.

For the second evaluation method, an adaptive threshold is adopted. The saliency map is first over-segmented using mean-shift, and then the average saliency is calculated for every region and for the entire image. Regions with saliency more than twice the average image saliency are set as foreground. Average precision, recall, and F-measure are compared over the entire ground-truth database. F-measure is defined as  $F = ((1 + \beta^2) \text{precision} * \text{recall}) / (\beta^2 \text{precision} + \text{recall})$ , where  $\beta^2 = 0.3$  as in [7,15].

**Comparisons with different components.** As mentioned above, our final saliency map is a combination of two feature maps, cluster spatial compactness and region contrast. The precision-recall curves with a fixed threshold for different components are shown in Fig. 8; the performance of color spatial distribution in [13] and the RC method without spatial weighting (NRC) [7] are also included. For the NRC, we use the authors' implementation, while for the spatial distribution in [13], we re-implemented the algorithm, and the pixels' probabilities of different components are computed by  $p(z|w)$  as is the case in our model. The results show that our cluster spatial compactness, measured by relative spread, performs better than the approach used in [13]. In addition, we compared our color name regional contrast method, CNRC, with NRC. CNRC yields higher precision than NRC, which coincides with the comparison between the color descriptors CN and the RGB histograms. As for image segmentation before region contrast computation, we attempted the graph-based segmentation method as in [7]; however, this resulted in worse performance when combined with a CN descriptor. The segmentation method based on pixel assignment is more consistent with CN and generally yields better results.

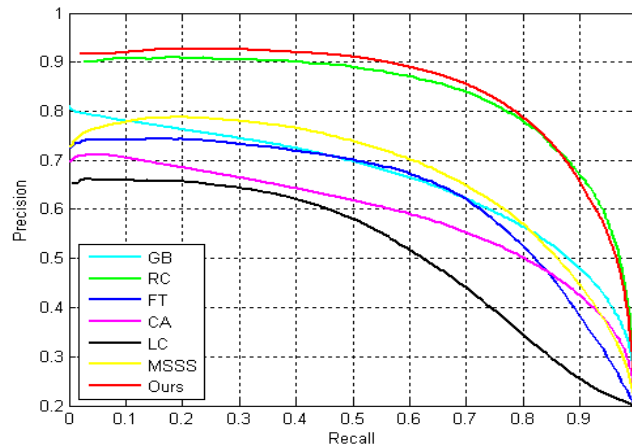


**Fig. 8.** Precision-recall curves of our method compared with different components on the 1000-image dataset. The performance of color spatial distribution in [13] and NRC in [7] are also included.

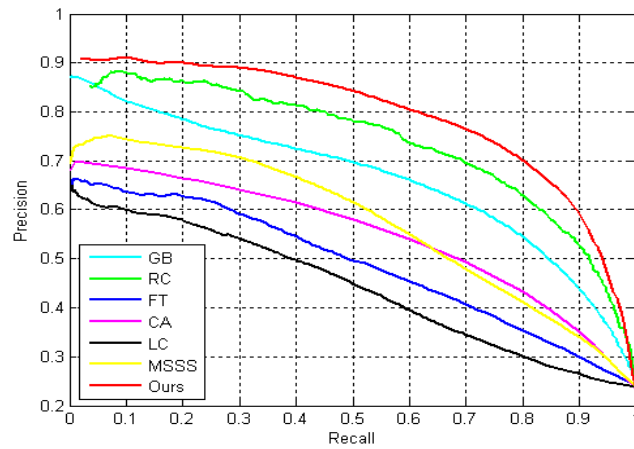
**Comparisons with state-of-the-art methods.** We compared our method with the following 6 state-of-the-art methods: GB[16], FT[15], CA[14], LC[17], MSSS[18] and RC[7]. All of these methods have been recently published and are cited frequently; moreover, RC is related to our method.

The precision-recall curves on the 1000-image dataset are shown in **Fig. 9**. Our method has a better precision-recall curve than GB, FT, CA, LC and MSSS. Compared to RC, our method achieves higher accuracy with a high threshold and becomes lower when the recall is above 0.85. This is because the salient pixels from our method fall well within the salient regions and have near-uniform values, but don't cover the entire salient object such as RC when the threshold is too small. When the image scenes are more complex, our method results in better performance than RC. **Fig. 10** shows the precision-recall curves of different methods on the CSSD dataset. All the methods result in inferior performance on this dataset than on the 1000-image dataset except GB. However, our method consistently outperforms existing saliency methods on the CSSD dataset, yielding higher precision and better recall rates. Both our method and RC require over-segmentation and calculate saliency by regional contrast; however, our method is based on the color name descriptor, which is more discriminative and compact than RGB histograms. Moreover, we take the color clusters' spatial compactness into account and combine it with the regional contrast method. Both improvements contribute to the high precision of our method.

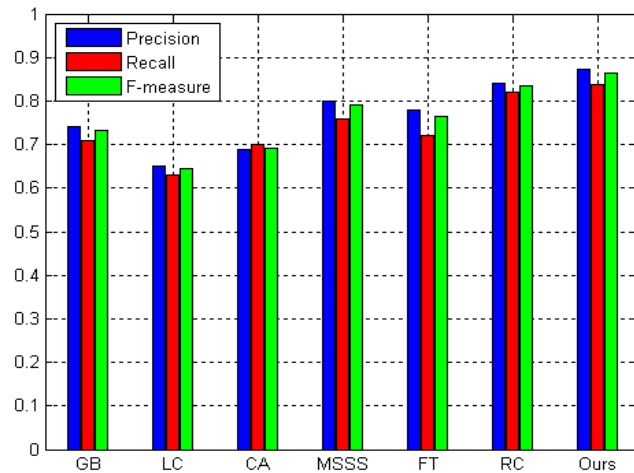
As shown in **Fig. 11** and **Fig. 12**, our method also obtains the best precision, recall and F-measure on both the 1000-image and the CSSD datasets. Saliency maps of our method contain more pixels with higher saliency value. In other words, our saliency maps can more easily be segmented with the sample algorithm. The F-measure for the RC method seems to be different from the one that is demonstrated in the original paper. The reason is that saliency maps of RC can be well-segmented by a complicated segmentation method, such as GrabCut [29], which is iteratively applied to improve the saliency cut. However, we use sample adaptive-thresholding segmentation, which only considers the saliency values' relation between segments and the entire image. Thus, the F-measure of the RC method obtained by our method will be lower than the one in the original paper.



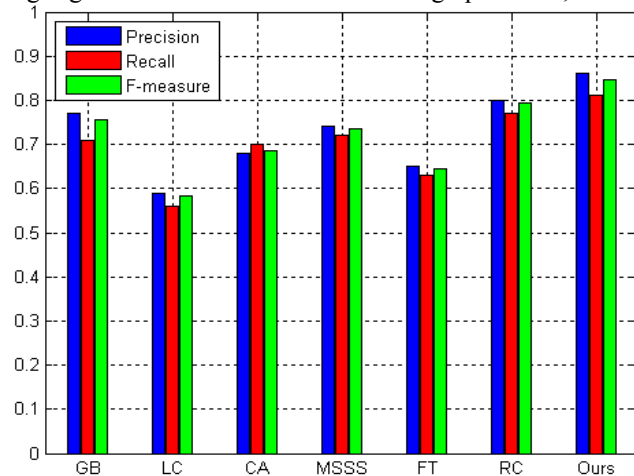
**Fig. 9.** Precision-Recall curves on the 1000-image dataset compared with GB[16], FT[15], CA[14], LC[17], MSSS[18] and RC[7].



**Fig. 10.** Precision-Recall curves on the CSSD dataset compared with GB[16], FT[15], CA[14], LC[17], MSSS[18] and RC[7].

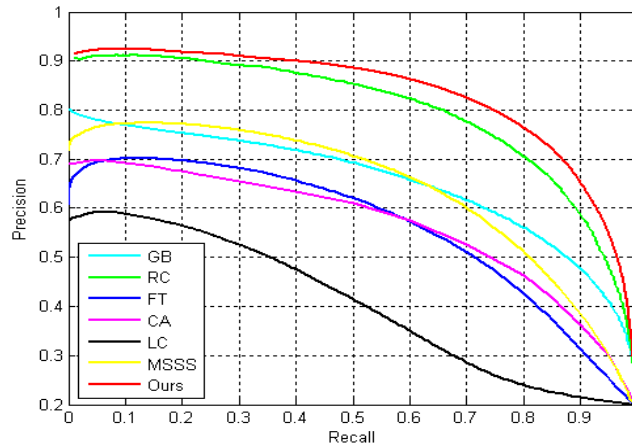


**Fig. 11.** Precision-recall bars over the 1000-image dataset for different methods using adaptive-thresholding segmentation. Our method shows high precision, recall, and F-measure.



**Fig. 12.** Precision-recall bars over the CSSD dataset for different methods using adaptive-thresholding segmentation.

**Performance evaluation on noisy dataset.** The proposed method works properly for noise free images, its performance is still not investigated for noisy images. So, we compare our method with other state-of-the-art methods under the noise corruption images. We add Gaussian noise of mean 0 and variance 0.05 to the 1000-image dataset, and then test the detection performance. **Fig. 13** shows the precision-recall curves of all the methods. **Fig. 14** shows a sample of visual comparison in noisy image.

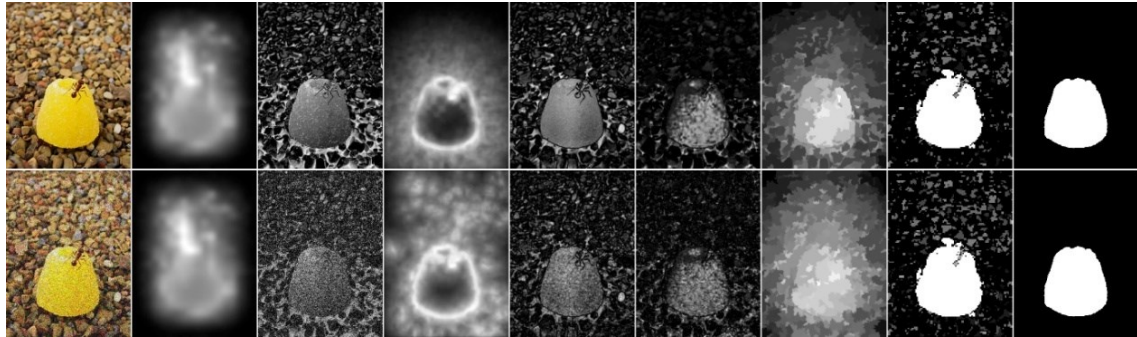


**Fig. 13.** Precision-Recall curves on the noise corruption images compared with GB[16], FT[15], CA[14], LC[17], MSSS[18] and RC[7].

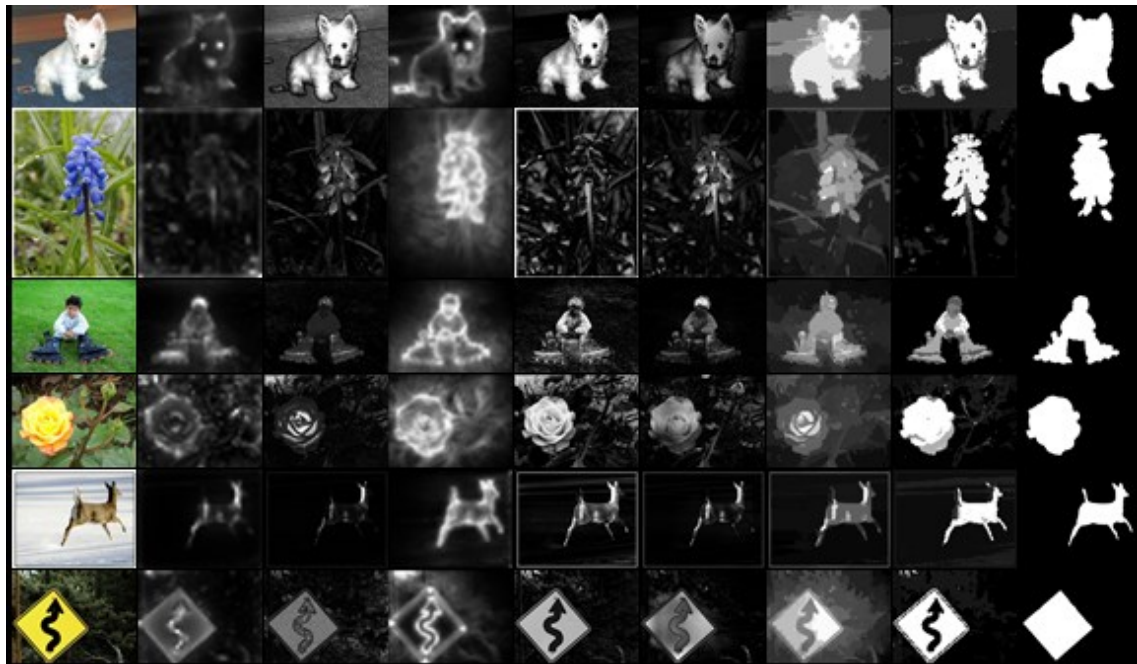
Some visual comparisons of saliency maps produced by different methods are shown in **Fig. 15**. Experiments show that our method is more effective for objects with vivid colors, such as flowers and road signs, or for objects with high contrast with backgrounds in color. Moreover, our method works well for noisy and complex scene images. For images in which color plays less of a role in saliency, our method does not provide that much benefit, as shown in **Fig. 16**.

## 5. Conclusions

In this study, we propose a novel unified framework for saliency object detection. In the model, an image is treated as a composition of eleven basic color names. Every pixel belongs to a color name in a probabilistic manner. Based on that idea, the images are segmented into different color clusters, and each cluster includes some regions. The spatial compactness of clusters is measured by the relative spread of the color names, and the feature map of the spatial distribution can be obtained. Then, we consider the regional contrast in which regional saliency is quantified by inter-region distances between the color name descriptors. Finally, we take the spatial distribution of the color names into account and modulate the region's saliency value by the corresponding cluster compactness. Our approach achieves the best results compared with some state-of-the-art methods when applied to the public datasets. Furthermore, it will obtain better performance when considering task-dependent applications such as road sign detection, which requires further study.



(a) Original (b) GB (c) LC (d) CA (e) MSSS (f) FT (g) RC (h) Ours (i) GT  
**Fig. 14.** Salient objects extracted using different methods for noisy image. The saliency maps created by noise free image are shown in the first row, and the saliency maps created by Gaussian noise image are in the second row.



(a) Original (b) GB (c) LC (d) CA (e) MSSS (f) FT (g) RC (h) Ours (i) GT  
**Fig. 15.** Examples of salient objects extracted using different methods. (a) Original images, saliency regions produced by (b) GB, (c) FT, (d) CA, (e) LC, (f) MSSS, (g) RC and (h) our method. (i) Human labeled ground truth.



**Fig. 16.** Example images in which color plays a smaller role in saliency. The original images are in the first row, and the saliency maps are in the second row.

## References

- [1] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000Still Image Coding System: An Overview," *IEEE Trans. Consumer Electronics*, 46(4), pp.1103-1127, (2000). [Article \(CrossRef Link\)](#)
- [2] C. Guo and L. Zhang, "A Novel Multi resolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," *IEEE Trans. Image Processing*, 19(1), pp.185-198,(2010). [Article \(CrossRef Link\)](#)
- [3] B.C. Ko and J.-Y. Nam, "Object-of-Interest Image Segmentation Based on Human Attention and Semantic Region Clustering," *J. Optical Soc. of Am. A*, 23(10), pp.2462-2470, (2006). [Article \(CrossRef Link\)](#)
- [4] J. Han, et al., "Unsupervised Extraction of Visual Attention Objects in Color Images," *IEEE Trans. Circuits and Systems for Video Technology*, 16(1), pp.141-145, (2006). [Article \(CrossRef Link\)](#)
- [5] L. Itti, C. Gold, and C. Koch, "Visual attention and target detection in cluttered natural scenes," *Optical Engineering*, 40(9), pp. 1784-1793, (2001). [Article \(CrossRef Link\)](#)
- [6] D. Walther et al., "Selective Visual Attention Enables Learning and Recognition of Multiple Objects in Cluttered Scenes," *Computer Vision and Image Understanding*, vol. 100, nos. 1/2, pp.41-63,(2005). [Article \(CrossRef Link\)](#)
- [7] M.M. Cheng, G.X. Zhang, N. J. Mitra, X. Huang, and S.M. Hu, "Global contrast based salient region detection," *CVPR*, pp. 409–416, (2011). [Article \(CrossRef Link\)](#)
- [8] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.* 34(10), pp. 1915-1926, (2012) [Article \(CrossRef Link\)](#)
- [9] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," *ICCV*, pp. 2232 - 2239, (2009). [Article \(CrossRef Link\)](#)
- [10] A.Toet, "Computational versus Psychophysical Bottom-Up Image Saliency: A Comparative Evaluation Study," *IEEE Trans. Pattern Anal. Mach. Intell.* 33(11), pp. 2131-2146, (2011). [Article \(CrossRef Link\)](#)
- [11] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun, "Salient Object Detection by Composition," *ICCV*, pp. 1028-1035, (2011). [Article \(CrossRef Link\)](#)
- [12] J. van de Weijer, C. Schmid, J. J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Trans. Image Processing*, 18(7), pp. 1512–1524, (2009). [Article \(CrossRef Link\)](#)
- [13] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," in *Proc. of CVPR*, 2007. [Article \(CrossRef Link\)](#)
- [14] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. of CVPR*, 2010. [Article \(CrossRef Link\)](#)

- [15] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. of CVPR*, pp. 1597–1604,(2009). [Article \(CrossRef Link\)](#)
- [16] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. of NIPS*, (2007). [Article \(CrossRef Link\)](#)
- [17] Y. Zhai and M. Shah. "Visual attention detection in video sequences using spatiotemporal cues," *ACM Multimedia*, pp. 815–824, (2006). [Article \(CrossRef Link\)](#)
- [18] R.Achanta and S. Süsstrunk. "Saliency detection using maximum symmetric surround," *ICIP*, pp. 2653-2656, (2010). [Article \(CrossRef Link\)](#)
- [19] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* 20(11), pp. 1254–1259, (1998). [Article \(CrossRef Link\)](#)
- [20] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.* 33(2), pp. 353–367,(2011). [Article \(CrossRef Link\)](#)
- [21] Rahtu E, Kannala J, Salo M, and Heikkilä J, "Segmenting salient objects from images and videos," *Computer Vision–ECCV 2010*, 2010: 366-379. [Article \(CrossRef Link\)](#)
- [22] F. S. Khan, R.M. Anwer, J. van de Weijer, A. D. Bagdanov, M. Vanrell, and A. M. Lopez, "Color Attributes for Object Detection," in *Proc. of CVPR*, (2012). [Article \(CrossRef Link\)](#)
- [23] B. Berlin and P. Kay, "Basic Color Terms: Their Universality and Evolution," University of California Press, Berkeley, CA, 1991.
- [24] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 50–57,(1999). [Article \(CrossRef Link\)](#)
- [25] Q. Zhang, H. Liu, J. Shen, G. Gu, and H. Xiao, "An Improved Computational Approach for Salient Region Detection," *Journal of Computers*, 5(7), pp. 1011-1018, (2010). [Article \(CrossRef Link\)](#)
- [26] K.E.A. van de Sande, T. Gevers, and C.G.M. Snoek, "Evaluating Color Descriptors for Object and Scene Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* 32(9), pp. 1582-1596,(2009). [Article \(CrossRef Link\)](#)
- [27] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient Region Detection by Modeling Distributions of Color and Orientation," *IEEE Trans on Multimedia*, 11(5), pp. 892-905, (2009). [Article \(CrossRef Link\)](#)
- [28] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. "Saliency Filters: Contrast Based Filtering for Salient Region Detection," in *Proc. of CVPR*, pp. 733–740, (2012). [Article \(CrossRef Link\)](#)
- [29] C. Rother, V. Kolmogorov, and A. Blake. "Grabcut"– Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004. [Article \(CrossRef Link\)](#)
- [30] Q. Yan, L.Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. of CVPR*, pp. 1–8, (2013).
- [31] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. of ICCV*, volume 2, pages 416–423, July 2001. [Article \(CrossRef Link\)](#)





**Dan Xu** received her M.S. degree in computer application technology from Jiangsu University of Science and Technology in Zhenjiang, China, in 2006. She is currently pursuing her Ph.D. degree in the Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. Her research interests include road sign detection and recognition and visual attention in computer vision.



**Zhenmin Tang** received his M.S. and Ph.D. degree in computer application technology and pattern recognition from Nanjing University of Science and Technology in 1988 and 2002 respectively. He works as a professor in the same university from 1999. His researches include image processing, pattern recognition and intelligent robot.



**Wei Xu** received his bachelor's degree in computer science and technology from Nanjing University of Science and Technology in Nanjing, China, in 2009. He is currently pursuing his Ph.D. degree in the same university. His research interests include salient object detection, color transfer between images, and pavement distress recognition in the intelligent transportation systems.