

Salient Region Detection by Modeling Distributions of Color and Orientation

Viswanath Gopalakrishnan, *Student Member, IEEE*, Yiqun Hu, and Deepu Rajan, *Member, IEEE*

Abstract—We present a robust salient region detection framework based on the color and orientation distribution in images. The proposed framework consists of a color saliency framework and an orientation saliency framework. The color saliency framework detects salient regions based on the spatial distribution of the component colors in the image space and their remoteness in the color space. The dominant hues in the image are used to initialize an expectation-maximization (EM) algorithm to fit a Gaussian mixture model in the hue-saturation (H-S) space. The mixture of Gaussians framework in H-S space is used to compute the inter-cluster distance in the H-S domain as well as the relative spread among the corresponding colors in the spatial domain. Orientation saliency framework detects salient regions in images based on the global and local behavior of different orientations in the image. The oriented spectral information from the Fourier transform of the local patches in the image is used to obtain the local orientation histogram of the image. Salient regions are further detected by identifying spatially confined orientations and with the local patches that possess high orientation entropy contrast. The final saliency map is selected as either color saliency map or orientation saliency map by automatically identifying which of the maps leads to the correct identification of the salient region. The experiments are carried out on a large image database annotated with “ground-truth” salient regions, provided by Microsoft Research Asia, which enables us to conduct robust objective level comparisons with other salient region detection algorithms.

Index Terms—Image modeling, salient regions, visual attention.

I. INTRODUCTION

THE explosive growth of multimedia content has thrown open a variety of applications that calls for high-level analysis of the multimedia corpus. Information retrieval systems have moved on from traditional text retrieval to searching for relevant images, videos and audio segments. Image and video sharing is becoming as popular as text blogs. Multimedia consumption on mobile devices requires intelligent adaptation of the content to the real estate available on terminal devices. Given the huge amount of data to be processed, it is natural to explore ways by which a subset of the data can be judiciously selected so that processing can be done on this subset to achieve the objectives of a particular application.

Manuscript received August 23, 2008; revised February 21, 2009. First published May 02, 2009; current version published July 17, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jie Yang.

The authors are with the Centre for Multimedia and Network Technology, School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: visw0005@ntu.edu.sg; yqhu@ntu.edu.sg; asdrajan@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2009.2021726

Visual attention is the mechanism by which a vision system picks out relevant parts of a scene. In the case of the human visual system (HVS), the brain and the vision system work in tandem to identify the relevant regions. Such regions are indeed the salient regions in the image. Extraction of salient regions facilitate further processing to achieve high-level tasks, e.g., multimedia content adaptation for small screen devices [3], image zooming and image cropping [34], and image/video browsing [6]. The zooming application can be implemented by identifying the centroid of the saliency map and placing the zoom coordinates accordingly. We have described an example of image cropping for display on small screen devices in [13]. Saliency maps can also be used for salient object segmentation by using the segmentation information of the original image [1], [11]. The accuracy of saliency maps plays a crucial role in such segmentation tasks.

While the answer to “what is salient in an image” may be loaded with subjectivity, it is nevertheless agreed upon that certain common low-level processing governs the task independent visual attention mechanism in humans. The process eventually results in a majority agreeing to a region in an image as being salient. The issue of subjectivity is partly addressed in [22] where a large database of images (about 5000) is created and nine users annotate what they consider to be salient regions. The ground truth data set, on which our own experiments are based, provides a degree of objectivity and enables the evaluation of salient region extraction algorithms.

Visual attention analysis has generally progressed on two fronts: bottom-up and top-down approaches. Bottom-up approaches resemble fast, stimulus driven mechanisms in pre-attentive vision for salient region selection and are largely independent of the knowledge of the content in the image [9], [16], [25], [30]. On the contrary, top-down approaches are goal oriented and make use of prior knowledge about the scene or the context to identify the salient regions [7], [8], [28]. Bottom-up approaches are fast to implement and require no prior knowledge of the image, but the methods may falter in distinguishing between salient objects and irrelevant clutter which are not part of salient object when both regions have high local stimuli. Top-down methods are task dependent and demands more complete understanding of the context of the image which results in high computational costs. Integration of top-down and bottom-up approaches has been discussed in [27].

Some research on salient region detection attempt to first define the salient regions or structures before detecting them. In [21], scale space theory is used to detect salient blob like structures and their scales. Reisfeld *et al.* introduce a context free attention operator based on an intuitive notion of symmetry [32]. They assume symmetric points are more prone to be attentive

and develop a symmetry map based on gradients on grey level images. In [12], the idea is extended to color symmetry. The method focuses on finding symmetric regions on color images which could be missed by the grey scale algorithms. The underlying assumption in these methods is that symmetric regions are attention regions. However, for general salient region extraction, this is a very tight constraint.

Kadir *et al.* retrieve “meaningful” descriptors from an image from the local entropy or complexity of the regions in a bottom-up manner [19]. They consider saliency across scales and automatically select the appropriate scales for analysis. A self-information measure based on local contrast is used to compute saliency in [2]. In [8], a top-down method is described in which saliency is equated to discrimination. Those features of a class which discriminate it from other classes in the same scene are defined as salient features. Gao *et al.* extend the concept of discriminant saliency to bottom-up methods inspired by “center-surround” mechanisms in pre-attentive biological vision [9]. Zhiwen *et al.* discuss a rule based algorithm which detects hierarchical visual attention regions at the object level in [40]. By detecting salient regions at object level they try to bridge the gap between traditional visual attention regions and high-level semantics. In [14], a new linear subspace estimation method is described to represent the 2-D image after a polar transformation. A new attention measure is then proposed based on the projection of all data to the normal of the corresponding subspace. Liu *et al.* propose to extract salient regions by learning local, regional and global features using conditional random fields [22]. They define a multiscale contrast as the local feature, center-surround histogram as the regional feature and color spatial variance as the global feature.

In this paper we analyze the influence of color and orientation in deciding the salient regions of an image in a novel manner. The definition of saliency revolves around two main concepts in the proposed paper. 1) *Isolation (rarity)* and 2) *compactness*. For colors, isolation refers to the rarity of a color in the color space while compactness refers to the spatial variance of the color in the image space. For orientations, isolation refers to the rarity of the complexity of orientations in the image space in a local sense, while compactness refers to the spatial confinement of the orientations in the image space. The goal of the proposed method is to generate the **saliency map**, which indicates the salient regions in an image. It is not the objective here to segment a salient object, which can be viewed as a segmentation problem. As mentioned earlier the saliency maps can be used for zooming, cropping and segmentation of images.

The role of color and orientation in deciding the salient regions has been studied previously also as can be seen in early works of Itti *et al.* [16]. Itti’s works on color and orientation are motivated by the behavior of neurons in the receptive fields of human visual cortex [20]. The incomplete understanding of human visual system brings limitations to such HVS based models. In the proposed work, we look at the distribution of color and orientations in the image separately, propose certain saliency models based on their respective distributions and design a robust framework to detect the final salient regions. The proposed framework can be divided into a color saliency framework (CSF) and an orientation saliency framework (OSF). The final saliency map is selected from either one of these models in a systematic manner. Robust objective measures are employed on a large user anno-

tated image database provided by Microsoft Research Asia to illustrate the effectiveness of the proposed method.

The CSF is motivated by the fact that the colors in an image play an important role in deciding salient regions. It has been shown in [33] that a purely chromatic signal can automatically capture visual attention. In [18], the contribution of color in visual attention is objectively evaluated by comparing the computer models for visual attention on gray scale and color images with actual human eye fixations. It has been observed that the similarity between the model and the actual eye fixation is improved when color cues are considered in the visual attention modeling. Weiher *et al.* exploit the possibilities of color distinctiveness in addition to shape distinctiveness for salient point detection in [36]. Methods relying on shape distinctiveness use gradient strength of luma difference image to compute salient points. In contrast, [36] uses the information content of the color differential images to compute salient points in the color image.

We believe that a color-based saliency framework should consider the global distribution of colors in an image as opposed to localized features described in [16] and [23]. In [16], a center-surround difference operator on Red-Green and Blue-Yellow colors that imitates the color double opponent property of the neurons in receptive fields is used to compute the color saliency map. Similarly in [23], the local neighborhood contrasts of the LUV image is used to generate the saliency map. A fuzzy region growing method is then used to draw the final bounding box on the saliency map. It can be seen that [16] and [23] use highly localized features and the global aspect of colors in the image has not been considered. For example, the saliency of a red apple in the background of green leaves is attributed more to the global distribution of color than to the local contrasts. The saliency of colors based on its global behavior in the image has been discussed in literature in the works of Liu *et al.* [22] and Zhai *et al.* [39]. In [22], it is assumed that lesser the spatial variance of a particular color, the more likely it is to be salient. This assumption may fail when there are many colors in an image with comparable spatial variances but with varying saliencies. Zhai *et al.* [39] compute the spatial saliency by looking at the 1-D histogram of a specific color channel (e.g., R channel) of the image and computing the saliency values of different amplitudes in that channel based on their rarity of occurrence in the histogram. This method focuses on a particular color channel and does not look at the relationship between various colors in the image. In [38], a region merging method is introduced to segment salient regions in a color image. The method focuses on image segmentation so that the final segmented image contains regions that are salient. However, their notion of saliency is related simply to the size of the region—if the segmented region is not large enough, it is not salient. Similarly in [31], a nonparametric clustering algorithm that can group perceptually salient image regions is discussed. The proposed work differs from [38] and [31], as ours is a saliency map generation method while [38] and [31] essentially focus on image segmentation so that the final segmented image contains different regions that are salient.

In the proposed color saliency framework we look at the saliency of colors rather than the saliency of regions. We introduce the concepts of “compactness” of a color in the spatial domain and is “isolation” in the H-S domain to define the saliency of the color. Compactness refers to the relative

spread of a particular color with respect to other colors in the image. *Isolation* refers to the distinctiveness of the color with respect to other colors in the color space. We propose a measure to quantify compactness and isolation of various competing colors and probabilistically evaluate the saliency among them in the image. The saliency of the pixels is finally evaluated using probabilistic models. Our framework considers salient characteristics in the feature space as well as in the spatial domain, simultaneously.

The proposed OSF identifies the salient regions in the image by analyzing the global and local behavior of different orientations. The role of orientation and of spectral components in an image to detect saliency have been explored in [16] and [37], respectively. In [16], the orientation saliency map is obtained by considering the local orientation contrast between center and surround scales in a multiresolution framework. However, a nonsalient region with cluttered objects will result in high local orientation contrast and mislead the saliency computation. In [37], the gist of the scene is represented with the averaged Fourier envelope and the differential spectral components are used to extract salient regions. This method works satisfactorily for small salient objects but fails for larger objects since the algorithm interprets the large salient object as part of the gist of the scene and consequently fails to identify it.

In [19], Kadir *et al.* consider the local entropy as a clue to saliency by assuming that flatter histograms of a local descriptor correspond to higher complexity (measured in terms of entropy) and hence, higher saliency. This assumption is not always true since an image with a small smooth region (low complexity) surrounded by a highly textured region (high complexity) will erroneously result in the larger textured region as the salient one. Hence, it is more useful to consider the contrast in complexity of orientations and to this end, we propose a novel orientation histogram as the local descriptor and determine how different its entropy is from the local neighborhood, leading to the notion of orientation entropy contrast. In the proposed orientation saliency framework, global orientation information is obtained by determining the spatially confined orientations and local orientation information is obtained by considering their orientation entropy contrast.

Each of CSF and OSF creates a saliency map such that salient regions are correctly identified by the former if color is the contributing factor to saliency and by the latter if saliency in the image is due to orientation. From a global perspective, saliency in an image as perceived by the HVS is derived either from color or orientation. In other words, the visual attention region pops out due to its distinctiveness in color or texture. The final module in our framework predicts the appropriateness of the two saliency maps and selects the one that leads to the correct identification of the salient region. Even if a region in an image is salient due to both color and texture, our algorithm will detect the correct salient region since both the saliency maps will indicate them. Fig. 1 shows the flowchart of the proposed framework for salient region detection using color and orientation distribution information.

The paper is organized as follows. Section II describes the color saliency framework which computes the saliency of component colors of the image and the color saliency map. In Section III, the saliency detection using the orientation distribution and the computation of orientation saliency map is

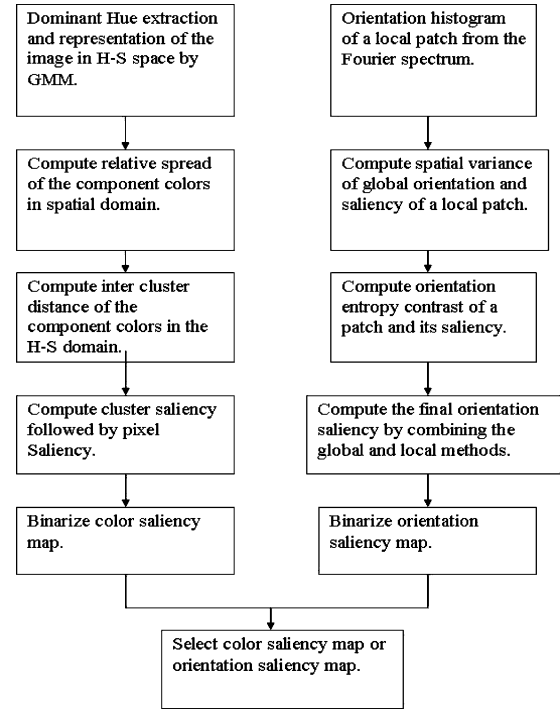


Fig. 1. Proposed framework for salient region detection in an image.

discussed. Section IV discusses the final saliency map selection from the color and orientation saliency maps. Experimental results and comparisons are briefed in Section V. Finally, conclusions are given in Section VI.

II. COLOR SALIENCY FRAMEWORK

This section describes the general framework used to compute color cluster saliency to extract the attention regions on color images. The first step involves identifying the dominant hue and saturation values in the image. The information from the dominant hues provides a systematic way to initialize the expectation-maximization (EM) algorithm for fitting a Gaussian mixture model (GMM) in the hue-saturation (H-S) space. Using the mixture of Gaussians model, the compactness or relative spread of each color in the spatial domain as well as the isolation of the respective cluster in the H-S domain is calculated. These two parameters determine the saliency of each cluster and finally, the color saliency of each pixel is computed from the cluster saliency.

A. Dominant Hue Extraction

The colors represented in the H-S space are modeled as a mixture of Gaussians using the EM algorithm. The critical issue in any EM algorithm is the initialization of the parameters, which in this case includes the number of Gaussians, their corresponding weights, and the initial mean and covariance matrices. If the parameters are not judiciously selected, there is a likelihood of the algorithm getting trapped in a local minimum. This will result in suboptimal representation of the colors in the image.

The initialization problem is solved by identifying the dominant hues in the image using a method similar to [26]. The number of bins in the hue histogram is initially set to 50. This is

a very liberal initial estimate of the number of perceptible and distinguishable hues in most color images. In order to include the effect of saturation, the hue histogram is defined as

$$H(b) = \sum_{(x,y) \in P_b} S(x,y) \quad (1)$$

where $H(b)$ is the histogram value in the b th bin, P_b is the set of image coordinates having hues corresponding to the b th bin, and $S(x,y)$ is the saturation value at location (x,y) . The histogram is further smoothened using a sliding averaging filter that extends over five bins, to avoid any secondary peaks. **The maxima in the final histogram are the dominant hues and the number of dominant hues is set as the number of Gaussian clusters for the EM algorithm.** The range of hues around a peak in the hue histogram constitutes a *hue band* for the corresponding hue. The local minimum (valley) between two dominant hues in the hue histogram is used to identify the boundary of the hue band. The saturation values pertaining to the dominant hue belonging to the k th hue band is obtained as

$$S_k = \{S(x,y) : (x,y) \in D_k\} \quad (2)$$

where D_k is the set of pixels with hues from the k th hue band and $S(x,y)$ is the saturation at (x,y) . The histogram of S_k contains a maximum which is taken as the dominant saturation value of the k th dominant hue. For each hue band, the covariance matrix of the hue-saturation pair is obtained as

$$C_k = \sum_{(x,y) \in D_k} V_k V_k^T \quad (3)$$

where

$$V_k = \begin{bmatrix} H(x,y) \\ S(x,y) \end{bmatrix} - \begin{bmatrix} H_k^{dom} \\ S_k^{dom} \end{bmatrix} \quad (4)$$

$H(x,y)$ and $S(x,y)$ is the hue and saturation at (x,y) , respectively and H_k^{dom} and S_k^{dom} are the maximum hue and saturation values, respectively, in the k th band. **The histogram strengths of dominant hue values normalized between 0 and 1 are selected as the initial weights of the Gaussian clusters.** Thus, we develop a systematic way to initialize the mixture of Gaussians for the EM algorithm. The number of clusters is the number of dominant hues, the initial mean of the Gaussians are the dominant hue and saturation values, the co-variances are determined from (3), and the weights are relative strengths of dominant hues.

Fig. 2 shows an example of the dominant hues extracted and the corresponding Gaussians that initialize the EM algorithm. Fig. 2(a) is the original image. Fig. 2(b) shows the palette of dominant hues in the image and the Gaussian clusters that initialize the EM algorithm are shown in Fig. 2(c). In [38], a color region segmentation algorithm for the YUV image is initialized by identifying the dominant Y, U and V values of the image leading to the processing of a 3-D histogram. The proposed method builds the color model for the image around the 1-D histogram of hues. The method described in [38] can also be

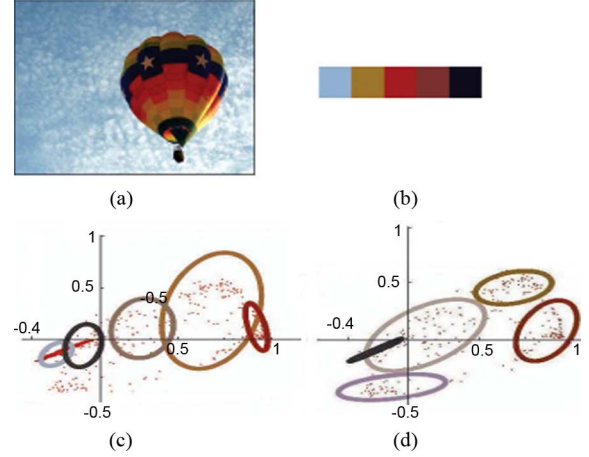


Fig. 2. Initialization and fitting of Gaussian clusters. (a) Original image. (b) Dominant hue palette. (c) Initial Gaussian clusters in hue-saturation space. (d) Gaussian clusters after EM based fitting. Note that in (c) and (d), data points lie in hue-saturation circle, where hue is represented by the positive angle between the line joining the origin (0,0) to the data point and the X-axis and saturation is represented by the radial distance from the origin (0,0).

used to initialize our color saliency framework. However, since our probabilistic framework for color saliency is robust to small errors in the initial identification of dominant colors, we have used a simple method based on the 1-D hue histogram.

B. Model Fitting in H-S Domain

The EM algorithm is performed to fit the GMM to the H-S space representation of the image. The dominant color extraction in the proposed framework is intended to make a reasonable initialization of the Gaussian clusters. Even if the number of clusters (colors) are estimated to be more than what is required, e.g., two clusters for different shades of green, it must be noted that each pixel in the image belongs to all clusters in a probabilistic sense. This will ensure that the final pixel saliency values computed in our framework will be still correct with the extra clusters present. However, the extra clusters increase the computational complexity of the framework. To reduce the computational complexity, we eliminate some of the extra clusters during EM fitting of the Gaussian model. **This is done by inspecting the determinant of the covariance matrix of a cluster. If it falls below a predetermined threshold, the corresponding cluster is removed and the fitting process is continued.** Hence the dominant color module only needs to supply a reasonable estimate of colors in the image. Fig. 2(d) shows the final fitting of the GMM for the original image in Fig. 2(a).

When the modeling is complete, the probability of a pixel at location (x,y) belonging to the i th color cluster is found as [22]

$$p(i|I_{x,y}) = \frac{w_i \mathcal{N}(I_{x,y}|\mu_i, \Sigma_i)}{\sum_i w_i \mathcal{N}(I_{x,y}|\mu_i, \Sigma_i)} \quad (5)$$

where w_i , μ_i , and Σ_i are the weight, mean value and covariance matrix respectively, of the i th cluster.

The spatial mean of the i th cluster is

$$\mu_i^{sp} = \begin{bmatrix} M_x(i) \\ M_y(i) \end{bmatrix} \quad (6)$$

where

$$M_x(i) = \frac{\sum_{(x,y) \in A_{xy}} p(i|I_{x,y}) \cdot x}{\sum_{(x,y) \in A_{xy}} p(i|I_{x,y})} \quad (7)$$

and

$$M_y(i) = \frac{\sum_{(x,y) \in A_{xy}} p(i|I_{x,y}) \cdot y}{\sum_{(x,y) \in A_{xy}} p(i|I_{x,y})} \quad (8)$$

where $I_{x,y}$ is the H-S vector at location (x, y) and A_{xy} is the set of 2-D vectors representing the integer pixel locations of the image.

As noted earlier, color saliency in a region is characterized by its compactness and isolation. Next, we define distance measures that quantify these two parameters.

C. Cluster Compactness (Spatial Domain)

The colors corresponding to the dominant hues and their respective saturation values will appear in the spatial domain as clusters that are spread to different extents over the image. It should be noted that these clusters of pixels in the image belong to different component colors in a probabilistic manner. The less the colors are spread, the more will be its saliency. The colors in the background have larger spread compared to the salient colors. This spread has to be evaluated by the distance in the spatial domain between the clusters rather than by intra color spatial variance alone, as in [22]. The reason is that there can be colors with similar intra spatial variance, but their compactness or relative spread may be different. The relative spread of a cluster is quantified by the mean of the distance of the component pixels of the cluster to the centroids of the other clusters. With the calculation of relative spread, we can not only eliminate the background colors of large spatial variance, but also bias the colors occurring towards the center of the image with more importance, in case the other competing colors in the image possess similar spatial variance. The distance, d_{ij} , between cluster i and cluster j quantifies the relative compactness of cluster i with respect to cluster j , i.e.,

$$d_{ij} = \frac{\sum_{X \in A_{xy}} \|X - \mu_j^{sp}\|^2 \cdot P_i(X)}{\sum_{X \in A_{xy}} P_i(X)} \quad (9)$$

where μ_j^{sp} is the spatial mean of the j th cluster, A_{xy} is the set of 2-D vectors representing the integer pixel locations of the image, $P_i(X)$ is the probability that the H-S value in the spatial location $X \in A_{xy}$ falls into the cluster i .

Finally, the compactness $COMP_i$ of cluster i is calculated as the inverse of the sum of the distances of cluster i from all the other clusters, i.e.,

$$COMP_i = \left(\sum_j d_{ij} \right)^{-1}. \quad (10)$$

Next we look at the clusters in H-S domain representation of the image to calculate the isolation or remoteness of the component colors.

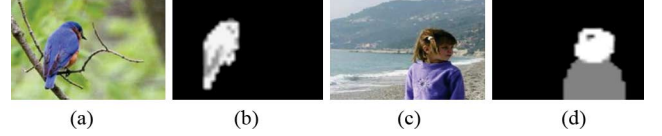


Fig. 3. (a), (c) Original images. (b), (d) Color saliency maps.

D. Cluster Isolation (H-S Domain)

Isolation of a color refers to how distinct it is with respect to other colors in the image. The distinction naturally contributes to the saliency of a region. The inter-cluster distance in H-S space quantifies the isolation of a color. As the image is represented as a mixture of Gaussians in the H-S space, the most isolated Gaussian is more likely to capture the attention. The inter-cluster distance between clusters i and j that quantifies the isolation of a particular color is defined as

$$h_{ij} = \frac{\sum_{Z \in S_{xy}} \|Z - \mu_j^{hs}\|^2 \cdot P_i(Z)}{\sum_{Z \in S_{xy}} P_i(Z)} \quad (11)$$

where μ_j^{hs} is the mean of the H-S vector of the j th cluster, S_{xy} is the set of 2-D vectors representing the hue and the corresponding saturation values of the image, and $P_i(Z)$ is the probability that the H-S value falls into the cluster i .

The isolation ISO_i of cluster i is calculated as the sum of the distances of cluster i from all other clusters in the H-S space, i.e.,

$$ISO_i = \sum_j h_{ij}. \quad (12)$$

E. Pixel Saliency

The saliency for a particular color cluster is computed from its compactness and isolation. A more compact and isolated cluster is more salient. Hence, the saliency Sal_i of a color cluster i is

$$Sal_i = ISO_i \times COMP_i. \quad (13)$$

The saliency values of the color clusters are normalized between 0 and 1. The top down approach, taking the global distribution of colors into account, results in propagating the cluster saliency values down to the pixels. The final color pixel saliency Sal_{Cp} , on pixel p in location $X \in A_{xy}$ is computed as

$$Sal_{Cp} = \sum_i Sal_i \cdot P_i(X) \quad (14)$$

where $P_i(X)$ is the probability that the H-S values at $X = (x, y)$ falls into the cluster i . The pixel saliencies are once again normalized to obtain the final saliency map.

Fig. 3 shows preliminary results to demonstrate the effectiveness of the proposed color saliency framework, when color is indeed the contributing factor to saliency. Fig. 3(a) and (c) shows original images and Fig. 3(b) and (d) shows their corresponding color saliency maps. The color saliency maps are in agreement with what can be considered as visual attention regions for the HVS.



Fig. 4. Example images in which color plays no role in saliency.

III. ORIENTATION SALIENCY FRAMEWORK

From a global perspective, certain colors or orientation patterns provide important cues to saliency in an image. It is this aspect of saliency that we wish to capture in the proposed salient region detection framework. In cases where color is not useful to detect salient regions, it is some property of the dominant orientation in the image that allows such regions to be identified. For example, in the images in Fig. 4, there is no salient color, whereas the wings of the bird in Fig. 4(a) and the orientation of the buildings in the background in Fig. 4(b) act as markers to aid in the detection of salient objects. Note that in the second case, the dominant orientation actually contributes to the nonsalient region so that the structure in the foreground is detected as the salient region.

Determination of saliency in an image due to orientation has been previously studied in works like [16] and [17]. However, these methods consider orientation only as a local phenomenon. In the proposed orientation saliency framework, we model both the local as well as the global manifestations of orientation. In particular, we propose a novel orientation histogram of image patches that characterizes local orientations and their distribution over the image is used to model the global orientation. Furthermore, the orientation entropy of each patch is compared against the neighborhood patches and a high contrast is deemed to be more salient. The final orientation saliency map is obtained by a combination of the global orientation information obtained from the distributions and the local orientation information obtained from the entropy contrasts.

A. Local Orientation Histogram

Local gradient strengths and Gabor filters are two commonly used methods for constructing the orientation histogram of an image [4], [5]. We introduce a new and effective method to compute the orientation histogram from the Fourier spectrum of the image as a prelude to our orientation saliency framework. The absolute values of the complex Fourier coefficients indicates the presence of the respective spatial frequencies in the image and their strengths in a specific direction of the center shifted Fourier transform is an indication of the presence of orientations in a perpendicular direction [10]. Hence, the Fourier transform of

an 8×8 patch in an image is computed and the orientation histogram is found as

$$H(\theta_i) = \sum_{\tan^{-1}(m^*/n^*) \in \theta_i} \log(|F(m,n)| + 1) \quad (15)$$

where $|F(m,n)|$ is the absolute Fourier magnitude in location (m,n) , (m^*, n^*) are shifted locations with respect to the center of Fourier window for the purpose of calculating the orientation of the spectral components and are obtained by subtracting the center of the Fourier window from (m,n) , and θ_i is the set of angles in the i th histogram bin. The orientations ranging from -90° to 90° are divided into bins of 10° each. Logarithm of the Fourier transform is used to reduce the dynamic range of the coefficients. Since the Fourier transform is symmetric over 180° , only the right half of the center shifted Fourier transform is used to build the orientation histogram.

In order to remove the dependency of the high frequency components on the patch boundary, the orientation histogram is calculated as the average histogram on the patch over three different scales of the image, $\sigma_i \in \{0.5, 1, 1.5\}$. The scale space is generated by convolving the image with Gaussian masks derived from the three scales. Fig. 5 shows a simple example in which Fourier transform is used to obtain the orientation histogram of the image. Fig. 5(a) shows an image from the Brodatz texture album. Its center shifted Fourier transform is shown in Fig. 5(b), and the orientation histogram is shown in Fig. 5(c). The orientation histogram peaks at two different angles with one direction being dominant over the other as it is evident from the original image. It is noted that for a patch containing a smooth region there will be no dominant orientation and the histogram will be almost flat. Hence the orientation histogram is modified by subtracting the average orientation magnitude followed by normalization between zero and one. Thus, in the strict sense, the orientation histogram loses the property of a histogram, but it does provide a useful indication of the dominant orientations present in the image.

B. Spatial Variance of Global Orientation

The local orientation histogram is used to analyze the global behavior of the various orientations in the histogram. The orientations that exhibit large spatial variance across the images are less likely to be salient. For example, in the images of Fig. 6, the orientations in the salient regions are confined in a local spatial neighborhood as compared to the rest of the image. The dominant orientations in the background contributed by the door and by the ground in the images in Fig. 6(a) have a large spatial variance and hence are not salient. This property can be objectively measured by analyzing the global behavior of the different orientations ranging from -90° to 90° . The global spatial mean of orientation θ_i is computed as

$$M_x(\theta_i) = \frac{\sum_{(x,y)} x \cdot H(\theta_i)}{\sum_{(x,y)} H(\theta_i)} \quad (16)$$

$$M_y(\theta_i) = \frac{\sum_{(x,y)} y \cdot H(\theta_i)}{\sum_{(x,y)} H(\theta_i)} \quad (17)$$

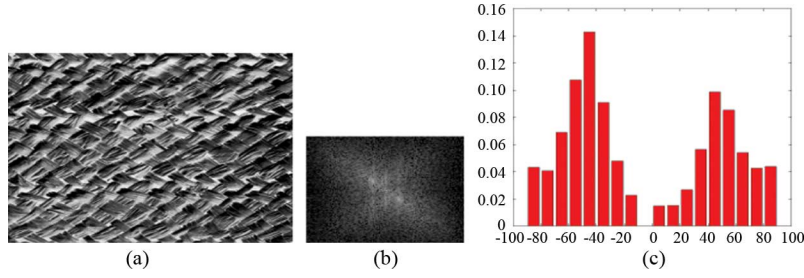


Fig. 5. Orientation histogram from Fourier spectrum. (a) Original image. (b) Center-shifted Fourier transform. (c) Orientation histogram.

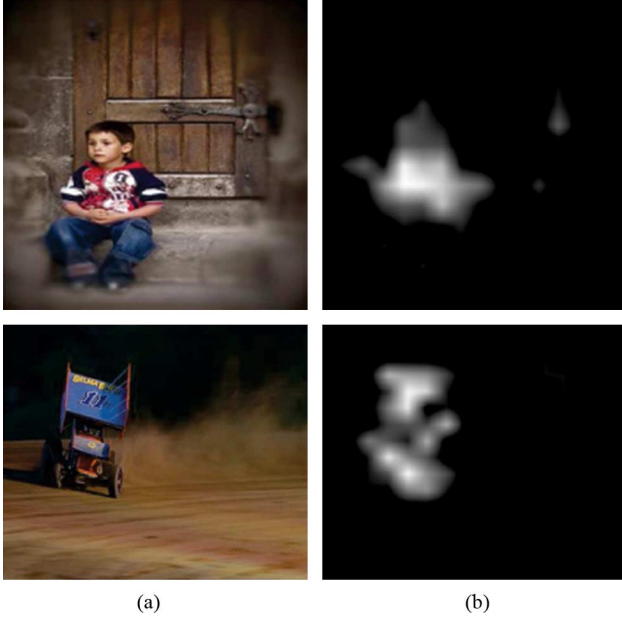


Fig. 6. (a) Original images. (b) Saliency maps extracted by computing spatial variance of global orientation.

where x and y are the image coordinates. The spatial variance of orientation θ_i in the x and y directions are then computed as

$$Var_x(\theta_i) = \frac{\sum_{(x,y)} (x - M_x(\theta_i))^2 \cdot H(\theta_i)}{\sum_{(x,y)} H(\theta_i)} \quad (18)$$

and

$$Var_y(\theta_i) = \frac{\sum_{(x,y)} (y - M_y(\theta_i))^2 \cdot H(\theta_i)}{\sum_{(x,y)} H(\theta_i)} \quad (19)$$

and the total spatial variance $Var(\theta_i)$ is calculated as

$$Var(\theta_i) = Var_x(\theta_i) + Var_y(\theta_i). \quad (20)$$

$Var(\theta_i)$ is a direct measure of the spatial confinement of the i th orientation and hence, its relative strength among the various orientations is used as a direct measure of saliency of the respective orientations. In other words, the saliency of an orientation θ_i , denoted by $Sal(\theta_i)$, is its total spatial variance $Var(\theta_i)$ and

it is normalized so that the orientation having least spatial variance receives a saliency of one and that with maximum variance receives a saliency of zero. Finally, the saliency of a local patch P is calculated as

$$Sal_{gP} = \sum_i Sal(\theta_i) \cdot H_P(\theta_i) \quad (21)$$

where $H_P(\theta_i)$ is the contribution of orientation θ_i in the patch P .

We evaluate the effectiveness of the proposed global salient orientation algorithm on the images in Fig. 6(a). As expected, the first saliency map in Fig. 6(b) shows the boy as the salient region ignoring the background orientation that has a larger spatial variance while the second saliency map in Fig. 6(b) shows the vehicle as the saliency region while excluding the ground in the background.

C. Orientation Entropy Contrast

The spatial variance of the orientations helps in detecting salient objects by virtue of their confinement in the spatial domain and the exclusion of other dominant orientations due to their large variance. However, images in which there are several dominant orientations from the foreground and the background that are confined and hence result in low spatial variance, will compete for saliency according to the method in the previous section. In this case, it is required to determine which among those competing orientations has higher saliency. We next introduce the orientation entropy contrast to address this issue.

The entropy of the orientation histogram of a patch is considered to be a measure of the complexity of the orientation in that patch. Higher entropy indicates more uncertainty in the orientations; this can happen in smooth as well as in highly textured regions with dominant orientations in several directions. Similarly, low entropy is an indication of strong orientation properties contributed by edges and curves. The orientation entropy E_P of patch P with orientation histogram H_P is calculated as

$$E_P = - \sum_{\theta_i} H_P(\theta_i) \cdot \log(H_P(\theta_i)). \quad (22)$$

The orientation complexity is also calculated at three different scales, $\sigma_i \in \{0.5, 1, 1.5\}$, to reduce the dependencies on the selected patch size. Hence the 3-D vector $E_P = [E_{\sigma 1}, E_{\sigma 2}, E_{\sigma 3}]$ represents the final entropy vector on patch P of the image,

where E_{σ_i} represents the orientation entropy calculated at scale σ_i .

Saliency of a local patch is now calculated based on the norm of the vector difference between its entropy and the entropy of the neighboring patches. Such a measure of entropy contrast is better than [16] since it is not biased towards edges, smooth regions or textured regions. Any region with such characteristics can be salient depending on the entropy of the neighboring regions. Hence, the final criterion of salient region selection is how discriminating a particular region is, in terms of its orientation complexity. Furthermore, in the spatial domain, as the Euclidian distance between the patches increases, the influence of the entropy contrast decreases. Integrating the distance factor into the saliency computation, we obtain the saliency of a patch P , Sal_{IP} , based on local orientation entropy contrast as

$$Sal_{IP} = \sum_{Q, Q \neq P} \frac{\|E_P - E_Q\|}{\|D_{PQ}\|} \quad (23)$$

where E_Q is the orientation entropy vector of patch Q and $\|D_{PQ}\|$ is the Euclidian distance between the centers of patches P and Q . For a particular patch of size 8×8 , we consider all the remaining patches in the image as neighborhood patches. However, their contribution to the orientation saliency is inversely weighted by the Euclidean distance between the neighborhood patch and the particular patch under consideration. In this way, the major contribution to the entropy contrast will come from spatially neighboring patches. The saliency of a patch is normalized similar to Section III-B.

Fig. 7(a) shows two images in which there are dominant textures in the background that are confined to a spatial region so that their spatial variance is small. However, the contrast in entropy of such regions with the surrounding regions is not very high. On the other hand, the foreground objects exhibit high distinctiveness in its orientation complexity with the surrounding regions. These objects are rightly determined as salient regions in the salient maps shown in Fig. 7(b).

D. Final Orientation Saliency

The two approaches adopted for orientation based saliency complement each other. The first method using spatial variance of global orientation searches for spatially confined orientations and in the process detects salient objects as well as cluttered nonsalient regions. The cluttered regions will have low contrast of orientation entropy and hence will be eliminated by the second method that used orientation entropy contrast. At the same time, long nonsalient edges in the images like horizon, fence etc. can possess high orientation entropy contrast depending on the neighborhood. But those orientations will have high spatial variance and hence will be eliminated by the global orientation variance method. For example, in the boy image in Fig. 6(a), the latch and the planks on the wooden door has high orientation entropy contrast, but is eliminated by the first method. Similarly, in the images in Fig. 7(a) the background is rich in orientations that are confined in their spatial neighborhood. This misleads the global variance computation, but

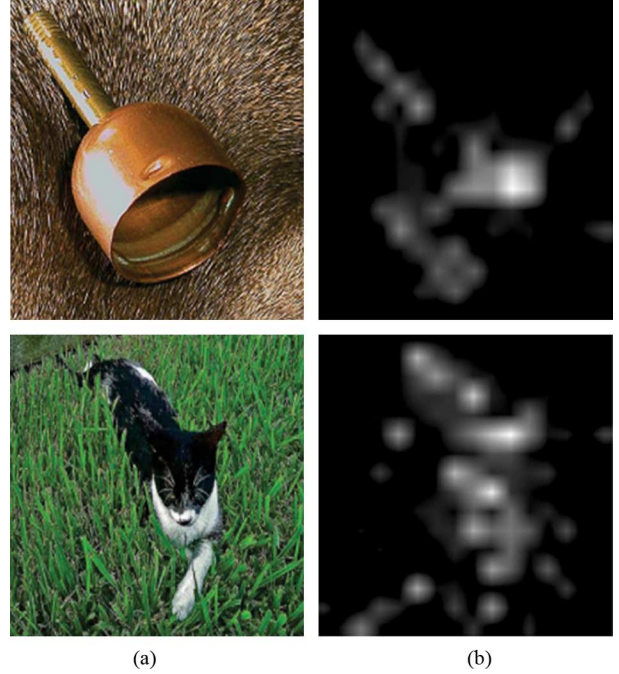


Fig. 7. (a) Original images. (b) Saliency map based on orientation entropy contrast method.

the orientation entropy contrast rightly eliminates such regions. The final orientation saliency of patch P is computed as

$$Sal_{OP} = Sal_{gP} \times Sal_{IP}. \quad (24)$$

It will have a maximum value on locations where both the methods agree and vary proportionately in regions that favor only one of the methods.

IV. SALIENCY MAP SELECTION

The color saliency framework and orientation saliency framework produce two different saliency maps. Previous methods such as [15] and [16] fuse the individual saliency maps corresponding to each feature based on weights assigned to each map. Our premise is that *salient regions in most images can be globally identified by certain colors or orientations*. Since saliencies occur at a global level rather than at a local level, it is not feasible to combine the saliency maps through weights, for this may result in a salient region detection with less accuracy than when individual maps are considered. We introduce a systematic way by which the correct saliency map can be estimated so that the attribute due to which saliency is present—color or orientation—is faithfully captured.

The saliency maps are binarized by Otsu's thresholding method [29]. The saliency map in which the salient regions are more *compact* and form a *connected* region will be the one that better identifies the salient region in the image. The *compactness* of a region is objectively evaluated by considering the saliency map as a probability distribution function and evaluating its variance as

$$SalMap_{Var} = SalMap_{VarX} + SalMap_{VarY} \quad (25)$$

where

$$SalMap_{VarX} = \frac{\sum_{(x,y)} (x - Sal_x)^2 \cdot Sal(x,y)}{\sum_{(x,y)} Sal(x,y)}$$

and

$$SalMap_{VarY} = \frac{\sum_{(x,y)} (y - Sal_y)^2 \cdot Sal(x,y)}{\sum_{(x,y)} Sal(x,y)} \quad (26)$$

are the spatial variances of the binary salient regions in the x and y directions, respectively, and

$$Sal_x = \frac{\sum_{(x,y)} x \cdot Sal(x,y)}{\sum_{(x,y)} Sal(x,y)}$$

and

$$Sal_y = \frac{\sum_{(x,y)} y \cdot Sal(x,y)}{\sum_{(x,y)} Sal(x,y)} \quad (27)$$

are the spatial means of the salient regions in the x and y directions, respectively, with $Sal(x,y)$ denoting the saliency value at location (x,y) .

The *connectivity* of the saliency map is measured by considering the number of salient pixels in the 8-point neighborhood of a salient pixel p in the binary saliency map and is evaluated as

$$SalMap_{Con} = \frac{\sum_{p \in ISal} \sum_{(x,y) \in N_p} \mathbf{I}_i(p_{xy})}{|\{ISal\}|} \quad (28)$$

where $\{ISal\}$ is the set of salient pixels in the saliency map, $|\{ISal\}|$ is the cardinality of the set, $\mathbf{I}_i(p_{xy})$ is the indicator function denoting whether the pixel p at location (x,y) is a salient pixel, N_p is the set of coordinates in the 8-point neighborhood of pixel p .

The saliency map with high connectivity and less spatial variance is estimated to represent a meaningful salient object. Hence, the final measure, *SalIndex*, used to select between color and orientation saliency maps is

$$SalIndex = \frac{SalMap_{Con}}{SalMap_{Var}}. \quad (29)$$

The saliency map with higher *SalIndex* is chosen as the final saliency map. The reason for binarizing the saliency maps was to facilitate the selection of either the color saliency map or the orientation saliency map through efficient computation of the spatial variance and the connectivity of each saliency map. This computation will not be reasonable on the original (nonbinary) saliency maps since the saliencies take on continuous values in the greyscale and the notion of connectivity of a region is not clear in such a case. Moreover, the saliency values in both the maps have different dynamic ranges and therefore it is better to binarize the maps in order to compare the spatial variance. As demonstrated in the next section, our experiments on a large database of images annotated with “ground truth,” validate the above method.

V. EXPERIMENTS

The experiments are conducted on a database of about 5000 images provided by Microsoft Research Asia [22]. As described previously, the MSRA database used in this paper contains the ground truth of the salient region marked as bounding boxes by nine different users. Similar to [22], a saliency probability map (SPM) is obtained by averaging the nine users’ binary masks. The performance of a particular algorithm is objectively evaluated by its ability to determine the salient region as close as possible to the SPM. We carry out the evaluation of the algorithms based on Precision, Recall, and F-Measure. *Precision* is calculated as ratio of the total saliency (sum of intensities in the saliency map) captured inside the SPM to the total saliency computed for the image. *Recall* is calculated as the ratio of the total saliency captured inside the SPM to the total saliency of the SPM. *F-Measure* is the overall performance measurement and is computed as the weighted harmonic mean between the precision and recall values. It is defined as

$$F-Measure_\alpha = \frac{(1 + \alpha) \cdot Precision \cdot Recall}{(\alpha \cdot Precision + Recall)} \quad (30)$$

where α is real and positive and decides the importance of precision over recall while computing the *F-Measure*. While absolute value of precision directly indicates the performance of the algorithms compared to ground truth, the same cannot be said for recall. This is because, for recall we compare the saliency on the area of the salient object inside the SPM to the total saliency of the SPM. However, the salient object need not always fill the SPM. Yet, calculation of recall allows us to compare our algorithm with others in a relative manner. Under these circumstances, the improvement in precision is of primary importance. Therefore, while computing the F-measure, we weight precision more than recall by assigning $\alpha = 0.3$. The following subsections discuss the various intermediate and final results of the proposed framework and present comparisons with some existing popular techniques for salient region detection.

A. Experiment 1: Color Saliency

First, we evaluate the color saliency framework, preliminary results for which were presented in Section II-E. Fig. 8(a) and (c) shows some of the original images selected from the MSRA database of 5000 images. Fig. 8(b) and (d) shows the color saliency maps. It can be observed that color plays a crucial role in deciding salient regions in the images shown in Fig. 8(a) and (c). The color saliency maps in Fig. 8(b) and (d) are shown without any threshold. We see that the salient regions are very accurate in relation to what the HVS would consider to be salient.

In Fig. 9, a comparison with the color spatial variance feature described in [22] is illustrated. As shown in Fig. 9(b), there are many clusters that have comparable variances for the original images in Fig. 9(a). Hence, if intra-color spatial variance is used as feature to model the color saliency, the neighboring regions of the salient regions are also extracted. The proposed color saliency framework detects a more accurate salient region as shown in Fig. 9(c). Thus, if the saliency in an image is dominated by color content, the global color modeling proposed in this paper detects salient regions with sufficient accu-

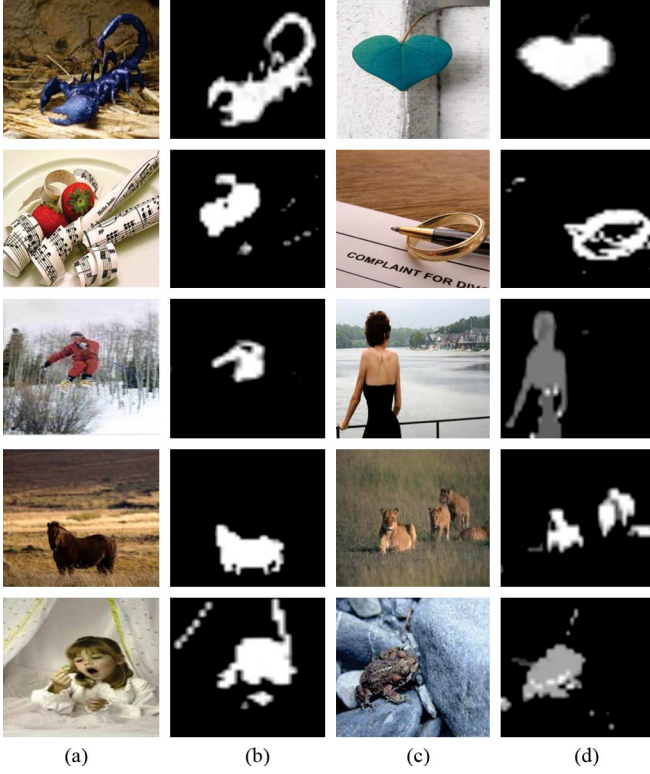


Fig. 8. Salient region detection by the color saliency framework. (a), (c) Original images. (b), (d) Color saliency maps.

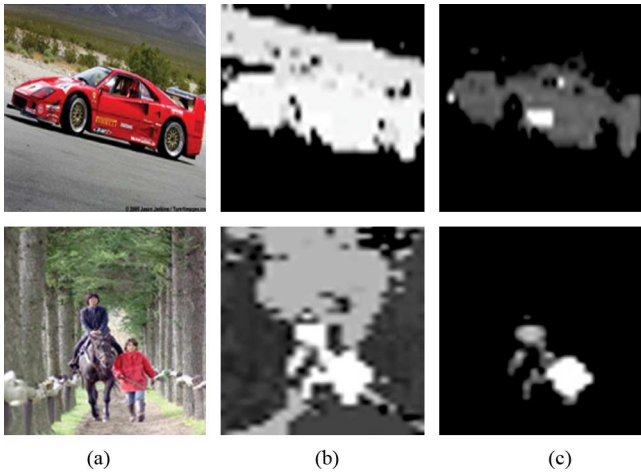


Fig. 9. (a) Original images. Color saliency maps using (b) intra-color spatial variance feature [22], and (c) the proposed color saliency framework.

racy. A center-weighted spatial variance feature is also introduced in [22] to eliminate regions with small variances close to the boundaries of the image since they are less likely to be the part of a salient region. But this assumption may not be true in all cases since salient regions can be present along the image borders. In our method, the color regions occurring near the image boundary having less color spatial variance may have large inter-cluster distance in the spatial domain and will be automatically eliminated without any center-weighting. However, if they are isolated in the hue-saturation space, they are more likely to be salient and the proposed color saliency framework will detect such regions.

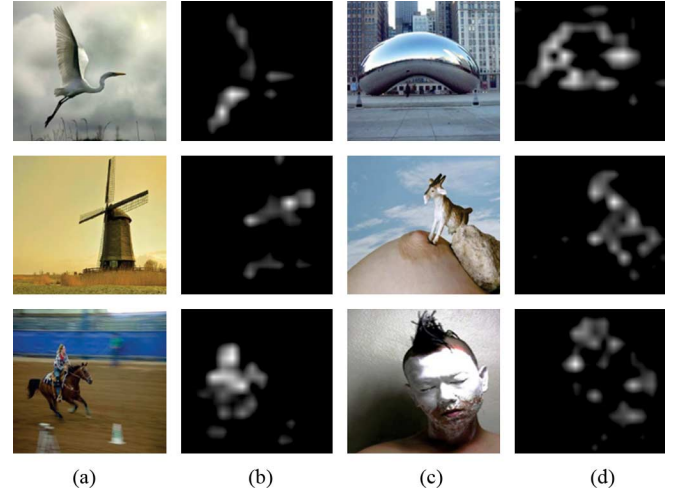


Fig. 10. (a), (c) Original images. (b), (d) Saliency map based on spatial variance of global orientation.

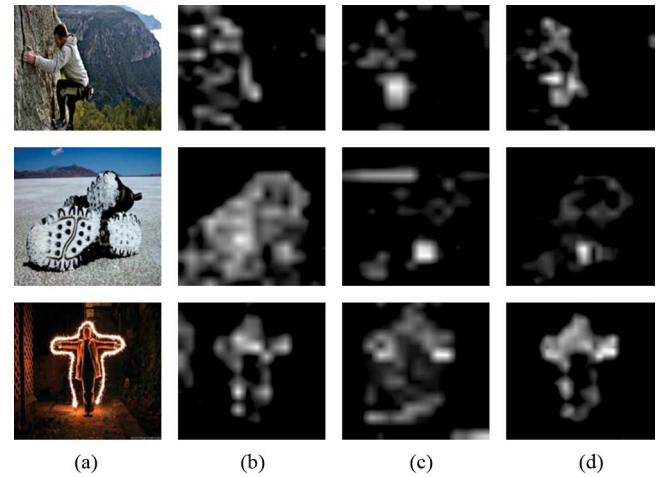


Fig. 11. (a) Original image. Saliency map based on (b) spatial variance of global orientation (Sal_{gP}) and (c) orientation entropy contrast (Sal_{IP}). (d) Final orientation saliency (Sal_{OP}).

B. Experiment 2: Orientation Saliency

We use the same database of images as was used in the previous experiment to test the effectiveness of the orientation saliency framework. Initially, we consider the methods involving spatial variance of the orientations and the orientation entropy contrast separately in order to ascertain their individual performance. Fig. 10 shows some images in which the salient regions are extracted based on the spatial variance of the dominant orientations. Fig. 10(a) and (c) shows the original images that contain salient objects having compact orientations and the extracted salient regions are shown in Fig. 10(b) and (d). In all of these images, the salient region cannot be distinguished by color since the contrast between the salient object and the background is too little for it to pop out.

Although the contrast in orientation entropy has been shown to be useful for detection of salient regions in some images in Fig. 7, we observe that it serves more as a critical complement to the spatial variance of orientation in order to determine the final orientation saliency. We illustrate this fact through results in Fig. 11. The original images are shown in Fig. 11(a), and the



Fig. 12. (a) Original images. (b) Binary color saliency map. (c) Binary orientation saliency map. (d) Final saliency map, which corresponds to the binary color or orientation saliency map with higher *SalIndex*.

orientation saliency maps obtained by the spatial variance of the orientation and the contrast in orientation entropy methods are shown in Fig. 11(b) and (c). Individually, each of the approaches returns saliency maps that are different; however, when they are combined according to (24), those regions in which both the approaches are in agreement are reinforced so that the final orientation saliency map contains the accurate salient region, as shown in Fig. 11(d).

The improvement obtained by combining the two orientation saliency determination frameworks is quantified by computing the precision values achieved for the MSRA database. The average precision obtained by the global variance method, the entropy contrast method and the final orientation saliency obtained by combination through (24) are 0.61, 0.46, and 0.67, respectively. As expected, the combination of the two methods results in the highest precision.

C. Experiment 3: Final Saliency Map Selection

The previous experiments showed the results obtained when color saliency and orientation saliency frameworks are applied separately to the selected image data set. Now we move on to the final stage of our proposed framework. We convert the color and orientation saliency maps to binary maps for the purpose of final saliency calculation as described in Section IV. The saliency map with higher *SalIndex* as defined in (29) is selected as the final saliency map. Fig. 12(b) and (c) shows the binary color and orientation saliency maps, respectively, of the original images shown in Fig. 12(a). The final saliency map in Fig. 12(d) is selected based on the *SalIndex* of the binary color and orien-

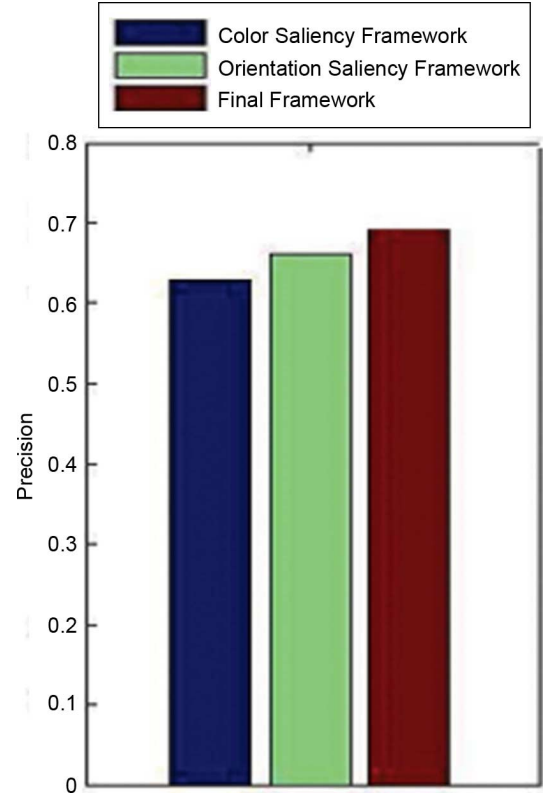


Fig. 13. Comparison of average precision values of color saliency framework, orientation saliency framework, and the combined framework.

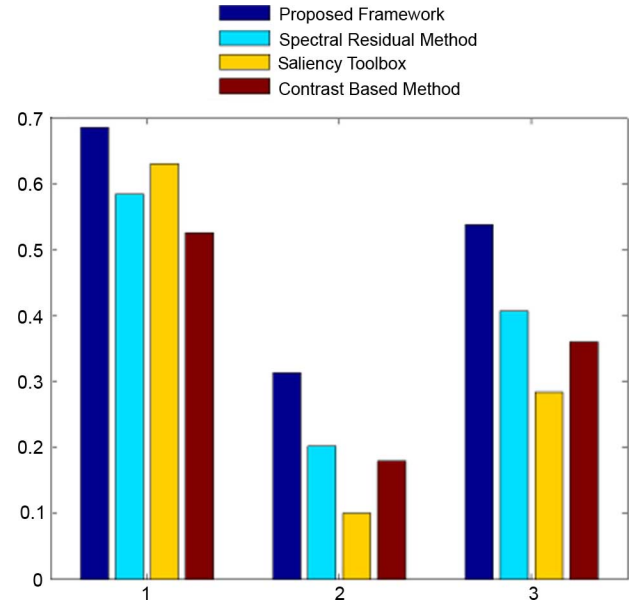


Fig. 14. Comparison of average precision, recall, and f-measure values of proposed framework, spectral residual method [37], saliency tool box [35], and contrast method [23]. Horizontal axis shows (1) Precision, (2) Recall, and (3) F-Measure.

tation maps. It can be verified from the images in Fig. 12(d) that the final saliency map selection based on *SalIndex* is justified. The first image shown in Fig. 12(a) is salient both in color as well as in orientation. However, based on the *SalIndex* value, the orientation saliency map is selected as the final saliency map in the first image. The next two images are color salient while

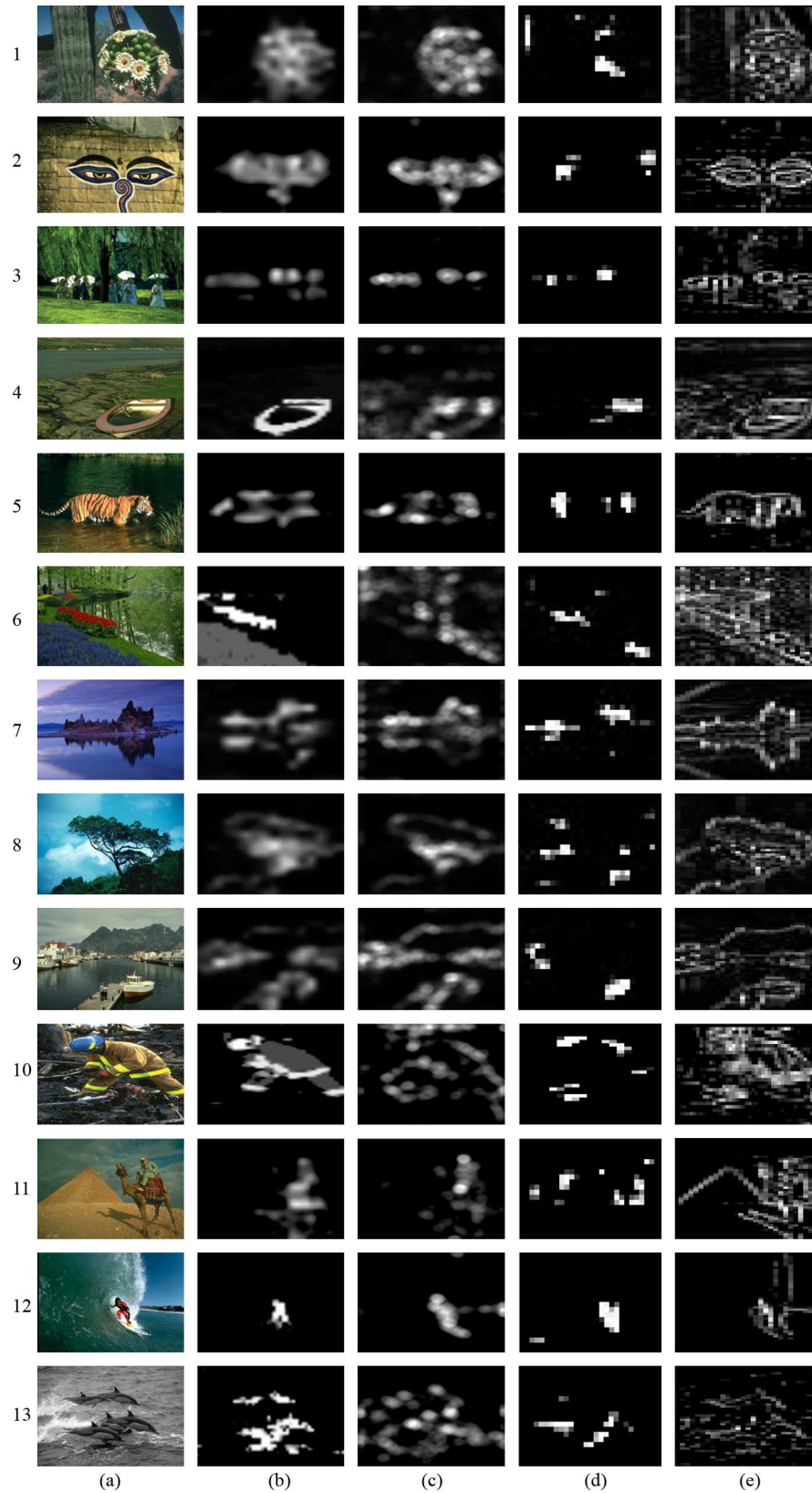


Fig. 15. Saliency map comparisons in Berkeley image database (a) original image. Saliency maps generated by (b) proposed method. (c) Spectral residual method [37]. (d) Saliency tool box [35]. (e) Contrast-based method [23].

the last two are orientation salient. The color and orientation saliency maps of the images detect the salient regions accordingly. Recall that we model the global saliency according to the premise that salient regions in most images are either

color salient or orientation salient. Even if the saliency is due to both color and orientation, the proposed method will indicate the correct salient region as shown in the first image of Fig. 12(a).

Fig. 13 shows the comparison of average precision for images in MSRA database obtained with the color saliency framework, orientation saliency framework and when the final saliency map is selected to be one of color or orientation. It is clear that the selection results in better performance than the individual methods. This is due to the fact that the color saliency framework and the orientation saliency framework complement each other quite well. The final map is selected mostly from the framework that returns the true salient region, as can be seen in the examples in Fig. 12.

D. Comparison With Other Saliency Detection Methods

The performance of the proposed framework is compared with Saliency Tool Box (STB) (latest version available at <http://www.saliencytoolbox.net>) [35], contrast based method [23] and spectral residual method [37]. The user annotation in the MSRA database facilitates objective comparisons of different algorithms. The precision, recall and F-measure of various methods computed for the MSRA database is shown in Fig. 14. The recall of all methods keep generally lower values due to the reasons explained at the beginning of this section. It can be seen from the figure that the proposed method performs consistently better with respect to precision, recall and F-measure. We have also compared the performance of the proposed method with the other existing methods in the Berkeley image database [24] as well. The saliency map comparisons of the proposed method and other methods on selected images from Berkeley database is shown in Fig. 15. It is evident that the proposed method outperforms the other methods. For example, in row (6), our method detects the salient flower regions correctly while the spectral residual method does not take color into account and fails. The proposed method has rightly picked up the red color region as the salient region. Similarly, in row (13), disconnected salient regions corresponding to the dolphins have been accurately detected. It is not possible to compute the precision of saliency maps for the Berkeley image database since no ground truth is available for the salient region; instead the database contains the ground truth only for image segmentation results. Note that the proposed method is independent of the size of the salient object as it looks for salient features. The regions containing the salient features get more saliency in the saliency map. If such regions are small as a consequence of a small salient object, the saliency map will indicate that. The same logic holds true for large salient objects.

E. Failure Cases and Analysis

As stated previously, the proposed framework detects salient regions under the assumption that from a global perspective, certain colors or orientations contribute to defining a region as salient. This assumption may fail in those images in which saliency is decided by more complex psychovisual aspects influenced by prior training in object recognition. Fig. 16 shows two such examples. The bird in Fig. 16(a) is not salient due to its color or orientation as shown in the corresponding maps in Fig. 16(b) and (c), but it is still a salient object due to the object recognition capability of the trained human brain. For similar reasons, the reptile in Fig. 16(a) is salient; the color and orientation maps do not indicate it as a salient region.

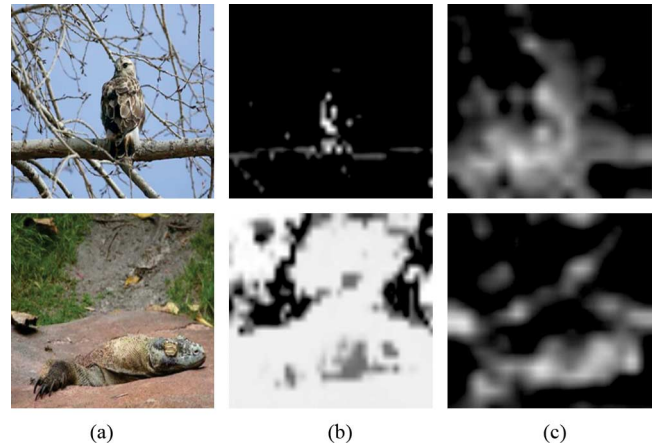


Fig. 16. Failure cases. (a) Original image. (b) Color saliency map. (c) Orientation saliency map.

VI. CONCLUSION

In this paper, we have presented a framework for salient region detection which consists of two parts—color saliency and orientation saliency. For most images, a region turns out to be salient due to these two important factors and this fact is captured in the proposed framework. The properties of the color and orientation distributions of the image are explored in the image domain as well as in the feature domain. This is in contrast to previous works which consider either one of the spaces. The new notion of isolation of a color in the color space of the image and its relation to saliency has been explored. In terms of orientation, we propose a new mechanism to determine local and global salient orientations and combine them to yield an orientation salient region. As opposed to previous works which fuse saliency maps through weights, we propose a simple selection of either color or orientation saliency map. The robustness of the proposed framework has been objectively demonstrated with the help of a large image data base.

REFERENCES

- [1] R. Achanta, F. J. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proc. ICVS*, 2008, pp. 66–75.
- [2] N. D. Bruce and J. K. Tsotsos, "Saliency based on information maximization," in *Proc. NIPS*, 2005, pp. 155–162.
- [3] L. Q. Chen, X. Xie, X. Fan, W. Y. Ma, H. J. Zhang, and H. Q. Zhou, "A visual attention model for adapting images on small displays," *Multimedia Syst.*, vol. 9, no. 4, pp. 353–364, 2003.
- [4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, 2005, pp. 886–893.
- [5] A. G. Dugue and A. Oliva, "Classification of scene photographs from local orientations features," *Pattern Recognit. Lett.*, vol. 21, no. 13–14, pp. 1135–1140, Dec. 2000.
- [6] X. Fan, X. Xie, W. Y. Ma, H. J. Zhang, and H. Q. Zhou, "Visual attention based image browsing on mobile devices," in *Proc. 2003 Int. Conf. Multimedia and Expo (ICME '03)*, Washington, DC, 2003, pp. 53–56.
- [7] S. Frintrop, G. Backer, and E. Rome, "Goal-directed search with a top-down modulated computational attention system," in *Proc. DAGM05*, 2005, pp. 117–124.
- [8] D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," in *Proc. NIPS*, 2004, pp. 481–488.
- [9] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," in *Proc. ICCV*, 2007.
- [10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.
- [11] J. Han, K. N. Ngan, M. Li, and H. J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 141–145, Jan. 2006.

- [12] G. Heidemann, "Focus-of-attention from local color symmetries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 817–830, Jul. 2004.
- [13] Y. Hu, L. T. Chia, and D. Rajan, "Region-of-interest based image resolution adaptation for MPEG-21 digital item," in *Proc. 12th Annu. ACM Int. Conf. Multimedia (MULTIMEDIA'04)*, New York, 2004, pp. 340–343.
- [14] Y. Hu, D. Rajan, and L. T. Chia, "Robust subspace analysis for detecting visual attention regions in images," in *Proc. ACM Multimedia*, 2005, pp. 716–724.
- [15] Y. Hu, X. Xie, W. Y. Ma, L. T. Chia, and D. Rajan, "Salient region detection using weighted feature maps based on the human visual attention model," in *Proc. 5th Pacific-Rim Conf. Multimedia*, 2004, pp. 993–1000.
- [16] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [17] J. S. Joseph and L. M. Optican, "Involuntary attentional shifts due to orientation differences," *Percept. Psychophys.*, vol. 58, pp. 651–665, 1996.
- [18] T. Jost, N. Ouerhani, R. Wartburg, R. Müri, and H. Hügli, "Assessing the contribution of color in visual attention," in *Proc. CVIU*, 2005, vol. 100, no. 1–2, pp. 107–123.
- [19] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, 2001.
- [20] A. G. Leventhal, *The Neural Basis of Visual Function: Vision and Visual Dysfunction*. Boca Raton, FL: CRC, 1991, vol. 4.
- [21] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention," *IJCV*, vol. 11, no. 3, pp. 283–318, 1993.
- [22] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," in *Proc. CVPR*, 2007.
- [23] Y. F. Ma and H. J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. 11th ACM Int. Conf. Multimedia (MULTIMEDIA'03)*, 2003, pp. 374–381.
- [24] 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. 8th Int. Conf. Computer Vision*, Jul. 2001, vol. 2, pp. 416–423.
- [25] O. L. Meur, P. L. Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 802–817, May 2006.
- [26] B. S. Morse, D. Thornton, Q. Xia, and J. Uibel, "Image based color schemes," in *Proc. ICIP*, 2007, vol. 3, pp. III-497–III-500.
- [27] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimizing detection speed," in *Proc. CVPR06*, 2006, pp. II:2049–II:2056.
- [28] A. Oliva, A. Torralba, M. S. Castelhana, and J. M. Henderson, "Top-down control of visual attention in object detection," in *Proc. ICIP*, 2003, pp. I:253–I:256.
- [29] N. Otsu, "A threshold selection method from grey-level histograms," *SMC*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [30] S. J. Park, J. K. Shin, and M. Lee, "Biologically inspired saliency map model for bottom-up visual attention," in *Proc. BMCV02*, 2002, pp. 418–426.
- [31] E. J. Pauwels and G. Frederix, "Finding salient regions in images: Non-parametric clustering for image segmentation and grouping," *Comput. Vis. Image Understand.*, vol. 75, no. 1–2, pp. 73–85, 1999.
- [32] D. Reissfeld, H. Wolfson, and Y. Yeshurun, "Context free attentional operators: The generalized symmetry transform," *Int. J. Comput. Vis.*, vol. 26, pp. 119–130, 1995.
- [33] R. J. Snowden, "Visual attention to color: Parvocellular guidance of attentional resources?," *Psychol. Sci.*, vol. 13, no. 2, pp. 180–184, Mar. 2002.
- [34] F. Stentiford, "Attention based auto image cropping," in *Proc. 5th Int. Conf. Computer Vision Systems*, Bielefeld, Germany, 2007.
- [35] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, pp. 1395–1407, 2006.
- [36] J. Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 150–156, Jan. 2006.
- [37] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. CVPR*, Jun. 2007.
- [38] C. M. Kuo, Y. H. Kuan, and N. C. Yang, "Color-based image salient region segmentation using novel region merging strategy," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 832–845, Aug. 2008.
- [39] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM Multimedia*, 2006, pp. 815–824.
- [40] Y. Zhiwen and H. S. Wong, "A rule based technique for extraction of visual attention regions based on real-time clustering," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 766–784, Jun. 2007.



Viswanath Gopalakrishnan (S'09) received the B.Tech degree in electronics and communication engineering from National Institute of Technology, Calicut (formerly REC Calicut), India, in 2001. He is currently pursuing the Ph.D. degree in the School of Computer Engineering, Nanyang Technological University, Singapore.

From 2002 to 2006, he worked in Texas Instruments and Emuzed India, Bangalore, India, in the domain of image and video compression. His research interests include visual attention in images/videos and graph-based approaches in computer vision.



Yiqun Hu received the B.S. degree in computer science from Xiamen University, Xiamen, China, in 2002 and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in 2008.

He is currently the research fellow at Nanyang Technological University. His research interest include pattern analysis, machine learning, and their applications in computer vision and multimedia, especially computational attention detection.



Deepu Rajan (M'02) received the B.Eng. degree in electronics and communication engineering from Birla Institute of Technology, Ranchi, India, the M.S. degree in electrical engineering from Clemson University, Clemson, SC, and the Ph.D. degree from the Indian Institute of Technology, Bombay, India.

He is an Assistant Professor in the School of Computer Engineering at Nanyang Technological University, Singapore. From 1992–2002, he was a Lecturer in the Department of Electronics at Cochin University of Science and Technology, Cochin, India. His research interests include image processing, computer vision, and multimedia signal processing.