Random Walks on Graphs to Model Saliency in Images

Viswanath Gopalakrishnan Yiqun Hu Deepu Rajan
School of Computer Engineering
Nanyang Technological University, Singapore 639798
visw0005@ntu.edu.sg yqhu@ntu.edu.sg asdrajan@ntu.edu.sg

Abstract

We formulate the problem of salient region detection in images as Markov random walks performed on images represented as graphs. While the global properties of the image are extracted from the random walk on a complete graph, the local properties are extracted from a k-regular graph. The most salient node is selected as the one which is globally most isolated but falls on a compact object. The equilibrium hitting times of the ergodic Markov chain holds the key for identifying the most salient node. The background nodes which are farthest from the most salient node are also identified based on the hitting times calculated from the random walk. Finally, a seeded salient region identification mechanism is developed to identify the salient parts of the image. The robustness of the proposed algorithm is objectively demonstrated with experiments carried out on a large image database annotated with 'ground-truth' salient regions.

1. Introduction

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. Detection of salient regions can be effectively utilized to automatic zooming into 'interesting' regions [3] or for automatic cropping of 'important' regions in an image [14, 15]. Object recognition algorithms can use the results of saliency detection, which identifies the location of the object as a pop-out from other regions in the image. Salient region detection also reduces the influence of cluttered background and enhances the performance of image retrieval systems [17].

One of the early attempts made towards detecting visual attention regions in images is the bottom-up method proposed by Itti et al [10] focusing on the role of color and orientation. Itti et al. use a center-surround difference operator on Red-Green and Blue-Yellow colors that imitate the color double opponent property of the neurons in receptive fields of human visual cortex to compute the color saliency map. The orientation saliency map is obtained by considering the local orientation contrast between centre and surround scales in a multiresolution framework. Walther and Koch extended Itti's model to infer the extent of a proto-object from the feature maps in [16], leading to the creation of a saliency toolbox.

Researchers have also attempted to define saliency based on information theory. A self-information measure based on local contrast is used to compute saliency in [2]. In [5], a top-down method is described in which saliency is equated to discrimination. Those features of a class that 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 [6]. Liu et al. locate salient regions with a bounding box obtained by learning local, regional and global features using conditional random fields [12]. They define a multi-scale contrast as the local feature, center-surround histogram as the regional feature and color spatial variance as the global feature.

The role of spectral components in an image to detect saliency has been explored in [18] in which 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 [8], Guo et al. argue that the phase spectrum of the Fourier transform of the image is more effective and computationally more efficient than the spectral residue (SR) method of [18]. In [11], 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 com-

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

plexity) will erroneously result in the larger textured region as the salient one.

In this paper, we consider the properties of random walks on image graphs and its relationship to saliency in images. Previous works on detecting salient regions from images represented as graphs include [4] and [9]. In [4], Costa presents two models in which random walks on graphs enable the identification of salient regions by determining the frequency of visits to each node at equilibrium. While some results are presented on only two synthetic images, there is no evaluation of how the method will work on real images. A similar approach in [9] uses edge strengths to represent the dissimilarity between two nodes; the strength between two nodes decreases as the distance between them increases. Here, the most frequently visited node will be most dissimilar in a local context. A major problem when looking for local dissimilarities of features is that cluttered backgrounds will yield higher saliencies as such backgrounds possess high local contrasts. The proposed method differs from that in [9] in the following way. We evaluate the 'isolation' of nodes in a global sense and identify nodes corresponding to 'compact' regions in a local sense. A robust feature set based on color and orientation is used to create a fully connected graph and a k-regular graph to model the global and local characteristics, respectively, of the random walks. The behavior of random walks on the two separate graphs are used to identify the most salient node of the image along with some background nodes. The final stage of seeded salient region identification uses information about the salient and the background nodes in an accurate extraction of the salient part in the image. Lastly, it is to be noted that the main objective in [9] is to predict human fixations on natural images as opposed to identifying salient regions that correspond to objects, as illustrated in this paper.

The remainder of the paper is organized as follows. Section 2 reviews some fundamental properties of ergodic Markov chains. Section 3 details the feature set used to create the graphs and presents the analysis of random walks on global and k-regular graphs. Further in Section 4 we describe the method of finding the most salient node and the background nodes based on Markov random walks. Section 5 describes the method for seeded salient region extraction. Experimental results are presented in Section 6 and conclusions are given in Section 7.

2. Ergodic Markov chain fundamentals

In this section, we review some of the fundamental results from the theory of Markov chains [1], [7], [13].

A Markov chain having N states is completely specified by the $N \times N$ transition matrix $\mathbf{P}$, where p_{ij} is the probability of moving from state i to state j and an initial probability distribution on the states. An ergodic Markov chain is one in which it is possible to go from every state to every state, not necessarily in a single step. A Markov chain is called a regular chain if some power of the transition matrix has only positive elements. Hence every regular chain is ergodic but the converse is not true. In this paper, the Markov chains are modeled as ergodic but not necessarily regular.

A random walk starting at any given state of an ergodic Markov chain reaches equilibrium after a certain time; this means that the probability of reaching a particular state from any state is the same. This equilibrium condition is characterized by the equilibrium probability distribution of the states, π , which satisfies the relation

$$\pi.\mathbf{P} = \pi \tag{1}$$

where π is the $1 \times N$ row vector of the equilibrium probability distribution of N states in the Markov chain. In fact, π is the left eigen vector of the stochastic matrix $\mathbf{P}$ corresponding to the eigenvalue one and hence is easy to compute.

The matrix $\mathbf{W}$ is defined as the $N \times N$ matrix obtained by stacking the row vector π, N times. For regular Markov chains, $\mathbf{W}$ is the limiting case of the matrix $\mathbf{P}^n$ as n tends to infinity. The fundamental matrix $\mathbf{Z}$ of an ergodic Markov chain is defined as

$$\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{W})^{-1} \tag{2}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. The fundamental matrix $\mathbf{Z}$ can be used to derive a number of interesting quantities involving ergodic chains including the hitting times. We define $E_i(T_i)$ as the expected number of steps taken to return to state i if the Markov chain is started in state i at time $t = 0$. Similarly, $E_i(T_j)$ is the expected number of steps taken to reach state j if the Markov chain is started in state i at time $t = 0$. $E_i(T_j)$ is known as the *hitting time* to state j from state i. $E_\pi(T_i)$ is the expected number of steps taken to reach state i if the Markov chain is started in the equilibrium distribution π at time $t = 0$, i.e., the hitting time to state i from the equilibrium condition is $E_\pi(T_i)$.

The three quantities $E_i(T_i)$, $E_i(T_j)$ and $E_\pi(T_i)$ can be derived from the equilibrium distribution π and the fundamental matrix $\mathbf{Z}$ as [1]

$$\begin{aligned} E_i(T_i) &= \frac{1}{\pi_i} \\ E_i(T_j) &= E_j(T_j) \times (Z_{jj} - Z_{ij}) \\ E_\pi(T_i) &= E_i(T_i) \times Z_{ii} \end{aligned} \tag{3}$$

where π_i is the i^{th} element of the row vector π and Z_{ii}, Z_{jj}, Z_{ij} are the respective elements of the fundamental matrix $\mathbf{Z}$. For detailed proofs of the results in eq. (3), please refer to [1].

3. Graph representation

We represent the image as a graph $G(V, E)$, where V is the set of vertices or nodes and E is the set of edges. The

vertices (nodes), $v \in V$, are patches of size 8×8 on the image while the edges, $e \in E$, represent the connection between the nodes. An edge between node i and node j is represented as e_{ij}, while $w(e_{ij})$ or simply w_{ij} represents the weight assigned to the edge e_{ij} based on the similarity between the feature set defined on node i and node j.

The important roles played by color and orientation in deciding the salient regions in an image has been well documented [10]. Hence, we consider these two features to be defined on each node. The image is represented in YCbCr domain and the Cb and Cr values on each node are taken as the color feature. The motivation for choosing YCbCr is that is perceptually uniform and is a better approximation of the color processing in the human visual system. As for orientation, we consider the complexity of orientations in a patch than the orientations itself. As stated in section 1, a salient region can be highly textured or smooth and the saliency is indicated by how its complexity is different with respect to the rest of the image. Hence, it is more useful to consider the contrast in complexity of orientations and to this end, we propose a novel feature derived from the orientation histogram entropy at different scales. Recall that in [11], only the local entropy is computed. We calculate the orientation histogram of the local patch and the complexity of the patch is calculated as the entropy of histogram. The orientation entropy E_P of the patch (or node in the graph) P having the orientation histogram H_P is calculated as

$$E_P = -\sum_i H_P(\theta_i) \log H_P(\theta_i) \tag{4}$$

where $H_P(\theta_i)$ is the histogram value of the i^{th} orientation bin corresponding to the orientation θ_i. We calculated the orientation entropy at five different scales $\rho_i \in \{0.5, 1, 1.5, 2, 2.5\}$ to capture the multi-scale structures in the image. The scale space is generated by convolving the image with Gaussian masks derived from the five scales. When we consider the orientation complexity at different scales, the dependency of the feature set on the selected patch size is reduced. Hence, the seven dimensional vector $x = [Cb, Cr, E_{\rho_1 \ldots \rho_5}]$ represents the feature vector associated with a node on the graph. Here Cb and Cr are calculated as the average values over the 8×8 image patch and E_{ρ_i} represents the orientation entropy calculated at scale ρ_i.

The weight w_{ij} of the edge connecting node i and node j is

$$w_{ij} = e^{\frac{-||(x_i - x_j||^2}{\sigma^2}}. \tag{5}$$

where x_i and x_j are the feature vectors attributed to node i and node j respectively. The value of σ is fixed to unity in our experiments.

In the proposed framework, the detection of salient regions is initiated by the identification of the most 'salient' node in the graph. In doing so, we wish to incorporate both global as well as local information into salient node identification mechansim. Clearly, such a technique is better compared to considering either global or local information only. The image is represented as a fully connected (complete) graph and a k-regular graph to capture the global and local characteristics, respectiely, of the random walk. In the complete graph, every node is connected to every other node so that there is no restriction on the movement of the random walker as long as the strength of the edge allows it. Hence, it is possible for the random walker to move from one corner of the image to the other corner in one step depending on the strength of the edge strength between the nodes. Spatial neighborhood of the nodes is given no preference and this manifests the global aspect of features in the image. The $N \times N$ affinity matrix, A^g, that captures the global aspects of the image features is defined as

$$A^g_{ij} = \begin{cases} w_{ij}, & i \neq j \\ 0, & i = j. \end{cases} \tag{6}$$

The degree d^g_i of a node i, is defined as the sum of all weights connected to node i and the degree matrix D^g is defined as the diagonal matrix with the degrees of the node in its main diagonal, i.e.,

$$d^g_i = \sum_j w_{ij},$$
$$D^g = diag\left\{d^g_1, d^g_2 \ldots d^g_N\right\}. \tag{7}$$

The transition matrix for the random walk on the fully connected graph, $\mathbf{P}^g$, is given as

$$\mathbf{P}^g = D^{g^{-1}} A^g. \tag{8}$$

The equilibrium distribution π^g and the fundamental matrix $\mathbf{Z}^g$ can be obtained from the transition matrix $\mathbf{P}^g$according to equations (1) and (2). This leads to the computation of the hitting times of the complete graph, viz., $E^g_i(T_i)$, $E^g_i(T_j)$ and $E^g_\pi(T_i)$, according to equation (3).

The characteristics of the features in a local area in the image are encoded into the k-regular graph where every node is connected to k other nodes. We choose a particular patch and the 8 patches in its spatial neighborhood to study the local properties of the random walk so that $k = 8$. Thus, the random walker is restricted to a local region in the image while its path is determined by the features in that region. In this configuration, therefore, a random walker at one corner of the image cannot make jump to the opposite corner, but has to traverse through the image according to the strengths of the edges. Such random walks capture the properties of the local features of the image. The $N \times N$ affinity matrix, A^l, that captures the local aspects of the image features is defined as

$$A^l_{ij} = \begin{cases} w_{ij}, & j \in N(i) \\ 0, & otherwise \end{cases} \tag{9}$$

where $N(i)$ denotes the nodes in the spatial neighborhood of node i. The degree matrix D^l and the transition matrix $\mathbf{P}^l$ of the k-regular graph can be obtained from the affinity matrix A^l similar to equations(7) and (8). Further we calculate the equilibrium distribution π^l, fundamental matrix $\mathbf{Z}^l$ and the hitting times $E_i^l(T_i)$, $E_i^l(T_j)$, $E_\pi^l(T_i)$ for the k-regular graph in a similar manner as that for the complete graph.

4. Node selection

Having represented the image as a graph, the next task is to identify the node that corresponds to a patch that most likely belongs to the salient region of the image - this node is called the most salient node. The selection of the most salient node is based on the hitting times of the random walker computed on both the complete graph as well as the k-regular graph. This is followed by identification of a few nodes that correspond to the background of the image. The most salient node and the background nodes together enable a seeded extraction of the salient region.

4.1. Most salient node

The most salient node in the image should globally pop-out in the image when compared to other competing nodes. At the same time it should fall on a compact object in the image in some local sense. The global pop-out and compactness properties are reflected in the random walks performed on the complete graph and the k-regular graph, respectively. When a node is globally a pop-out node, what it essentially means is that it is isolated from the other nodes so that a random walker takes more time to reach such a node. On the other hand, if a node is to lie on a compact object, a random walker should take less time to reach it on a k-regular graph. We now elaborate on these concepts.

We calculate the global isolation of a node by the time taken to access the node when the Markov chain is in equilibrium. In [9], Harel et al. identify the most frequently visited node as the most salient node. This is directly measured from π_i or from the first return time $E_i(T_i)$ and it characterizes the dissimilarity of the node in a local sense, i.e., their activation map encodes how different a particular location in the image is, compared to its neighborhood. As mentioned earlier, we believe that the isolation of a node should be computed in the *global* sense and hence, the measure in [9] will not characterize the global isolation of node i. A better characterization will be by the sum of hitting times from all other nodes to node i on a complete graph, i.e.,

$$H_i = \Sigma_j E_j^g(T_i). \tag{10}$$

Since the edge strengths represent the similarity between nodes, a higher value of H_i indicates higher global isolation of the node. However, in the computation of H_i, the hitting times from all the other nodes to node i are given equal preference. The measure can be further improved if the hitting times from the most frequently visited nodes are given priority over hitting times from less frequently visited nodes. This property is inherent in $E_\pi^g(T_i)$, which is the time taken to access the node i when the Markov chain starts from equilibrium. The equilibrium distribution directly depends on the frequency of access to different nodes, and hence $E_\pi^g(T_i)$ gives priority to hitting times from most frequently visited nodes over hitting time from less frequently visited nodes. Hence, the gloal isolation of a node is measured by $E_\pi^g(T_i)$.

Next, we ensure that the most isolated node falls on a compact object by looking at the equilibrium access time of nodes in the k-regular graph. The local random walk under Markov equilibrium will reach the nodes corresponding to compact structures faster as it is guided by strong edge strengths from the neighborhoods. Hence a low value of local random walk equilibrium access time, $E_\pi^l(T_i)$ ensures that the respective node falls on a compact surface. Considering both the global and local aspects, saliency of node i, $NSal_i$, is defined as

$$NSal_i = \frac{E_\pi^g(T_i)}{E_\pi^l(T_i)}. \tag{11}$$

The most salient node can be identified as the node that maximizes $NSal_i$, i.e.,

$$N_s = \arg\max_i \quad NSal_i. \tag{12}$$

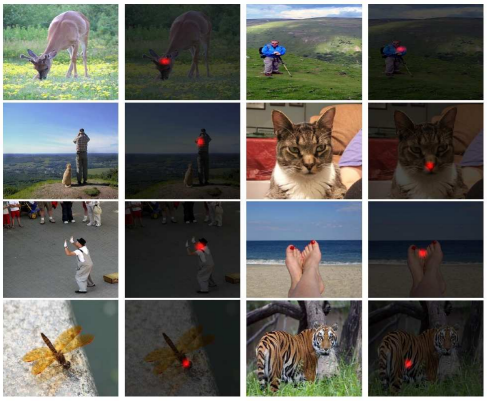
Figure 1 compares the detection of the most salient node between [9] and the proposed method.While the method in [9] generates fixations produced on the image by the algorithm, for comparison with our method, we use their definition of the least visited node indicated by the first return time of the random walk as the most salient node. However, it must be noted that in [9], the edge strengths represent the dissimilarity between nodes as opposed to similarity in our case. Hence, it turns out that the random walker's most frequently visited node in [9] is actually the least frequently visited node using our definition of edge strength. Figures 2(a) and 2(c) show some more example images from the database and figures 2(b) and 2(d) show the most salient node marked with a red region in the respective images.

4.2. Background nodes

After identifying the most salient node in the image we move on to identify certain nodes in the background. This is to facilitate the extraction of the salient regions using the most salient node and the background nodes as seeds. The most important feature of a background node is obviously the less saliency of the node as calculated from equation (11). Moreover, the background nodes have the property that it is at a large distance from the most salient node, N_s.

(a) **(b)** **(c)**

Figure 1. Comparison of [9] and proposed method for detection of most salient node. (a) Original images, (b) most salient node detected by proposed method and (c) most salient node detected by [9].



(a) **(b)** **(c)** **(d)**

Figure 2. Detection of the most salient node. (a, c) Original images and (b, d) Most salient node spotted with the red region.

The distance from node N_s to some node j is measured as the hitting time, $E^g_{N_s}(T_j)$, which is the average time taken to reach node j if the random walk starts from node N_s. We use the complete graph to calculate the hitting times since a global view of the image has to be taken in order to identify the background. Hence the first background node, N_{b1}, is calculated as

$$N_{b1} = \arg\max_j \quad \frac{E^g_{N_s}(T_j)}{NSal_j} \tag{13}$$

The background in an image is, more often than not, inhomogeneous, e.g. due to clutter or due to regions having different feature values. In the cat image in figure 2(c), for example, the background consists of regions with different colors. Our goal is to capture as much of these variations as possilbe by locating at least one background node in each of such regions. Hence, while maximizing the distance of a node to the maximum salient node, we impose an additional condition of maximizing the distance to all background nodes identified so far. This will ensure that the newly found background node falls on a new region. Even if there are no multiple backgrounds, i.e. the background is relatively homogeneous, the algorithm will only place the new node in the same background region. This does not affect the performance of the algorithm. Thus, the n^{th} background node, N_{bn}, is identified as

$$N_{bn} = \arg\max_j \quad \frac{E^g_{N_s}(T_j).E^g_{N_{b1}}(T_j)....E^g_{N_{b(n-1)}}(T_j)}{(NSal_j)^n} \tag{14}$$

The above equation can be viewed as a product of n terms, where each term is of the form $\frac{E^g_\alpha(T_j)}{NSal_j}$. In our experiments the value of n is fixed to 4 but it can be increased to improve the accuracy of the algorithm at the cost of increased computational complexity.

(a) **(b)** **(c)** **(d)**

Figure 3. Detection of the background nodes. (a, c) Original images and (b, d) background node spotted with the green region.

Figure 3 shows examples of the background nodes as green spots extracted using the proposed method. The most salient node is also marked by the red spot. As expected, the background nodes are pushed away from the most salient node as well as from the previous background nodes that are detected. We note that the background nodes are placed such that they represent as much of the inhomogeneity in the background as possible. For example in the first row of Figure 3(d), the background nodes are placed in the foreground water area, the tree and the sky regions in the original image shown in the first row of figure 3(c). Similar observations can be made for the rest of the images.

5. Seeded salient region extraction

The identification of the most salient node and the background nodes enables the extraction of the salient regions. The most salient node and the background nodes act as seeds and the problem now is to find the most probable seed that can be reached for a random walk starting from a particular node. In other words, we need to determine the seed with the least hitting time when the random walker starts from a particular node. If the hitting time from a node to the most salient node is less compared to hitting times to all the background nodes, then that node is deemed to be part of the salient object. This process is repeated for the rest of

the nodes in the graph so that at the end of the process, the salient region is extracted.

In the above process, it might seem obvious that the random walk should be performed at a global level. However, in a global random walk it may turn out that a node that is far from a salient node in the spatial domain, but close to it in the feature domain (as indicated by the edge weights) may be erroneously classified as belonging to the salient region. On the other hand, a local random walk may treat a background region that is spatially close to a salient node as part of the salient object, since the random walk is restricted to a smaller area. Hence, we propose a linear combination of the global and local attributes of a random walk by defining a new affinity matrix for the image given by

$$A^c = \lambda \times A^g + A^l. \tag{15}$$

where A^c is the combined affinity matrix and λ is a constant that decides the mixing ratio of global and local matrices. The values of the equilibrium distribution π^c, the fundamental matrix $\mathbf{Z}^c$, the hitting times $E_i^c(T_i)$, $E_i^c(T_j)$, and $E_\pi^c(T_i)$, follow from the definition of the combined affinity matrix as described in section 3. A particular node k is regarded as part of the salient region if the hitting time, $E_k^c(T_{N_s})$ to the most salient node N_s is less than the hitting times to other background nodes N_{b1} ,N_{b2},...and N_{bn}. We fix the value of λ to 0.05 in our experiments.

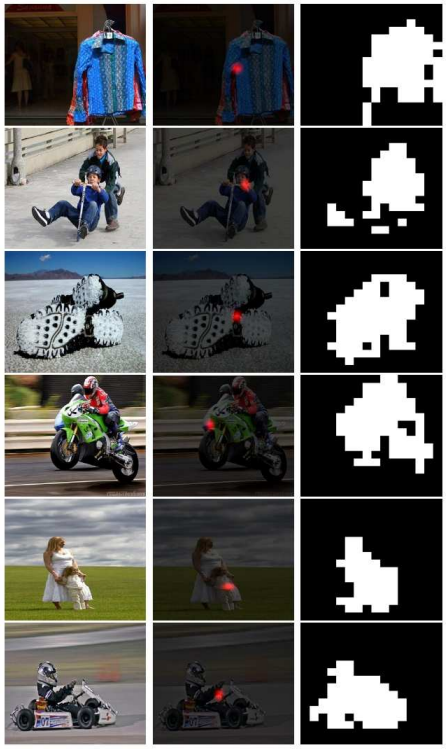
6. Experimental results

The experiments are conducted on a database of about 5,000 images available from [12]. The database used in this work contains the ground truth of the salient region marked as bounding boxes by nine different users. The median of the nine boxes is selected as the final salient region.

We have shown some results of identifying the most salient node in Figure 2. In order to evaluate the robustness of the detection of the most salient node, we calculate the percentage of images in which it falls on the user annotated salient object. On a database of 5000 images, we obtained an accuracy of 89.6%.

Figure 4 shows examples of the saliency map extracted using the proposed algorithm. Figure 4(a) shows the original images and figure 4(b) shows the most salient node marked as the red region in the respective images. The result of the seeded salient region extraction is shown in figure 4(c). Note that we directly obtain a binary saliency map unlike previous methods like [10],[12] and [18], in which a saliency map has to be thresholded to obtain the bounding box. In our case, since the saliency map is already binary, the bounding box to denote the salient region can be easily obtained.

In Figure 5, we show further examples of the proposed salient region extraction algorithm with the salient image marked with a red bounding box. The original images are in figures 5(a) and 5(c) and the corresponding salient regions are marked in figures 5(b) and 5(d). The final bounding box over the salient region can be used in applications like cropping and zooming for display on small screen devices.



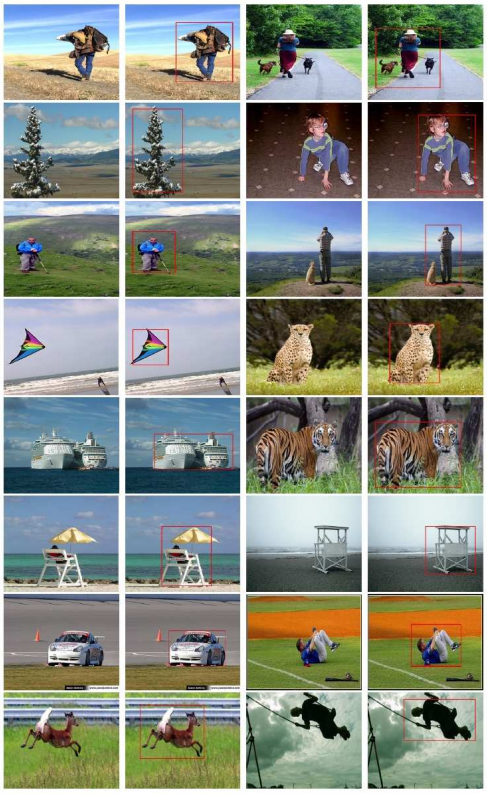
(a) **(b)** **(c)**

Figure 4. Results of seeded salient region extraction. (a) Original image (b) Most salient node marked with a red spot (c) The final binary saliency map.

We also show some failure examples of the proposed method for salient region detection in Figure 6. The underlying reason for the failures seem to be the similarity of features on the salient object with the background which affects the random walk, e.g., the branches and the bird in the first row and the fish and the rock in the second row. However, as noted earlier, the general framework of the algorithm allows for more robust features to be utilized. In the third row, the features caused the building in the background to be detected as the salient region; however, this failure has opened up the question of what effect, if any, does depth have on sailency since it is evident that there is a large variation in the depth field of the image.

6.1. Comparison with other saliency detection methods

We have compared the saliency map of the proposed random walk based method with the saliency maps generated by the saliency toolbox (STB)[10], the spectral residual method based on [18] and the phase spectrum based on



(a) **(b)** **(c)** **(d)**

Figure 5. Bounding box on salient region. (a, c) Original images and (b, d) Red bounding box over the salient region.

(a) **(b)** **(c)**

Figure 6. Failure examples. (a) Original image, (b) most salient node and (c) the corresponding bounding box.

[8]. The evaluation of the algorithms is carried out based on Precision, Recall and F-Measure. *Precision* is calculated as ratio of the total saliency, i.e., sum of intensities in the saliency map captured inside the user annotated rectangle to the total saliency computed for the image. *Recall* is calculated as the ratio of the total saliency captured inside the user annotated window to the area of the user annotated window. *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

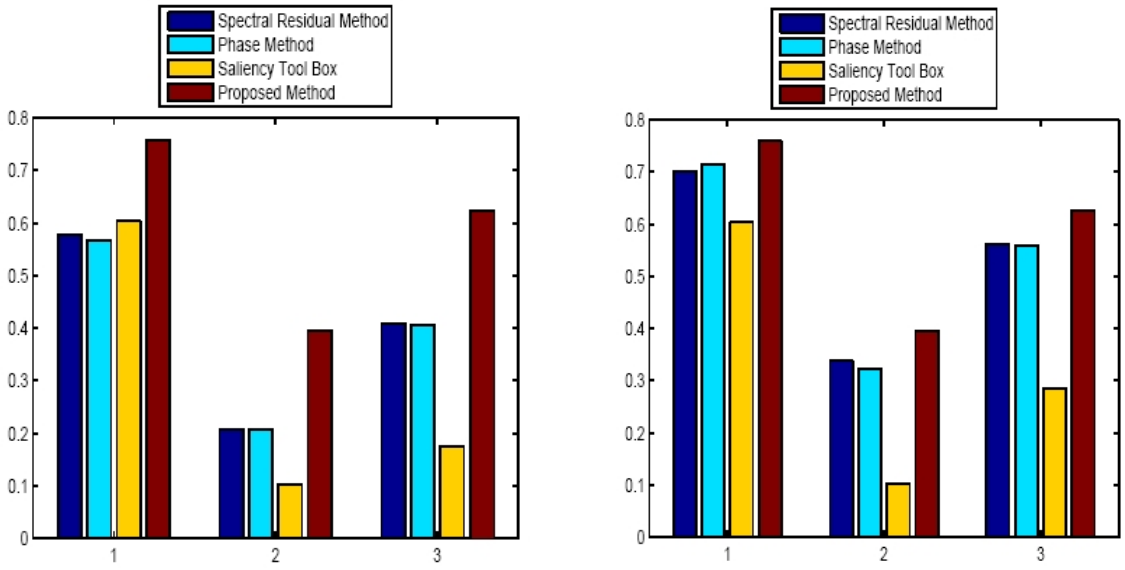
$$F\text{-}Measure_{\alpha} = \frac{(1+\alpha).Precision.Recall}{(\alpha.Precision + Recall)}, \qquad (16)$$

where α is real and positive and decides the importance of precision over recall. While absolute value of precision directly indicates the performance of the algorithms compared to ground truth, the same cannot be said for recall. In computing recall, we compare the saliency on the area of the salient object inside the user bounding box to the area of the user bounding box. However, the salient object need not always fill the user annotated bounding box completely. Even so, the calculation of recall allows us to compare our algorithm with other algorithms. 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 α = 0.3.

As noted, the intensities of the saliency map are used in the computation of precision and recall, The final saliency maps obtained in our case and in the salient tool box are binary; however, in the spectral residue method and in the phase method, the saliency maps are not binary. We compare the proposed method with the other methods in two ways - in the first method, we do not binarize the saliency maps of the spectral residue and phase methods. Figure 7(a) shows the precision, recall and F-measure marked as 1, 2 and 3 on the x-axis, respectively, for the proposed method as well as for the other methods. In the second method, we binarize the saliency maps of the spectral residue and phase spectrum method using Otsu's thresholding technique, so that the effect of intensity values is applied equally on all the saliency maps. Figure 7(b)shows the results in this latter case. The precision and recall values and hence the f-measure of the spectral residue and phase spectrum methods have increased although the proposed method still outperforms the rest. In any case, the advantage of our approach is that we directly obtain the binary saliency map without requiring a thresholding stage. The recall of all the methods has low values due to the reason explained in the previous paragraph.

7. Discussion and Conclusion

It is observed that the detection of the most salient node in the proposed random walk is quite robust. This could be effectively used for zooming, where the co-ordinate of the zoom fixation is directly available from the most salient node. The background node detection also performs quite well unless the image has a complex background consisting of many regions. In such cases a consensus on salient region is in any case difficult. The main limitation of the proposed work is in determining the mixing ratio of the local and the global affinity matrices to facilitate seeded salient region extraction. Currently, the ratio is empirically decided to give



(a) (b)

Figure 7. Comparison of precision, recall and f-Measure values of Spectral Residual Method [18], Saliency Tool Box [10], Phase spectrum method [8] and the proposed method. Horizontal axis shows 1) Precision 2) Recall 3) F-Measure. (a) Without binarizing (b) After binarizing saliency maps of [18] and [8].

best results on the test data base. However, a more reliable way would be to employ image-specific mixing ratios based on certain properties of the random walk.

We have presented an algorithm to extract salient regions in images through random walks on a graph. It provides a generic framework that can be enriched with more robust feature sets. The proposed method captures saliency using both global and local properties of a region by carrying out random walks on a complete graph and a k-regular graph, respectively. This also allows computations for both types of features to be similar in the later stages. The robustness of the proposed framework has been objectively demonstrated with the help of a large image data base and comparisons with existing popular salient region detection methods demonstrate its effectiveness.

References

[1] D. Aldous and J. A. Fill. Reversible markov chains and random walks on graphs, http://stat-www.berkeley.edu/users/aldous/RWG/book.html.

[2] N. D. Bruce and J. K. Tsotsos. Saliency based on information maximization. In *NIPS*, pages 155–162, 2005.

[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.*, 9(4):353–364, 2003.

[4] L. da Fontoura Costa. Visual saliency and attention as random walks on complex networks, arXiv preprint, 2006.

[5] D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In *NIPS*, pages 481 – 488, 2004.

[6] D. Gao and N. Vasconcelos. Bottom-up saliency is a discriminant process. In *ICCV*, 2007.

[7] C. M. Grinstead and L. J. Snell. Introduction to probability, American Mathematical Society, 1997.

[8] C. Guo, Q. Ma, and L. Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *CVPR*, 2008.

[9] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, pages 545–552, 2006.

[10] 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):1254–1259, 1998.

[11] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[12] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Y.Shum. Learning to detect a salient object. In *CVPR*, 2007.

[13] J. Norris. Markov chains, Cambridge University Press, Cambridge, 1997.

[14] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. F. Cohen. Gaze-based interaction for semi-automatic photo cropping. In *CHI*, pages 771–780, 2006.

[15] F. Stentiford. Attention based auto image cropping. In *The 5th International Conference on Computer Vision Systems, Bielefeld*, 2007.

[16] D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Network*, 19:1395–1407, 2006.

[17] X. J. Wang, W. Y. Ma, and X. Li. Data-driven approach for bridging the cognitive gap in image retrieval. In *2004 IEEE International Conference on Multimedia and Expo,*, pages 2231–2234, 2004.

[18] X.Hou and L.Zhang. Saliency detection: A spectral residual approach. In *CVPR*, June 2007.