Salient Region Detection in Natural Images using EM-Based Structure-Guided Statistical Textural Distinctiveness

Srujy Krishna A U
PG Student
Department of Computer Science and Engineering
Federal Institute Of Science and Technology

Shimy Joseph
Assistant Professor
Department of Computer Science and Engineering
Federal Institute Of Science and Technology

Abstract

The objective of salient region detection is to separate salient region from entire image. This salient region detection framework consists of a structure-guided statistical textural distinctiveness approach. This approach includes the five main stages: i) image decomposition ii) textural representation, iii) texture modeling, iv) matrix construction, and v) saliency map computation. In the image decomposition stage, decomposition of image into structural image elements to learn a structure-guided texture model. In second stage, define a rotational-invariant neighborhood based texture feature model that represents the underlying textural characteristics of regions in a local manner. In texture modeling stage, Sparse texture modeling is done using structure-guided texture learning. In matrix construction stage, characterize all pair-wise statistical textural distinctiveness within the sparse texture model of the image and construct a textural distinctiveness matrix. In the final stage, the saliency of a region can be computed as the expected statistical textural distinctiveness of the region in the given image .The proposed approach has been extensively evaluated on images from MSRA-1000 datasets.

Keywords- Saliency, Salient region detection, Structure-guided

I. INTRODUCTION

Human eye is perceptually more sensitive to certain colors and intensities and objects with such features are considered more salient. Visual saliency is the perceptual quality that makes an object, person, or pixel stand out relative to its neighbors and thus capture our attention. Visual attention results both from fast, pre-attentive, bottom-up visual saliency of the retinal input, as well as from slower, top-down memory and volition based processing that is task-dependent. The term saliency was used by Itti et al [1] in their work on rapid scene analysis. Saliency has also been referred to as visual attention, unpredictability, rarity, or surprise. The salience or saliency of an item be it an object, a person, a pixel, etc.- is the state or quality by which its stands out relative to its neighbors. Detecting salient regions in natural images aims to identify and localize regions of interest which are distinct in their attributes when compared to the rest of the scene. For example, Fig. 1 shows examples of natural images, where the flower is visually unique and draw a viewer's attention from the surrounding environment.

Fig. 1: From left to right: Salient object computed saliency map and ground truth mask.

In the proposed method, the advantage of textural characteristics is explicitly taken in a quantitative manner to detect saliency objects of interest within a scene. In contrast to early methods, a the proposed approach to salient region detection avoids pre-processing and attempts to address the issue of incorporating texture by making use of a neighborhood-based texture feature The main contribution of this proposed method is the introduction of a simple yet effective approach to salient region detection based on the concept of EM based structure-guided statistical textural distinctiveness (EMSGTD). Here a multilayer approach is used.

The rest of the report organized as follows. In section II, an overview of the related works. The existing methodology and the proposed method are described in the section III. Experimental results and analysis is described in section IV. Finally, conclusion and future work are drawn in section V.

II. RELATED WORKS

In the literature, many salient region detection methods are explained. In [1] L.Itti et al.proposes a visual attention system where multistage image features are combined into a single topographical saliency map. A dynamic neural network then selects attended locations in order to decreasing saliency. The system breaks down the complex problem of scene understanding by rapidly selecting, in a computationally efficient manner, conspicuous locations to be analyzed in detail. The model uses feature integration theory. Which is proposed to explain human visual search strategies? First a set of topographic feature maps are extracted by decomposing the visual input. All feature maps feed into a master saliency map in a bottom up manner.

In [2] C. Koch et al. proposes a method for bottom up visual saliency. A Graph-Based Visual Saliency (GBVS) is proposed. It consists of two steps: first forming activation maps on certain feature channels, and then normalizing them in a way which highlights conspicuity and admits combination with other maps. Feature map is computed by linear filtering followed by some elementary nonlinearity. Activation map is computed using markovian approach. Normalization is done through concentrating mass on activation maps

The model is simple and biologically plausible insofar as it is naturally parallelized. This model powerfully predicts human fixations on 749 variations of 108 natural images, achieving 98% of the ROC area of a human-based control. That is GBVS predicts human fixations more reliably than the standard algorithms. The model exploiting the computational power, topographical structure, and parallel nature of graph algorithms to achieve natural and efficient saliency computations. The model is more reliable and Highlights salient region away from object border. But Object boundaries are not clear in this methodology, Reduce spatial frequencies in the original image and computationally quite expensive

In [3] L. Zhang et al. propose A Spectral Residual Approach for saliency detection. This is a simple method. The model is independent of features, categories, or other forms of prior knowledge of the objects. By analyzing the log-spectrum of an input image, we extract the spectral residual of an image in spectral domain, and propose a fast method to construct the corresponding saliency map in spatial domain. In which a front-end method to simulate the behavior of pre-attentive visual search is used. In this model analyze the log spectrum of each image and obtain the spectral residual. Then transform the spectral residual to spatial domain to obtain the saliency map. This is used to find the positions of proto-objects.

From the perspective of information theory, effective coding decompose the image information image into two parts, innovation and prior knowledge, Innovation denotes the novelty part, and prior knowledge is the redundant information that should be suppressed by a coding system. This method will demonstrate a method to approximate the innovation part of an image by removing the statistical redundant components. The spectral residual can be obtained by removing the shape information from the original log spectrum. The saliency map can be constructed by using Inverse Fourier Transform. This saliency map will be an explicit representation of proto-objects. Proto-objects are detected from the object map. To construct object map a simple threshold segmentation is used. In the saliency map, if the salient value greater than the given threshold then that region is object, otherwise not. While the object map is generated, proto-objects can be easily extracted from their corresponding positions in the input image. The main advantage of spectral residual approach is its generality. The prior knowledge required for saliency detection is not necessary in this system. The spectral residual resolves the problem of weighting features from different channels. But the computational cost is large. And the processing is limited to static images.

In [4]R. Achanta et al. proposes a method frequency-tuned salient region detection. Which is a salient region detection method that outputs full resolution saliency maps with well-defined boundaries of salient objects? This method exploits features of color and luminance. This model proposes an automatic visual salient region detection system. Which is useful in applications such as adaptive content delivery, adaptive region-of-interest based image compression, image segmentation, object recognition, and content aware image resizing? The method introduce a a frequency-tuned approach to estimate center-surround contrast using color and luminance features that overs three advantages over existing methods: uniformly highlighted salient regions with well de_ned boundaries, full resolution, and computational efficiency.

There are five requirements for a saliency detector. They are: emphasize the largest salient objects, uniformly highlight whole salient regions, establish well defined boundaries of salient objects disregard high frequencies arising from texture, noise and blocking artifacts, efficiently output full resolution saliency maps. To highlight large salient objects, consider very low frequencies from the original image. This also helps highlight salient objects uniformly. Retain high frequencies from the original image to establish well-defined boundaries. To avoid noise, coding artifacts, and texture patterns, the highest frequencies need to be disregarded. Combining the outputs of several band pass filters with contiguous pass bands is appropriate for saliency map containing a wide range of frequencies. To obtain a full resolution saliency map, DoG filter is choose for band pass filtering. A DoG filter is a simple band-pass filter whose pass band width is controlled by the ratio $\sigma 1$: $\sigma 2$. Saliency map can be computed by taking L_2 norm of deference between mean image feature vector and corresponding image pixel vector value in the Gaussian blurred version of the original image. When objects contain salient small-scale patterns, saliency could generally be misled by their complexity. This is the main disadvantage of this method.

In [5] M M Cheng et al. propose a global contrast based salient region detection algorithm. Which is a regional contrast based saliency extraction algorithm. The algorithm simultaneously evaluates global contrast differences and spatial coherence. It is a histogram-based contrast method. Which maps assign pixel-wise saliency values based simply on color separation from all other image pixels to produce full resolution saliency maps? It also uses region-based contrast maps. Where first segment the input image into regions, and then assign saliency values to them. The saliency value of a region is then calculated using a global contrast score, measured by the regions contrast and spatial distances to other regions in the image. In histogram-based contrast, color statistics of the input image is used to define saliency values for image pixels. Specifically, the saliency of a pixel is defined using its color contrast to all other pixels in the image. Uniformly divide each color channel into 12 different levels. It will help to quantize the color space into 123 different colors. For best result, quantization of colors is performed in the RGB color space and color differences are measured in the Lab color space. In region-based contrast, first segment the input image into regions, then compute color contrast at the region level, and define the saliency for each region as the weighted sum of the regions contrasts to all other regions in the image. The algorithm simultaneously evaluates global contrast differences and spatial coherence and generate full resolution saliency map. But when objects contain salient small-scale patterns, saliency could generally be misled by their complexity.

In [6] F Perazzi et al. propose an algorithm based on filtering, Saliency Filters: Contrast Based Filtering for Salient Region Detection. This algorithm consists of four basic steps. First, decomposes a given image into compact, perceptually homogeneous elements that abstract unnecessary detail. Based on this abstraction compute two measures of contrast that rate the uniqueness and the spatial distribution of these elements. From the element contrast then derive a saliency measure that produces a pixel-accurate saliency map which uniformly covers the objects of interest and consistently separates fore- and background. The complete contrast and saliency estimation can be formulated in a unified way using high dimensional Gaussian filters. The algorithm first decomposes the input image into basic elements. Based on these elements define two measures for contrast that are used to compute per-pixel saliency. Algorithm consists of four steps, abstraction, element uniqueness, element distribution, and saliency assignment. In abstraction, to decompose the image into basic elements that preserves relevant structure, but abstract undesirable detail. Specifically, each element should locally abstract the image by clustering pixels with similar properties into perceptually homogeneous regions. One approach to achieve this type of decomposition is an edge-preserving, localized over segmentation based on color. In element uniqueness, evaluate how different each respective element is from all other elements constituting an image, essentially measuring the rarity of each element. Colors belonging to the background will be distributed over the entire image exhibiting a high spatial variance, whereas foreground objects are generally more compact. This principle is used in element distribution stage. Contrast measures are defined on a per-element level. In the final step, assign the actual saliency values to the input image to get a pixel-accurate saliency map. The algorithm is efficient implementation with linear complexity. But saliency estimation based on color contrast may not always be feasible.

In [7] X Shen et al. propose a unified model to incorporate traditional low-level features with higher-level guidance to detect salient objects. In this model, an image is represented as a low-rank matrix plus sparse noises in a certain feature space, where the non-salient regions can be explained by the low-rank matrix, and the salient regions are indicated by the sparse noises. In this model first decompose an image into small regions by image segmentation after multi-scale feature extraction. The mean of the feature vectors in a segment is treated as the feature of that segment. A linear feature transformation is further trained from labeled data to ensure that the matrix representing the background has a low rank in the learned feature space. Different higher-level information is fused into a prior map, which is then incorporated to the objective function. By utilizing higher-level priors, the salient regions are further highlighted and the saliency detection performance is significantly improved. This model provides solutions to potential task-dependent saliency extraction by incorporating different task-dependent and volition-driven priors. To ensure the model to be valid for visual saliency, a linear transformation of the feature space is introduced and learned. Higher-level priors can be naturally integrated into this model. Saliency is then jointly determined by low-level and high-level cues in a unified way. This model used as a prototype model in task-dependent saliency applications by integrating different high-level guidance. But it requires High computational cost.

In [8]J Wang et al. propose a method which uses a discriminative regional feature integration approach. Where saliency map computation is regarded as a regression problem. This method is based on multi-level image segmentation, uses the supervised learning approach to map the regional feature vector to a saliency score, and finally fuses the saliency scores across multiple levels, yielding the saliency map. This approach integrates the regional contrast, regional property and regional backgroundness descriptors together to form the master saliency map. This approach consists of three main steps. The first one is multi-level segmentation, which decomposes the image to multiple segmentations from a fine level to a coarse one. Region saliency computation that maps the features extracted from each region to a saliency score with a random forest regressor, and multi-level saliency fusion that combines the saliency maps over all the levels of segmentations to get the final saliency map. The most time consuming step is the feature extraction on the multi-level segmentation. This method introduces a novel backgroundness descriptor. But the system is time consuming.

In [9] L Xu et al. propose a multi-layer approach to analyze saliency cues. Analyze saliency cues from multiple levels of structure, and then integrate them to infer the final saliency map. The model is able to deal with salient small-scale structure, so that salient objects are labeled more uniformly. This model contributes a new measure of region scales, which is compatible with human perception on object scales, and construction of a new scene dataset, which contains challenging natural images for saliency detection. In this method, three image layers of different scales are extracted from the input. Saliency cues are computed

for each layer. They are finally fused into one single map using a graphical model. But it is hard to determine which layer is the best by heuristics.

In [10] H Lu et al. proposes method for saliency detection via graph-based manifold ranking. This method considers both foreground and background cues in a different way. Here rank the similarity of the image elements with foreground cues or background cues via graph based manifold ranking. The image is represented as a close-loop graph with super pixels as nodes. These nodes are ranked based on the similarity to background and foreground queries, based on affinity matrices. Saliency detection is carried out in a two-stage scheme to extract background regions and foreground salient objects efficiently. In the first stage of this method, exploit the boundary prior by using the nodes on each side of image as labeled background queries. From each labeled result, compute the saliency of nodes based on their relevances to those queries as background labels. The four labeled maps are then integrated to generate a saliency map. In the second stage, apply binary segmentation on the resulted saliency map from the first stage, and take the labeled foreground nodes as salient queries. The saliency of each node is computed based on its relevance to foreground queries for the final map. Construct four saliency maps using boundary priors and then integrate them for the final map, which is referred as the separation/combination (SC) approach. The system requires high computational cost.

In [11] C Scharfenberger et al. proposes an approach that uses statistical textural distinctiveness. There are four main stages in this approach. i) rotational-invariant neighborhood-based textural representation, ii) sparse texture modeling via representative texture atom learning, iii) statistical textural distinctiveness graphical model construction, and iv) saliency map computation based on occurrence probabilities of representative texture atoms, statistical textural distinctiveness, and general visual attentive constraints. Rotational invariant neighborhood-based texture representations are extracted and used to learn a set of representative texture atoms for defining a sparse texture model for the image. Based on the learnt sparse texture model, a statistical textural distinctiveness graphical model is constructed to characterize the distinctiveness between all texture atom pairs. Finally, the saliency of each pixel in the image is computed based on the probability of occurrence of the representative texture atoms within the image, their respective statistical textural distinctiveness based on the constructed graphical model, and general visual attentive constraints. This approach makes limited use of one prior only and achieves comparable performance. And does not rely on any previously trained data at all. This makes this method more robust and suitable for applications where a priori background information is not available. But the approach fails for images containing finer structure.

In [12] C Scharfenberger et al. proposes an approach that uses structure-guided statistical textural distinctiveness. This is an extension of [11]. This method uses a multilayer approach to analyze the structural and textural characteristics of natural images as important features for salient region detection from a scale point of view. This approach to salient region detection that includes the five main stages: i) image decomposition into structural image elements, ii) rotational-invariant neighborhood-based textural representation, iii) structure guided sparse texture modeling via representative texture atom learning, iv) statistical textural distinctiveness matrix construction, and v) saliency map computation based on occurrence probabilities of representative texture atoms and statistical textural distinctiveness. First stage is to decompose the image into compact image elements to preserve the local image structure. Segmenting the image in the CIELabXY space using k-means clustering and limiting the search space to a region proportional to the element size. The neighborhood-based texture feature model is then used to build a sparse texture model to represent the global textural characteristics of an image. Employing a sparse texture model for representing the heterogeneous textural characteristics of the entire image reduces the computational and memory requirements for representing and quantifying the relationships between each texture pair since only the representative texture atoms need to be analyzed. In fourth stage define the statistical textural distinctiveness between two texture patterns. Then constructed a textural distinctiveness matrix. Using the statistical textural distinctiveness matrix, compute the saliency map of an image. Finally do all the steps in each layer.

III. PROPOSED WORK

The two-layer EM based structure-guided statistical textural distinctiveness approach (EMSGTD) is designed to quantify the saliency of regions by taking advantage of textural and structural characteristics. Fig. 2 shows the overall architecture of the proposed approach to salient region detection that includes the five main stages: i) structural image elements extraction, ii) textural representation, iii) structure guided sparse texture modeling via representative texture atom learning, iv) statistical textural distinctiveness matrix construction, and v) saliency map computation.

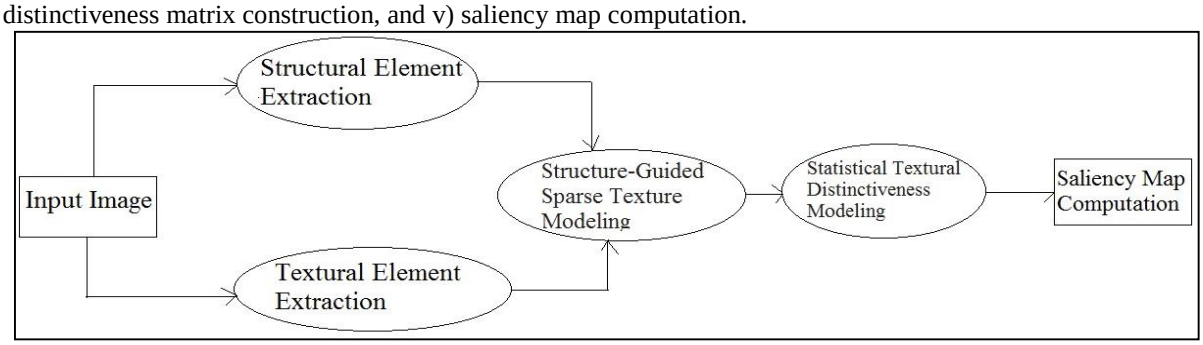

Fig. 2: Proposed methodology

A. Structural Image Elements Extraction

To extract the structural elements from the image, define a model that represents the underlying structural characteristics of the image and can guide texture modeling. In this model, a color-based structural model is implemented based upon the work of Achanta et al. [13] that aims to decompose the image into compact image elements to preserve the local image structure, i.e., region contours, as sharp boundaries between the elements. Such a structural model *l*:

$$l = \{\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots \varepsilon_J\}$$

Containing J image elements ε_v can be extracted from an image by segmenting the image in the CIELabXY space using EM clustering.

B. Textural Representation

From the segmented image from the first stage, textural features are extracted. First, extract the image into 5 X 5 patches. Then sort the patches into ascending order. The sorting beneficial in striking the balance between robustness to distortional variations and preservation of spatial-intensity context, making it well-suited for local textural representation in the proposed work. The sparsified radially-sorted textural representation can be described as follows. Let the image has size M X N that we wish to analyze, and I(x) the intensity of a pixel x in image. Given a neighborhood *N* centered at pixel location x in the image, we define the corresponding local textural representation $h_c(x)$ for each color channel c as:

$$h^c(x) = \langle \; I^c(x) \text{sort}_{\uparrow}\{I_c(x_1, j)\} sort_{\uparrow}\{I_c(x_2, j)\} \ldots sort_{\uparrow}\{I_c(x_n, j)\}\} \; \rangle$$

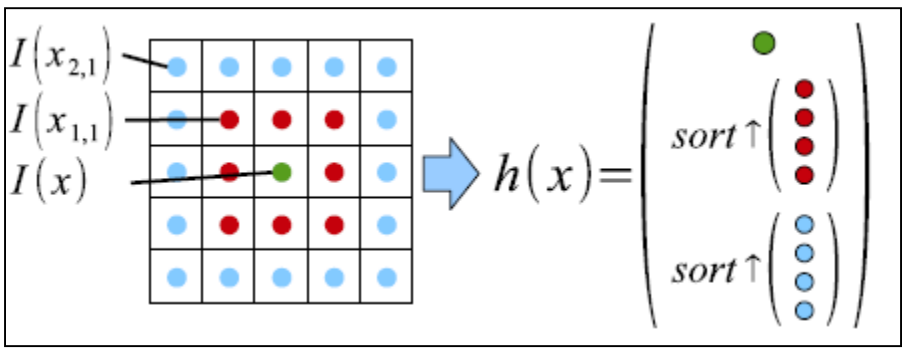

Fig. 3: Layer-based sorted textural representation

where $x_{i,j}$ denotes the pixel in the jth position of the ith radial layer about pixel location x, and $\text{sort}_{\uparrow}$ denotes sorting in ascending order. Fig. 3 shows a sorted textural representation for individual layers $I_c(x_g, j)$, and Fig. 4 shows a globally sorted textural representation of all pixels within a neighborhood *N*.

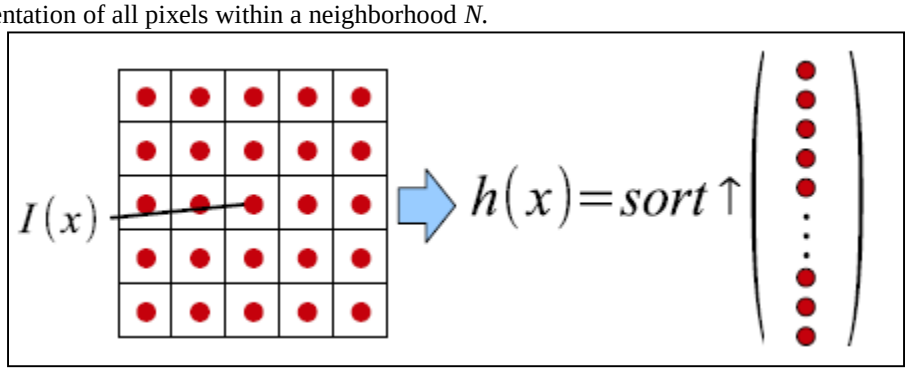

Fig. 4: Globally sorted textural representation

Given the local textural representation h(x), e.g., $h(x) = \langle h_L(x), h_a(x), h_b(x) \rangle$ with 363 element for Lab images, to increase the variance between the elements and produce a com- pact and sparsified textural representation t_x by taking the u principal components of the local textural representation h(x) with the highest variance using PCA:

$$t_x = \langle \phi_i(h(x)) \rangle$$

where ϕ_i is the ith principal component of h(x). The choice of u is based on a selection criteria related to a variance compaction. Selected the u principal components of h(x) that represent 95 % of the variance of all textural representations.

C. Structure Guided Textural Representation

A natural image as being composed of a set of image elements, each characterized by an individual local texture pattern, and where the number of image elements J is much smaller than the total number of pixels within the image. As such, here define a set T^{ε} of J element based local textural representations t_{ε_v} as

$$T^{\varepsilon} = \left\{ t_{\varepsilon_1}, t_{\varepsilon_2}, t_{\varepsilon_3}, \ldots t_{\varepsilon_j} \right\}$$

We determined the element-based local textural representation t_{ε_v} for each image element by computing the element-wise average over all U local textural representations t_x within an image element ε_v, i.e.,

$$t_{\varepsilon_v} = \frac{1}{U} \sum_{x}^{U} t_x, x \in \varepsilon_v$$

After extracting the underlying image structure, further generalize a natural image as being composed of a set of just a few regions S_i, where a particular texture pattern described by a representative texture atom t_l^r is repeated over each region, and where the number of regions with unique texture patterns is much smaller than the total number of image elements or pixels within the image. This generalization allows to describe the global textural characteristics of the entire image by a small set of compact distinctive textural Representations t_l^r, with each textural representation t_l^r associated with a region S_i. This representation of the global textural characteristics allows for the use of a texture model T^r that can be defined as a set of m representative texture atoms t_l^r,

$$T^r = \left\{ t_i^r \right\}$$

Given the small set of compact distinctive textural representations t_l^r describing the textural characteristics of an entire image, T^r is the sparse texture model. Employing a sparse texture model for representing the heterogeneous textural characteristics of the entire image reduces the computational and memory requirements for representing and quantifying the relationships between each texture pair since only the representative texture atoms need to be analyzed.

D. Statistical Textural Distinctiveness Matrix Construction

Salient regions are regions that have highly unique and distinctive textural characteristics when compared to the rest of the scene. Here introduce a metric for quantifying the distinctiveness of texture patterns within an image relative to each other using the concept of statistical textural distinctiveness. From the learned sparse texture model, first define the statistical textural distinctiveness between two texture patterns. Let t_l^r and t_j^r denote a pair of representative texture atoms in the sparse texture model, where texture atom t_l^r is a noisy observation of another atom t_j^r, and t_l^r and t_j^r differ only in an additive random component:

$$t_j = t_i^r + \eta_{i,j}$$

$\eta_{i,j}$ is a random field with probability distribution P($\eta_{i,j}$), which models differences between t_l^r and t_j^r due to noise and due to inherent texture variability. A more meaningful metric for quantifying textural distinctiveness is the probability of t_l^r not being a noisy observation of t_j^r

$$\beta_{i,j} = \exp\left(-\sum_k \frac{\left(t_{i,k}^r - t_{j,k}^r \right)^2}{\sigma^2} \right)$$

E. Saliency Map Computation

Saliency map is computed using the statistical textural distinctiveness matrix constructed in the previous step. Consider the rarity of regions based on their probability of occurrence within an image and use it to weight the statistical textural distinctiveness between regions. In the proposed work, the probability of occurrence is defined as the number of pixels in the corresponding region. Given these assumption, the saliency α_i of a region S_i can be computed as the expected statistical textural distinctiveness of S_i given the image:

$$\alpha_i = \sum_{j=1}^{m} |S_i| . \beta_{i,j}$$

where $|S_i|$ is the number of pixels in the corresponding region S_i in the image. Then compute the saliency $\Psi(x)$ for each pixel x in the image based on the representative texture atom in the sparse texture model that the pixel maps to:

$$\Psi(x) = \alpha_i , x \in S_i$$

The saliency map computed for each layer and then combine to form the final saliency map.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The method was implemented in a MATLAB 2014 prototype and tested with randomly selected images from MSRA-1000 dataset. The image processing was performed on a desktop PC with the following characteristics: Intel Core i3 CPU, 2.30 GHz, 2 GB RAM. The dataset was formed by randomly selecting 26 images from MSRA-1000 dataset and 5 images that dos not contain any salient region.

The salient region detection is demonstrated by an example selected from MSRA-1000 database. In the first stage the input image is convert into Lab color space. The remaining process are taken in this image.

(a)

(b)

Fig. 5: (a) and (b) contain input image, after converted into Lab color space, and final saliency

To measure the performance of the proposed method three metrics were used: precision, recall and accuracy. Precision is related to the detection exactness, recall refers to the detection completeness and accuracy refers to the average correctness of the process.

Precision = TP/(TP + FP)
Recall = TP/(TP + FN)
Accuracy = (TP + TN)/(TP + FP + TN + FN)

The salient region detection method is tested using 31 images where 26 images contain salient region. But the remaining 5 does not. The preliminary results were promising providing 95.45% precision, 84.0% recall and 83.87% accuracy. Table 1 provides a summary of the detection method results.

Table 1: Performance method for saliency detection method

Performance metrices	
Total TP	*21*
Total FP	*1*
Total TN	*5*
Total FN	*4*
Precision	*95.45 %*
Recall	*84.0 %*
Accuracy	*83.87 %*

V. CONCLUSION

The goal of salient region detection is to identify salient region in an image and separate it from the background. Approach based on the concept of EM based structure-guided statistical texture distinctiveness was presented. The two-layer EM based structure-guided statistical textural distinctiveness approach (EMSGTD) is designed to quantify the saliency of regions by taking advantage of textural and structural characteristics. Proposed approach to salient region detection that includes the five main stages: i) structural image elements extraction, ii) textural representation, iii) structure guided sparse texture modeling via representative texture atom learning, iv) statistical textural distinctiveness matrix construction, and v) saliency map computation. Future work involves investigating alternative sparse textural representation and textural models to evaluate whether improvements in salient region detection can be achieved. Also investigating method to reduce the computational time.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,", in Proc. IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 12541259, Nov. 1998.
[2] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,", in Advances in Neural Information Processing Systems 19. Cambridge, MA, USA: MIT Press, 2007.
[3] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach", in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2007, pp. 18.
[4] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,", in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 15971604.
[5] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-H. Hu, “Global contrast based salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 409416.
[6] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, “Saliency filters: Contrast based filtering for salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2012, pp. 733740.
[7] X. Shen and Y. Wu, “A uni_ed approach to salient object detection via low rank matrix recovery," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2012, pp. 853860.
[8] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, “Salient object detection: A discriminative regional feature integration approach," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 2083 2090.
[9] Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical saliency detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 1155 1162.
[10] C. Yang, L. Zhang, H. Lu, K. Ruan, and M.-H. Yang, “Saliency detection via graph-based manifold ranking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 31663173.
[11] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, “Statistical textural distinctiveness for salient region detection in natural images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 979986.
[12] Christian Scharfenberger, Alexander Wong, Member, IEEE, and David A. Clausi,Senior Member, IEEE “Structure-Guided Statistical Textural Distinctiveness for Salient Region Detection in Natural Images" IEEE Transactions On Image Processing, VOL. 24, NO. 1, January 2015
[13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Suesstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 22742282, Nov. 2012.