SALIENT OBJECT DETECTION USING NORMALIZED CUT AND GEODESICS

Keren Fu[A,B] *Chen Gong*[A] *Irene Y.H. Gu*[B] *Jie Yang*[A*] *Pengfei Shi*[A]

[A] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China
[B] Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

Recently the Normalized cut (Ncut) has been introduced to salient object detection [1, 2]. In this paper we validate that instead of proposing new detection models that leverage the Ncut, the previous geodesic saliency detection model which computes shortest paths on a graph can be adapted to eigenvectors of the Ncut to produce superior performance. Since the Ncut partitions a graph in a normalized energy minimization fashion, resulting eigenvectors contain decent cluster information that can group visual contents. Combining it with the existing geodesic saliency detection is conducive to highlighting salient objects uniformly, yielding to improved detection accuracy. Experiments by comparing with 12 existing methods on four benchmark datasets show the proposed method significantly outperforms the original geodesic saliency model and achieves comparable performance to state-of-the-art methods.

Index Terms— Salient object detection, Normalized cut, Geodesic saliency, Saliency map

1. INTRODUCTION

A new sub-field in saliency detection called salient object/region detection has recently drawn a lot of research attentions. It aims at compensating the drawback of previous eye fixation prediction models [3, 4, 5, 6] on enhancing holistic and entire objects. A huge variety of models have been proposed in the past decade, mainly including heuristic color contrast-based models [7, 8, 9, 10, 11, 12, 13, 14], learning-based models [15, 16], hierarchical/multi-scale segmentation assisted approaches [17, 18, 19, 1, 2], and graph-based saliency modeling [20, 21, 22, 23, 24]. Other notable works include: sparse and low rank matrix recovery [25], Bayesian framework [26] and data-driven saliency aggregation [27].

Recently the Normalized cut (Ncut) has been introduced to salient object detection [1, 2] and shown its success. [2] employs the 2-way Ncut to recursively partition a graph and produce a coarse-to-fine segmentation tree that separates salient objects from background as early as possible. [1] leverages the first smallest eigenvectors of the Ncut by an adaptive region merging strategy. Generally, in order to benefit from the Ncut, both [2] and [1] try to convert the cluster information from the Ncut into regions which better segregate objects and background. Some saliency criteria are then applied to these regions to compute regional saliency followed by across-scale integration. Different from the above works [2, 1], this paper gives some new insights towards using the Normalized cut (Ncut) for saliency detection. Instead of applying region-based saliency measure, we propose to adapt a previous geodesic saliency model [21] to the Ncut. The model in [21] measures connectivity of image patches/superpixels boundary of a given image by computing shortest paths on a graph. By assisting geodesic saliency detection through the Ncut, our proposed method leverages the cluster information from the Ncut in a novel way. Using the Ncut has been shown useful for improving detection performance and has the potential to be considered by future saliency models.

The rest paper is organized as follows. Section 2 reviews some fundamentals of the Normalized cut. Section 3 describes the proposed method. Section 4 shows the comparisons and evaluations. The conclusion is given in Section 5.

2. THE NORMALIZED CUT: REVIEW

The normalized graph cut (Ncut) proposed by Shi *et al* [28] normalizes the graph cut cost by using the total edge connections towards all nodes in a graph. Given a similarity graph G=(V, E) (a graph whose edges measure similarity between vertices), let $\mathbf{W}$ be its adjacency matrix, $\mathbf{D}$ be its degree matrix (a diagonal matrix with diagonal entry $d_i = \sum_j w_{ij}$, where w_{ij} is entry of $\mathbf{W}$), and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ be its Laplacian matrix. The cut cost between two subsets A and B is defined as $cut(A, B) := \sum_{i \in A, j \in B} w_{ij}$. Let $\bar{A}$ be the complement of A. For a given number k of subsets, the Ncut aims to choose a partition $A_1, ..., A_k$ ($\bigcup_{i=1}^{k} A_i = V$ and $A_i \bigcap A_j = \emptyset$) which minimizes:

$$Ncut(A_1, ..., A_k) = \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{assoc(A_i, V)} \qquad (1)$$

where $assoc(A_i, V) := \sum_{i \in A, j \in V} w_{ij}$ indeed is a measurement for set size, i.e. the larger $|A_i|$ is, the higher $assoc(A_i, V)$ will be. The exact solution to (1) is NP hard.

*This research is partly supported by National Science Foundation, China (No: 61273258, 61105001), Ph.D. Programs Foundation of Ministry of Education of China (No. 20120073110018). Jie Yang is the corresponding author (email: jieyang@sjtu.edu.cn).

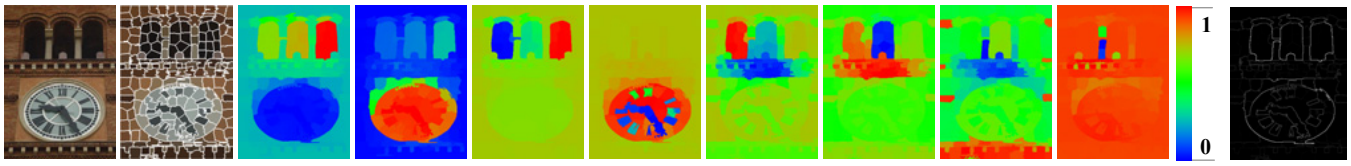

Fig. 1. Given an input image, the top eight smallest eigenvectors ($\mathbf{v}_1$ to $\mathbf{v}_8$) of the Ncut are shown from left to right. Eigenvectors are visualized in pseudo colors. The last image visualizes the reconstructed graph edge weight from the eigenvectors of the Ncut.

However, by first defining a hard indicating vector for each A_i and then relaxing the hard constraints [29], the continuous indicating vectors for multi-cluster Ncut (1) can be obtained from the first k smallest eigenvectors of $\mathbf{D}^{-1}\mathbf{L}$, or the first k smallest eigenvectors of the below eigen-system:

$$(\mathbf{D} - \mathbf{W})\mathbf{v} = \lambda \mathbf{D}\mathbf{v}. \tag{2}$$

where $\mathbf{v}$ and λ denote some eigenvector and eigenvalue. Indeed the Normalized cut is tightly related to spectral clustering. Since the continuous indicating vectors of multiple cluster Ncut contain cluster information, k-means clustering is often applied to those eigenvectors to get cluster labels, known as spectral clustering [29].

Unique benefit: The Ncut normalizes each cut cost as a fraction of the total edge connections to all the nodes in the graph. Due to this normalization, Ncut biases cut of fairly large set of vertices. It right fits our goal to find good grouping of visual contents for salient object detection, which are usually large objects, meanwhile prevent grouping of small clusters, which are usually noise. An example of Ncut eigenvectors generated by our method is shown in Fig.1. It implies that Ncut eigenvectors contain decent cluster information that well group visual contents together, i.e. objects versus background. Contents with the similar pseudo colors in each eigenvector in Fig.1 are likely to belong to the same cluster, inferring the same saliency level.

3. METHODOLOGIES FOR SALIENCY DETECTION

3.1. Graph Construction for the Ncut

An undirected graph is first constructed from a moderate number (≈ 200) of SLIC superpixels [30]. Let the ith superpixel be R_i and the corresponding average CIELab color and spatial location be $\mathbf{c}_i$ and $\mathbf{p}_i$. We define a graph $G = (V, E)$ whose vertices V are superpixels and edges E are to be detailed later. Thereby only a small set of graph nodes need to be considered.

Construct graph edge connection: The 2-ring graph topology is used. We connect superpixels R_i and R_j that satisfy either $R_j \in N_i$ or $\{R_j \in N_k, R_k \in N_i\}$. N_i denotes the neighborhood (spatial adjacency) of superpixel R_i. Note $R_j \in N_i$ yields $R_i \in N_j$ symmetrically. Besides, we associate superpixels located at the image boundary (called boundary superpixels in the followings) with each other so that they can be clustered together when belonging to the same background (somewhat shown in Fig.1 $\mathbf{v}_7$ by the red colors).

Construct graph edge weight: The graph edge weight of the Ncut should encode the similarity between nodes. Given two connected superpixels R_i and R_j in the graph, we suggest a combination of appearance (color) cue and intervening contour cue derived from a weighted combination:

$$d_{ij}^{app+edge} = (1-\alpha)d_{ij}^{app} + \alpha d_{ij}^{edge} \tag{3}$$

where d_{ij}^{app} is the appearance difference and d_{ij}^{edge} is the intervening contour magnitude. The appearance difference d_{ij}^{app} between two superpixels is similar to many previous works (e.g. [1, 2, 22]) and is defined as:

$$d_{ij}^{app} = ||\mathbf{c}_i - \mathbf{c}_j||_2 \tag{4}$$

As an improvement upon (4), in this paper the intervening contour cue is incorporated in (3). We define the intervening contour term d_{ij}^{edge} as:

$$d_{ij}^{edge} = \max_{\mathbf{p} \in l(\mathbf{p}_i, \mathbf{p}_j)} \mathbb{E}(\mathbf{p}) \tag{5}$$

where $l(\mathbf{p}_i, \mathbf{p}_j)$ is the straight line connecting the locations of superpixels R_i and R_j. $\mathbf{p}$ runs over every pixel location on the line and $\mathbb{E}(\mathbf{p})$ is the corresponding edge probability on an edge map $\mathbb{E}$. (5) is known as *intervening contour cue* [31]. In certain cases, using an edge map provides better delineation between objects and background than (4). Note any edge detector that outputs boundary probability map can be employed. Here we use the structured random forest edge detector [32] which works on multiscales and has state-of-the-art performance with decent processing speed.

Specially, (5) should not hold for boundary superpixels since the spatial distance between two boundary superpixels could be large. For example one superpixel is on the left border while the other is on the right border of an image. So d_{ij}^{edge} can be extremely large as there often exists strong edges on the straight line connecting them, which prevents clustering between potential background nodes. Hence for d_{ij}^{edge} among all boundary superpixels, we let $d_{ij}^{edge} = d_{ij}^{app}$ to alleviate this problem. To make the two terms addable in (3), d_{ij}^{edge} and d_{ij}^{app} are normalized to the same scale [0,1] by dividing their global maximum, respectively. In (3) α specifies the relative importance of the edge detector. We did not tune this parameter carefully and empirically set it as 0.5 to render them with equal weight, which works well in our tests.

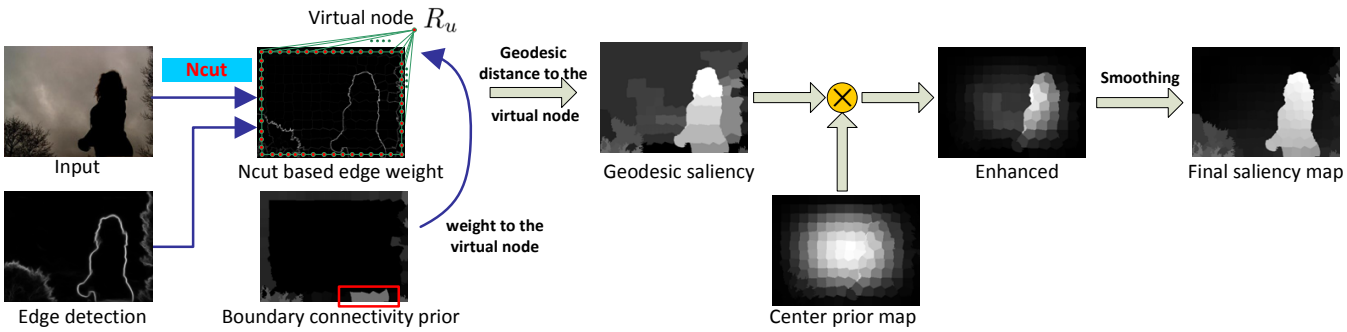

Fig. 2. The block diagram of the proposed method. Boundary superpixels are connected with each other and acquire connection to a virtual node R_u. Note in the boundary connectivity prior map, boundary superpixels belonging to a salient object obtain high prior energy (highlighted by the red box), contributing to larger weight from boundary nodes towards the virtual node. Non-boundary superpixels are shown in black in this prior map.

We define entries of matrix $\mathbf{W}$ by using feature difference computed in (3) as:

$$w_{ij} = \begin{cases} \exp(-\frac{d_{ij}^{app+edge}}{\sigma^2}) & \text{if } R_i \text{ and } R_j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

σ^2 is empirically set as 0.1 for all experiments. Diagonal entries of $\mathbf{W}$ are set to zeros to prevent self-loops in the graph. w_{ij} measures the affinity (similarity) between vertices and is close to 1 if two connected vertices are deemed similar.

3.2. Geodesic Saliency Detection based on the Ncut

We solve for the generalized eigenvectors $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{nvec}$ (correspond to $nvec+1$ smallest eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{nvec}$) of system (2). Note the resulting eigenvectors of the Ncut are indeed the soft indicator vectors of different clusters [29]. In the following, we develop a new method to leverage these eigenvectors. Similar to [1], we first use the inter-class distance to reconstruct the graph edge weight as:

$$\widehat{e}_{ij} = \sum_{k=1}^{nvec} \frac{1}{\sqrt{\lambda_k}} |\mathbf{v}_k(R_i) - \mathbf{v}_k(R_j)| \quad (7)$$

where $\mathbf{v}_k(R_i)$ indicates the value in eigenvector $\mathbf{v}_k$ corresponding to superpixel R_i. $\widehat{e}_{ij}$ is again normalized to [0,1] by dividing its global maximum. In practice, $nvec = 8$ suffices while further increasing it introduces extra noise. Such reconstructed edge weight is visualized in both examples in Fig.1 and Fig.2 as the boundary between superpixels. The larger weight yields to higher boundary intensity. Note only edge weight between adjacent superpixels is visualized though also there is weight between non-adjacent superpixels due to the 2-ring graph topology and the boundary connection.

Instead of region merging in [1], we find that a straightforward way is to compute the geodesic distance based saliency [21]. Since the inter-class distance (7) indicates how much superpixels belong to different clusters, clusters that could easily connect to the image boundary are more likely to be background [21]. To alleviate cases where objects are slightly cropped by image boarders, we add a virtual background node R_u to the graph (Fig.2). The virtual node R_u only acquires connection to all boundary vertices. Suppose R_k denotes a boundary superpixel. Instead of computing 1D contrast saliency on boundary superpixels as the weight [21], we use a more robust boundary connectivity prior [35] as the weight $\widehat{e}_{ku}$ (Fig.2). We essentially compute a soft region $\mathcal{A}_k$ associated with each boundary node R_k, and the prior is defined as $exp(-\frac{Ratio^2(R_k)}{2\sigma_{bc}^2})$, where $Ratio(R_k) = Len(\mathcal{A}_k)/\sqrt{Area(\mathcal{A}_k)}$. $Len(\mathcal{A}_k)$ is the cropped length of $\mathcal{A}_k$ by image boundary and $Area(\mathcal{A}_k)$ is the area of $\mathcal{A}_k$. The prior is close to 1 if the boundary node is deemed belonging to a salient object that is slightly cropped on image boundary. On the other hand, it is close to 0 if it is deemed as background which is largely cropped by the image boundary. With the reconstructed graph edge weight and the virtual node, we define the geodesic saliency of an arbitrary superpixel R_i as the shortest graph path towards R_u:

$$GeoSal(R_i) = \min_{P_1=i,P_2,...,P_n=u} \sum_{k=1}^{n-1} \widehat{e}_{P_k P_{k+1}} \quad (8)$$

s.t.: R_{P_k} and $R_{P_{k+1}}$ are adjacent vertices on the graph

where the shortest path $\{P_1, \cdots, P_n\}$ that minimizes (8) is found by *Dijkstra's* algorithm. Similar to [21], before computing the shortest paths, we perform small edge clip on $\widehat{e}_{ij}$ (0.05 is set as a reference after normalizing $\widehat{e}_{ij}$ into [0,1]).

By considering the center bias, $GeoSal(R_i)$ is enhanced by a 2D Gaussian distribution $\mathbb{G}(\mathbf{p})$ which is located at the image center (Fig.2):

$$Sal(R_i) = GeoSal(R_i) \times \mathbb{G}(\mathbf{p_i}) \quad (9)$$

where $\mathbb{G}(\mathbf{p}_i)$ corresponds to Gaussian value of superpixel location $\mathbf{p}_i$. Although it has been argued in [21] that boundary hypothesis is more generic than center prior, we still find that combining the latter is useful when there are multiple regions disconnected from image boundary but scattered in the whole image. The final saliency value $\overline{Sal}(R_i)$ is formulated by applying the manifold ranking [22] to smooth the combined map

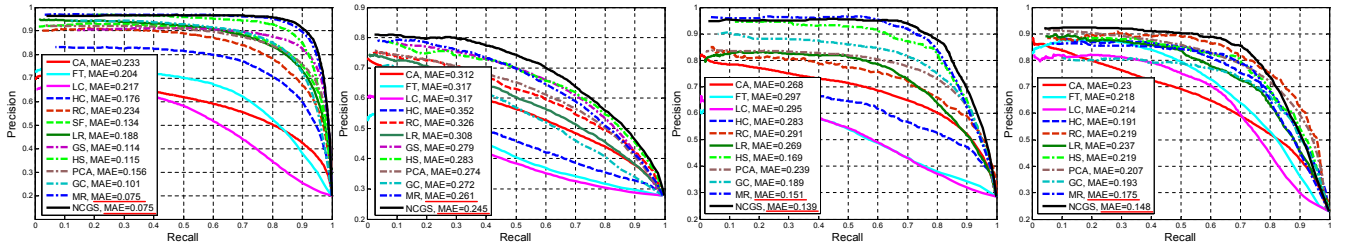

Fig. 3. Quantitative comparisons on four benchmark datasets with 12 existing methods. From left to right are respectively: MSRA-1000 [8], SOD [33], SED1, and SED2 [34]. Both the precision-recall curve and the mean absolute error (MAE) are used as criteria. The proposed method (NCGS) generates black curves that are competitive to the best state-of-the-art methods GS [21] (pink dot-dash), HS [17] (green dot-dash) and MR [22] (blue dot-dash). The first two lowest MAE scores among methods are underlined by red color. NCGS consistently achieves the lowest MAE on all four datasets. Note because SF only provides results on MSRA-1000 while GS only provides results on MSRA-1000 and SOD. We can not compare with them on the rest sets.

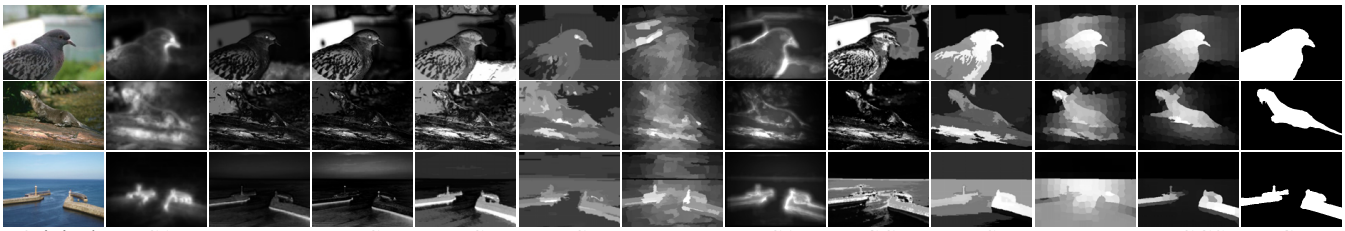

Fig. 4. Visual comparisons on several challenging examples, where NCGS achieves visually the best results.

in (9). The affinity matrix for the manifold ranking is the same as $\mathbf{W}$ in (6) and the query values are $Sal(R_i)$. Finally $\overline{Sal}(R_i)$ is normalized to the interval [0,1] to render a final saliency map. The processing pipeline of our Normalized cut based geodesic saliency detection is illustrated in Fig.2.

4. EXPERIMENTS AND RESULTS

In this section, we compare our scheme with state-of-the-art methods. Four benchmark datasets for evaluation include commonly used MSRA-1000 [8] (1000 images), SOD [33] (300 images), and SED [34] which consists of two subsets, i.e. SED1 (one object set) and SED2 (two objects set) each containing 100 images. We compare the proposed Normalized cut geodesic saliency technique (termed as NCGS) with 12 state-of-the-art methods: CA (Context Aware) [9], FT (Frequency Tuned) [8], LC (Luminance Contrast) [7], HC (Histogram Contrast) [10], RC (Regional Contrast) [10], SF (Saliency Filter) [11], LR (Low Rank) [25], GS (Geodesic Saliency) [21], HS (Hierarchical Saliency) [17], PCA [13], GC (Global Cue) [18], MR (Manifold Ranking) [22]. We use the precision-recall curve [8, 11, 22, 18] and the mean absolute error (MAE) [11, 18] for evaluation. A good saliency detection method should achieve high precision-recall curves meanwhile maintain low MAE scores. During all experiments, we keep parameters of NCGS the same as mentioned in Section 3.

Precision-recall curves generated by using fixed threshold from 0 to 255 are shown in Fig.3. The performance of NCGS is comparable to the best state-of-the-art techniques including GS, HS and MR. Our method significantly outperforms the original GS on the first two sets as well as HS on all datasets in terms of precision-recall curves. Besides, significantly better performance than MR of NCGS is observed on SOD and SED2 (Fig.3). For MSRA-1000 and SED1, only marginal improvement is observed. Another encouraging observation is from MAE scores. NCGS consistently achieves the lowest MAE scores on all datasets, indicating robustness on varied datasets. It has provided further support to our technique on achieving both high precision-recall curves with low MAE. Fig.4 show visual comparisons. Our method effectively suppresses background clutter and uniformly emphasizes foreground objects. Better delineation between object and background can be observed in complex scenes, such as the second row in Fig.4. Using the Ncut can effectively group the visual contents and provide ability to delineate entire objects from the background. By incorporating it with the geodesic saliency detection, superior performance is obtained.

5. CONCLUSION

This paper proposes a novel scheme to leverage the eigenvectors of the Normalized cut for salient object detection. We adapt the geodesic saliency detection to this problem. Our method computes geodesic distance (shortest pathes) to image boundary on the graph whose edge weight is reconstructed via the inter-class distance derived from the Ncut eigenvectors. Prior to the Ncut, the graph is constructed with a moderate number of superpixels as vertices. Graph affinity is computed by considering both superpixel appearance difference and intervening contour cue. Experiments on four public datasets have shown our method achieves state-of-the-art performance by providing competent precision and recall, meanwhile maintaining the lowest MAE.

6. REFERENCES

[1] K. Fu, C. Gong, Y. Yu, Y. Li, I. Gu, J. Yang, and J. Yu, "Adaptive multi-level region merging for salient object detection," in *BMVC*, 2014.

[2] K. Fu, C. Gong, I. Gu, J. Yang, and X. He, "Spectral salient object detection," in *ICME*, 2014.

[3] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *TPAMI*, vol. 35, no. 1, pp. 185–207, 2013.

[4] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *CVPR*, 2007.

[5] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *NIPS*, 2005.

[6] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *TPAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.

[7] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," *ACM Multimedia*, pp. 815–824, 2006.

[8] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009.

[9] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *CVPR*, 2010.

[10] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *CVPR*, 2011.

[11] F. Perazzi, P. Krahenbul, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*, 2012.

[12] K. Shi, K. Wang, J. Lu, and L. Lin, "Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors," in *CVPR*, 2013.

[13] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct," in *CVPR*, 2013.

[14] K. Fu, C. Gong, J. Yang, and Y. Zhou., "Salient object detection via color contrast and color distribution," in *ACCV*, 2012.

[15] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *CVPR*, 2013.

[16] T Liu, Z Yuan, J Sun, J Wang, and N Zheng, "Learning to detect a salient object," *TPAMI*, vol. 33, no. 2, pp. 353–367, 2011.

[17] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *CVPR*, 2013.

[18] M. Cheng, J. Warrell, W. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *ICCV*, 2013.

[19] J. Lafferty, A. McCallum, and F. Pereira, "Submodular salient region detection," in *CVPR*, 2013.

[20] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *TIP*, vol. 19, no. 12, pp. 3232–3242, 2010.

[21] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *ECCV*, 2012.

[22] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang, "Saliency detection via graph-based manifold ranking," in *CVPR*, 2013.

[23] B. Jiang, L. Zhang, H. Lu, C. Yang, and M. Yang, "Saliency detection via absorbing markov chain," in *ICCV*, 2013.

[24] K. Fu, C. Gong, I. Gu, and J. Yang, "Geodesic saliency propagation for image salient region detection," in *ICIP*, 2013.

[25] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *CVPR*, 2012.

[26] Y. Xie and H. Lu, "Visual saliency detection based on bayesian model," in *ICIP*, 2011.

[27] L. Mai, Y. Niu, and F. Liu, "Saliency aggregation: A data-driven approach," in *CVPR*, 2013.

[28] J. Shi and J. Malik, "Normalized cuts and image segmentation," *TPAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[29] U. von Luxburg, "A tutorial on spectral clustering. statistics and computing," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[30] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[31] P. Arbelaez, M. Maire, and C. Fowlkes et al, "Contour detection and hierarchical image segmentation," *TPAMI*, vol. 33, no. 5, pp. 898–916, 2010.

[32] P. Dollár and C.L. Zitnick, "Structured forests for fast edge detection," in *ICCV*, 2013.

[33] V. Movahedi and J. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *IEEE Computer Society Workshop on Perceptual Organization in Computer Vision*, 2010.

[34] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *CVPR*, 2007.

[35] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *CVPR*, 2014.