2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

SALIENCY DRIVEN CLUSTERING FOR SALIENT OBJECT DETECTION

Lei ZHOU[1], *YiJun LI*[1], *YiPeng SONG*[1], *Yu QIAO*[1] *and Jie YANG*[1]*

[1] Institute of Image Processing and Pattern Recognition Shanghai Jiaotong University,China.

ABSTRACT

In this work, a novel salient object detection method is proposed based on the saliency driven clustering. To capture visual patterns of an image, the color contrast prior and boundary prior are utilized to generate the image clusters automatically. Then, a simple operation like regional saliency computation is applied to refine the saliency maps generated by two priors. The final saliency map are obtained by combining the refined contrast prior saliency and boundary prior saliency. Extensive experiments show that our proposed model achieves better performance on salient region detection against the state-of-the-art methods.

Index Terms— Saliency driven clustering, regional saliency computation, boundary prior, color contrast prior.

1. INTRODUCTION

Visual saliency detection is a hot research topic and has wide applications, such as content-based image retrieval [1], image/video compression and coding [2], object recognition and scene understanding [3] in areas of computer vision and computer graphics. In natural scene, salient regions generally stand out relative to its surroundings. This mechanism can be explained by a center-surround difference model [4], which is implemented in the feature spaces of luminance, color and orientation.

In recent years, salient object detection has aroused researches' interest and the related work has been divided into two categories, i.e, approaches of bottom-up category and approaches of top-down category respectively. In bottom-up visual saliency, previous research [5, 6] revealed that *contrast* is the most influential factor in low-level visual saliency. By computing the contrast over a pixel domain or region domain, many visual saliency models have been proposed over the past year [7, 8, 9, 10, 11, 12, 13]. The existing saliency models based on color contrast can simultaneously compute global contrast of an image and spatial coherence between regions and have displayed impressive results. The contrast-based models tell "what the objects look like" by highlighting the pixels with great center-surround difference. However, the performance of the saliency maps that only rely on color contrast will degrade when the images are of confusing pattern or complex scene.

Different from the contrast prior, background prior tackles the salient object detection problem by asking the question "what the background should look like". In [14], two priors, boundary and connectivity prior were used as the priors about common backgrounds in natural images. The boundary prior was discovered from the observations that "the image boundary is mostly background" and "the salient objects seldom touch the image boundary". Even if the boundary priors work for most images, it may fail when objects significantly touching the image boundary or the images are of complex background.

Intuitively, the combination of priors about "what the salient objects look like" and "what the background should look like" may exhibit diverse and meaningful visual pattern information of natural images. In this paper, we present a method for efficiently creating saliency maps by capturing visual patterns of images. To extract the structural information of an image, we propose a saliency driven clustering method based on the combination of contrast prior saliency and boundary prior saliency. The cluster number and clusters' centers are determined automatically by analyzing the histogram of combined saliency, which are used for generating image clusters applying k-means. Then, average contrast prior saliency and average boundary prior saliency are computed to demonstrate the role of clustering in improving the quality of saliency map. Our experimental results demonstrate that meaningful visual structure information can be extracted from contrast prior and boundary prior by clustering. Then, a simple operation like regional saliency computation can improve the performance of the two priors greatly.

2. SALIENCY DRIVEN CLUSTERING

2.1. Definition of Priors for Saliency Computation

Color Contrast Prior Saliency: Color contrast is inspired by the observation that color components of a salient object may have a strong contrast to their surroundings. Assume that an image with size N is divided into regions (or superpixels) $R_i, i \in \{1, 2, 3, ..., M\}$. Then, region R_i's color contrast saliency S_i^{con} is computed according to the definition in [15]:

$$S_i^{Rcon} = \sum_{j \neq i} D_c(R_i, R_j) D_s(R_i, R_j), \tag{1}$$

where $D_c(R_i, R_j)$ is the color distance between the two regions and $D_s(R_i, R_j)$ in 1 stands for the spatial distance between regions R_i and R_j. In the experiment, 300 superpixels are generated using SLIC [16]. Finally, the pixel level color contrast saliency is given as $S_i^{con} = S_j^{Rcon}, i \in R_j$. We normalize S^{con} to the range [0,1] $S^{con} = (S^{con} - min(S^{con}))/(max(S^{con}) - min(S^{con}))$.

Boundary Prior Saliency: As stated in [17] that boundary prior is an important and helpful measure for salient object

Corresponding author: Jie Yang, jieyang@sjtu.edu.cn

detection. For an image, the pixel-level undirected weighted graph is represented as $G = \{V, \varepsilon\}$. In the graph, the boundary nodes (Ω_B) are selected using the strategy similar with [17]. Then, the geodesic distance of a pixel m to the boundary nodes is computed as the nearest distance to all the background nodes

$$g(m) = \min_{s \in \Omega_B} d_g(s, m), \tag{2}$$

where $d_g(a, b)$ is the geodesic distance between two nodes a and b, which is computed based on the length of a discrete path:

$$d_g(a, b) = \min_{\Gamma \in P_{a,b}} L(\Gamma), \tag{3}$$

where $P_{a,b}$ stands for the set of paths between node a and b, and the length L of a discrete path Γ is defined as:

$$L(\Gamma) = \sum_{i=1}^{n-1} \sqrt{(1-\gamma_g) d(\Gamma^i, \Gamma^{i+1})^2 + \gamma_g \parallel \nabla(\Delta^i) \parallel^2}, \tag{4}$$

where Γ is an arbitrary discrete path composed with n pixels $\{\Gamma^1, ..., \Gamma^n\}$. $d(\Gamma^i, \Gamma^{i+1})$ is the Euclidean distance between two pixels Γ^i and Γ^{i+1}) and $\parallel \nabla(\Delta^i) \parallel$ is the a finite difference approximation of the image gradient between Γ^i and Γ^{i+1}. We use the parameter γ_g to weight two kinds of distances: the Euclidean distance and the geodesic distance. The role of γ_g has been studied in [18] and we set $\gamma_g = 0.2$ in the experiments. The pathes in eq. (4) are computed using fast marching algorithm [19]. Then, the boundary prior saliency is defined as $S_i^{Boundary} = g(i), i \in [1, ..., N]$ and it is normalized to the range [0,1] as well.

2.2. Saliency Driven Clustering

To integrate the complementary strength of two kinds of saliency maps, the contrast saliency and boundary saliency are combined nonlinearly,

$$S_i^{cb} = S_i^{con} * S_i^{Boundary}. \tag{5}$$

Then, the image space is separated into several regions by K-means clustering [20] based on S_i^{cb}. We run K-means to generate K clusters and the number K and the initial positions of centroids are determined automatically using the Hill Climbing algorithm [21]. The procedure is described below:

1. To construct the histogram of combined saliency map S^{cb}. An example is displayed in Fig. 1 and the histogram are displayed in Fig. 1 (c).

2. We construct a search window of size 30, the center of the search window will moves from 0 to 255 to search for the peak of histogram. The number of pixels of current bin is compared with the neighbouring bins' pixel numbers, and the current bin will be selected as a peak if its pixel number is the largest in the search window. We take the number of peaks as the clusters number K and the set of peak bins Pb are recorded as the starting centroids for classification. The search window is illustrated in Fig. 1 (d).

3. K-means clustering is applied with the adaptive numbers K and selected starting centroids set Pb. Then K clusters $RS_1, ..., RS_K$ are generated. An example is displayed in Fig. 2, the image space is separated into nine regions according to histogram analysis.

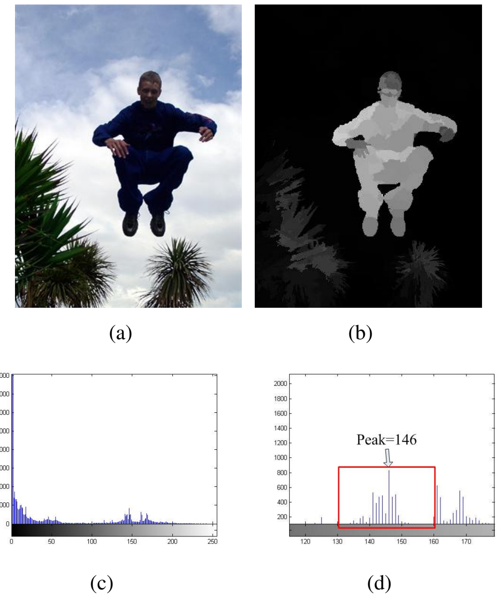


Fig. 1: An example illustrates the procedure of peak selection by histogram analysis. Seven peaks are located [0, 46, 89, 115, 146, 192, 255]. (a) The source image; (b) S^{cb}; (c) Histogram ranges from 0 to 255; (d) The selected Peak 146. The red rectangle is the search window;

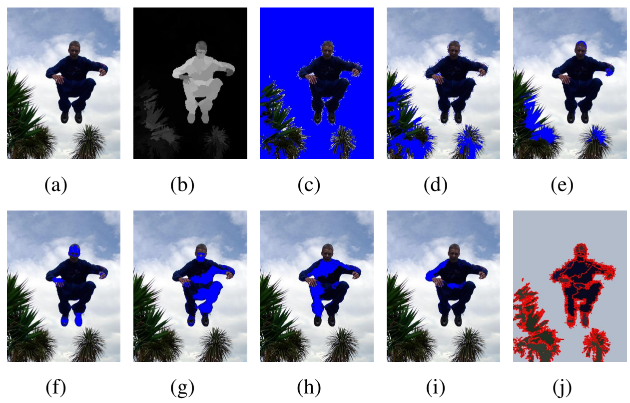


Fig. 2: Illustration of the process of clustering using adaptive K-means. (a) original image; (b) the combined saliency map; (c)-(i) are the separated nine regions (marked in blue); (j) Clustering map, different clusters are labeled different colors;

3. REGIONAL SALIENCY COMPUTATION

To incorporating the visual pattern information into the saliency detection model, the operation of regional saliency computed is introduced. The average color contrast saliency values and boundary prior values are computed for the generated cluster. Then, the regional saliency values are combined effectively to generate the final pixel level saliency map.

The Average Color Contrast Saliency: The average color saliency value for region RS_i is defined as

$$color(i) = \frac{\sum_{x \in RS_i} S_x^{con}}{|RS_i|}, i \in \{1, ..., K\}. \tag{6}$$

The regions with higher average saliency values are more likely to be contained in a salient object.

The Average Boundary Prior Saliency: The average boundary saliency of region RS_i is defined as:

$$bound(i) = \frac{\sum_{x \in RS_i} S_x^{Boundary}}{|RS_i|}, i \in \{1, ..., K\}. \quad (7)$$

The average saliency values *color* and *bound* of regions are projected onto the pixels contained. The pixel level refined saliency values are define as

$$\begin{aligned} M_{color}(x) &= color(i), x \in RS_i, x = [1, 2, .., , N], \\ M_{bound}(x) &= bound(i), x \in RS_i, x = [1, 2, .., , N], \end{aligned} \quad (8)$$

where N is the size of image. The refined saliency map is normalized to the range [0,1], $M = \frac{M - min(M)}{max(M) - min(M)}$. The final saliency value at pixel x is defined as:

$$S^{final}(x) = M_{color}(x) * M_{bound}(x), x = [1, 2, .., , N], \quad (9)$$

where the S^{final} is normalized to the range [0,1] as well. Some visual results of the regional saliency maps and combined saliency maps are displayed in Fig. 3 and we can see that the regional saliency maps perform well in suppressing background noises and highlighting the salient objects. The structural information of images reflected by clusters is useful for generating saliency maps of high quality. In the example listed in the 3rd row, the sky and grass near the boundary are classified into one cluster. Even if the saliency values for sky in S^{con} are high, the regional saliency value are low by computing the average saliency value on a larger area.

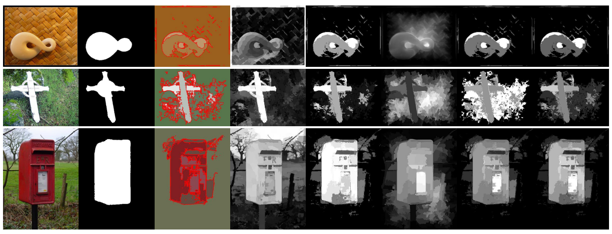


Fig. 3: Illustration of regional saliency values. From left to right: original image, ground truth, the generated clusters, color contrast saliency (S^{con}), average color saliency (M_{color}), boundary prior saliency ($S^{boundary}$), average boundary prior saliency (M_{bound}) and the combined final saliency map S^{final}. 3 clusters, 4 clusters and 4 clusters are generated for the three examples listed respectively.

4. EXPERIMENT

The empirical analysis is implemented on two popular saliency databases: MSRA-1000 [22] and Berkley-300 database [23]. For the first quantitative comparison, the precision and recall rates of various models are computed and the average precision-recall (PR) curve is obtained by averaging the results from all the testing iamges. In addition to PR curves, for each image, we follow [22, 24] to segment a saliency map by adaptive threshold

$$T_s = \min\{2 \times \frac{\sum_i^N V_i}{N}, T_{max}\}, \quad (10)$$

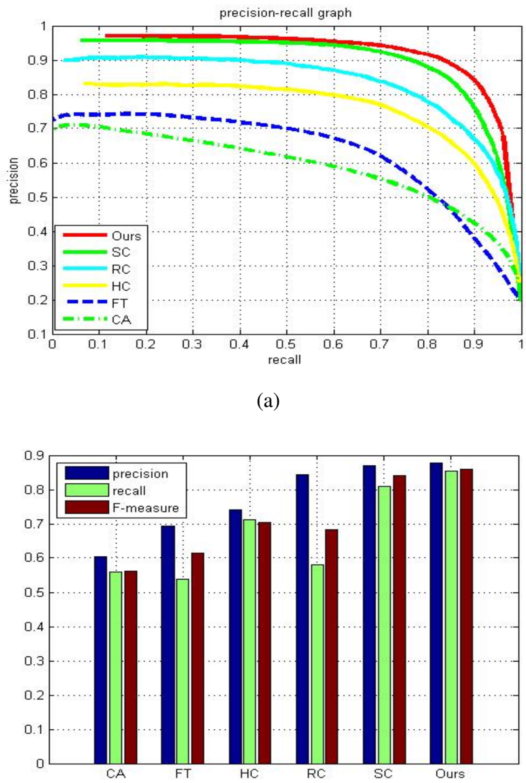
where N denotes the number of pixels in the saliency map and i is the pixel index. V_i is the saliency value on pixel i. T_{max} is the upper bound for the saliency value. In the experiment, the saliency values are projected into the range of [0,255] and we set $T_{max} = 255$ in the experiment. Then the precision, recall and F-Measurement values are computed over the ground truth maps, where F-Measurement is defined as $F_\beta(T) = \frac{(1+\beta^2)Precision(T) \times Recall(T)}{\beta^2 \times Precision(T) + Recall(T)}$, where $\beta = 0.3$.

4.1. MSRA-1000 Dataset

First, we generate the saliency maps for all 1000 testing images using the proposed saliency model S^{final} (eq. 9). The saliency detection performance of the proposed saliency model is compared with five state-of-the-art saliency models, that are FT [22], HC [25], RC [25], CA [13] and SC [15].

(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4: Subjective comparison of our saliency map with five stat-of-the-art methods on the MSRA-1000 database. (a) original images, saliency maps of (b) CA, (c) FT, (d) HC, (e)RC, (f)SC, (g) the proposed method, (h)the ground truth.



(a)

(b)

Fig. 5: Experiment results on MSRA dataset. (a) the PR curves for approaches comparison. (b) the results of adaptive segmentation. The picture in the third row displays the PR curves for refined and non-refined results.

The visual results of our method and the other five compared approaches are listed in Fig. 4. In Fig. 5 (a), we compare our precision-recall curve with other methods. The proposed saliency map demonstrates the highest recall rate and precision rate compared with other methods. Fig. 5 (b) displays the average precision, recall and F-measure values in the adaptive threshold experiment. Among all the approaches,the proposed method achieves higher precision, recall and F-measure values compared with other approaches.

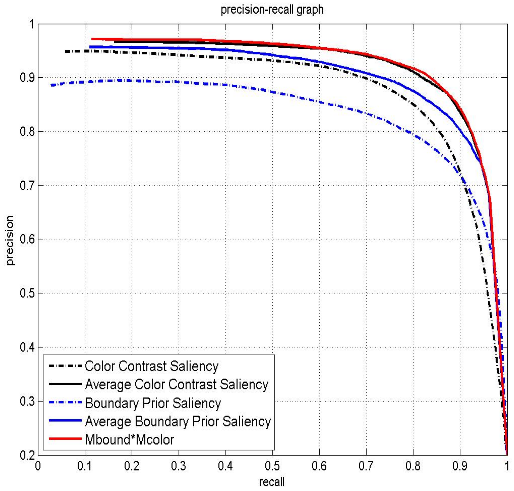


Fig. 6: Experiment results on MSRA dataset. The picture displays the PR curves for refined and non-refined results.

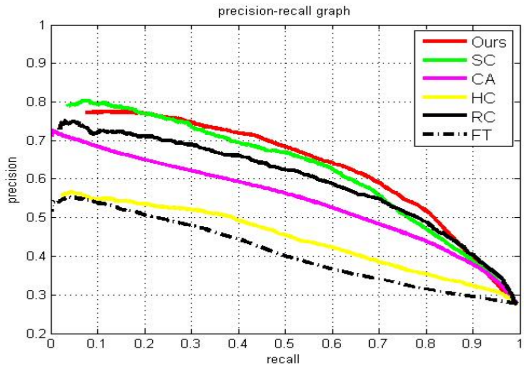
The operation of clusters based refinement can improve the performance greatly. It is observed that both refined contrast based saliency and refined boundary prior saliency achieve dramatic performance improvement compared with the original curves.

4.2. Berkley-300 Dataset

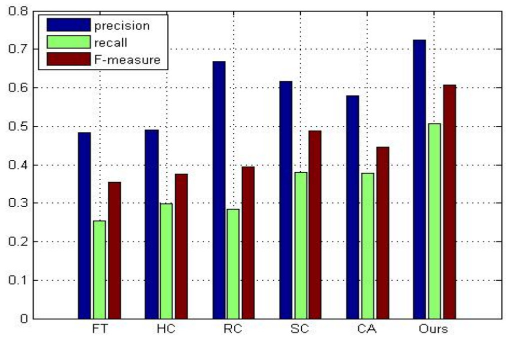
(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 7: Subjective comparison of our saliency map with five stat-of-the-art methods on the Berkley-300 database. (a) original images, saliency maps of (b) CA, (c) FT, (d) HC, (e)RC, (f)SC, (g) the proposed method, (h)the ground truth.

We compare the curve of our approach with FT [22], HC [25], RC [25], CA [13] and SC [15]. The PR curves are shown in Fig. 8 and the average precision, recall and F-measure values is displayed in the second row of Fig. 8. Our method achieves the best performance both in the PR curve and the adaptive segmentation experiment. The visual results for comparison are listed in Fig. 7.



(a)



(b)

Fig. 8: Experiment results on Berkley dataset. (a) the PR curves. (b) the results of adaptive segmentation.

5. CONCLUSION

In this paper, we have presented a salient object method by combining the priors about "what the salient objects look like" and "what the background should look like". To exhibit diverse and meaningful visual pattern information of natural images, we propose a saliency driven clustering method based on the combination of contrast prior saliency and boundary prior saliency. Experimental results on the most popular datasets demonstrate that computing regional saliency values based on the generated clusters can improve the performance of color contrast prior or boundary prior greatly. The comparison of experimental results also indicate the advantages of building a saliency model based on the high-order visual information of images, such as clusters. Since a simple clustering method based on histogram analysis is used for saliency detection, we will exploit a more effective clustering strategy for generating semantic regions in our future work.

Acknowledgements: This research is partly supported by NSFC, China (No:61375048,31100672), Ph.D. Programs Foundation of Ministry of Education of China (No.20120073-110018)

6. REFERENCES

[1] Hong Fu, Zheru Chi, and Dagan Feng, "Attention-driven image interpretation with application to image retrieval," *Pattern Recognition*, vol. 39, no. 9, pp. 1604–1621, 2006.

[2] Laurent Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *Image Processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1304–1318, 2004.

[3] Paul L Rosin, "A simple method for detecting salient regions," *Pattern Recognition*, vol. 42, no. 11, pp. 2363–2371, 2009.

[4] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1254–1259, 1998.

[5] Wolfgang Einhäuser and Peter KoÈnig, "Does luminance-contrast contribute to a saliency map for overt visual attention?," *European Journal of Neuroscience*, vol. 17, no. 5, pp. 1089–1097, 2003.

[6] Derrick Parkhurst, Klinton Law, Ernst Niebur, et al., "Modeling the role of salience in the allocation of overt visual attention," *Vision research*, vol. 42, no. 1, pp. 107–124, 2002.

[7] Laurent Itti, Nitin Dhavale, and Frederic Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in *Optical Science and Technology, SPIE's 48th Annual Meeting*. International Society for Optics and Photonics, 2004, pp. 64–78.

[8] Olivier Le Meur, Patrick Le Callet, Dominique Barba, and Dominique Thoreau, "A coherent computational approach to model bottom-up visual attention," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 5, pp. 802–817, 2006.

[9] Harel J, Koch C, and Perona P, "Graph-based visual saliency," in *Neural Information Processing Systems (NIPS)*, 2006.

[10] Umesh Rajashekar, Ian van der Linde, Alan C Bovik, and Lawrence K Cormack, "Gaffe: A gaze-attentive fixation finding engine," *Image Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 564–573, 2008.

[11] Jie Feng, Yichen Wei, Litian Tao, Chao Zhang, and Jian Sun, "Salient object detection by composition," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1028–1035.

[12] Yao Lu, Wei Zhang, Hong Lu, and Xiangyang Xue, "Salient object detection using concavity context," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 233–240.

[13] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal, "Context-aware saliency detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 10, pp. 1915–1926, 2012.

[14] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *Computer Vision–ECCV 2012*, pp. 29–42. Springer, 2012.

[15] Keren Fu, Chen Gong, Jie Yang, and Yue Zhou, "Salient object detection via color contrast and color distribution," in *Computer Vision (ACCV), 2012 IEEE 11th Asian Conference on*. IEEE, 2012.

[16] William F Noh and Paul Woodward, "Slic (simple line interface calculation)," in *Proceedings of the Fifth International Conference on Numerical Methods in Fluid Dynamics June 28–July 2, 1976 Twente University, Enschede*. Springer, 1976, pp. 330–340.

[17] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *Computer Vision, 2012 Europe Conference on*, pp. 29–42. Springer, 2012.

[18] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3129–3136.

[19] Liron Yatziv, Alberto Bartesaghi, and Guillermo Sapiro, "O(n) implementation of the fast marching algorithm," *Journal of computational physics*, vol. 212, no. 2, pp. 393–399, 2006.

[20] John A Hartigan and Manchek A Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[21] T Ohashi, Z Aghbari, and A Makinouchi, "Hill-climbing algorithm for efficient color-based image segmentation.," in *IASTED International Conference on Signal Processing, Pattern Recognition, and Applications*, 2003, pp. 17–22.

[22] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk, "Frequency-tuned salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2009 IEEE Conference on*. IEEE, 2009, pp. 1597–1604.

[23] Vida Movahedi and James H Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 49–56.

[24] Federico Perazzi, Philipp Krahenbuhl, Yael Pritch, and Alexander Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 733–740.

[25] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 409–416.