# GESTALT SALIENCY: SALIENT REGION DETECTION BASED ON GESTALT PRINCIPLES

*Jie Wu and Liqing Zhang*

MOE-Microsoft Laboratory for Intelligent Computing and Intelligent Systems
Dept. of CSE, Shanghai Jiao Tong University

## ABSTRACT

Salient region detection is of great significance in computer vision such as object recognition, image segmentation and image retrieval. However, low-level saliency has certain limitations due to lack of object level information. In this paper, we propose a saliency detection method based on Gestalt principles in which we introduce mid-level Gestalt concepts for low-level saliency. We propose an algorithm based on Gestalt principles of *similarity & anomaly* to select and suppress the similar background regions, using variance of clusters of image regions. Moreover, we propose two smoothing procedures based on Gestalt principles of *similarity & proximity* to group near and similar regions and therefore uniformly highlight the salient object. Experimental results on public data set show that our method performs well compared with state-of-the-art approaches.

***Index Terms*—** Image saliency, Salient region

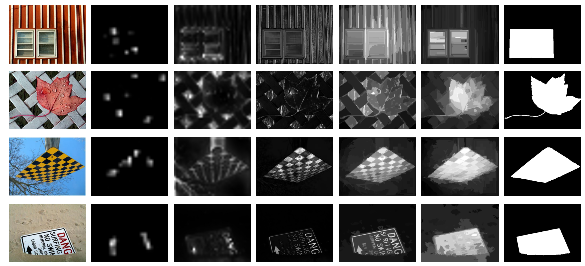



**Fig. 1**. Saliency maps of previous and our approaches. (a) original image, (b-f) saliency maps of IT [9], GB [19], FT [14], RC [18] and our approach, (h) ground truth.

## 1. INTRODUCTION

Visual saliency is the perceptual quality which makes some items in the scene pop out from their neighbours and immediately grab our attention. Though visual saliency is a purely scientific issue, recently saliency detection methods have raised much interest in many applications such as image and video compression [1], image segmentation [2], and object recognition [3]. Saliency detection methods are roughly divided into two categories. One is to predict human fixation [4, 5], the other is object level saliency detection which aims at detecting salient objects [6, 7, 8]. In this work, we mainly focus on the *object level* saliency detection.

### 1.1. Previous Works in Low-level Saliency

In general, due to lack of high-level knowledge, bottom up saliency detection methods rely on low-level features such as intensity, orientation, color, etc to determine contrast of image regions relative to their surroundings. Various low-level processing methods have been used in saliency detection. Some methods utilize purely low-level features from local neighbourhoods, which is known as local methods [9, 10, 11, 6, 12, 13], others combine low-level processing and the consideration of property of entire image, which are known as global methods [14, 15, 16, 17, 18, 19].

Low-level features are essential in saliency detection methods, but still have certain limitations. Firstly, some pixels would be wrongly regarded as salient because they have strong local low-level features. These pixels are locally salient, but in object level, these pixels are not part of salient object. Top two rows in Figure 1 show that for images with cluttered or complex background, low-level methods cannot effectively inhibit background regions of images. This is because regions in cluttered background usually have very strong center-surround local contrast. Secondly, low-level methods cannot uniformly highlight entire object. This is understandable because due to lack of high-level knowledge, low-level features cannot detect entire object in object level. Last two rows in figure 1 show that saliency value of different part of salient object vary significantly. We can also observe a phenomenon that different methods use different low-level features which result in various saliency maps, but the common limitation is that approaches relying only on low-level processing are insufficient for object level saliency detection. These two problems are severe in local methods. Though local methods are able to find some human fixation points, it's difficult for them to find salient regions in object level. Global methods also suffer from these problems even if they take properties in entire image into account.

### 1.2. Gestalt principles for Low-level Saliency

To solve the problems mentioned above in low-level saliency, we introduce *mid-level* Gestalt concepts for *low-level* saliency detection. Gestalt principles refer to theories of visual perception which attempt to describe how people tend to organize visual elements into groups or unified wholes when certain principles are applied [20, 21]. Main Gestalt principles include similarity, continuation, closure, anomaly and proximity. In this work, we propose a saliency measure based on Gestalt principles, called *Gestalt saliency*. Firstly we propose an algorithm based on Gestalt principles of *similarity & anomaly* to select and suppress the similar background regions, using variance of clusters of image regions. Secondly, we propose two smoothing procedures in region level based on Gestalt principles of *similarity & proximity* to group near and similar regions together and uniformly highlight entire object.

Our idea of Gestalt saliency is not entirely new. Some models [22, 23] *explicitly* use Gestalt laws. Some global mod-

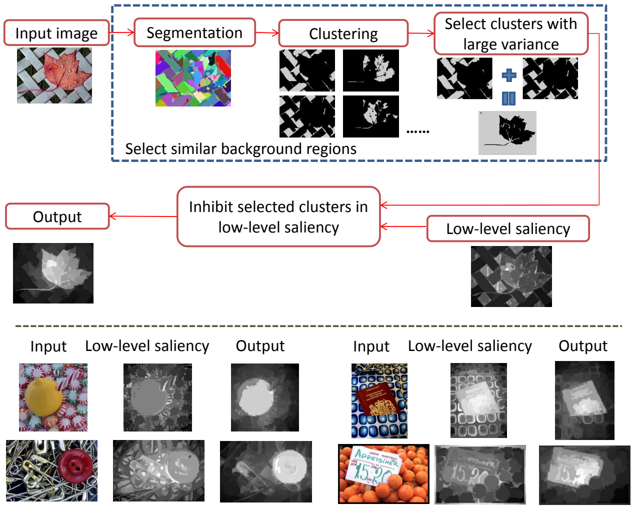



**Fig. 2**. Top: flow chart of Gestalt principle of similarity & anomaly. Bottom: original input, low-level saliency map and final saliency map after procedure based on similarity & anomaly.

els [24, 17] *implicitly* use Gestalt principles, because some laws such as similarity, proximity are widely used in saliency models or segmentation. Our method differs from these methods because 1) most of them perform Gestalt laws in pixel level, while we perform Gestalt descriptors in region level, which is a key to capture middle-level features. 2) they focus more on low-level features and consider less in Gestalt laws, while our method gives a stronger assumption in Gestalt laws, resulting in better performance in object level saliency detection.

## 2. PROPOSED ALGORITHM

In this section, we describe the two steps of our method: 1) Inhibiting similar background regions based on Gestalt principles of similar & anomaly. 2) Smoothing based on Gestalt principles of similar & proximity.

### 2.1. Similarity & Anomaly

To effectively suppress saliency values in complex background of image, we propose a method inspired by Gestalt principles of similarity and anomaly. *Similarity* occurs when objects look similar to one another. People often perceive them as a group or pattern. When similarity occurs, an object that is extremely dissimilar to the others is emphasized. This is called *anomaly*. Gestalt principles of similarity and anomaly indicate that human tends to *inhibit* similar background regions when there *exists* out-standing regions. For the input image in Figure 2, people tend to focus on the red leaf (out-standing region) but not the black blocks (similar background regions) because of similarity and anomaly. If we can *select* similar background regions and *suppress* the saliency values of them, foreground salient object can be emphasized more clearly. Figure 2 shows the flow chart of this procedure.

#### *2.1.1. Selecting Similar Background Regions*

Given an input image, we first perform a image segmentation method [25] to group similar pixels into regions. However, the pixels grouped by segmentation are restricted to be connected in spatial domain. For images with cluttered or complex background, similar background regions can not be grouped together. To group similar but scattered regions together, we perform a clustering algorithm to select similar background regions based on color distance between regions. We choose Partitioning Around Medoids (PAM) algorithm [26] which is a most common and simple realisation of $k$-medoids clustering. We fix number of clusters K = 8 in experiments.

Intuitively, similar background regions have a more scattered spatial distribution, so after grouping similar regions into clusters, we use *spatial variance* of cluster to separate foreground clusters and background clusters. Mathematically, for a cluster $R$ containing similar regions, variance of the cluster is defined as

$$Var(R) = \frac{1}{|R|} \sum_{r_i \in R} \|CG(r_i) - center(R)\| \tag{1}$$

where $CG(r_i)$ is the center of gravity (CG) of region $r_i$, $center(R) = \frac{1}{|R|} \sum_{r_i \in R} CG(r_i)$ is the center of cluster $R$. Note that before calculate CG of each region $r_i$, we normalize the spatial coordinate of each pixel in image to $[0..1] \times [0..1]$ to ensure the same weights of horizontal and vertical coordinates of pixels.

Bigger regions should have a higher weight, so we add area of region to refine the variance of cluster $R$:

$$Var'(R) = \frac{1}{|R|} \sum_{r_i \in R} \|CG(r_i) - center'(R)\| Area(r_i) \tag{2}$$

where $center'(R) = \frac{1}{|R|} \sum_{r_i \in R} CG(r_i) Area(r_i)$, and $Area(\cdot)$ is pixel number of region to give higher weights to bigger region.

Intuitively, clusters with relatively larger spatial variance should be chosen as background regions we want to inhibit. Moreover, anomaly occurs when there *exists* out-standing regions besides similar background regions, so it's necessary to ensure there remains salient (out-standing) regions after selecting the similar background regions. We employ low-level saliency to ensure whether the rest clusters contain certain portion of salient regions. Specifically, we sort clusters according to their variance from large to small and get the sequence $R_1, R_2 ... R_K$. We take the first $t$ clusters with largest variances, $R_1, R_2 ... R_t$, to form the similar and scattered background regions $R_{final}$:

$$\begin{aligned} R_{final} = \{ & R_1 \cup R_2 ... \cup R_t | t = \max t', \text{satisfing} \\ & var(R_{t'}) \geq \alpha, \sum_{i=1}^{t'} S(R_i) < \beta \sum_{i=1}^{K} S(R_i) \} \end{aligned} \tag{3}$$

where $S(R_i)$ is sum of low-level saliency of each pixel in $R_i$. We set $\alpha = \frac{1}{4}$ and $\beta = 95\%$. Since contrast is the most influential factor in low-level saliency, we use a contrast-based saliency detection [18] as our low-level saliency.

### 2.1.2. *Inhibiting Selected Regions in Low-level Saliency*

After we get similar and scattered background regions $R_{final}$, we consider methods to inhibit saliency values of $R_{final}$ in low-level saliency. There are several methods to reduce saliency of given regions. The easiest way is to directly set the saliency of given regions to 0. However, $R_{final}$ is still an estimation and it is possible that there exists foreground region in $R_{final}$ or background region not in $R_{final}$, so simply setting $R_{final}$ to 0 is likely to hurt the performance.

We reduce saliency of $R_{final}$ in a softer way. In our low-level saliency, the contrast-based saliency of a region $r_k$ in [18] is defined as

$$S(R_k) = \sum_{r_k \neq r_i} w(r_i) D_c(r_k, r_i) \tag{4}$$

where $w(r_i)$ is the weight of region $r_i$ and $D_c(\cdot,\cdot)$ is the color distance metric between the two regions. We reduce the color distances between any two regions in $R_{final}$. Specifically, for any two regions $r_a, r_b \in R_{final}$, $D'_c(r_a, r_b) = D_c(r_a, r_b)/\gamma$. Then we re-compute the contrast-based saliency based on refined color distances $D'_c(\cdot)$ using equation (4). Since we find that results change little when $\gamma$ is in certain range, we fix $\gamma = 5$ in experiment.

## 2.2. Similarity & Proximity

*Proximity* occurs when elements are placed close together. Visual system tends to group close or similar regions together, which can be explained by Gestalt law of proximity and similarity. We propose two smoothing methods. One is based on Gestalt law of *proximity*, the other is based on both *similar and proximity*. Note that our methods differ from [22] which also uses proximity and similarity because our methods are in region level, thus can group close or similar regions together to uniformly highlight entire object, therefore alleviate the problem that *low-level* saliency often fails to assign same (or similar) saliency values to entire salient object in image.

### 2.2.1. *Smoothing Based on Proximity*

According to law of proximity, if a region with low saliency value is all surrounded with regions with high saliency values, then the region and its surrounding regions tend to form an unified object and should be assigned same saliency values. We replace the saliency of each region by the weighted average of saliency of its adjacent regions. Therefore, saliency value of region $r_i$ is defined as

$$S'(r_i) = \frac{1}{(m-1)T} \sum_{r_j \in N(r_i)} (T - log(Area(r_j)))S(r_j) \tag{5}$$

$N(r_i)$ is the *directly adjacent* regions (neighbours) of $r_i$, $T = \sum_{r_j \in N(r_i)} log(Area(r_j))$ is the sum of $log$ of area of adjacent regions $r_j$ of $r_i$. We use $log$ to reduce the weight of very big regions. The normalization term $(m-1)T = \sum_{r_j \in N(r_i)} (T - log(Area(r_j)))$. Similar to [18], we use a linear-varying smoothing weight $(T - log(Area(r_j)))$ to give smaller weights to big regions. Note that $S(r_i)$ is updated by $S'(r_i)$ only when $S'(r_i)$ is larger than $S(r_i)$. We also restrict the chosen regions only to neighbours of $r_i$, because proximity occurs only when regions have a very close spatial distance. In middle row of Figure 3, by smoothing based on proximity, saliency values of some blocks are emphasized due to high saliency values of their neighbour regions.

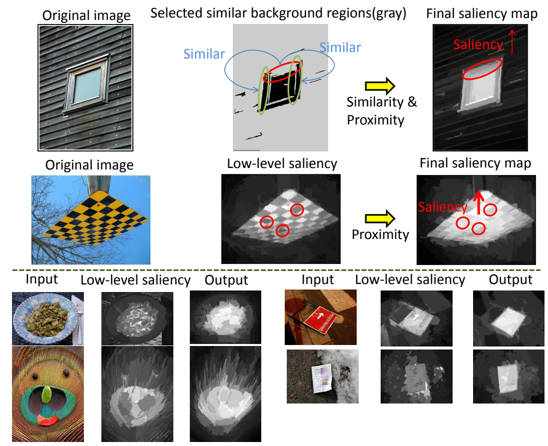



**Fig. 3**. Top: procedure of smoothing based on similarity & proximity. Middle: procedure of smoothing based on proximity. Bottom: original input, low-level saliency map and final saliency map after above two smoothing procedures.

### 2.2.2. *Smoothing Based on Similarity & Proximity*

Besides the smoothing based purely on spatial distance (proximity), we can also take color distance (similarity) into account to give higher weights to more similar regions. Typically we choose $k = |R|/8$ *spatially closest* regions $N'(r_i)$ of region $r_i$ to refine the saliency value of $r_i$ by

$$S'(r_i) = \frac{1}{(m-1)T} \sum_{r_j \in N'(r_i)} (T - D_c(r_i, r_j) D_s(r_i, r_j)) S(r_j) \tag{6}$$

where $D_c(r_i, r_j)$ and $D_s(r_i, r_j)$ relatively represent color distance and spatial distance of two regions $r_i$ and $r_j$, and similarly $T = \sum_{r_j \in N'(r_i)} D_c(r_i, r_j) D_s(r_i, r_j)$. Compared to smoothing based on proximity, the participation of similarity enables us to relax the chosen regions from directly adjacent regions to $k = |R|/8$ spatially closest regions. Note that in top row of Figure 3, gray regions in image of middle column represent selected similar background regions we select. Although foreground regions in red circle are wrongly predicted as similar background regions, they achieve high saliency values in smoothing part because the smoothing process group similar and close regions and assign same saliency to them based on similarity and proximity.

## 3. EXPERIMENTAL COMPARISON

We compare our method (GP) with several (eleven) state-of-the-art methods on a database of 1000 images provided by [14]. The database contains ground truth in the form of accurate human-marked labels for salient regions.

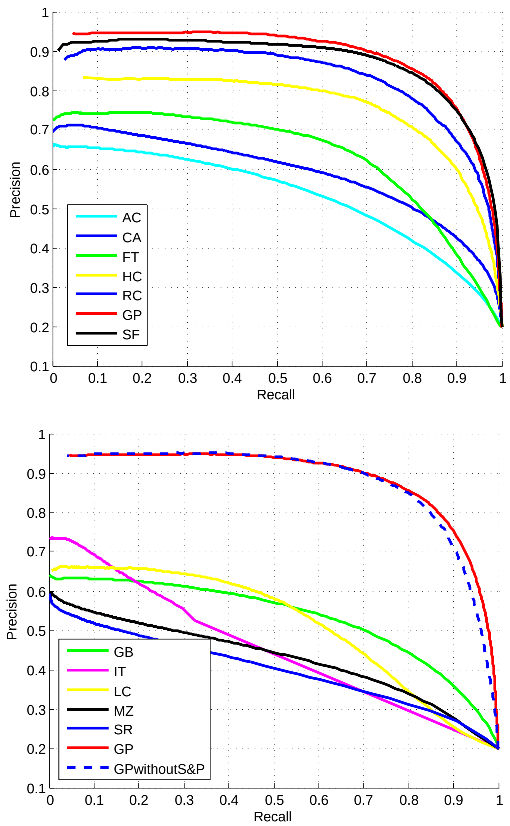
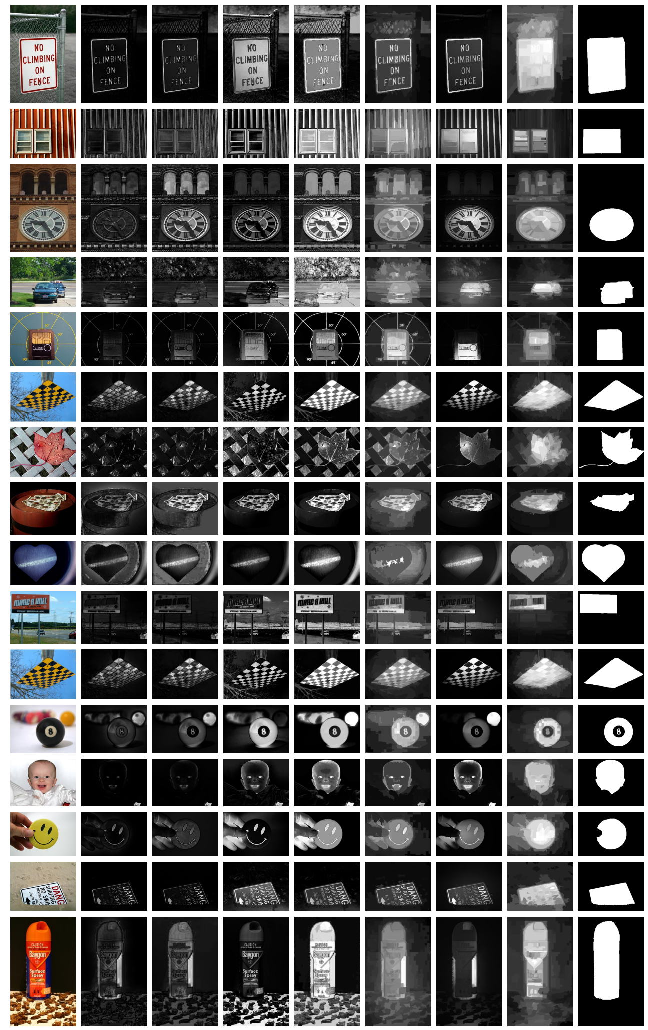



**Fig. 4**. Precision-recall curve of state-of-the-art methods as well as our method. We compare our method(GP) with GB [19], MZ [11], FT [14], IT [9], SR [16], AC [14], CA [27], LC [12], HC [18], RC [18], SF [24]. GPwithoutS&P donates our method after removing the smoothing procedures based on similarity & proximity.



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

**Fig. 5**. Visual comparison of saliency maps. (a) original image, (b) AC [14], (c) FT [14], (d) LC [12], (e) HC [18], (f) RC [18], (g) SF [24], (h) our method (GP), (i) ground truth.

We use measurements of precision and recall curve to evaluate each method. To segment salient objects and calculate precision and recall curves, we binarize the saliency map using every possible fixed threshold, similar to the fixed threshold experiment in [14, 18]. Figure 4 shows that precision and recall curves of our method (GP) outperform other methods. *GPwithoutS&P* in Figure 4 represents our method after removing the smoothing procedures based on similarity & proximity, which shows that only using similarity & anomaly operator is still competitive. After adding the similarity & proximity operator, the performance gets better.

Visual comparisons of saliency maps obtained by the various methods are illustrated in Figure 5. For images with repetitive patterns in background, our method is able to consistently inhibit the repetitive patterns in background of images. For images with cluttered background, our method achieves better results in suppressing cluttered background of images. The good performance in handling complex backgrounds are not surprising because compared to other low-level methods, our method adds mid-level similarity & anomaly concept. Also, due to smoothing based on mid-level proximity & similarity, example results show that our method can group similar and near regions and uniformly emphasize the entire salient object better.

## 4. CONCLUSION

We propose Gestalt saliency, a saliency detection method based on Gestalt principles, in which we introduce mid-level Gestalt concepts for low-level saliency. Our method consistently inhibits similar background regions of images based on Gestalt principles of similarity & anomaly. We also refine the saliency map using two smoothing methods based on Gestalt principles of similarity & proximity. Experimental results indicate that our method outperforms state-of-the-art methods on public database[14].

For future work, we believe that introducing more mid-level Gestalt principles for low-level saliency such as closure, continuation, symmetry will improve the performance of saliency estimation. Furthermore, more advanced segmentation and clustering algorithms can be used in our framework for better performance and efficiency such as spectral clustering [28], approximate clustering, etc.

## 5. REFERENCES


[1] L. 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.

[2] J. Han, K.N. Ngan, M. Li, and H.J. Zhang, "Unsupervised extraction of visual attention objects in color images," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 1, pp. 141–145, 2006.

[3] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, no. 9, pp. 1395–1407, 2006.

[4] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th international conference on*. IEEE, 2009, pp. 2106–2113.

[5] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli, and T.S. Chua, "An eye fixation database for saliency detection in images," *Computer Vision–ECCV 2010*, pp. 30–43, 2010.

[6] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.Y. Shum, "Learning to detect a salient object," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 2, pp. 353–367, 2011.

[7] D.A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2214–2219.

[8] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2185–2192.

[9] L. Itti, C. Koch, and E. 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.

[10] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry.," *Hum Neurobiol*, vol. 4, no. 4, pp. 219–27, 1985.

[11] Y.F. Ma and H.J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003, pp. 374–381.

[12] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th annual ACM international conference on Multimedia*. ACM, 2006, pp. 815–824.

[13] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," *Advances in neural information processing systems*, vol. 20, pp. 1–8, 2007.

[14] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1597–1604.

[15] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[16] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. Ieee, 2007, pp. 1–8.

[17] Z. Ren, Y. Hu, L.T. Chia, and D. Rajan, "Improved saliency detection based on superpixel clustering and saliency propagation," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1099–1102.

[18] M.M. Cheng, G.X. Zhang, N.J. Mitra, X. Huang, and S.M. Hu, "Global contrast based salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 409–416.

[19] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in neural information processing systems*, vol. 19, pp. 545, 2007.

[20] A. Desolneux, L. Moisan, and J.M. Morel, "Computational gestalts and perception thresholds," *Journal of Physiology-Paris*, vol. 97, no. 2-3, pp. 311–324, 2003.

[21] K. Koffka, "Principles of gestalt psychology," 1935.

[22] Z. Wang and B. Li, "A two-stage approach to saliency detection in images," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. Ieee, 2008, pp. 965–968.

[23] G. Kootstra, N. Bergström, and D. Kragic, "Gestalt principles for attention and segmentation in natural and artificial vision systems," in *Semantic Perception, Mapping and Exploration (SPME), ICRA 2011 Workshop*. eSMCs, 2011.

[24] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. 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] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[26] S. Theodoridis and K. Koutroumbas, "Pattern recognition," *Academic, San Diego*, 1999.

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

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