Extraction of Salient Object Regions by Virtual Super Pixel and Global Feature

Fumihiko Mori, Naotoshi Sugano
Tamagawa University/Japan
6-1-1Tamagawagakuen Machida, Tokyo, Japan 194-8610
morif@lab.tamagawa.ac.jp; sugano@eng.tamagawa.ac.jp

Abstract - A new salient object extraction method is constructed by a virtual super pixel (VSP) and global features to solve the following crucial problems of the conventional methods: (a) although a threshold to obtain the salient region close to ground truth (GT) is set for each image, the rule for setting the threshold is not explicit, and (b) although "contrast" is used as an element of the saliency calculation, the contrast does not compare features in adjacent regions in spite of the weak effect of a non-adjacent region. In this paper, the salient object region is obtained through the following process: (1) extraction of the background regions, (2) calculation of the saliency of the segmented region, and (3) unification of low-saliency regions into high-saliency ones. The method is applied to the extraction of object regions including important information such as characters and instruction icons in the living environment.

Keywords**:** salient object region, virtual super pixel, global feature, adjacent region, unification

1. Introduction

Region segmentation and salient object extraction are difficult but fundamental technologies for image analysis and understanding. Therefore, many methods have been proposed, as reported by Cheng D.H. et al. (2001), Ren et al. (2003), Lowe et al. (2004), Itti et al. (1998), and Goferman et al. (2010). Even now, these methods are actively studied, as reported by Cheng M.M. et al. (2011), Jiang H. et al. (2013), Jiang Z. et al. (2013), Mai et al. (2013), Scharfenberger et al. (2013), Shi et al. (2013), Siva et al. (2013), Yan et al. (2013), and Yang et al. (2013).

Many objects include important information such as characters and instruction icons in the living environment. Artificial visual aids such as optical character recognition (OCR), robots, and cameras mounted on cars help detect and understand such information.

Recently, many methods for extracting a region close to ground truth (GT) have been presented, as follows.

(1) A saliency aggregation method in which many saliency maps are combined by a machine learning method was shown by Shi et al. (2013).
(2) Region segmentation based on high histogram bin colours and saliency calculations by the colour contrast and distance between regions was shown by Cheng M.M. et al. (2011).
(3) A saliency map that considers background regions was presented in Jiang Z. et al. (2013), Jiang H. et al. (2013),Yang et al. (2013), and Mori et al. (2009, 2011a, 2012).
(4) Region segmentation using colour and orientation histograms in a block (nxn size) and a multi-scale watershed method were presented by Yan et al. (2013).
(5) Region segmentation using top-down information was presented by Jiang Z. et al. (2013).

However, two crucial problems need to be solved:

(a) Although a threshold to obtain the salient region close to GT is set for each image, the rule for setting the threshold is not explicit.

(b) Although "contrast" is used as an element of the saliency calculation, the contrast should be obtained by comparing features in adjacent regions because of the weak effect of a non-adjacent region, as shown in Figure 1(a). This effect is similar to a picture frame (a border line) that isolates a region from its surrounding. Contrast is calculated as the sum of contrasts from all regions multiplied by a distance function, except in papers by Jiang H. et al. (2013) and Mori et al. (2011a, 2012).

Although Jiang H. et al. (2013) regard the immediately neighbouring regions (adjacent regions), the adjacent regions are treated as a single region. That is, the difference between the mean feature value of the combined regions and the value of the target region is used as the contrast descriptor. The length contacting the adjacent regions is not considered. Therefore, the situation in Figure 1(a) happens when a big region has a small contact length, as shown in Figure 1(b). Mori et al. (2011a, 2012) reported a related concept, but it was incomplete.

The purpose of this paper is to propose a new method close to the abovementioned problems (Section 2.4.3 for the threshold problem and Sections 2.3 and 2.4.2 for the contrast problem) and to apply the method to detect regions including information such as characters and instruction icons.

Edge information, which has been presented, for example, by Kundu (1990), Canny (1986), Sobel (1990), Fukui (1995), and Yakimovsky (1976), is necessary not only for object recognition, but also for saliency calculation because the contrast around the border between regions is one of the main factors of saliency. Recently, a new concept named “virtual edge” (VE), which is obtained by separating a block, was proposed by Mori et al. (2011b). In this paper, we introduce a new concept named “virtual super pixel” (VSP), which is obtained as a by-product of the VE. The VSP is the small region obtained by separation of a block. Therefore, the VSP keeps both edge and region characteristics and enables simple, high-speed processing without much loss of detail.

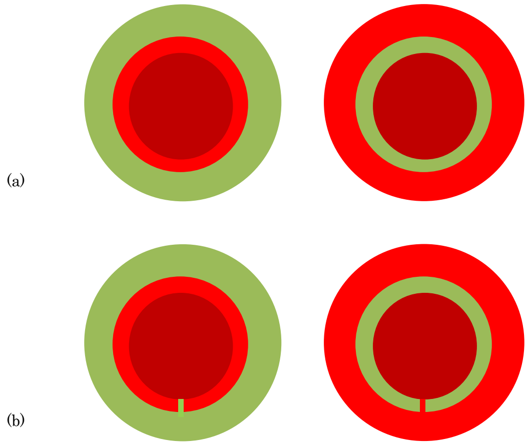

Fig. 1. Saliency effect of the adjacent region.
(The left central red circle is not salient.)

2. Salient Object Extraction System

The process of salient object extraction is shown in Figure 2. Each process is explained in the following sections.

2. 1. Extraction of VE and VSP

An image is first divided into (nxn)-size blocks. Mean values and standard deviations (S.D.) of the colours are calculated and colour α, which has the biggest S.D., is selected. Then each block is divided into two small regions, R_1 and R_2, based on the mean value μ_α. In the dividing process, the number of pixels N_i, the mean value $\mu_{i\xi}$, and standard deviations $\sigma_{i\xi}$ are calculated in region R_i (VSP), where $\xi \in \{x, y, u, v, R, G, B, r, g, b, I\}$, $I=R+G+B$, $r=R/I$, $g=G/I$, and $b=B/I$. The coordinates (u, v) are obtained by 90° rotation of the (x, y) coordinate system.

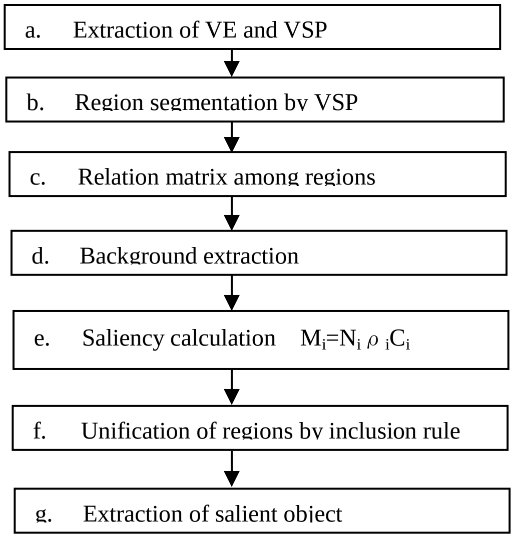
a. Extraction of VE and VSP
b. Region segmentation by VSP
c. Relation matrix among regions
d. Background extraction
e. Saliency calculation $M_i=N_i \rho_i C_i$
f. Unification of regions by inclusion rule
g. Extraction of salient object

Fig. 2. Process of salient object extraction.

VE is defined as a set of edge-like features composed of separability (η), colour difference (Δ), boundary location (e_x, e_y), boundary orientation (θ_x, θ_y) and distance (d) between R_1 and R_2. These are defined by equations (1)–(5).

$$\eta = 1 - \frac{N_1\sigma_1 + N_2\sigma_2}{n^2\sigma_0} \tag{1}$$

$$\Delta = |\mu_{2R} - \mu_{1R}| + |\mu_{2G} - \mu_{1G}| + |\mu_{2B} - \mu_{1B}| \tag{2}$$

$$(e_x, e_y) = \frac{N_1(\mu_{2x}, \mu_{2y}) + N_2(\mu_{1x}, \mu_{1y})}{n^2} \tag{3}$$

$$(\theta_x, \theta_y) = \frac{(-\mu_{2y} + \mu_{1y}, \mu_{2x} - \mu_{1x})}{\|(-\mu_{2y} + \mu_{1y}, \mu_{2x} - \mu_{1x})\|} \tag{4}$$

$$d = \left\|(\mu_{2x} - \mu_{1x}, \mu_{2y} - \mu_{1y})\right\| \tag{5}$$

The VSP is defined as a set of domain-like features {N_i, $\mu_{i\xi}$, $\sigma_{i\xi}$} added to the VE. Only the simple sum ΣX and the square sum ΣX^2 of feature X are kept temporally, because the mean value and the S.D. are directly calculated from these sums.

The concepts of the VE and the VSP are shown in Figure 3.

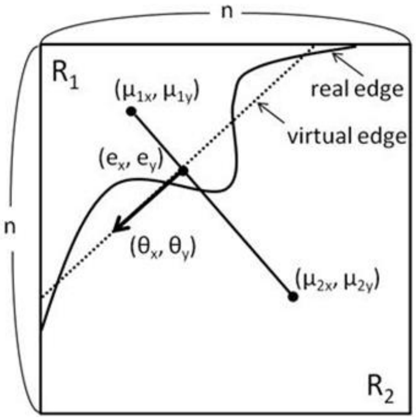

Fig. 3. Image of VE and VSP (Mori et al. (2011b, 2012)).

2. 2. Region segmentation by VSP

The process of region segmentation by the VSP is shown in Figure 4.

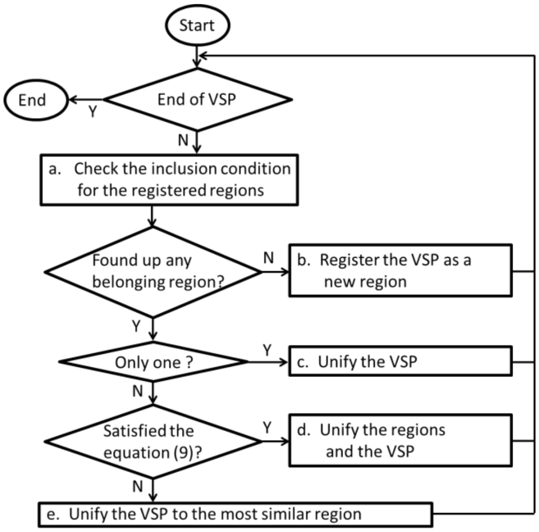
Start
End
Y
End of VSP
N
a. Check the inclusion condition for the registered regions
Found up any belonging region?
N
b. Register the VSP as a new region
Y
Only one ?
Y
c. Unify the VSP
N
Satisfied the equation (9)?
Y
d. Unify the regions and the VSP
N
e. Unify the VSP to the most similar region

Fig. 4. Process of region segmentation by VSP.

The region segmentation process is executed through one sequential sweep by the VSP. In this process, the VSPs of small regions are combined into a region or separated into different regions according to an integration rule that includes the global features of location and colour. The first VSP is registered as a seed region and the segmentation process is started. In process (a) of Figure 4, the inclusion condition of the VSP for the registered regions (equations (6)–(8)) is checked. Equations (6) and (7) use the min/max and the standard deviation to check whether the location of the VSP is included in the sphere of influence of the registered region. Equation (8) checks whether the colour of the VSP is included in the sphere of influence of the registered region. The corresponding other processes are as follows. (b) When no registered region satisfies the inclusion condition, the VSP is added as a new registered region. (c) When only one registered region satisfies the inclusion condition, the VSP is integrated into the registered region by updating the number of pixels, the min/max values, the mean values, and the standard deviations. (d) When multiple registered regions satisfy the inclusion condition, the unification condition is tested by applying the similarity described by equation (9). In the case that the condition is satisfied, the registered regions are unified. (e) When equation (9) is not satisfied, the VSP is integrated into the most similar registered region.

$$\rho_{min} - \lambda \leqq \rho \leqq \rho_{max} + \lambda \,, \qquad (6)$$
$$\mu_\rho - \kappa\sigma_\rho - \lambda \leqq \rho \leqq \mu_\rho + \kappa\sigma_\rho + \lambda \qquad (7)$$

Where $\rho \in \{x, y, u, v\}$, $k=2$, and λ is an expansion factor for the location.

$$\frac{(\mu_r,\mu_g,\mu_b)\cdot(b_r,b_g,b_b)}{\|(\mu_r,\mu_g,\mu_b)\|\|(b_r,b_g,b_b)\|} > cos\theta_0 \quad \text{and} \quad \mu_I - \lambda_I < b_I < \mu_I + \lambda_I \qquad (8)$$

Where • is the inner product, cos θ_0 is the similarity limit for the normalized colour, and λ_I is an expansion factor for the brightness.

$$\frac{(\mu_r,\mu_g,\mu_b)\cdot(\mu_r',\mu_g',\mu_b')}{\|(\mu_r,\mu_g,\mu_b)\|\|(\mu_r',\mu_g',\mu_b')\|} > cos\theta_0 \quad \text{and} \quad \mu_I - \lambda_I < \mu_I' < \mu_I + \lambda_I \qquad (9)$$

Where μ_r’, μ_g’, μ_b’, and μ_I’ are the mean values of another registered region.

2. 3. Relation matrix between regions

The adjacent situation between regions is very important information for calculating the saliency and recognizing a salient object from the contour form. An algorithm to obtain the number of adjacent VSPs among regions is presented here.

In the process of region segmentation, the region number including a VSP is stored as a feature of the VSP. Therefore, we know the adjacency between regions by comparing the region numbers in a pair of VSPs in a block. The adjacent VSP number is obtained by counting according to the following rules.

(Rule 1) Case a≠b (“a” and “b” are the region number stored in VSP in a block)

S(a, b)++ and S(b, a)++

(Rule 2) Case a=b & a*≠b* & a≠a* & a≠b* (“*” means a block left, right, up or down)

S(a, a*)++, S(a, b*)++, S(a*, a)++, and S(b*, a)++

(Rule 3) Case a=b & a# = b# & a≠a#   (“#” means right of or under a block)

S(a, a#)++ and S(a#, a)++

The VSP numbers of regions adjacent to the first region in the Data #1 image (see Figure 6) are shown in Figure 5.

The adjacent situation among the segmented regions is easily obtained in the proposed method and reflected in equation (11), which defines the contrast value of a region.

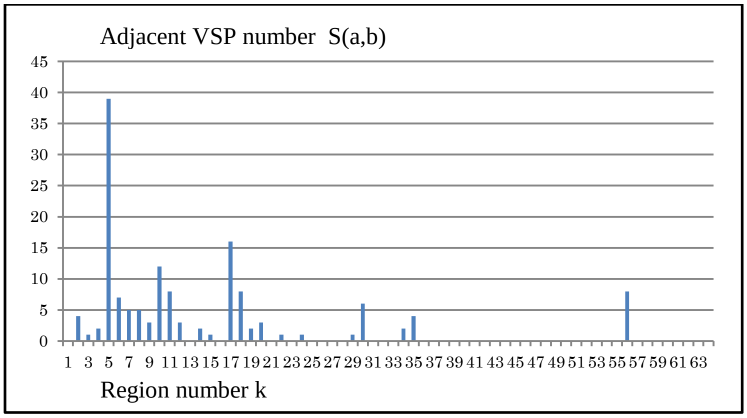

Fig. 5. Example of adjacent VSP numbers (the ordinate is adjacent VSP number S(a,b)).

2. 4. Background, Saliency, and Unification

2. 4. 1. Background

The four end regions of the top, bottom, left, and right ends are defined first. The number of VSPs included in each end region for each segmentation region is counted. The background region is defined according to the following rule: (1) the number of VSPs in the top end is greater than two, (2) the number in the right and left ends is greater than two, or (3) the number in the left, right, or bottom ends is greater than two and the region centre is not in the central region.

2. 4. 2. Saliency

Saliency M_i of a segmentation region i is defined by equation (10).

$$M_i = N_i \rho_i C_i \qquad (10)$$

Where N_i is the number of pixels included in the region, ρ_i is the density, and C_i is the contrast defined by equation (11).

$$C_i = \frac{\sum_{k=1}^{m} \frac{l_{ik}\Delta I_{ik}}{I_i + I_k}}{L_i} \qquad (11)$$

where l_{ik} is the number of VSPs in contact between regions i and k, L is the total number of VSPs in contact with region i ($S(i,k)=l_{ik}$) and m is the number of regions contacting region i.

Equation (11) is composed of a popular contrast used in psychology and includes the adjacent length l_{ik}. It is clear that the abovementioned problem (b) of conventional method, Jiang H. et al. (2013) is solved.

2. 4. 3. Region unification and extraction rule of salient object region

The regions except the background are arranged in the order of saliency M. The low-saliency regions are unified into the higher saliency region according to the inclusion rule (Region unification).

The following extraction rule of the salient-unified-object-region is used in this paper:

(1) Saliency of the most salient region is $M_0 > \lambda_M$. The smallest value of saliencies of the most salient object in an image ("wolf" in the tenth image in Figure 6) is used as the threshold λ_M.

(2) $M_i > M_0/5$

(3) $i < 5$ (The magical number 5 or 7 is considered.)

This rule answers the abovementioned problem (a) of the conventional method.

3. Results

Examples of extracted salient objects are shown in Figure 6. Regions whose saliency M_i is greater than one-fifth of the maximum saliency M_0 and a threshold λ_M (=50), were extracted for up to five regions and marked by the colors of white, red, orange, yellow, and green according to the saliency order.

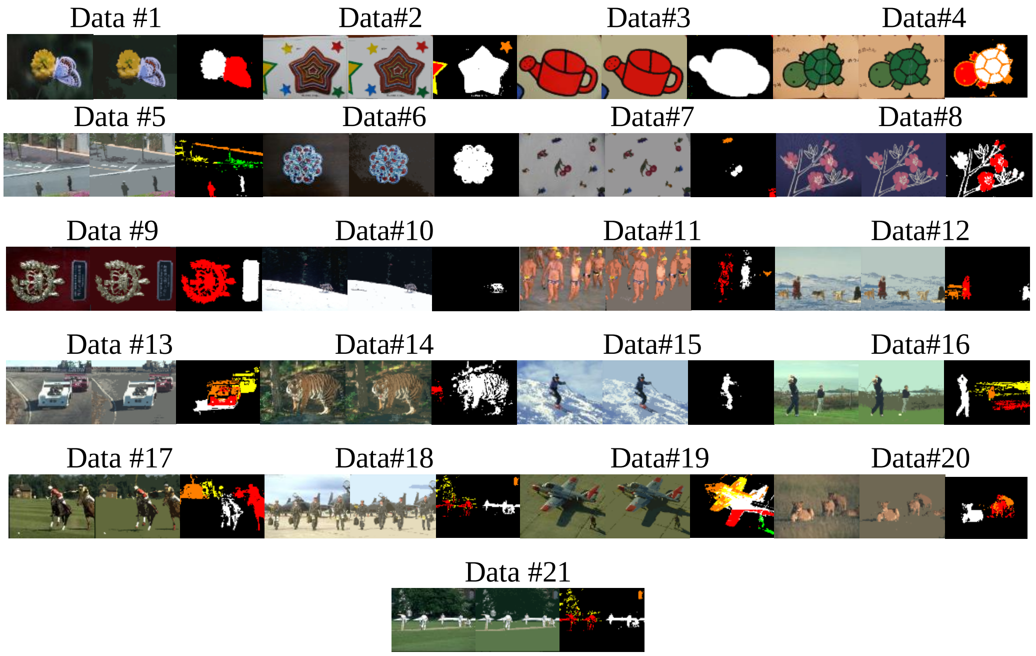

Fig. 6. Examples of the salient object regions.
[Input] [regions] [salient objects]

A comparison with conventional approaches is shown in Figure 7. The presented method was also applied to the database "Image-MSRA5000" and some results are shown in Figure 8. The method was also applied to the database CSSD composed of 200 images and the residual area of the background, which means no threshold, was compared with the GT. The score of the F-measure was about the same as those in the case of the best threshold shown in Yan et al. (2013).

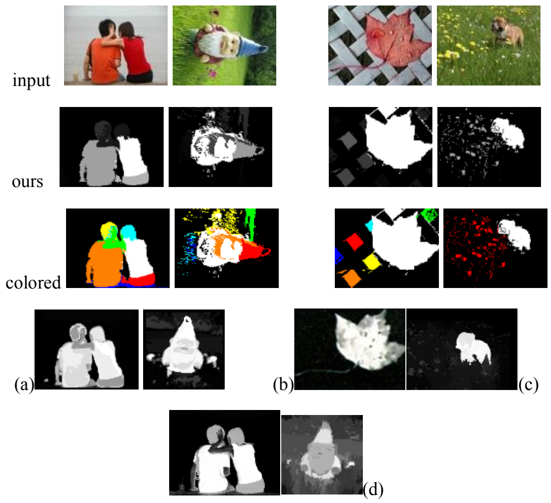

Fig. 7. Comparison of the conventional methods.
(a)Jiang H et al.(2013) (b) Shi et al.(2013) (c)Yan et al.(2013) (d) Cheng M.M.(2011)

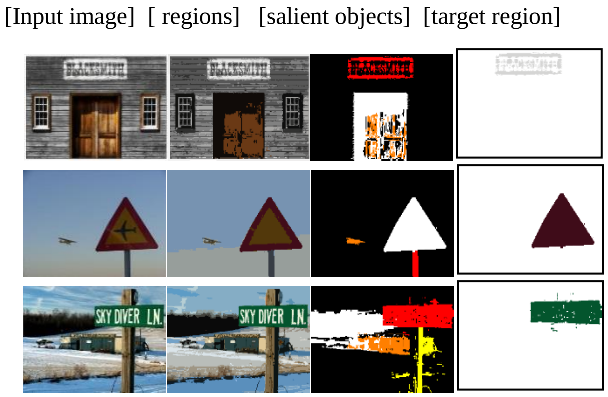

Fig. 8. Extraction of characters or instruction icon region in the living environment.

4. Conclusion

A new method is proposed to extract salient objects that include important information such as characters and instruction icons. The success of our approach is primarily due to the concept of a VSP, the contrast based on the adjacent matrix among regions, and the global features in the region segmentation process. Object recognition based on the VE and a histogram of the VE is intended as future work.

References

Canny, J. (1986). Computational Approach to Edge Detection. *IEEE Trans. PAMI, PAMI-8, 6*, 679-698.

Cheng, D.H., Jiang, X.H, Sun, Y., & Wang, J.L. (2001). Colour Image Segmentation: Advances and Prospects. *Pattern Recognition*, *34*, 2259- 2281.

Cheng, M.M., et al. (2011). Global Contrast Based Salient Region Detection. *Proc. CVPR2011*, 409-416.

Fukui, K. (1995). Edge Extraction Method Based On Separability of Image Features. *IEICE Trans. INF. & SYST. E78-D*, *12*, 1533-1538.

Goferman, S., et al. (2010). Context-Aware Saliency Detection. *Proc. CVPR2010*, 2376-2383.

Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. PAMI*, *20*, 1252-1259.

Jiang, H., et al. (2013). Salient Object Detection: A Discriminative Region Feature Integration Approach. *CVPR 2013*, 2083-2090.

Jiang, Z., & Davis, L.S. (2013). Submodular Saliency Region Detection. *CVPR 2013*, 2043-2050.

Kundu, A. (1990). Robust Edge Detection. *Pattern Recognition*, *23*(5), 423-440.

Lowe, D.C. (2004). Distinctive Image Features from Scale- Invariant Key-Points. *IJCV*, *60*, 91-110.

Mai, L., et al. (2013). Saliency Aggregation: A Data-Driven Approach, *CVPR 2013*, 1131-1138.

Mori, F., & Mori, T. (2012). Region Segmentation and Object Extraction Based On Virtual Edge and Global Features. *Proc. ACCV2012 Workshop on Computational Photography and Low-Level Vision*, WS-M2-3-1－WS-M2-3-8.

Mori, F., Yamada, H., Mizuno, M., & Sugano, N. (2011a). Colour Image Segmentation Based on Statistics of Location and Feature Similarity. *IEEJ trans. On Electronics, Information and Systems C*, *131*(11), 2022-2029.

Mori, F., Yamada, H., Mizuno, M., & Sugano, N. (2011b). Virtual Edge Extraction Method Based on New Separability. *Trans. IEICE*, *J94-95D*, *12*, 2105-2114.

Mori, F., et al. (2009). Two Steps Region Segmentation Method Using Colour and Location. *IEICE Tech. Report*.

Ren, X., & Malik, J. (2003). Learning a Classification Model for Segmentation. *Proc. 9th ICCV*, *1*, 10-17.

Shi, K., et al. (2013). PISA: Pixel Wise Image Saliency By Aggregating Complementary Appearance Contrast Measures With Spatial Priors. *CVPR 2013*, 2115-2122.

Scharfenberger, C. et al. (2013). Statistical Textural Distinctiveness for Salient Region Detection in Natural Images. *CVPR 2013*, 979-986.

Siva, P., et al. (2013). Looking Beyond the Image: Unsupervised Learning for Object Saliency and Detection. *CVPR 2013*, 3238-3245.

Sobel, I. (1990). An Isotropic 3x3 Image Gradient Operator. In H. Freeman (Ed.), *Machine Vision for Three-Dimensional Scenes* (pp. 376-379). Academic Press.

Yakimovsky, Y. (1976). Boundary and Object Detection in Real World Images. *J. ACM*, *23*(4), 599-608.

Yan, Q., et al. (2013). Hierarchical Saliency Detection. *CVPR 2013*, 1155-1162.

Yang, C., et al. (2013). Saliency Detection via Graph-Based Manifold Ranking. *CVPR 2013*, 3166-3173.