A New Saliency Detection Method for Stereoscopic Images Using Contrast and Prior Knowledge

Sang-Hyun Cho[1] and Hang-Bong Kang[2]

[1] Dept. of Computer Engineering, The Catholic University of Korea,
#43-1 Yeokgok 2-dong, Wonmi-Gu, Bucheon, Gyeonggi-do, Korea
cshgreat@catholic.ac.kr
[2] Dept. of Digital Media, The Catholic University of Korea,
#43-1 Yeokgok 2-dong, Wonmi-Gu, Bucheon, Gyeonggi-do, Korea
hbkang@catholic.ac.kr

Abstract. In this paper, we propose a new visual saliency detection method, which is effective regardless of unreliable disparity information, by using contrast and prior knowledge. Our proposed method consists of two phases. In the first phase, we used region based contrast information to compute the saliency of an input image. We consider not only global but also local contrast in color and disparity information to efficiently extract salient regions in a stereoscopic image. In addition, we introduce a confidence measure to handle unreliable disparity information. In the second phase, we used region based prior knowledge existent in a stereoscopic image. The region based prior knowledge is constructed from low-level features such as color, frequency, location and disparity in the stereoscopic image. Finally, we integrate contrast-based and prior knowledge-based saliency to accurately detect saliency from input stereoscopic image. Experimental results show that our method efficiently detects salient contents in stereoscopic images.

Keywords: Saliency, visual attention, stereoscopic.

1 Introduction

Considerable research efforts have been devoted over the last few years to detect salient regions, because saliency analysis can be applied to many computer vision fields, such as object detection, object recognition, and image retrieval. Recently, various saliency detection methods from monoscopic image [1-5] and stereoscopic image and video [6-13] have been investigated. In previous researches, most saliency detection methods for stereoscopic images require an accurate disparity map to obtain reliable saliency detection. Even though a dense stereo matching method has been improved for the past few years, it remains a challenging problem. As a result, a saliency detection method to effectively exploit unreliable disparity information is necessary to compute desirable saliency from complex stereoscopic images.

In this paper, we focus on bottom-up data driven saliency detection using adaptive disparity cue depending on the quality of the disparity map. Our main contributions in

E. Bayro-Corrochano and E. Hancock (Eds.): CIARP 2014, LNCS 8827, pp. 778–786, 2014.

this paper are as follows. First, we introduce a confidence measure to handle reliability issues of disparity maps in saliency analysis of stereoscopic images. If disparity quality is low, disparity related components are less well reflected in the confidence measure of the disparity map. Second, our contrast based analysis deals with global and local contrast in both the color and the disparity domains. Finally, we apply prior knowledge such as frequency, color, size, location and disparity to obtain accurate saliency for the given image. Prior knowledge helps us to detect saliency without the context information of the image.

2 Stereoscopic Saliency Detection

2.1 Region-Based Contrast from Stereoscopic Image

Given one side (left image) of an input stereoscopic image, we first segment an image into regions, using graph based image segmentation method [14]. To reflect disparate qualities in the input stereoscopic image, we use the curvature of cost curve metric [15] as a disparity attribute confidence measure for each region $\mathbf{R}$ in the stereoscopic images. The curvature of cost curve metric, $C_f(\mathbf{x})$, is defined as

$$C_f(\mathbf{x}) = -2c(\mathbf{x},d) + c(\mathbf{x},d-1) + c(\mathbf{x},d+1), \tag{1}$$

where $c(\mathbf{x},d) = \frac{1}{\|\mathbf{W}(\mathbf{x})\|}\sum_{\mathbf{x}\in\mathbf{W}(\mathbf{x})} e(\mathbf{x},d)$, $e(\mathbf{x},d) = \sum_{cl\in\{R,G,B\}} \left|\mathbf{I}_{ch}^{L}(x) - \mathbf{I}_{ch}^{R}(x-d)\right|$, d is disparity, $\mathbf{W}(\mathbf{x})$ is local window at $\mathbf{x}$ and $\mathbf{I}_{ch}^{L}$, $\mathbf{I}_{ch}^{R}$ are the left and right normalized image at channel ch, respectively.

Then, the disparity attribute confidence measure of a region $\mathbf{R}$, $\lambda_{\mathbf{R}}$, is computed as follows.

$$\lambda_{\mathbf{R}} = \frac{1}{n_{\mathbf{R}}} \sum_{\mathbf{x}=(x,y)\in\mathbf{R}} 1 - \exp\left(-\left|C_f(\mathbf{x})\right| / \sigma_{cf}\right), \tag{2}$$

where $n_{\mathbf{R}}$ is the number of pixel in $\mathbf{R}$ and σ_{cf} is a parameter. We set $\sigma_{cf} = 0.35$.

The contrast based global saliency value of a region is computed as follows.

$$S_g(\mathbf{R}_i) = \sum_{\mathbf{R}_i\neq\mathbf{R}_k} (1-\lambda_{\mathbf{R}_i}) D_c(\mathbf{R}_i,\mathbf{R}_k) + \lambda_{\mathbf{R}_i} D_d(\mathbf{R}_i,\mathbf{R}_k), \tag{3}$$

where $D_c(\cdot,\cdot)$ is the color distance, $D_d(\cdot,\cdot)$ is the disparity distance metric between two regions, and $\lambda_{\mathbf{R}}$ is the confidence measure of region $\mathbf{R}$.

The color distance between two regions, $D_c(\cdot,\cdot)$ is defined by the Bhattacharyya distance between the color distributions of two regions.

Then, a region $\mathbf{R}$ in color space is defined by a color distribution as follows.

$$p(\mathbf{R}) = \left\{p_{\mathbf{R}}^{(u)}\right\}_{u=1,\cdots,m}, \tag{4}$$

where $p_{\mathbf{R}}^{(u)} = N_c \sum_{i=1}^{m} k(\|\mathbf{x}_i - \mathbf{x}_c\|)\delta[b_c(\mathbf{x}_i) - u]$, $k(r) = 1 - r^2$, $\mathbf{x}_i$ is normalized pixel location in $\mathbf{R}$, $\mathbf{x}_c$ is normalized center location of $\mathbf{R}$, $b_c(\cdot)$ is the mapping function from the pixel location to bin index in the quantized color space, δ is the Kronecker delta function and N_c is normalization factor used to impose the condition $\sum_{u=1}^{m} p_{\mathbf{R}}^{(u)} = 1$.

We compute the color distance between two regions using the Battacharyya distance between each color distribution of the two regions.

$$D_c(\mathbf{R}_i, \mathbf{R}_j) = \sqrt{1 - \rho[p(\mathbf{R}_i), p(\mathbf{R}_j)]} \quad \text{where} \quad \rho[p(\mathbf{R}_i), p(\mathbf{R}_j)] = \sum_{u=1}^{m} \sqrt{p_{\mathbf{R}_i}^{(u)} \cdot p_{\mathbf{R}_j}^{(u)}} \tag{5}$$

The disparity distance is defined in a similar way as the color distance. We used a disparity distribution to represent a region $\mathbf{R}$. This region $\mathbf{R}$ in disparity space is represented by disparity distribution as follows,

$$q(\mathbf{R}) = \{q_{\mathbf{R}}^{(u)}\}_{u=1,\cdots,m}, \tag{6}$$

where $q_{\mathbf{R}}^{(u)} = N_d \sum_{i=1}^{m} k(\|\mathbf{x}_i - \mathbf{x}_c\|)\delta[b_d(\mathbf{x}_i) - u]$, $k(r) = 1 - r^2$, $\mathbf{x}_c$ is center of $\mathbf{R}$, $b_d(\cdot)$ is the mapping function from the pixel location to bin index in the quantized disparity space, δ is the Kronecker delta function, and N_d is a normalization factor to impose the condition $\sum_{u=1}^{m} q_{\mathbf{R}}^{(u)} = 1$.

The disparity distance $D_d(\cdot,\cdot)$ between two regions is computed using the Battacharyya distance between disparity distributions of two regions.

$$D_d(\mathbf{R}_i, \mathbf{R}_j) = \sqrt{1 - \rho[q(\mathbf{R}_i), q(\mathbf{R}_j)]} \tag{7}$$

Since human visual systems tend to group similar or neighboring regions together, local contrast is also an important factor to determine the saliency of a region. Thus, we compute color and disparity saliencies between a region and its adjacent neighbors. Therefore, the saliency value of the local contrast of a region is computed as follows.

$$S_l(\mathbf{R}_i) = \sum_{\mathbf{R}_j \in N(\mathbf{R}_i)} (1 - \lambda_{\mathbf{R}_i}) D_c(\mathbf{R}_i, \mathbf{R}_j) + \lambda_{\mathbf{R}_i} D_d(\mathbf{R}_i, \mathbf{R}_j), \tag{8}$$

where $N(\mathbf{R}_i)$ is the direct adjacent regions of $\mathbf{R}_i$.

Finally, the contrast based saliency value for each region is computed by,

$$S_{ct}(\mathbf{R}_i) = \alpha_1 S_g(\mathbf{R}_i) + \alpha_2 S_l(\mathbf{R}_i), \tag{9}$$

where α_1 and α_2 are weight factors for global and local contrast, respectively. We set α_1=0.7 and α_2 = 0.3.

Fig. 1 shows our contrast based saliency detection results for a stereoscopic image. Although the disparity map quality is not good, the saliency of each region is efficiently computed by our contrast based saliency detection method.

2.2 Prior Knowledge-Based Saliency from Stereoscopic Images

Although prior knowledge-based saliency has many limitations, it has still been found useful in analyzing the saliency of a scene because it is simple and does not require context information. The image based prior knowledge, $S_{img}(\mathbf{R})$, is defined as follows:

$$S_{img}(\mathbf{R}) = \left(\gamma_1 S_F(\mathbf{R}) + \gamma_2 S_C(\mathbf{R}) + \gamma_3 S_A(\mathbf{R})\right) \tag{10}$$

where S_F is frequency, S_C is color, S_A is size based prior knowledge, respectively, while γ_1, γ_2 and γ_3 are weight factors. We set γ_1=0.6, γ_2=0.2 and γ_3=0.2.

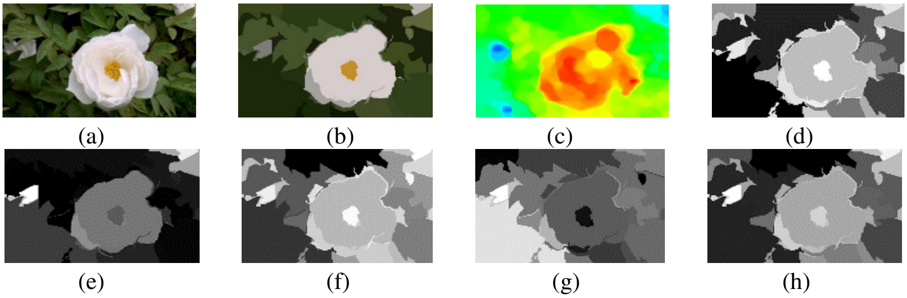

(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 1. Example of region contrast-based saliency. (a) Left image. (b) Segmentation. (c) Disparity. (d) Global color contrast. (e) Global disparity contrast. (f) Local color contrast. (g) Local disparity contrast. (h) Contrast based saliency.

We also combine location and disparity prior as spatial based prior knowledge, $S_{spa}(\mathbf{R})$, as follows:

$$S_{spa}(\mathbf{R}) = S_L(\mathbf{R}) \cdot S_D(\mathbf{R}) \tag{11}$$

where S_L is location, and S_D is the disparity-based prior knowledge, respectively.

Then, the final prior knowledge saliency, $S_{pk}(\mathbf{R})$, is defined by combining image and spatial-based prior knowledge.

$$S_{pk}(\mathbf{R}) = S_{img}(\mathbf{R}) \cdot S_{spa}(\mathbf{R}) \tag{12}$$

Fig. 2 shows our prior knowledge based saliency in a stereoscopic image. Although the context information of a stereoscopic image is not used, the saliency of each region in the input stereoscopic image is efficiently detected. The details of components of our prior knowledge are presented as follows.

2.2.1. Frequency-Based Prior Knowledge

The integrated band-pass filtering responses from each color channel, such as the CIELab color space, show good performance in salient analysis [16, 17]. Thus, we define the frequency based prior saliency of a region $\mathbf{R}$, $S_F(\mathbf{R})$, as

$$S_F(\mathbf{R}) = \frac{1}{2}\left(S_{DoG}(\mathbf{R}) + S_{lGb}(\mathbf{R})\right), \tag{13}$$

where

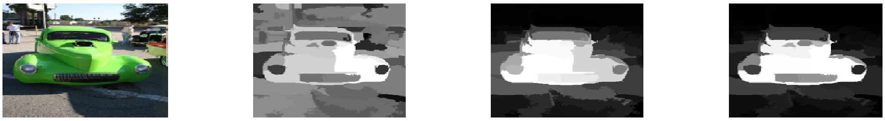
$$S_{DoG}(\mathbf{R}) = \frac{1}{n_R}\sum_{\mathbf{x}\in R}\left[\left(\mathbf{I}_L^{\mu} - \mathbf{I}_L^{G}(\mathbf{x})\right)^2 + \left(\mathbf{I}_a^{\mu}(\mathbf{x}) - \mathbf{I}_a^{G}(\mathbf{x})\right)^2 + \left(\mathbf{I}_b^{\mu}(\mathbf{x}) - \mathbf{I}_b^{G}(\mathbf{x})\right)^2\right]^{\frac{1}{2}}$$

$$S_{lGb}(\mathbf{R}) = \frac{1}{n_R}\sum_{\mathbf{x}\in R}\left[\left(\mathbf{I}_L(\mathbf{x}) * G_L\right)^2 + \left(\mathbf{I}_a(\mathbf{x}) * G_L\right)^2 + \left(\mathbf{I}_b(\mathbf{x}) * G_L\right)^2\right]^{\frac{1}{2}},$$

where * denotes convolution operator, $\mathbf{I}_L$, $\mathbf{I}_a$ and $\mathbf{I}_b$ as pixel values at the CIELab color channel of image, respectively. $\mathbf{I}_L^{\mu}$, $\mathbf{I}_a^{\mu}$ and $\mathbf{I}_b^{\mu}$ are arithmetic mean pixel values at the CIELab color channel of image, respectively. $\mathbf{I}_L^{G}$, $\mathbf{I}_a^{G}$ and $\mathbf{I}_b^{G}$ are the CIELab color channel of Gaussian blurred image, respectively,

$$G_L(u,v) = \exp\left(-\log\left(\frac{\sqrt{u^2+v^2}}{\omega_0}\right) / 2\sigma_F^2\right), \tag{14}$$

where (u,v) is the coordinate in the frequency domain, and σ_F are filer bandwidth and ω_0 is center frequency. We set $\sigma_F = 6.2$ and $\omega_0 = 0.002$.



(a) Left image (b) Image prior (c) Spatial prior (d) Prior saliency

Fig. 2. Example of prior knowledge-based saliency

2.2.2. Color-Based Prior Knowledge

It is known that warm colors such as red are more attraction-demanding for human visual perception than cold colors such as blue [17, 18]. Thus, for a given a pixel in the CIELab color space, we define the color based prior saliency, $S_C(\mathbf{R})$, as,

$$S_C(\mathbf{R}) = 1 - \exp\left(-\frac{C_a(\mathbf{R}) + C_b(\mathbf{R})}{\sigma_C^2}\right), \tag{15}$$

where $C_a(\mathbf{R}) = \frac{1}{n_R}\sum_{\mathbf{x}\in\mathbf{R}} a(\mathbf{x})$, $C_b(\mathbf{R}) = \frac{1}{n_R}\sum_{\mathbf{x}\in\mathbf{R}} b(\mathbf{x})$, $a(\cdot)$ is normalized a channel value, $b(\cdot)$ is normalized b channel value of CIELab color space of image and σ_C is a parameter. We set $\sigma_C = 0.2$.

2.2.3 Size-Based Prior Knowledge

Since a large-sized region is typically an important region in an image, and thus more demanding of attention, we computed the size of each region in image as the size based prior knowledge.

$$S_A(\mathbf{R}) = 1 - \exp\left(-\frac{A(\mathbf{R})}{\sigma_{area}^2}\right), \tag{16}$$

where σ_{area} is a parameter and $A(\cdot)$ is normalized area of a region. We set σ_{area} = 0.35.

2.2.4 Location-Based Prior Knowledge

Since, people have been found to pay more attention to objects located at the center of an image [17], we define the location-based prior saliency, $S_L(\mathbf{R})$, as

$$S_L(\mathbf{R}) = \frac{1}{n_{\mathbf{R}}} \sum_{\mathbf{x}\in\mathbf{R}} \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|_2^2}{\sigma_{loc}^2}\right), \tag{17}$$

where σ_{loc} is a parameter, and $\mathbf{c}$ is center of an image. We set σ_{loc} = 80.

2.2.5 Disparity-Based Prior Knowledge

People usually pay more attention to objects having large negative disparities. Thus, we define the disparity-based prior saliency, $S_D(\mathbf{R})$, as,

$$S_D(\mathbf{R}) = 1 - \exp\left(-\frac{\lambda_{\mathbf{R}} \cdot D(\mathbf{R})}{\sigma_{dis}^2}\right), \tag{18}$$

where $D(\mathbf{R}) = \frac{d_{\max} - d(\mathbf{R})}{d_{\max} - d_{\min}}$, $d(\mathbf{R}) = \frac{1}{n_R}\sum_{\mathbf{x}\in\mathbf{R}} d(\mathbf{x})$, σ_{dis} is a parameter, $d_{\max}$ and $d_{\min}$ are maximum and minimum disparity of the scene, respectively. $d(\cdot)$ is disparity value at location in a region, and $\lambda_{\mathbf{R}}$ is confidence measure of region $\mathbf{R}$'s disparity. We set σ_{dis} = 0.25.

2.3 Stereoscopic Saliency Detection

The stereoscopic saliency of a region is computed from both contrast-based saliency and prior knowledge-based saliency. It is computed as follows:

$$S(\mathbf{R}_i) = \begin{cases} S_{ct}(\mathbf{R}_i)\cdot S_{pk}(\mathbf{R}_i) & \text{if } S_{ct}(\mathbf{R}_i)\cdot S_{pk}(\mathbf{R}_i) > 0.2 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Note that each S_{ct} and S_{pk} is remapped to [0, 1] by simple linear mapping.

3 Experimental Results

We compare our method with six state-of-the-art saliency detection methods, including CA [19], GB [20], SR [21], FT [16], RC [3], HS [22] and SS [13] using the Stereo Saliency Benchmark Dataset introduced in [13]. The salient regions are detected in the stereoscopic image more accurately with our method than with the other methods compared as shown Fig. 3.

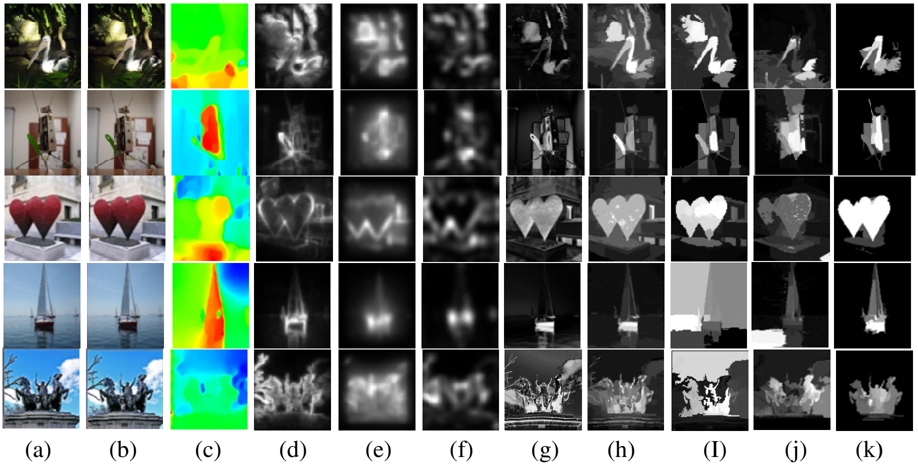
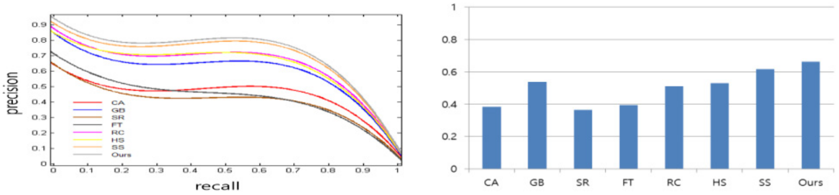


Fig. 3. Visual comparison results of various methods to our method. (a) Left (b) Right (c) Disparity (d) CA (e) GB (f) SR (g) FT (h) RC (i)HS (j) SS (k) Ours.



(a) Precision recall curves

(b) F-measure score comparison

Fig. 4. Performance comparison result

We evaluated the performance of a saliency detection method similar to previous studies [16, 17]. The average precision-recall curve is obtained by averaging the results from all of the test images. We plot the precision-recall curve for each method as in Fig. 4-(a). As shown in Fig. 4-(a), our method has better performance than other methods.

We also perform an evaluation by the adaptive thresholding method. Adaptive thresholding value T_a is determined by the mean saliency of the image. It is computed as follows.

$$T_a = \frac{2}{W \cdot H} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x, y), \tag{20}$$

where W and H are the width and height of the saliency map, respectively, and $S(x, y)$ is the saliency value at (x, y).

We obtain a binary image by the adaptive threshold value method and compute the F-measure, which is defined as

$$F = \frac{(1+\beta^2) P_r \cdot R_c}{\beta^2 P_r + R_c}, \tag{21}$$

where P_r is precision and R_c is recall. We set $\beta^2=0.3$ in our experiments, similar to previous studies [16, 17]. Fig. 4-(b) shows a comparison of F-measure scores resulting from various saliency detection methods.

4 Conclusion

In this paper, we proposed a novel regional saliency detection method by combining contrast and prior knowledge data with confidence measure to handle unreliable disparity information. We used not only global but also local contrast information while taking into account the Gestalt principle of human perception.

However, our method depends on the quality of region segmentation. Although image segmentation has been studied for many years, it is still a challenging problem.

References

1. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to Predict Where Humans Look. In: Proc. IEEE Int'l Conf. Computer Vision, pp. 2106–2113 (2009)
2. Itti, L., Koch, C., Niebur, E.: A model of saliency based visual attention for rapid scene analysis. IEEE Trans. PAMI, 1254–1259 (1998)
3. Cheng, M., Zhang, G., Mitra, N.J., Huang, X., Hu, S.: Global contrast based salient region detection. In: Proc. CVPR, pp. 409–416 (2011)
4. Kadir, T., Brady, M.: Saliency, scale and image description. Int. J. Comput. Vis., 83–105 (2001)
5. Wu, J., Zhang, L.: Gestalt Saliency: Salient Region Detection based on Gestalt Principles. In: Proc. IEEE ICIP, pp. 181–185 (2013)
6. Wang, J., Da Silva, M.P., Le Callet, P., Ricordel, V.: Computational Model of Stereoscopic 3D Visual Saliency. IEEE Transaction on Image Processing, 2151–2165 (2013)
7. Maki, A., Nordlund, P., Eklundh, J.: A computational model of depthbased attention. In: Proc. IEEE 13th Int. Conf. Pattern Recognit., pp. 734–739 (1996)
8. Zhang, Y., Jiang, G., Yu, M., Chen, K.: Stereoscopic visual attention model for 3D video. In: Boll, S., Tian, Q., Zhang, L., Zhang, Z., Chen, Y.-P.P. (eds.) MMM 2010. LNCS, vol. 5916, pp. 314–324. Springer, Heidelberg (2010)
9. Chamaret, C., Godeffroy, S., Lopez, P., Meur, O.L.: Adaptive 3D rendering based on region-of-interest. In: Proc. SPIE, pp. 75240–752412 (2010)
10. Ouerhani, N., Hugli, H.: Computing visual attention from scene depth. In: Proc. IEEE 15th Int. Conf. Pattern Recognit., pp. 375–378 (2000)
11. Dittrich, T., Kopf, S., Schaber, P., Guthier, B., Effelsberg, W.: Saliency Detection for Stereoscopic Video. In: Proc. ACM 4th Multimedia Systems Conf., pp. 12–23 (2013)
12. Kim, H., Lee, S., Bovik, A.C.: Saliency Prediction on Stereoscopic Videos. IEEE Trans on Image Processing, 1476–1490 (2014)
13. Niu, Y., Geng, Y., Li, X., Liu., F.: Leveraging stereopsis for saliency analysis. Proc. In: IEEE CVPR, pp. 454–461 (2012)
14. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph based image segmentation. Int. J. Comput. Vision, 167–181 (2004)

15. Egnal, G., Mintz, M., Wildes, R.: A stereo confidence metric using single view imagery with comparison to five alternative approaches. Image and Vision Computing, 943–957 (2004)
16. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: Proc. CVPR, pp. 1597–1604 (2009)
17. Zhang, L., Gu, Z., Li, H.: SDSP: A Novel Saliency Detection Method by Combining Simple Priors. In: Proc. IEEE ICIP, pp. 171–175 (2013)
18. Chen, X., Wu, Y.: A unified approach to salient object detection via low rank matrix recovery. In: Proc. CVPR, pp. 853–860 (2012)
19. Goferman, S., Zelnik-Manor, L., Tal., A.: Context-aware saliency detection. In: Proc. CVPR, pp. 2376–2383 (2010)
20. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. Adv. Neural Information Process. Syst., 545–552 (2007)
21. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: Proc. CVPR, pp. 1–8 (2007)
22. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical Saliency Detection. In: Proc. IEEE CVPR, pp. 1155–1162 (2013)