Leveraging Stereopsis for Saliency Analysis

Yuzhen Niu[†] Yujie Geng[†] Xueqing Li[‡] Feng Liu[†]
[†]Department of Computer Science
Portland State University
Portland, OR, 97207 USA
[‡]School of Computer Science and Technology
Shandong University
Jinan, Shandong, 250101 China
yuzhen@cs.pdx.edu yjgeng@cecs.pdx.edu xqli@sdu.edu.cn fliu@cs.pdx.edu

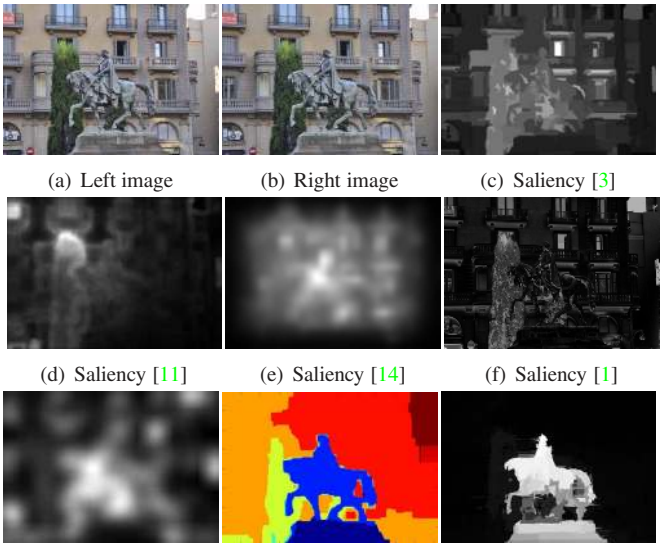
Abstract

Stereopsis provides an additional depth cue and plays an important role in the human vision system. This paper explores stereopsis for saliency analysis and presents two approaches to stereo saliency detection from stereoscopic images. The first approach computes stereo saliency based on the global disparity contrast in the input image. The second approach leverages domain knowledge in stereoscopic photography. A good stereoscopic image takes care of its disparity distribution to avoid 3D fatigue. Particularly, salient content tends to be positioned in the stereoscopic comfort zone to alleviate the vergence-accommodation conflict. Accordingly, our method computes stereo saliency of an image region based on the distance between its perceived location and the comfort zone. Moreover, we consider objects popping out from the screen salient as these objects tend to catch a viewer's attention. We build a stereo saliency analysis benchmark dataset that contains 1000 stereoscopic images with salient object masks. Our experiments on this dataset show that stereo saliency provides a useful complement to existing visual saliency analysis and our method can successfully detect salient content from images that are difficult for monocular saliency analysis methods.

1. Introduction

Image saliency refers to the distinct subjective perceptual quality that makes something in an image stand out from its neighbors and immediately catch a viewer's attention in the early stage of visual processing [16]. A rich literature exists on image saliency analysis in cognitive psychology [36,40], neuroscience [17,35], and vision perception [9,29].

These multi-disciplinary studies have shown that the human vision system is particularly sensitive to high-contrast stimulus (c.f. [17,29,31]). Accordingly, a wide variety of computational methods have been developed to estimate saliency from an image (c.f. [1,3–6,8,10,12,15,18,19,22–26,28,33,37–39,41]). Most of these methods compute saliency based on feature contrast with respect to color, shape, orientation, texture, curvedness, etc. Good results can be created when salient content has one or more of these distinguishable image attributes. However, when salient objects do not exhibit visual uniqueness with respect to one or more of these visual attributes, they become challenging for existing methods to detect, as shown in Figure 1.



(a) Left image (b) Right image (c) Saliency [3]
(d) Saliency [11] (e) Saliency [14] (f) Saliency [1]
(g) Saliency [15] (h) Disparity map (i) Stereo saliency

Figure 1. Stereo saliency. We only show the saliency maps for the left image. The performance of saliency analysis methods depends on feature contrast. When an object does not exhibit distinct visual features, such as color, it becomes challenging for saliency detection. This example shows that stereo saliency based on disparity analysis can be a useful complement to existing visual saliency.

In this paper, we explore stereopsis for saliency analysis. Stereopsis is a process in visual perception leading to the perception of depth from retinal disparities, which are created by projecting two slightly different images onto the retinas of the two eyes [30]. Stereopsis provides an additional depth cue and plays an important role in the human vision system. Stereoscopic depth often enables people to

identify objects from the background with similar visual attributes. Accordingly, we develop two methods for stereo saliency analysis as a useful complement to existing visual saliency analysis.

Our methods work on stereoscopic images as they become more and more popular these years. The recent advances in stereoscopic imaging and displaying technologies make stereoscopic content production easy for common users. For example, camera manufacturers now provide both professional HD stereoscopic camcorders and affordable consumer point-and-shoot stereoscopic cameras.

Our methods compute stereo saliency from the disparity map between the left and right view of a stereoscopic image. We explore both low-level disparity contrast and unique domain knowledge from stereoscopic photography and cinematography for stereo saliency analysis. Our first method computes stereo saliency based on disparity contrast analysis. We extend a recent global contrast-based saliency estimation method [3] to handle disparity map. For disparity contrast, the amount of change alone is often misleading in computing saliency. For example, while the ground in an image has a wide range of disparity values, the ground is not salient. We observe that the disparity (depth) change between an object and its background tends to be abrupt while the change inside the ground tends to be smooth. We then address this unique problem in disparity contrast analysis by treating abrupt change and smooth change differently.

Our second method explores the domain knowledge from stereoscopic photography for stereo saliency analysis. First, the *vergence-accommodation conflict* often causes 3D fatigue, as detailed in Section 2.2. A good stereoscopic image typically positions salient content inside the comfort zone. Based on this observation, our method considers image content inside the comfort zone more salient than the others. Moreover, objects that pop out from the screen tend to catch a viewer's attention [13], so our method considers these objects salient.

We build a stereo saliency benchmark dataset that contains 1000 stereoscopic images together with salient object masks. Our experiments on this benchmark dataset show that our method can successfully identify salient content from images that are difficult for existing visual saliency methods and stereo saliency can be a useful complement to existing visual saliency analysis.

This paper contributes to image saliency analysis with stereo saliency as a useful complement to existing methods. Our method not only estimates stereo saliency based on contrast analysis, but also explores unique domain knowledge in stereoscopic photography for assisting saliency estimation. Finally, this paper provides a photo collection with 1000 stereoscopic photos and manually segmented salient object masks. To our best knowledge, this is the first stereoscopic photo collection at this scale for saliency analysis.

2. Stereo Saliency

This paper presents two approaches to stereo saliency analysis. These approaches compute stereo saliency based on low-level disparity (depth)[1] contrast analysis and domain knowledge in stereoscopic photography and cinematography. While the role of stereoscopic disparity in the preattentive processing is still an ongoing research question, researchers consider stereoscopic disparity a probable attribute [40]. We consider stereo saliency more in a pragmatic way. When taking a photo, people tend to place an important object at a different depth level than the others. So our first method computes stereo saliency based on disparity contrast analysis. Moreover, a good stereoscopic image takes care of disparity distribution to avoid 3D fatigue [27]. Accordingly, our second method explores stereoscopic photography rules for stereo saliency analysis.

Stereo saliency analysis works on disparity map. Before we elaborate stereo saliency analysis, we first briefly describe how our method estimates a disparity map from the left and right view of a stereoscopic image. Dense disparity map is obtained from stereo matching, which is a classic computer vision problem. A good survey can be found in [34]. We apply the SIFT flow method to disparity estimation for its robustness [21]. An example of dense disparity estimation is shown in Figure 2 (c), with cool and warm color encoding small and big disparities, respectively. Objects with negative disparity values are perceived in front of the screen; vice versa, objects with positive disparity values are perceived behind the screen. We describe our methods for stereo saliency analysis in the following subsections.

2.1. Stereo Saliency from Disparity Contrast

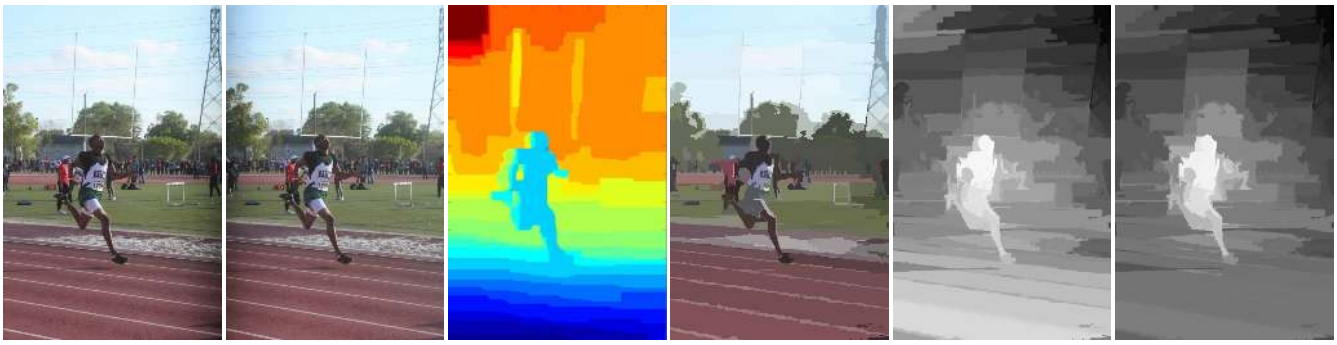
We extend a recent color contrast-based saliency detection method from Cheng *et al.* [3] for disparity contrast analysis. This method first segments an input image into regions using the graph-based image segmentation method [7]. Since our input has a disparity map, we adapt this method to consider both color and disparity during segmentation by treating the disparity value as a fourth channel. Then the saliency value for each region is computed based on its contrast with all the others in the image.

$$S_c(R_i) = \sum_{R_k \neq R_i} d(R_i, R_k) n_k \qquad (1)$$

where $S_c(R_i)$ is the saliency for region R_i, $d(R_i, R_k)$ is the disparity difference between R_i and R_k, and n_k is the size of R_k. This method considers a larger region R_k contributes more to the saliency of R_i than a smaller one.

We compute the disparity difference between R_i and R_k as the average disparity difference between each pixel in R_i

[1] Disparity and depth are two interdependent terms in stereo vision [34]. We use them interchangeably whenever appropriate.



(a) Left image (b) Right image (c) Disparity map (d) Segmentation (e) Saliency ($\beta = 0$) (f) Saliency ($\beta = 2$)

Figure 2. Contrast-based stereo saliency analysis. This example shows that the amount of disparity (depth) change alone is often insufficient for stereo saliency analysis (e). While the disparity changes in the ground (c), the ground is not salient. A better result (f) can be created by treating abrupt disparity change and smooth disparity change differently.

and R_k in a similar way to the color-based region difference [3]. However, the amount of disparity difference alone is often insufficient for disparity contrast-based saliency analysis. This is a unique problem of stereo saliency analysis. Consider an example shown in Figure 2. The ground in the image is not salient although it has a wide range of disparity values. We observe that the depth change between an object and its background tends to be abrupt while the depth change in the ground tends to be smooth. We then address this problem by considering the abruptness of disparity change in computing the disparity difference.

$$d(R_i, R_k) = \frac{\sum_{p \in R_i, q \in R_k} w(p,q)(d_v(p,q) + \beta d_m(p,q))}{n_i n_k} \quad (2)$$

where $d_v(p,q)$ is the disparity difference between pixel p and q, defined as $|d_q - d_p|$. $d_m(p,q)$ considers the abruptness of disparity change in a similar way that the intervening contour method considers the abruptness of intensity change [20]. Our method computes $d_m(p,q)$ as the maximal disparity change along the path between p and q. In this way, smooth disparity transition between two pixels contributes less than abrupt change. To cope with the noise in disparity estimation, our method selects the top 10% of the disparity changes and computes their average as $d_m(p,q)$. β is a weight with default value 2.0. $w(p,q)$ is a weight computed according to the spatial distance between p and q as $w(p,q) = e^{-\|p-q\|_2^2/\sigma^2}$, where the image coordinates are normalized to [0, 1] and σ^2 is a parameter with default value 0.4. We consider that a closer pixel influences saliency of its neighbor more than a farther one.

Figure 2 (e) shows the stereo saliency map computed according to Equation 1 and 2 without considering the disparity change abruptness (namely $\beta = 0$). We can see that the result is problematic as the ground is considered very salient. This problem is solved by considering the disparity change abruptness (with $\beta = 2$), as shown in Figure 2 (f).

2.2. Domain Knowledge Assisted Saliency Analysis

A good stereoscopic image needs to minimize 3D fatigue to viewers. One common source of 3D fatigue is the vergence-accommodation conflict [27]. This conflict increases as the perceived depth of an object becomes further away from the screen, as shown in Figure 3. A zone close to the screen is called the comfort zone. According to stereoscopic perception, an object with zero disparity is perceived on the screen and objects with small disparity magnitude are in the comfort zone. Image content with positive disparities is perceived behind the screen; vice versa, content with negative disparities is perceived in front of the screen.

In practice, content of interest is often given small or zero disparities to minimize the vergence-accommodation conflict. The disparity distribution can be adjusted by shifting the left and right view of a stereoscopic image either on the camera or by post-processing. Some consumer level cameras automate this adjustment online. Professional stereoscopic cameras also allow to toe-in the cameras to achieve an optimal disparity distribution. Moreover, studies have shown that an object that is perceived popping-out the screen often catches a viewer's attention [13].

These unique features of stereoscopic photography give useful cues for saliency analysis. We design the following two rules to compute knowledge-based stereo saliency.

1. Objects with small disparity magnitudes (e.g. in the comfort zone) tend to be salient;
2. Objects popping out from the screen tend to be salient.

We follow Rule 1 and compute a saliency map S_1 according to the disparity magnitude. Specifically, we assign big saliency values to regions with small disparity magnitudes.

$$S_1(R_i) = \begin{cases} \frac{d_{max} - \bar{d}_i}{d_{max}} & \text{if } \bar{d}_i \geq 0 \\ \frac{d_{min} - \bar{d}_i}{d_{min}} & \text{if } \bar{d}_i < 0 \end{cases} \quad (3)$$

where d_{max} and d_{min} are the maximal and minimal disparities. $\bar{d}_i$ is the average disparity in region R_i. Accord-

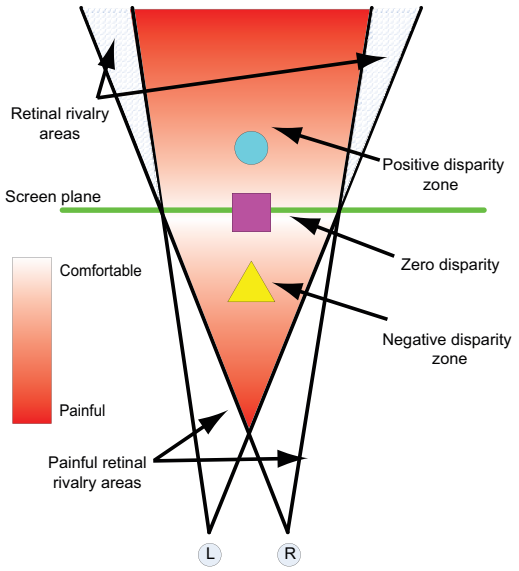


Figure 3. Illustration of stereoscopic comfort zone.

ing to this equation, regions perceived on the screen (e.g. zero disparity) are given the highest saliency values; while those perceived either furthest behind the screen (with the maximal positive disparities) or furthest front of the screen (with the minimal negative disparities) are assigned the lowest saliency values.

Objects with negative disparities are perceived popping out from the screen. This suggests that the more negative an object's disparity is, the more it is perceived popping out. Accordingly, we compute a new saliency map S_2 based on Rule 2 as follows.

$$S_2(R_i) = \frac{d_{max} - \bar{d}_i}{d_{max} - d_{min}} \tag{4}$$

We now further examine Rule 1 and Rule 2. As shown in Figure 4, these two rules are consistent in the positive disparity zone. In the negative disparity zone, they conflict with each other. According to Rule 1, an object with a small disparity magnitude is salient as it suffers less from the vergence-accommodation conflict. On the other hand, according to Rule 2, an object with a big disparity magnitude is salient as it pops out more from the screen.

How to combine saliency maps from Rule 1 and Rule 2 is a challenging problem. We tackle this problem by learning from the practice in stereoscopic photography. When an image only has negative disparities, the whole scene is perceived popping out from the screen. The image background that is furthest away from the camera has smallest disparity magnitude. Typically, the image background is not salient. We consider Rule 2 is more applicable than Rule 1. When an image has both negative and positive disparities, it is more likely that the disparity distribution of this image has been adjusted either manually or automatically by the camera to minimize the vergence-accommodation conflict.

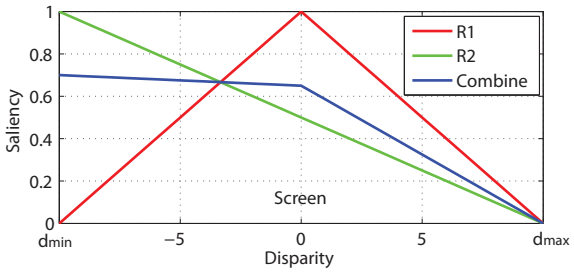


Figure 4. Illustration of R1 and R2.

Then, we consider Rule 1 is more applicable. For an image that only has positive disparities, Rule 1 and Rule 2 are consistent so they can be easily combined.

Based on the above analysis, we linearly combine saliency maps from Rule 1 and Rule 2 with a disparity map-dependent coefficient λ.

$$S_d(R_i) = (1-\lambda)S_1(R_i) + \lambda S_2(R_i) \tag{5}$$

$$\lambda = \gamma + \frac{n_{d_p<0}}{n}(1-\gamma) \tag{6}$$

where $S_1(R_i)$ and $S_2(R_i)$ are the saliency maps computed according to Rule 1 and Rule 2, respectively. $n_{d_p<0}$ is the number of pixels with a negative disparity and n is the total number of pixels in the image. γ is a parameter with default value 0.5.

Figure 5 shows examples of domain knowledge assisted saliency analysis. We find that our method works well for the first example. However, the second and third examples are less successful. For the third example, some of the background regions has the smallest positive disparities, so they are closest to the screen and considered the most salient. This actually is a common problem in a stereoscopic image: salient objects are always surrounded by the background; the background typically extends throughout the whole disparity range in the image. So some background regions will be detected as salient.

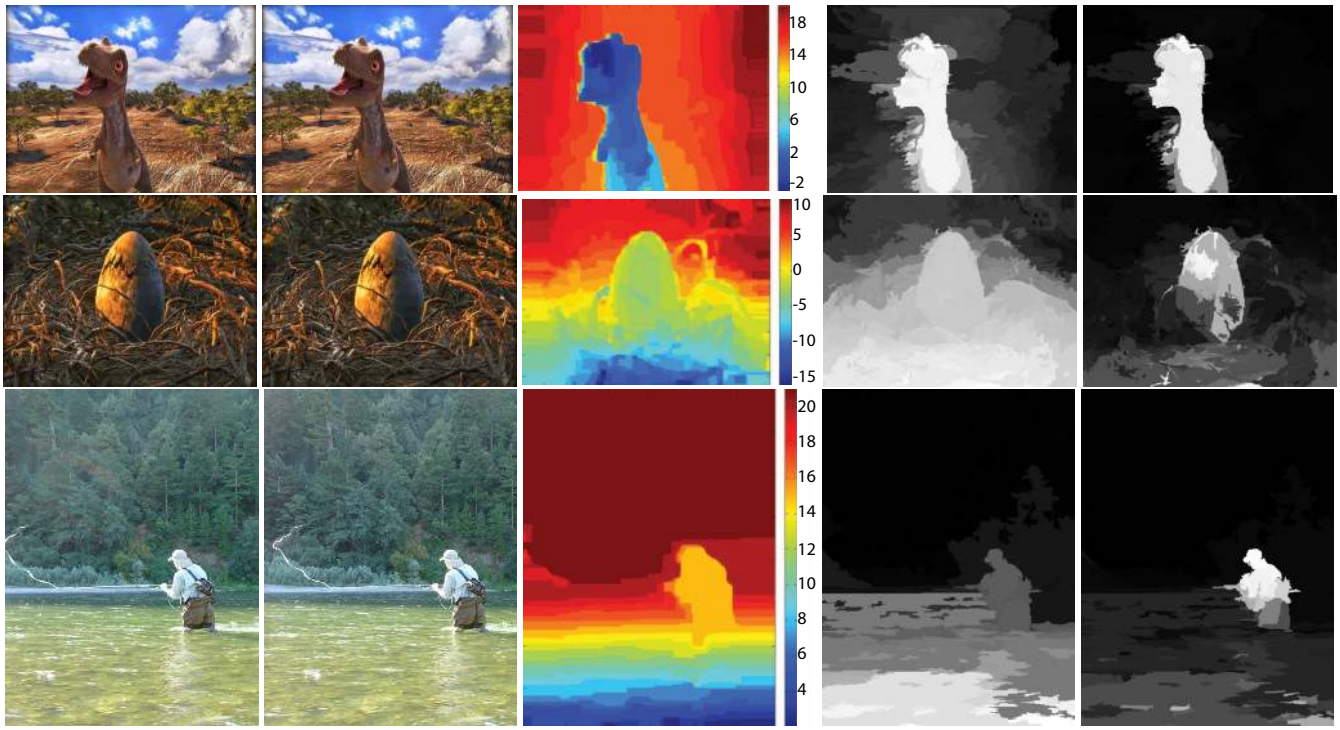
We address this problem by modulating the original saliency map with local contrast-based saliency analysis. We observe that disparities change little in each row in some of the background areas, such as the ground. So we compute the local disparity contrast along each row and use it to modulate the domain knowledge based saliency as follows.

$$S_r(R_i) = S_d(R_i)\frac{\sum_{p\in R_i}|\bar{d}_p^r - d_p|}{n_i} \tag{7}$$

where p is a pixel in region R_i, d_p is its disparity, and $\bar{d}_p^r$ is the average disparity of the row that contains p. As shown in Figure 5 (d), local disparity contrast modulation improves the original saliency map significantly.

2.3. Stereo Saliency Map

Our method multiplies the global disparity contrast based saliency S_c with the domain knowledge based



(a) Left image (b) Right image (c) Disparity map (d) Saliency before and after local contrast modulation

Figure 5. Domain knowledge assisted stereo saliency analysis.

saliency S_r as the final stereo saliency.

$$S_s(R_i) = S_c(R_i)S_r(R_i) \tag{8}$$

3. Experiments

We experiment with our stereo saliency detection methods on a collection of 1000 stereoscopic images. In this section, we first describe our stereoscopic image collection. Then we report the comparisons between our methods and a variety of saliency detection methods. We show that stereo saliency is a useful complement to existing visual saliency analysis and our methods can effectively detect stereo saliency.

3.1. Stereoscopic Image Database

Public benchmark datasets are available for image saliency analysis, such as Liu *et al*. [23] and Achanta *et al*. [1]. However, these datasets are dedicated to monocular images. To the best of our knowledge, there is no publicly available stereoscopic image database for saliency analysis. We build a stereoscopic image dataset with ground truth following the procedure that was designed by Liu *et al*. [23] to build their benchmark dataset.

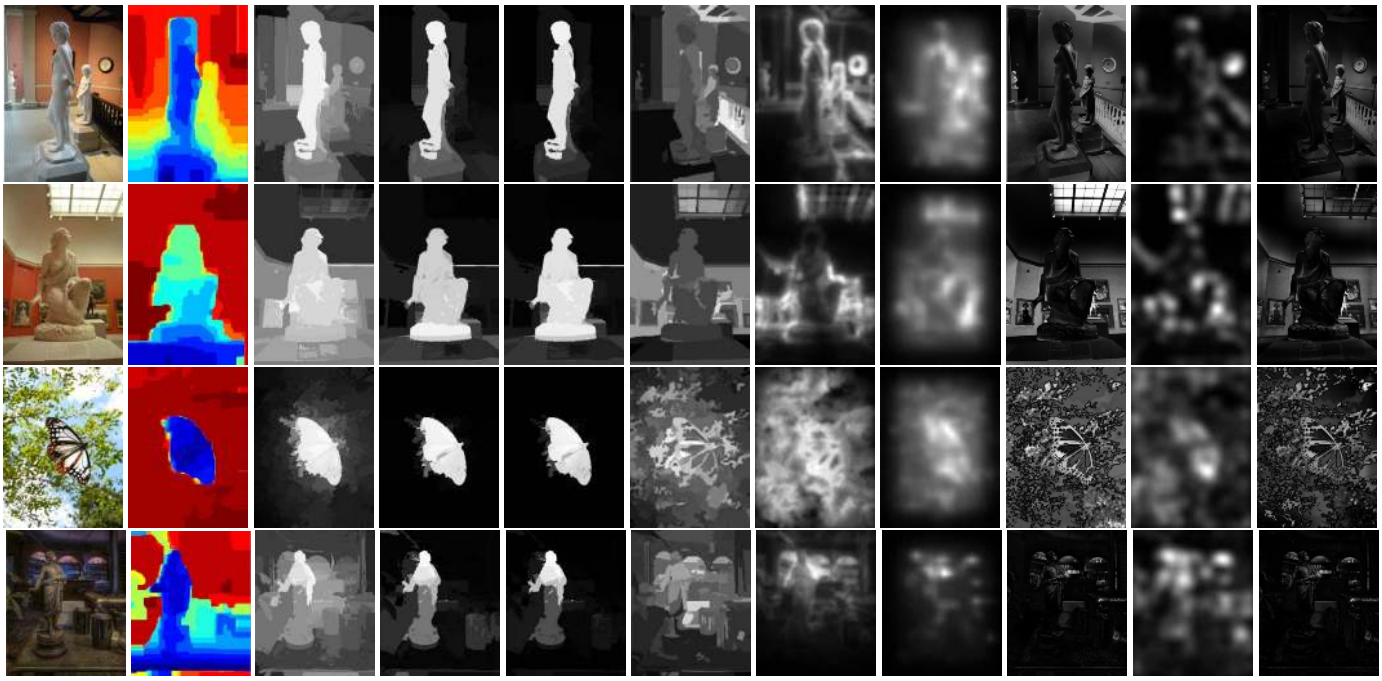
Figure 6. Benchmark examples. We only show the left views of stereoscopic images. The first image was consistently labeled by three users and was included into our dataset. Its salient object mask is shown in the middle. The last image was not consistently labeled and was removed from our dataset.

We first download 1250 stereoscopic images from *Flickr*[2], *Stereoscopic Image Gallery*[3], and *NVIDIA 3D Vision Live*[4]. Then, three users are asked to enclose the most salient object in each image with a rectangle. In this way, each image is labeled by three users. People may have different opinions on what a salient object is in the same image. We remove the images with the least consistent labels. Specifically, we sort the images according to the overlap between the labeled rectangles and select the top 1000 images. Then we ask a user to manually segment the salient object(s) in each of the 1000 images. It takes about 3 minutes on average for the user to segment an image using *Adobe Photoshop*. We show some examples in Figure 6.

3.2. Performance Evaluation

We evaluate the three saliency maps from our methods, namely global disparity contrast-based stereo saliency (CSS) in Section 2.1, knowledge-assisted stereo saliency (KSS) in Section 2.2, and their combination (SS) in Section 2.3. We also selected six state-of-the-art saliency detec-

[2]http://www.flickr.com/
[3]http://www.stereophotography.com/
[4]http://photos.3dvisionlive.com/



(a) Input (b) Disparity (c) CSS (d) KSS (e) SS (f) RC (g) CA (h) GB (i) FT (j) SR (k) MS

Figure 7. Stereo saliency maps (CSS, KSS, and SS) and the other saliency maps. The original images in the first and second rows are from http://www.flickr.com/photos/spacetime/3805384593/ and http://www.flickr.com/photos/spacetime/3806124429.

tion methods, including RC [3], CA [11], GB [14], FT [1], SR [15], and MS [2] and used these methods as well as ours to estimate saliency maps from our benchmark dataset. The implementations of all these six methods are downloaded from their authors' project websites. Since the saliency maps from CA and SR methods are smaller than the original images, we up-sample them to the same resolution as the original images. Figure 7 shows some visual comparison between our methods and all the other six methods.

In our experiments, we examine 1) how our stereo saliency methods compare to the state of the art and 2) whether stereo saliency analysis complements them.

For Question 1, we plot the precision-recall curve for each method. Specifically, we threshold the saliency map from each method with a gradually increasing value t and obtain a binary salient object map. We compare this object map with the ground truth object mask to compute the precision and recall curve. As shown in Figure 8, the disparity contrast based stereo saliency (CSS) is better than CA, SR, and FT for most recall values, comparable to MS, and worse than RC and GB. The knowledge-assisted stereo saliency (KSS) and the combination of CSS and KSS are consistently better than the other six methods.

For Question 2, we add each of the three stereo saliency maps to each of the six image saliency maps and plot the precision-recall curve for each combined saliency map. As shown in Figure 8, combining CSS with each of the six image saliency maps can improve both methods. This shows that CSS can effectively complement existing saliency analysis methods. Similarly, KSS can complement RC, GB, and MS well and CA slightly. Combining SR or FT with KSS does not improve KSS although improves SR and FT significantly. SS complements existing visual saliency analysis in a similar way to KSS. Overall, stereo saliency complements existing visual saliency analysis well. It at least will not hurt existing methods.

The computational cost of our methods consists of two parts. The first is disparity map estimation and pre-segmentation. It takes about 22 seconds for the SIFT flow method to estimate the disparity map from a 512×384 image and 0.2 seconds for pre-segmentation. The second is stereo saliency estimation. The global contrast-based stereo saliency takes 0.6 seconds and the domain knowledge assisted method takes 0.1 seconds. The performance data is obtained on a Desktop machine with an Intel Duo Core, 3.06GHz CPU and 4GB memory.

3.3. Automatic Salient Object Segmentation

We further examine the performance of our stereo saliency analysis within the application of automatic object segmentation. We adopted the same method of Cheng *et al*. [3]. This method initializes the segmentation with a binarized saliency map. The threshold which gives 95% recall rate in the previous experiment is used to threshold the saliency map and create the binarized map. Then the GrabCut method [32] is extended to take the binarized saliency

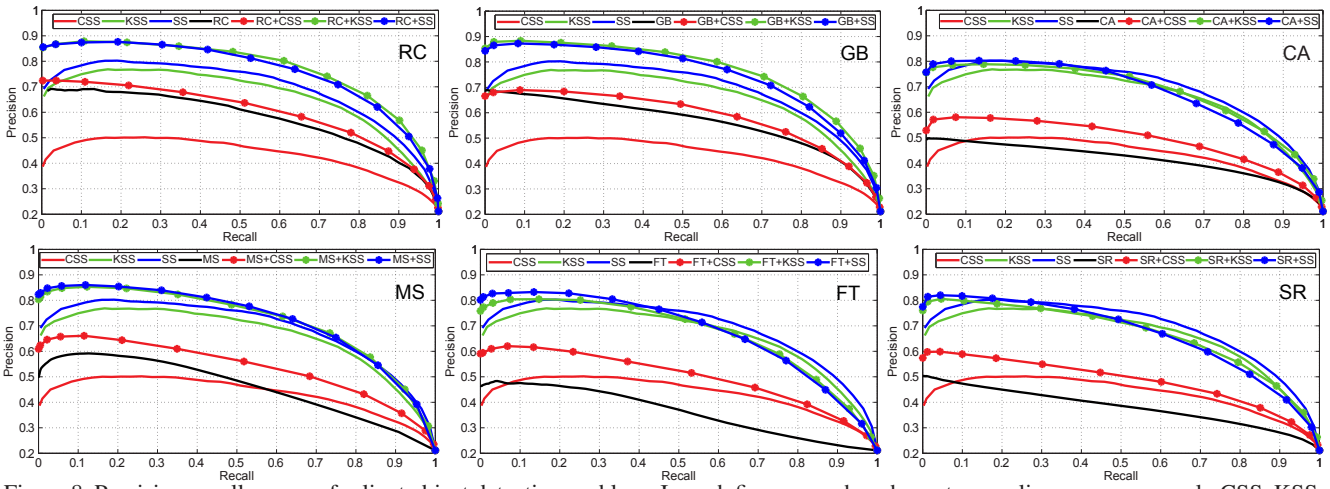


Figure 8. Precision-recall curves of salient object detection problem. In each figure, we show how stereo saliency maps, namely CSS, KSS, and SS, compare to and complement one of the existing saliency detection methods.

map as input and is iteratively applied to segmentation. As we have the disparity map, we use both color and disparity to compute the difference between two pixels in the GrabCut framework. We show the precision and recall for each method in Figure 9. Figure 10 shows some segmentation results from our stereo saliency maps (SS).

3.4. Limitations

The performance of our methods depend on the quality of disparity maps. While dense stereo matching has been making progress these years, it is still a challenging problem. Since our methods work on the granularity of regions, our methods do not require very accurate disparity map, as shown in Figure 7. However, when the disparity map is very misleading, our methods can still fail.

Our method explores stereoscopic photography rules for stereo saliency analysis. This is particularly useful for high-quality stereoscopic images as a good stereoscopic image needs to follow these rules to avoid 3D fatigue. However, as explained in this paper, these rules may conflict with each other for some images. While our method for combining these rules works well for most of images we experimented on, it can fail sometimes.

Finally, stereo saliency is useful only if a salient object stays at a different depth than its surroundings. This is the same as saliency analysis w.r.t other visual features. If an object cannot be separated from the background w.r.t one feature, the saliency analysis based on that feature fails.

4. Conclusion

This paper explored stereopsis for saliency analysis. We developed two methods for stereo saliency detection. The first method computes stereo saliency based on global disparity contrast. The second method makes use of the stereoscopic rules for saliency estimation. Our experiments

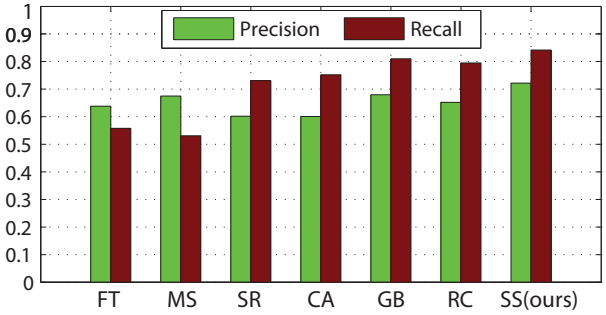
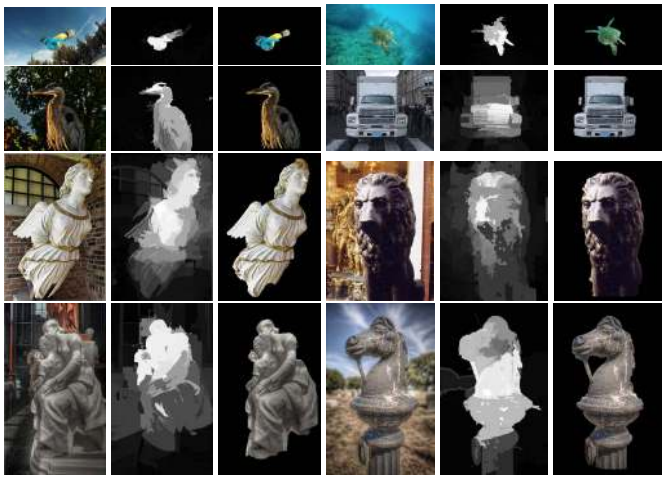


Figure 9. Automatic salient object segmentation.



(a) Input (b) Saliency (c) Segm. (d) Input (e) Saliency (f) Segm.

Figure 10. Object segmentation based on stereo saliency.

showed that stereo saliency is a useful complement to existing visual saliency analysis and our methods are able to achieve high-quality stereo saliency detection results.

Acknowledgements

We would like to thank reviewers for their constructive comments. We would also like to thank David Gadbois,

Luis A Ortiz, Tetsuya Shimizu, and Rob Crockett for providing their photos for our research and Flickr users, including -ytf-, Genista, Wagman, fossilmike, relaxednow, Balliolman, and Scott3DReilly, for letting us use their photos under a Creative Commons license or with their permissions. This work was supported by the Portland State University Faculty Enhancement Grant.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In *IEEE CVPR*, pages 1597–1604, 2009. 1, 5, 6

[2] R. Achanta and S. Süsstrunk. Saliency detection using maximum symmetric surround. In *IEEE ICIP*, 2010. 6

[3] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In *IEEE CVPR*, pages 409–416, 2011. 1, 2, 3, 6

[4] Y. Cohen and R. Basri. Inferring region salience from binary and gray-level images. *Pattern recognition*, 36:2349–2362, 2003. 1

[5] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu. Visual saliency detection by spatially weighted dissimilarity. In *IEEE CVPR*, pages 473–480, 2011. 1

[6] S. Fan and F. P. Ferrie. Structure guided salient region detector. In *BMVC*, 2008. 1

[7] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59:167–181, 2004. 2

[8] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun. Salient object detection by composition. In *IEEE ICCV*, 2011. 1

[9] D. Gao, V. Mahadevan, and N. Vasconcelos. On the plausibility of the discriminant center-surround hypothesis for visual saliency. *Journal of Vision*, 8:1–18, 2008. 1

[10] D. Gao and N. Vasconcelos. Bottom-up saliency is a discriminant process. In *IEEE ICCV*, 2007. 1

[11] S. Goferman, L. Zelnik-manor, and A. Tal. Context-aware saliency detection. In *IEEE CVPR*, 2010. 1, 6

[12] V. Gopalakrishnan, Y. Hu, and D. Rajan. Random walks on graphs to model saliency in images. In *IEEE CVPR*, pages 1698–1705, 2009. 1

[13] J. Häkkinen, T. Kawai, J. Takatalo, R. Mitsuya, and G. Nyman. What do people look at when they watch stereoscopic movies? In *Proc. SPIE 7524*, 2010. 2, 3

[14] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, pages 545–552, 2006. 1, 6

[15] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *IEEE CVPR*, 2007. 1, 6

[16] L. Itti. Visual salience. *Scholarpedia*, 2(9):3327, 2007. 1

[17] L. Itti and C. Koch. Computational modeling of visual attention. *Nature reviews neuroscience*, 2:194–203, 2001. 1

[18] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20:1254–1259, 1998. 1

[19] G. Kim, D. Huber, and M. Hebert. Segmentation of salient regions in outdoor scenes using imagery and 3-d data. In *IEEE WACV*, 2008. 1

[20] T. K. Leung and J. Malik. Contour continuity in region based image segmentation. In *ECCV*, pages 544–559, 1998. 3

[21] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):978–994, 2011. 2

[22] F. Liu and M. Gleicher. Region enhanced scale invariant saliency detection. In *IEEE ICME*, pages 1477–1480, 2006. 1

[23] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. In *IEEE CVPR*, 2007. 1, 5

[24] J. Luo and A. Singhal. On measuring low-level saliency in photographic images. In *IEEE CVPR*, pages 84–89, 2000. 1

[25] Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia*, pages 374–381, 2003. 1

[26] V. Mahadevan and N. Vasconcelos. Spatiotemporal saliency in highly dynamic scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32:171–177, 2010. 1

[27] B. Mendiburu. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, 2009. 2, 3

[28] N. Murray, M. Vanrell, X. Otazu, and C. Parraga. Saliency estimation using a non-parametric low-level vision model. In *IEEE CVPR*, pages 433–440, 2011. 1

[29] H. Nothdurft. Salience from feature contrast: additivity across dimensions. *Vision Research*, 40(10-12):1183–1201, 2000. 1

[30] G. Poggio and T. Poggio. The analysis of stereopsis. *Annual Review of Neuroscience*, 7(1):379–412, 1984. 1

[31] J. Reynolds and R. Desimone. Interacting roles of attention and visual salience in v4. *Neuron*, 37(5):853–863, 2003. 1

[32] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23:309–314, 2004. 6

[33] S. Sarkar. Learning to form large groups of salient image features. In *IEEE CVPR*, 1998. 1

[34] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47(1):7–42, 2002. 2

[35] A. Torralba. Contextual influences on saliency. *Neurobiology of attention*, pages 586–593, 2005. 1

[36] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980. 1

[37] R. Valenti, N. Sebe, and T. Gevers. Image saliency by isocentric curvedness and color. In *IEEE ICCV*, 2009. 1

[38] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. Rowley. Image saliency: From intrinsic to extrinsic context. In *IEEE CVPR*, pages 417–424, 2011. 1

[39] Z. Wang and B. Li. A two-stage approach to saliency detection in images. In *IEEE ICASSP*, 2008. 1

[40] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5:495–501, 2004. 1, 2

[41] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, pages 815–824, 2006. 1