Leveraging stereopsis for saliency analysis report

Paper's Authors

Yuzhen Niu, Yujie Geng, Xueqing Li, and Feng Liu

Advisor

Dr. Maryam Abedi

Student

Mohammad Shahpouri

November
2022

Contents

List of Figures

List of Tables

List of Equations

1 Stereo Saliency

Stereo saliency based on low-level disparity (depth)[1] contrast analysis and domain knowledge in stereoscopic photography and cinematography are computed to obtain stereo saliency analysis.

To obtain disparity map the SIFT flow method is applied to disparity estimation for its robustness [2].

1.1 2.1. Stereo Saliency from Disparity Contrast

A color contrast-based saliency detection method from Cheng et al. [3] is extended for disparity contrast analysis. This method first segments an input image into regions using the graph-based image segmentation method [4]. Then the saliency value for each region is computed based on its contrast with all the others in the image.

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

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. Disparity difference is calculated as follows:

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

where $d_v(p,q)$ is the disparity difference between pixel p and q, defined as $|d_q - d_p|$. $d_m(p,q)$ computes the maximal disparity change along the path between p and q. $\beta = 2.0$ is a weight. $w(p,q) = w(p,q) = e^{-\|p-q\|_2^2/\sigma^2}$ is a weight computed according to the spatial distance between p and q, where the image coordinates are normalized to $[0,1]$ and $\sigma^2 = 0.4$. It is supposed that a closer pixel influences saliency of its neighbor more than a farther one.

1.2 Domain Knowledge Assisted Saliency Analysis

To compute knowledge-based stereo saliency two rules are followed.

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.

Following Rule 1 and computing a saliency map S_1 according to the disparity magnitude. assigning big saliency values to regions with small disparity magnitudes.

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

where d_{max} and d_{min} are the maximal and minimal disparities. $\bar{d}_i$ is the average disparity in region R_i . According to this equation, regions perceived on the screen (e.g. zero disparity) are given the highest saliency values.

[1] Disparity and depth are two interdependent terms in stereo vision [1]

A new saliency map S_2 based on Rule 2 is computed as follows:

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

When an image has both negative and positive disparities, Rule 2 is more applicable than Rule 1. On the other hand, Rule 1 is more applicable when an image that only has positive disparities, Rule 1 and Rule 2 are consistent so they can be easily combined.

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

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

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. $\gamma = 0.5$.

To improve the domain knowledge based saliency the local disparity contrast along each row is applied 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.

1.3 Stereo Saliency Map

The final stereo saliency is calculated by multiplying the global disparity contrast based saliency S_c with the domain knowledge based saliency S_r.

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

2 Experimental results

They make a dataset with ground truth which consist 1000 stereoscopic images then evaluate their methods with RC [3], CA [5], GB [6], FT [7], SR [8], and MS [9].

2.1 Quantitative result

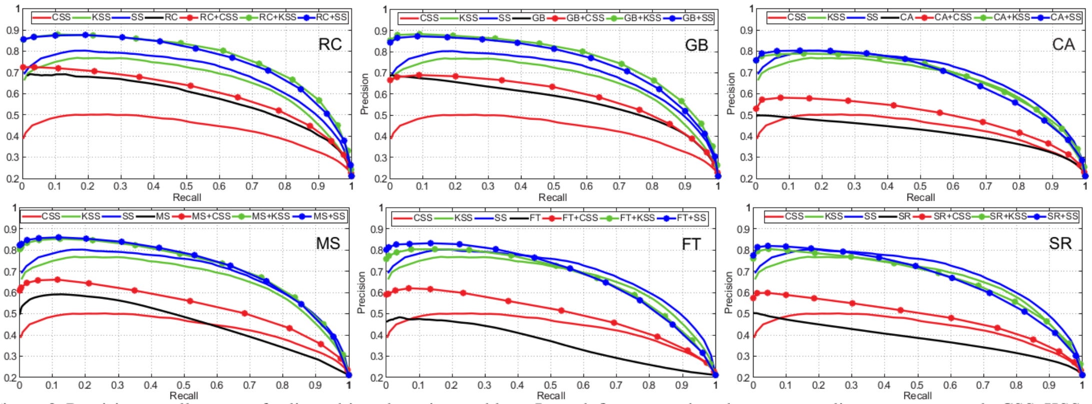
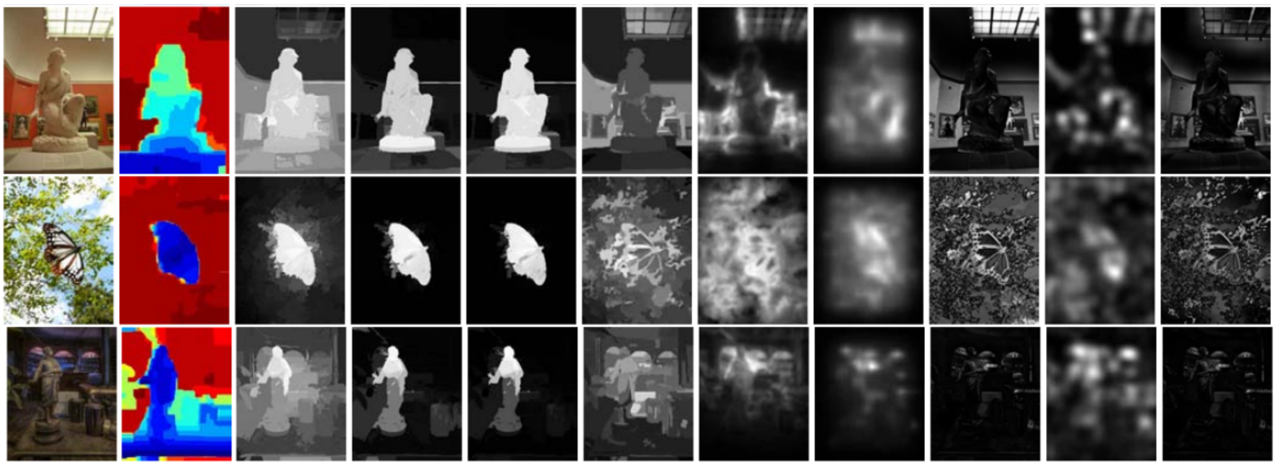


Figure 1. Precision-recall curves of salient object detection problem. each figure shows how stereo saliency maps, disparity contrast based stereo saliency (CSS), knowledge-assisted stereo saliency (KSS), and combination of CSS and KSS (SS), compare to and complement one of the existing saliency detection methods.

2.2 Qualitative result



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

Figure 2. Stereo saliency maps (CSS, KSS, and SS) and the other saliency maps.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, Apr 2002. 1

[2] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011. 1

[3] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *CVPR 2011*, pp. 409–416, 2011. 1, 3

[4] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, Sep 2004. 1

[5] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2376–2383, 2010. 3

[6] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems* (B. Schölkopf, J. Platt, and T. Hoffman, eds.), vol. 19, MIT Press, 2006. 3

[7] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, 2009. 3

[8] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007. 3

[9] R. Achanta and S. Süsstrunk, "Saliency detection using maximum symmetric surround," in *2010 IEEE International Conference on Image Processing*, pp. 2653–2656, 2010. 3