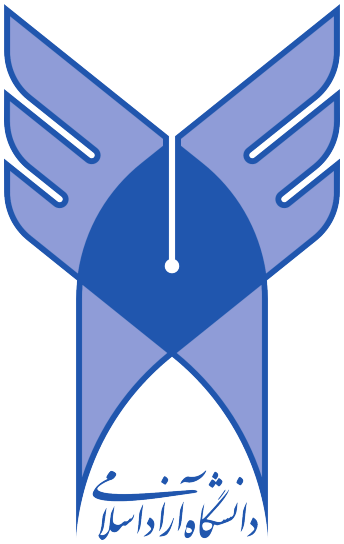

Salient Region Detection via High-Dimensional Color Transform report

Paper's Authors

Jiwhan Kim, Dongyoon Han, Yu-Wing Tai, Junmo Kim

Advisor

Dr. Maryam Abedi

Student

Mohammad Shahpouri

October
2022

Contents

List of Figures

List of Tables

List of Equations

1 Initial Salient Regions Detection

Superpixel Saliency Features They use SLIC superpixel [8], as its low computational cost and high performance, to over-segmentation image I to form superpixels $\mathbf{X} = \{X_1, \ldots, X_N\}$. N denotes the number of superpixels and it is set to 500. The location feature is superpixels' x and y locations because humans tend to

Table 1. Features which are used to compute feature vector for each superpixel.

Feature Descriptions	**Dim**
Location Features	
The average normalized x coordinates	1
The average normalized y coordinates	1
Color Features	
The average RGB values	3
The average CIELab values	3
The average HSV values	3
Color Histogram Features	
The RGB histogram	1
The CIELab histogram	1
The hue histogram	1
The saturation histogram	1
Color Contrast Features	
The global contrast of the color features	9
The local contrast of the color features	9
The element distribution of the color features	9
Texture and Shape Features	
Area of superpixel	1
Histogram of gradients (HOG)	31
Singular value feature	1

pay more attention to objects that are located around the center of an image [**?**].

Since certain colors tend to draw more attention than the others [**?**], colors are used as a feature.

The chi-square distance is calculated between histograms to measure histogram feature. It is defined as $D_{H_i} = \sum_{j=1}^{N} \sum_{k=1}^{b} [\frac{(h_{ik} - h_{jk})^2}{h_{ik} + h_{jk}}]$, where b is the number of histogram bins and it is set to 8.

Color contrast features are calculated by the global contrast of the color features, the local contrast of the color features, and the element distribution of the color features. The global contrast of the i^{th} superpixel is given by $D_{G_i} = \sum_{j=1}^{N} d(c_i, c_j)$ where $d(c_i, c_j$ denotes the Euclidean distance between the i^{th} and j^{th} superpixel's color value c_i and c_j. The local contrast of color features is defined as $D_{L_i} = \sum_{j=1}^{N} \omega_{i,j}^{p} d(c_i, c_j)$ where $\omega_{i,j}^{p} = \frac{1}{Z_i} \exp(-\frac{1}{2\sigma_p^2} \|\mathbf{p}_i - \mathbf{p}_j\|_2^2)$, in which $\mathbf{p}_i$ denotes the position of i^{th} superpixel, Z_i is the normalization term, and $\sigma_p^2 = 0.25$. Element distribution [**?**] measurs the compactness of colors in term of their spatial color variance.

For texture and shape features, area of superpixel, histogram of gradients (HOG) and the singular value feature are computed. HOG features implemented by Felzenszwalb et al [9]. The HOG provides appearance features by using around the pixels' gradient information at fast speed. The Singular Value Feature (SVF) [10] is utilized to determine the blurred region from a test image, because a blurred region often tends to be a background.

The aforementioned features are concatenated and will be used to estimate the initial saliency map. Hence, feature vectors consist of 75 dimensions.

Initial Saliency Map Estimation via Regression A random forest regressor is used to estimate each region's degree to be salient. The regressor has 200 trees, with no limit for the node size.

2 High-Dimensional Color Transform for Saliency Detection

Trimap Construction First, the initial saliency map, from previous section, is divided into 2×2, 3×3, and 4×4 regions. Otsu's multi-level adaptive thresholding [11] is adopted on each region. Then three different scale thresholded saliency maps are merged to obtain global thresholding as follows:

$$T(i) = \begin{cases} 1 & \text{if } T'(i) \geq 18 \\ 0 & \text{if } T'(i) \leq 6 \\ unknown & \text{else} \end{cases} \tag{1}$$

High-Dimensional Color Transform Applying gamma correction to each of the color coefficients, results

Table 2. Summary of color coefficients concatenated in our high-dimensional color transform space.

Color channel	**Gamma value γ_k**	k	**Dim**
RGB	$0.5k$	$1 \sim 4$	12
CIELab	$0.5k$	$1 \sim 4$	12
Hue	$0.5k$	$1 \sim 4$	4
Saturation	$0.5k$	$1 \sim 4$	4
Gradient of RGB	$0.5k$	$1 \sim 4$	12

in a $l = 44$ high-dimensional vector to represent the colors of an image:

$$\mathbf{K} = [R_S^{\gamma_k} \; G_S^{\gamma_k} \; B_S^{\gamma_k} \cdots] \in \mathbb{R}^{N \times l} \tag{2}$$

Saliency Map Construction via Optimal Linear Combination of Coefficients The problem is formulated as a least square problem which minimizes:

$$\min_{\alpha} \|(\mathbf{U} - \widetilde{\mathbf{K}}\alpha)\|_2^2 \tag{3}$$

where $\alpha \in \mathbb{R}^l$ is the coefficient vector that we want to estimate, $\widetilde{\mathbf{K}}$ is a $M \times l$ matrix with each row of $\widetilde{\mathbf{K}}$ corresponding to a color samples in the definite foreground/background regions, M is the number of color samples in the definite foreground/background regions ($M \ll N$), $\mathbf{U}$ is an M dimensional vector with its value equal to one if a color sample belongs to the definite foreground and 0 if a color sample belongs to the definite background.

After getting the optimal coefficient α, saliency map can be constructed as:

$$S_{LS}(X_i) = \textstyle\sum_{j=1}^{l} \mathbf{K}_{ij}\alpha_j, \quad i = 1, 2, \ldots, N \tag{4}$$

Spatial Refinement In this step pixels which are closer to the definite foreground region are given more weight and vice versa for pixels that are closer to the definite background region in the trimap. The spatial saliency

map is defined as

$$S_s(X_i) = \exp\left(-k\frac{\min_{j\in\mathcal{F}}(d(\mathbf{p}_i,\mathbf{p}_j))}{\min_{j\in\mathcal{B}}(d(\mathbf{p}_i,\mathbf{p}_j))}\right) \tag{5}$$

where $\min_{j\in\mathcal{F}}(d(\mathbf{p}_i,\mathbf{p}_j))$ and $\min_{j\in\mathcal{B}}(d(\mathbf{p}_i,\mathbf{p}_j))$ are the minimum Euclidean distance from i^{th} pixel to a definite foreground pixel and to a definite background pixel respectively. and $k = 0.5$.
The final saliency map is obtained adding the color-based saliency map and the spatial saliency map:

$$S_{final}(X_i) = S_{LS}(X_i) + S_s(X_i), \quad i = 1, 2, \ldots, N \tag{6}$$

To speed up the refinement processes, saliency map refinements is pefromed in superpixel level using the mean color of a superpixel as a pixel color, and the center location of a superpixel as a pixel location. Visual examples of estimating step-by-step saliency maps are presented in Figure 1.

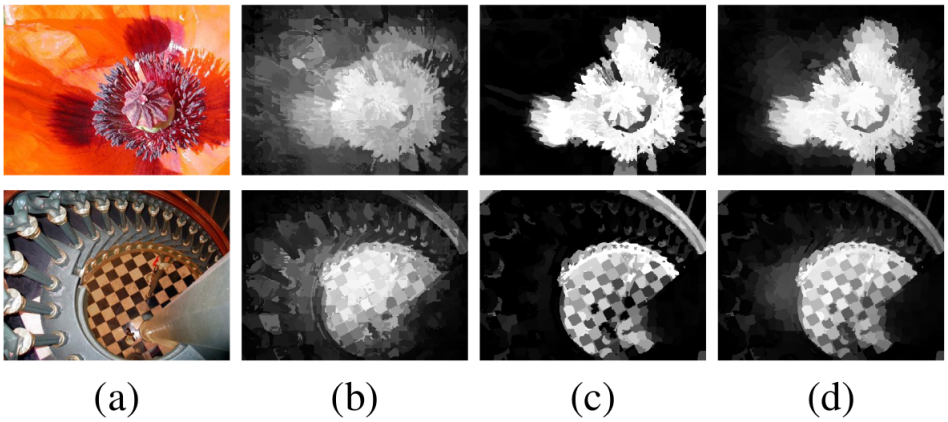

Figure 1. The visual examples of each step's result. (a) test images, (b) initial saliency map after Section 4, (c) refined saliency map S LS using high-dimensional color transform, and (d) our final saliency map after including spatial refinement.

3 Experiments

3.1 Quantitative comparison

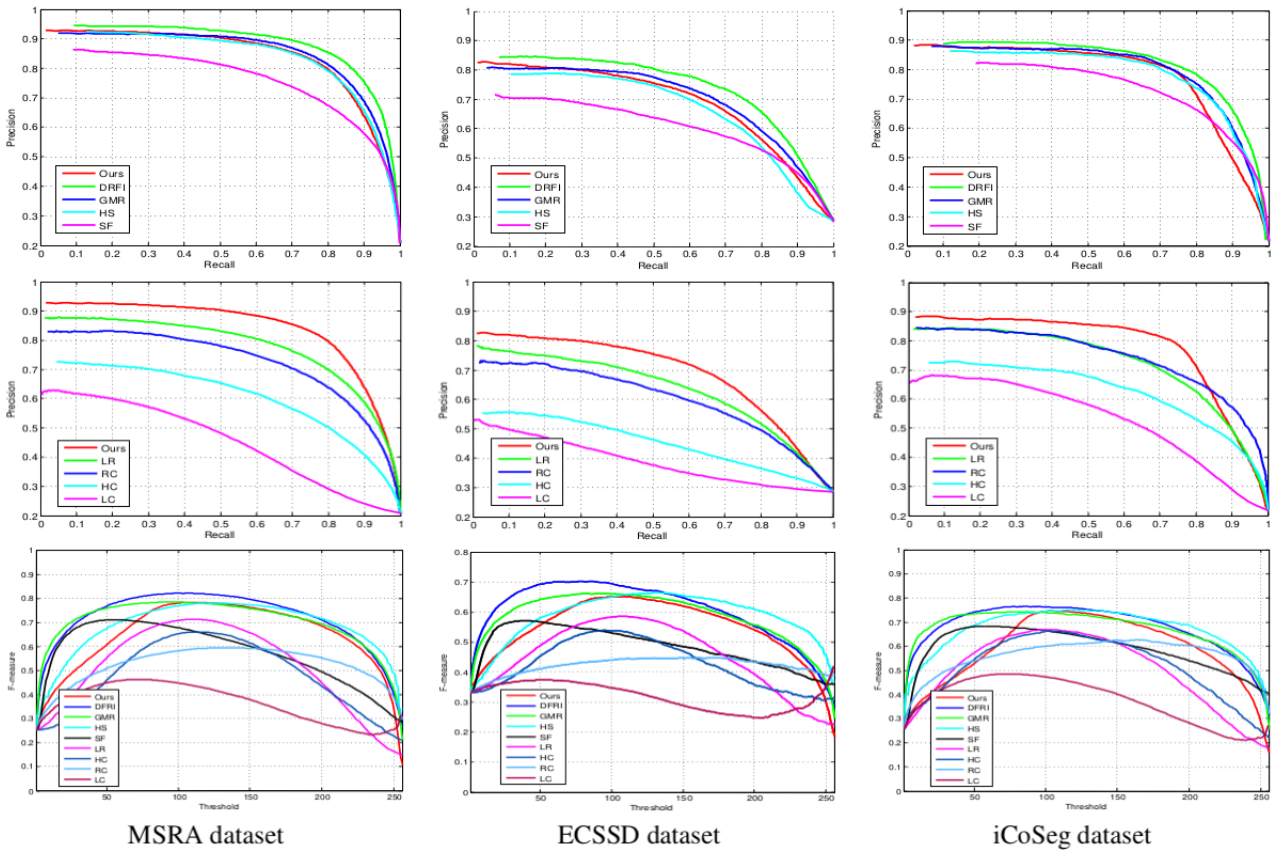

Figure 2. Comparison of the performance with eight state-of-the-art algorithms on three representative benchmark datasets: MSRA dataset, ECSSD dataset and iCoSeg dataset. The first and the second row is the PR curve, and the third row is the F-measure curve.

3.2 Qualitative comparison

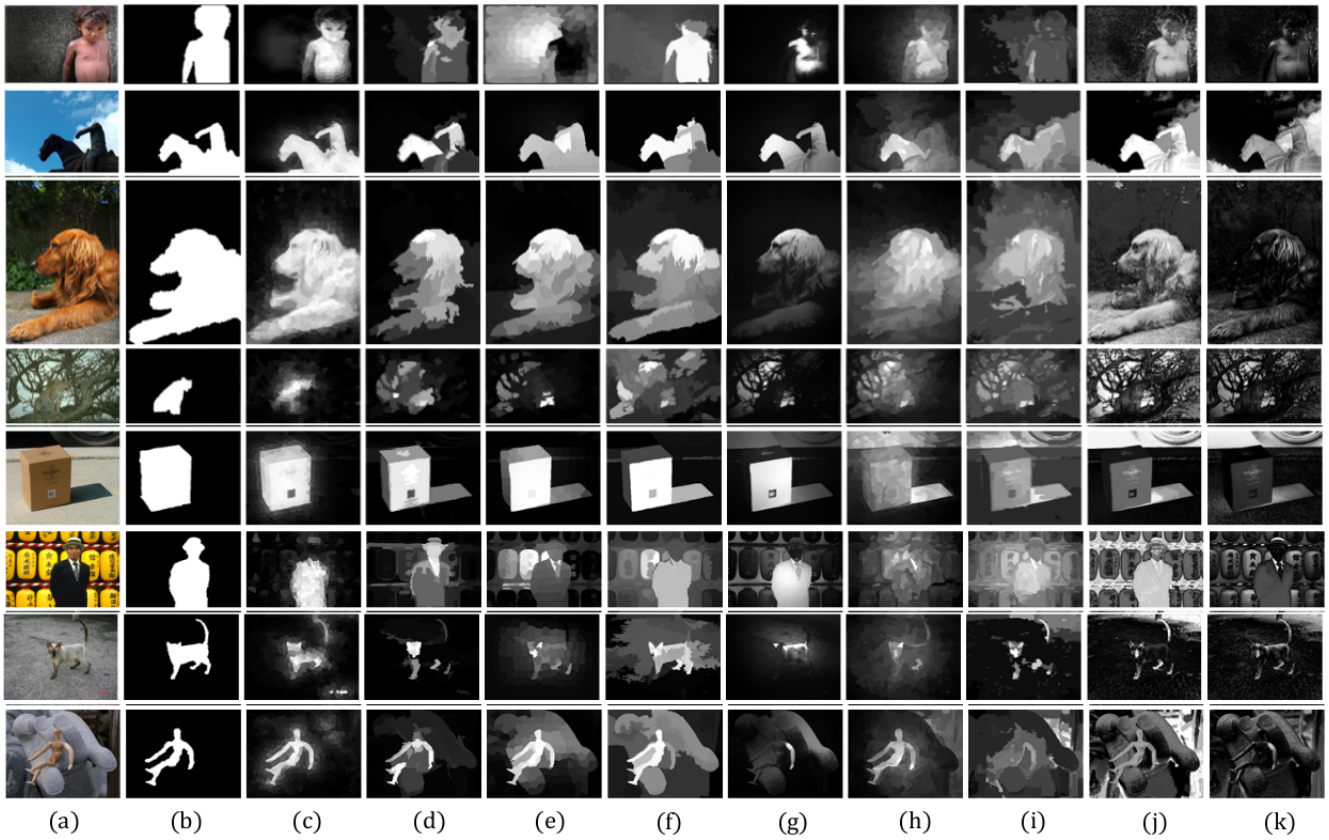

Figure 3. Visual comparisons of our results and results from previous methods. Each image denotes (a) test image, (b) ground truth, (c) our approach, (d) DRFI [1], (e) GMR [2], (f) HS [3], (g) SF [4], (h) LR [5], (i) RC [6], (j) HC [6], (k) LC [7].

References

[1] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2083–2090, 2013. ii, 6

[2] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166–3173, 2013. ii, 6

[3] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1155–1162, 2013. ii, 6

[4] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–740, 2012. ii, 6

[5] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 853–860, 2012. ii, 6

[6] 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. ii, 6

[7] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th ACM International Conference on Multimedia*, MM '06, (New York, NY, USA), p. 815–824, Association for Computing Machinery, 2006. ii, 6

[8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012. 1

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010. 2

[10] B. Su, S. Lu, and C. L. Tan, "Blurred image region detection and classification," in *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, (New York, NY, USA), p. 1397–1400, Association for Computing Machinery, 2011. 2

[11] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979. 2