Salient Region Detection via High-Dimensional Color Transform

Jiwhan Kim Dongyoon Han Yu-Wing Tai Junmo Kim
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology (KAIST)
{jhkim89, calintz}@kaist.ac.kr, yuwing@kaist.ac.kr, junmo@ee.kaist.ac.kr

Abstract

In this paper, we introduce a novel technique to automatically detect salient regions of an image via high-dimensional color transform. Our main idea is to represent a saliency map of an image as a linear combination of high-dimensional color space where salient regions and backgrounds can be distinctively separated. This is based on an observation that salient regions often have distinctive colors compared to the background in human perception, but human perception is often complicated and highly nonlinear. By mapping a low dimensional RGB color to a feature vector in a high-dimensional color space, we show that we can linearly separate the salient regions from the background by finding an optimal linear combination of color coefficients in the high-dimensional color space. Our high dimensional color space incorporates multiple color representations including RGB, CIELab, HSV and with gamma corrections to enrich its representative power. Our experimental results on three benchmark datasets show that our technique is effective, and it is computationally efficient in comparison to previous state-of-the-art techniques.

1. Introduction

Salient region detection is important in image understanding and analysis. Its goal is to detect salient regions, in terms of saliency map, from an image where the detected regions would draw the attentions of humans at the first sight of an image. As demonstrated in many previous works, salient region detection is useful in many applications including segmentation [20], object recognition [22], image retargetting [32], photo re-arrangement [25], image quality assessment [23], image thumbnailing [21], video compression [12], etc. The development of salient region detection is often inspired by the human visual perception concepts. One important concept is how "distinct to a certain extent" [9] the salient region is compared to other parts of an image. Since color is a very important visual cue to humans, many salient region detection techniques are built upon distinctive color detection from an image.

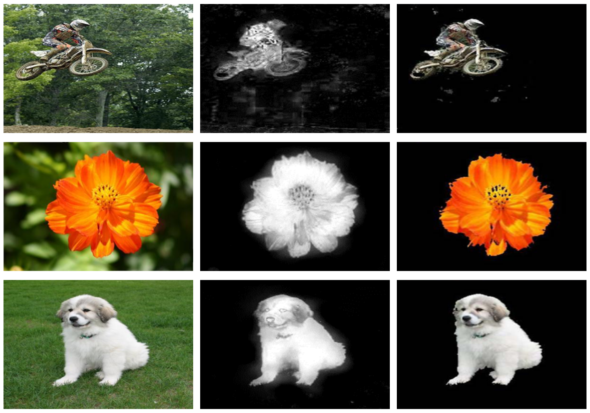

(a) Inputs (b) Saliency maps (c) Salient regions

Figure 1. Examples of our salient region detection.

In this paper, exploring the power of different color space representations, we propose high-dimensional color transform which maps a low dimensional RGB color tuple into a high-dimensional feature vector. Our high dimensional color transform combines several representative color spaces such as RGB, CIELab, HSV, together with different gamma corrections to enrich the representative power of our high-dimensional color transform space. Starting from a few initial color examples of detected salient regions and backgrounds, our technique estimates an optimal linear combination of color values in the high-dimensional color transform space that results in a per-pixel saliency map. As demonstrated in our experimental results, our per-pixel saliency map represents how distinctive the color of salient regions is compared to the color of the background. Note that a simple linear combination or transformation of the color space cannot achieve results similar to ours. Figure 1 shows examples of our detected saliency map and salient regions.

Assumptions Since our technique uses only color information to separate salient regions from the background, our technique shares a limitation when identically-colored objects are present in both the salient regions and the back-

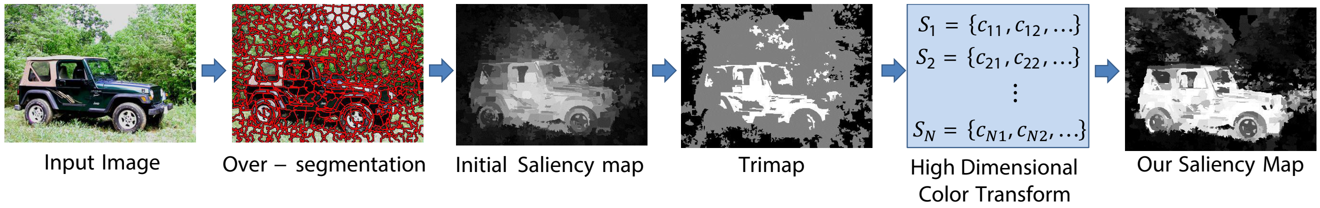

Input Image
Over – segmentation
Initial Saliency map
Trimap
$S_1 = \{c_{11}, c_{12}, \dots\}$
$S_2 = \{c_{21}, c_{22}, \dots\}$
⋮
$S_N = \{c_{N1}, c_{N2}, \dots\}$
High Dimensional Color Transform
Our Saliency Map

Figure 2. Overview of our method. Please refer to Section 3 for details.

ground. In such cases, utilizing high-level features, such as texture, is the only way to resolve this ambiguity. Nevertheless, we show that many salient regions can simply be detected using only color information via our high-dimensional color transform space, and we achieve high detection accuracy and better performance compared with many previous methods that utilizes multiple high-level features.

2. Related Works

Representative works in salient region detection are reviewed in this section. We refer readers to [30] for a more extensive comparative study of state-of-the-art salient region detection techniques.

There are many previous methods that utilize low-level features such as color and texture for saliency detection. Itti *et al.* [13] proposed a saliency detection method based on color contrast called “center-surround difference.” Harel *et al.* [11] suggested a graph-based visual saliency (GBVS) model based on the Markovian approach. Achanta *et al.* [1] viewed color and luminance in the frequency domain to compute an image's saliency. Goferman *et al.* [10] combined global and local contrast saliency to improve the detection performance. Klein and Frintrop [18] applied Kullback-Leibler divergence to the center-surround difference to combine different image features to compute saliency. Jiang *et al.* [14] performed salient object segmentation with multi-scale superpixel-based saliency and closed boundary prior. Perazzi *et al.* [26] used the uniqueness and distribution of the CIELab color to find the salient region. Shen and Wu [27] divided an image into a low-rank matrix and sparse noises to detect an object. Yan *et al.* [33] used a hierarchical model by computing contrast features of different scales of an image, and fused them into a single map using a graphical model.

Recently, statistical learning-based models have also been examined for saliency detection. Wang *et al.* [31] used a trained classifier called the auto-context model to enhance an appearance-based optimization framework for salient region detection. Jiang *et al.* [15] performed a regional saliency regressor by using regional descriptors. Borji and Itti [5] used patch-based dictionary learning for a rarity-based saliency model. Yang *et al.* [34] considered the foreground and background cues by using the graph-based method. Siva *et al.* [28] used an unsupervised approach to sample patches of an image which are salient by using patch features.

3. Overview

Figure 2 shows the overview of our method. First, we over-segment an image into super pixels and estimate an initial saliency map using existing saliency detection techniques. From the initial saliency map, we threshold it to obtain a trimap where pixel colors within the definite foreground and the definite background regions will be used as initial color samples of the salient regions and background. Then, we map the low-dimensional color into the high-dimensional color transform space and estimate an optimal linear combination of color channels subject to the color samples' constraints. Finally, we combine color values in the high-dimensional color space to obtain our saliency map.

4. Initial Salient Regions Detection

Superpixel Saliency Features As demonstrated in recent works [15, 26, 27, 34], features from superpixels are effective and efficient for salient object detection. For an input image I, we first perform an over-segmentation to form superpixels $\mathbf{X} = \{X_1, ..., X_N\}$. We use the SLIC superpixel [2] because of its low computational cost and high performance and we set the number of superpixels at $N = 500$.

To build feature vectors for saliency detection, we combine multiple information that are commonly used in saliency detection. We first concatenate superpixels' x- and y-locations into our feature vector. The location feature is used because humans tend to pay more attention to objects that are located around the center of an image [16]. Second, we concatenate the color features as this is one of the most important cues in the human visual system and certain colors tend to draw more attention than the others [27]. We compute the average pixel color and represent the color features using different color space representations.

Next, we concatenate the histogram feature since this is one of the most effective measurements for the saliency feature as demonstrated in [15]. The histogram feature is measured by using the chi-square distance between histograms. 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

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

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

is the number of histogram bins and we use eight bins for each histogram in our paper.

We have also used the global contrast and the local contrast as color features [1, 7, 26]. 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 $D_{L_i} = \sum_{j=1}^{N} \omega^{p}_{i,j} d(c_i, c_j)$ where $\omega^{p}_{i,j} = \frac{1}{Z_i} \exp(-\frac{1}{2\sigma_p^2} \|\mathbf{p}_i - \mathbf{p}_j\|_2^2)$, in which $\mathbf{p}_i$ denotes the position of i^{th} superpixel and Z_i is the normalization term. In our experiments, we use $\sigma_p^2 = 0.25$. In addition to the global and local contrast, we further evaluate the element distribution [26] by measuring the compactness of colors in term of their spatial color variance.

For texture and shape features, we utilize the area of superpixel, histogram of gradients (HOG) and the singular value feature. The HOG provides appearance features by using around the pixels' gradient information at fast speed. We use the HOG features implemented by Felzenszwalb *et al.* [8] which has 31 dimensions. The Singular Value Feature (SVF) [29] is used to detect the blurred region from a test image, because a blurred region often tends to be a background. The SVF is a feature based on eigenimages [3] which decompose an image by a weighted summation of a number of eigen-images, where each weight is the singular value obtained by SVD. The eigen-images corresponding to the largest singular values determine the overall outline of the original image and other smaller singular values depict detail information. Hence, some of the largest singular values occupy much higher weights for blurred images.

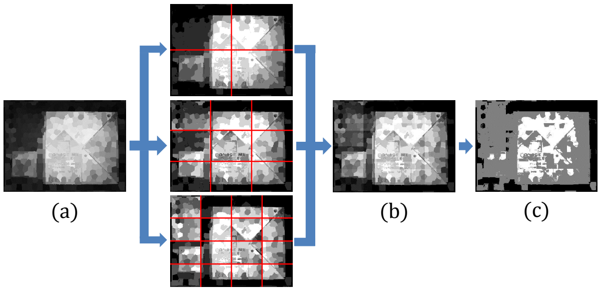


Figure 3. Our trimap construction processes. (a) Initial saliency map. We divide the initial saliency map into 2 × 2, 3 × 3, and 4 × 4 regions and apply adaptive thresholding algorithm for each region individually. After that, we sum up the thresholded saliency map to obtain a new saliency map in (b). This saliency map is then thresholded globally to obtain our trimap in (c).

The aforementioned features are concatenated and will be used to estimate our initial saliency map. Table 1 summarizes the features that we have used. In short, our superpixel feature vectors consist of 75 dimensions which combine multiple evaluation matrixes for saliency detection.

Initial Saliency Map Estimation via Regression After we calculate the feature vectors for every superpixels, we use a regression algorithm to estimate each region's degree to be salient. In this work, we use the random forest [6] regressor, because of its efficiency on large databases and generalization ability. We use 2,000 images from the MSRA-B dataset [19] which are selected as a training set from Jiang *et al.* [15] for training data, and we use annotated ground truth images for labels. We generate N feature vectors for each image, so that we train about one million vectors for the training data. We use 200 trees, with no limit for the node size. A visual example of an initial map is shown at Figure 2.

5. High-Dimensional Color Transform for Saliency Detection

In this section, we present our high-dimensional color transform and describe a step-by-step process to obtain our final saliency map starting from the initial saliency map from the previous section.

Trimap Construction The initial saliency map usually does not detect salient objects accurately and may contain many ambiguous regions. This trimap construction step is to identify very salient pixels from the initial saliency map that definitely belong to salient regions and backgrounds, and use our high-dimensional color transform method to resort the ambiguities in the unknown regions. In order to

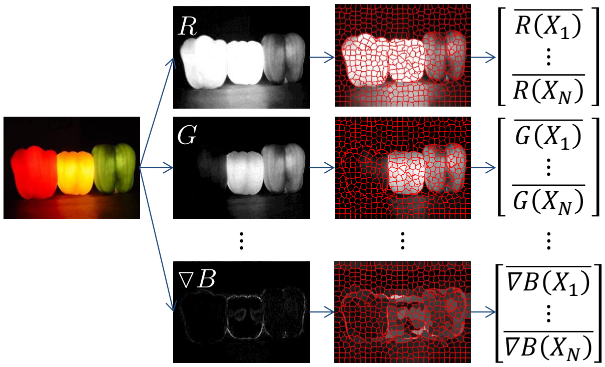


Figure 4. Our high-dimensional color transform space. We concatenate different nonlinear RGB transformed color space representations to form a high-dimensional feature vector to represent the color of a pixel.

catch the salient pixels more accurately, instead of using a single global threshold to obtain our trimap from the initial saliency map, we propose using a multi-scale analysis with adaptive thresholding. Our trimap construction process is described in Figure 3. First, we divide the initial map into 2×2, 3×3, and 4×4 regions and apply thresholding for each region individually. We apply Otsu's multi-level adaptive thresholding [24] to control the rate between the foreground, background and unknown regions in each subregion. In our experiments, we use a seven-level threshold in each subregion. After merging the three different scale thresholded saliency maps by summation, we obtain a locally thresholded 21-level map T'. This new saliency map has better local contrast than the initial saliency map. Therefore, it is able to locally capture very salient regions even though the local region might not be the most salient globally within the whole image. Finally, we obtain the trimap by global thresholding as follows:

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

High-Dimensional Color Transform Colors are important visual cues to our human visual system. Many previous studies [17] have discussed that the RGB color space does not fully correspond to the space where the human brain processes colors. It is also inconvenient to process colors in the RGB space since illumination and colors are nested here. For these reasons, many different color spaces have been introduced such as YUV, YIQ, CIELab, HSV, etc. Nevertheless, it is still unknown which color space is the best to process images, especially for applications like saliency detection that are tightly correlated to our human perception. Instead of picking a particular color space for processing, we introduce a high-dimensional color transform which unifies the strength of many different color representations. Our goal is to find a linear combination of color coefficients in the high dimensional color transform space such that colors of salient regions and colors of backgrounds can be distinctively separated.

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

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

To build our high-dimensional color transform space, we concatenate different *nonlinear* RGB transformed color space representations as illustrated in Figure 4. We concatenate only the nonlinear RGB transformed color space because the effects of the coefficients of linear transformed color space such as YUV/YIQ, will be cancelled when we linearly combine the color coefficient to form our saliency map. The color spaces we concatenated include the CIELab color space, and the hue and saturation channel in the HSV color space. We also include color gradients in the RGB space since our human perception is more sensitive to relative color differences instead of absolute color values. The different magnitudes in the color gradients can also handle cases when salient regions and backgrounds have different amount of defocus and different color contrast. In summary, 11 different color channel representations are used in our high-dimensional color transform space.

To further enrich the representative power of our high-dimensional color transform space, we apply gamma corrections to each of the color coefficients after normalizing the coefficient between $[0, 1]$. The gamma values we used range from 0.5 to 2 for each 0.5 interval. This results in a $l = 44$ high-dimensional vector to represent the colors of an image:

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

The nonlinear gamma correction takes into account that our human perception responds nonlinearly to incoming illumination. It also stretches/compresses the intensity contrast within different ranges of color coefficients. Table 2 summarizes the color coefficients concatenated in our high-dimensional color transform space. This process is applied to each superpixel in an input image individually.

A self-comparison of our high-dimensional color transform with other combinations of color channels is shown in Figure 7. The result shows that the performance is undesirable when only RGB is used, and using various nonlinear RGB transformed color spaces and gamma corrections helps to catch the salient regions more accurately.

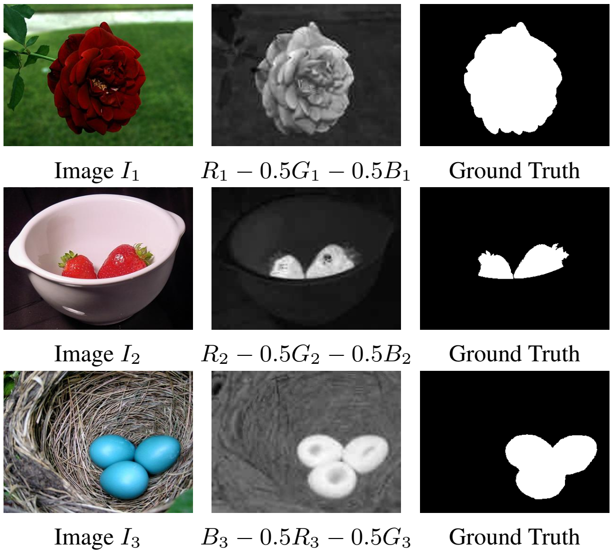


Figure 5. Illustrations of linear coefficient combinations for saliency map construction. (a) Input original images, (b) saliency maps are obtained by using a linear combination of RGB channels, and (c) Ground truth saliency map.

Saliency Map Construction via Optimal Linear Combination of Coefficients To obtain our saliency map, we utilize the definite foreground and definite background color samples in our trimap to estimate an optimal linear combination of color coefficients to separate the salient region color and the background color. We formulate this problem as a least square problem which minimizes:

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

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. Since we have a greater number of color samples than the dimensions of the coefficient vector, this least square problem is a well condition problem which can be solved easily using standard linear equation solvers.

After we get the optimal coefficient α, we can construct the saliency map as:

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

which denotes the linear combination of the color coefficient of our high-dimensional color transform space. Since the initial color samples may be limited, we repeat the processes described in this section to identify more reliable color samples and to further separate the color distance between the salient region and the background. In our experiments, we found that three iterations are sufficient to converge to a stable accurate saliency map. Figure 5 illustrates the idea of using linear combination of color coefficient for saliency detection.

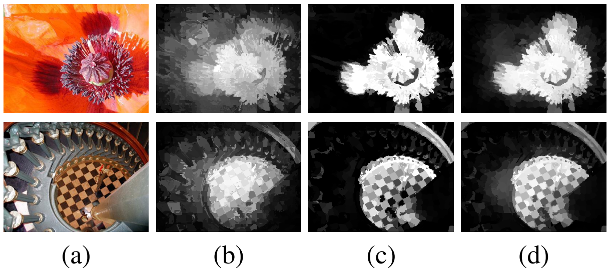


Figure 6. The visual examples of each step's result. (a) test images, (b) initial saliency map after Section 4, (c) refined saliency map S_{LS} using high-dimensional color transform, and (d) our final saliency map after including spatial refinement.

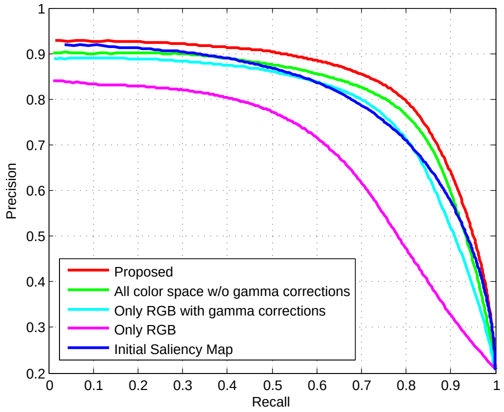


Figure 7. Comparison of precision-recall curves of our initial map and results based on different color transforms on the MSRA dataset.

Spatial Refinement In our final step, we further refine our saliency map using spatial information. The idea is to give more weight to pixels which are closer to the definite foreground region 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. In our experiment, we set the parameter $k = 0.5$. The final saliency map is obtained adding the color-based saliency map and the spatial saliency map:

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

Visual examples of our estimated step-by-step saliency maps are presented in Figure 6. To speed up our refinement processes,we perform saliency map refinements 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.

6. Experiments

We evaluate and compare the performances of our algorithm against previous algorithms on three representative benchmark datasets: the MSRA salient object dataset [19], the Extended Complex Scene Saliency Dataset (ECCSD) [33], and the Interactive cosegmentation Dataset (iCoSeg) [4]. The MSRA dataset contains 5,000 images with the pixel-wise ground truth by the authors provided by Jiang *et al.* [15]. This dataset contains comparatively obvious salient objects on the simple background and is considered as a less challenging dataset in saliency detection. The ECCSD dataset contains 1,000 images with multiple objects, which makes the detection tasks much more challenging. Unlike the MSRA dataset, test images in this dataset are more like real world images. So, this dataset can test the generalization ability of the salient region detection methods. Finally, the iCoSeg contains 643 images with multiple objects in a single image. This dataset is quite interesting because it contains many people which are relatively hard to detect as salient objects.

We compare eight state-of-the-art methods according to the evaluation criteria suggested by Achanta *et al.* [1]. The first evaluation compares the precision and recall rates. The first and second row of Figure 8 show the precision-recall curves for comparing our saliency method with the aforementioned state-of-the-art saliency detection methods, including those of Zhai *et al.* (LC) [35], Cheng *et al.* (HC,RC) [7], Shen and Wu (LR) [27], Perazzi *et al.*(SF) [26], Yan *et al.*(HS) [33], Yang *et al.*(GMR) [34], and Jiang *et al.* (DRFI) [15] for the datasets.

The second evaluation compares the F-measure rate. We compute the F-measure rates for the binarized saliency map as the threshold changes over the range $[0, 255]$, where the F-measure rate is given by

$$F_\beta = \frac{(1+\beta^2)\cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}. \quad (7)$$

As in previous methods [1, 27, 33], we use $\beta^2 = 0.3$. The third row of Figure 8 shows the F-measure curve for each of state-of-the-art methods.

The average run times of three state-of-the-art methods are compared in Table. 3. The run time is measured at a machine with an Intel Dual Core i5-2500K 3.30 GHz CPU. Considering that our method is implemented by using MATLAB with unoptimized code, the computational complexity of the proposed method is comparable to those of HS and GMR. Most of the time in our method is at the superpixel generation step (about 0.94s) and feature vector generation step (about 1.8s). Note that these steps are for initial saliency map estimation.

Method	Ours	DRFI[15]	HS[33]	GMR[34]
Time(s)	3.32	29.27	0.43	3.37
Code	Matlab	Matlab	C++	C++

Table 3. Comparison of average run time (seconds per image).

From the experiments' results, we find that our algorithm is effective and computationally efficient. Although our performance does not outperform the method of Yang *et al.* [34] and Jiang *et al.* [15], our algorithm's computational speed is much faster. Some visual examples of salient object detection on the MSRA dataset are presented in Figure 9 which demonstrate effectiveness of the proposed method.

7. Conclusions

We have presented a high-dimensional color transform-based salient region detection, which estimates foreground regions by using the linear combination of various color space. The trimap-based robust estimation overcomes limitations of inaccurate initial saliency map. As a result, our method achieves a fine performance and is computationally efficient in comparison to the other state-of-the art methods.

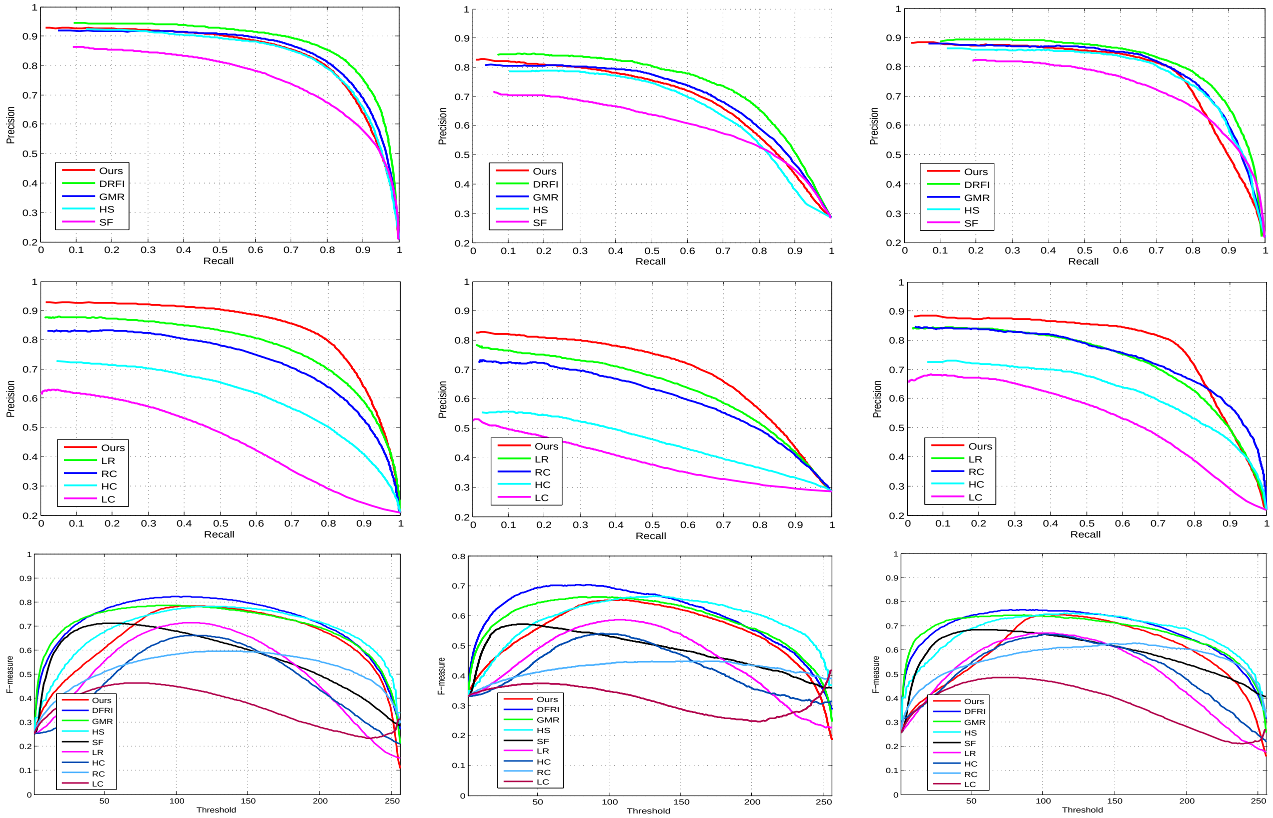
We note that our high dimensional color transform might not fully coincide with the human vision. However, it is effective in increasing the success of foreground and background color separation since the low dimensional RGB space is very dense where distributions of foreground and background colors are largely overlapped. We also note that if identical colors appear in both foreground and background or the initialization of color seed estimation is very wrong, our result is undesirable. In future, we plan to use more features to solve these limitations and improve the accuracy of saliency detection.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028680), and the Technology Innovation Program, 10045252, Development of robot task intelligence technology, funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea). Yu-Wing Tai was supported by the Basic Science Research Program through the NRF of Korea (NRF: 2011-0013349).

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009. 2, 3, 6



MSRA dataset ECSSD dataset iCoSeg dataset

Figure 8. 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.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE PAMI*, 34(11):2274–2282, 2012. 2

[3] H. C. Andrews and C. L. Patterson. Singular value decompositions and digital image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24:26–53, 1976. 3

[4] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *CVPR*, 2010. 6

[5] A. Borji and L. Itti. Exploiting local and global patch rarities for saliency detection. In *CVPR*, pages 478–485, 2012. 2

[6] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. 3

[7] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In *CVPR*, pages 409–416, 2011. 3, 6, 8

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE PAMI*, 32(9), 2010. 3

[9] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun. Salient object detection by composition. In *ICCV*, pages 1028–1035, 2011. 1

[10] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *CVPR*, pages 2376–2383, 2010. 2

[11] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, volume 19, pages 545–552, 2006. 2

[12] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE TIP*, 13(10):1304–1318, 2004. 1

[13] L. Itti, J. Braun, D. Lee, and C. Koch. Attentional modulation of human pattern discrimination psychophysics reproduced by a quantitative model. In *NIPS*, volume 2, pages 789–795, 1998. 2

[14] H. Jiang, J. Wang, Z. Yuan, T. Liu, and N. Zheng. Automatic salient object segmentation based on context and shape prior. In *Proceedings of the British Machine Vision Conference*, pages 110.1–110.12. BMVA Press, 2011. 2

[15] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. In *CVPR*, 2013. 2, 3, 6, 8

[16] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *ICCV*, pages 2106 –2113, 2009. 2

[17] P. Kaiser and R. M. Boynton. *Human Color Vision*, volume 2nd edition. Optical Society of America (Washington, DC), 1996. 4

[18] D. Klein and S. Frintrop. Center-surround divergence of feature statistics for salient object detection. In *ICCV*, pages 2214–2219, 2011. 2

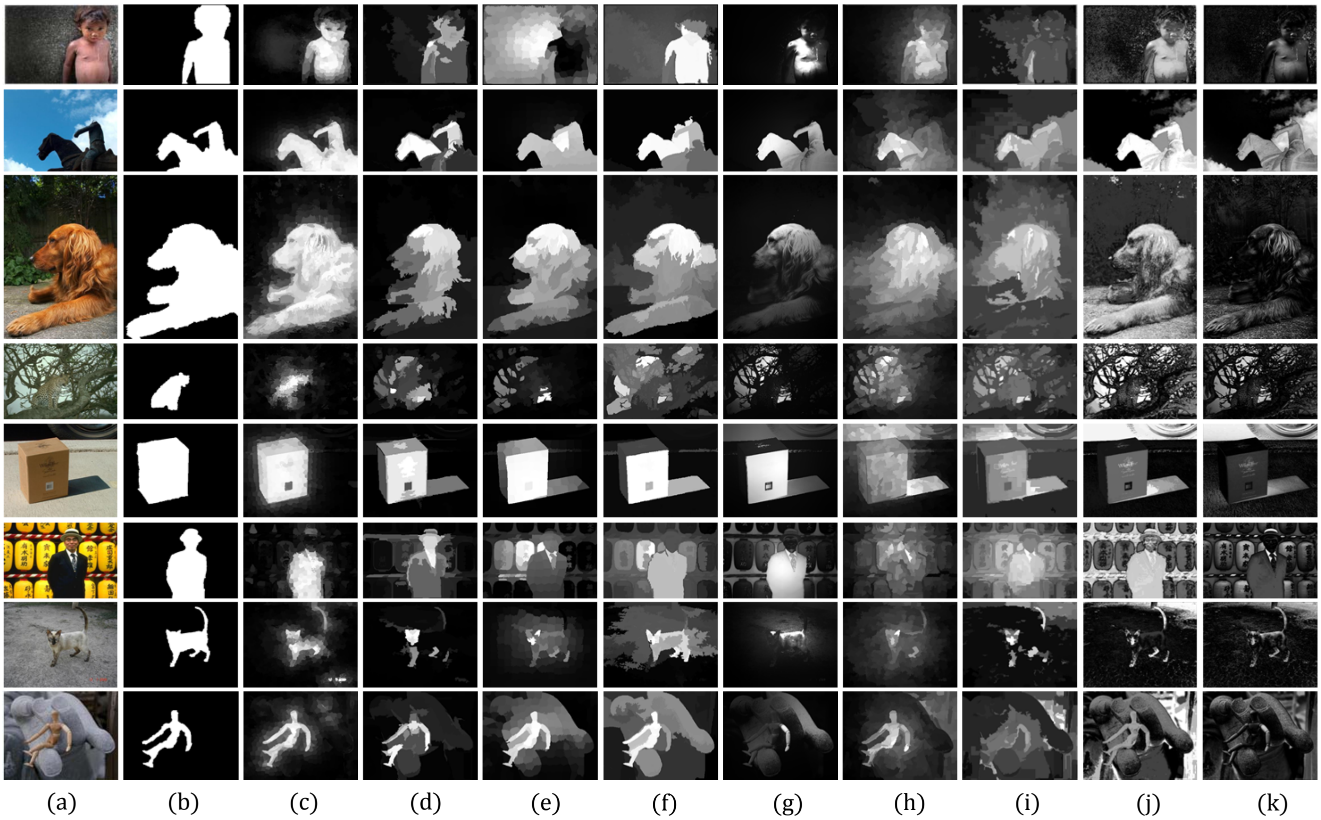


Figure 9. Visual comparisons of our results and results from previous methods. Each image denotes (a) test image, (b) ground truth, (c) our approach, (d) DRFI [15], (e) GMR [34], (f) HS [33], (g) SF [26], (h) LR [27], (i) RC [7], (j) HC [7], (k) LC [35].

[19] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum. Learning to detect a salient object. In *CVPR*, pages 1–8, 2007. 3, 6

[20] Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang. Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. *IEEE Transactions on Multimedia*, 14(4):1275–1289, 2012. 1

[21] L. Marchesotti, C. Cifarelli, and G. Csurka. A framework for visual saliency detection with applications to image thumbnailing. In *ICCV*, pages 2232–2239, 2009. 1

[22] V. Navalpakkam and L. Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. In *CVPR*, volume 2, pages 2049–2056, 2006. 1

[23] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barbba. Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric. In *ICCV*, volume 2, pages 169–172, 2007. 1

[24] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. 4

[25] J. Park, J.-Y. Lee, Y.-W. Tai, and I. Kweon. Modeling photo composition and its application to photo re-arrangement. In *ICIP*, 2012. 1

[26] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012. 2, 3, 6, 8

[27] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, pages 853–860, 2012. 2, 6, 8

[28] P. Siva, C. Russell, T. Xiang, and L. Agapito. Looking beyond the image: Unsupervised learning for object saliency and detection. In *CVPR*, 2013. 2

[29] B. Su, S. Lu, and C. L. Tan. Blurred image region detection and classification. In *ACM International Conference on Multimedia*, pages 1397–1400, 2011. 3

[30] A. Toet. Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. *IEEE PAMI*, 33(11):2131–2146, 2011. 2

[31] L. Wang, J. Xue, N. Zheng, and G. Hua. Automatic salient object extraction with contextual cue. In *ICCV*, pages 105–112, 2011. 2

[32] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.*, 27(5), 2008. 1

[33] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *CVPR*, 2013. 2, 6, 8

[34] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013. 2, 6, 8

[35] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, pages 815–824, 2006. 6, 8