Saliency Detection Using DCT Coefficients and Superpixel-Based Segmentation

Chi-Yu Hsu and Jian-Jiun Ding

Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan
r00942039@ntu.edu.tw, djj@cc.ee.ntu.edu.tw

Abstract. The salient region is the area of an image that attracts the attention of viewers. In this paper, a very effective saliency detection algorithm is proposed. Our algorithm is mainly based on two new techniques. First, the discrete cosine transform (DCT) is used for constructing the block-wise saliency map. Then, the superpixel-based segmentation is applied. Since DCT coefficients can reflect the color features of each block in the frequency domain and superpixels can well preserve object boundaries, with the two techniques, the performance of saliency detection can be significantly improved. The simulations performed on a database of 1000 images with human-marked ground truths show that our proposed method can extract the salient region very accurately and outperforms all of the existing saliency detection methods.

Keywords: saliency detection, saliency map, image segmentation, computer vision.

1 Introduction

Salient regions are parts of an image that a person pays more attention to. It can be used in a number of content-based image processing applications, such as adaptive image compression [1], similarity measurement, image retrieval [2], image segmentation [3][4], object recognition [5], graph cut [6], image resizing [7-9], etc. For example, when compressing an image adaptively, a large quantization step can be used in the non-salient region. When measuring the similarity of two images or two objects, more weights can be assigned to the salient region. In image resizing, saliency detection can be adopted to extract the low energy part of an image. In these applications, saliency map detection plays a very important role.

Based on the regional concept, we propose a framework that uses boundary scoring and the border measurement to construct the saliency map. Boundary scoring gives a region a higher saliency weight if it has a higher contrast in the boundary part. The border measurement counts the number of image border pixels in a region, and gives a higher saliency weight if the region has fewer image border pixels. Before applying them, the input image is segmented using our proposed superpixel-based segmentation method. We use different parameters to produce the segmentation results

B. Huet et al. (Eds.): PCM 2013, LNCS 8294, pp. 122–133, 2013.

with coarse and fine scales, for achieving the best performance in the two methods. In addition, a block-wise saliency map is generated using the DCT-based context aware saliency detection method. The flow chart of our framework is shown in Fig. 1.

This paper is organized as follows. In Section 2, the recent works of saliency detection are described. In Section 3, we present the three proposed techniques: (1) the DCT-based context aware saliency detection method, (2) the superpixel-based segmentation algorithm, and (3) precision-enhanced integration. In Section 4, several simulations are performed to compare the accuracies of the saliency map generated by our method and that of eleven state-of-the-art methods on the database provided by [11] that consists of 1000 images. The simulations show that our proposed approach outperforms all of the existing saliency detection methods.

2 Related Work

There are two types of computational models for saliency detection. The first one is the top-down saliency model, which uses high-level features based on the knowledge from neurosciences, biology, computer vision, etc. Which high-level feature is used is dependent on applications [10]. The second one is the bottom-up saliency model, which uses low-level stimulus such as intensity, color contrast, orientation, and motion. The bottom-up saliency model is more popular since it requires less computation time and is efficient in memory.

Most of the previous methods scale down the input image during saliency detection for the sake of efficient computation and generate spotlight saliency maps. However, the result has a lower resolution and does not well match object the boundaries. Spotlight saliency maps are useful for predicting eye fixation, but are not accurate enough for content-based applications, such as salient object segmentation and content-aware image retargeting. To overcome this problem, a pre-segmentation process is needed. In other words, saliency detection should be computed at the region-level instead of the pixel-level. In 2011, Cheng *et al.* [12] simply used a color contrast concept with graph-based segmentation [13] to compute saliency maps and achieved an excellent result with high precision and recall rates. Jiang *et al.* [14] adopted the similar concept and evaluated saliency scores in many scales to get more robust performance. In 2012, Perazzi *et al.* [15] segmented the input image into compact and homogeneous pieces and compute contrasts and saliency scores in a unified way using high-dimensional Gaussian filters. With their method, a very accurate pixel-wise saliency map can be obtained with linear complexity.

The context aware (CA) saliency detection method was proposed in [16]. Although its complexity is high since it computes the color differences between all blocks in three scales, it has very good performance especially on the boundaries of objects, since higher contrasts always occur in the region edges.

3 Proposed Methods

3.1 Block-Based Saliency Detection Using DCT Coefficient

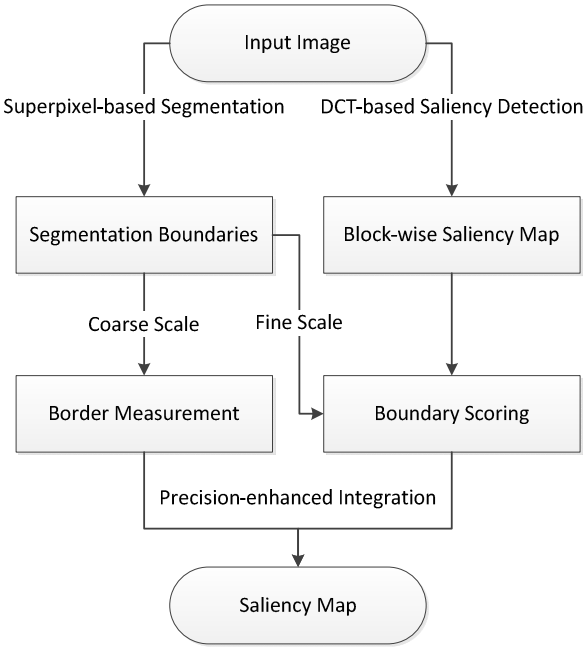


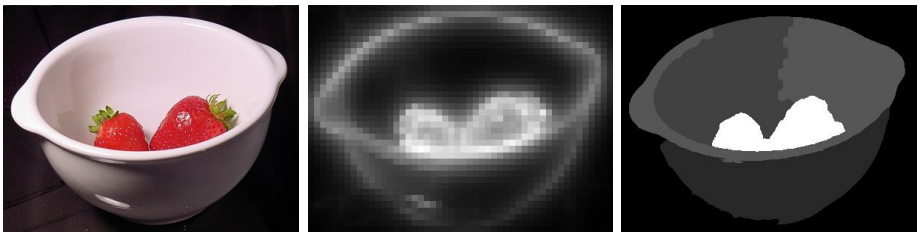
Fig. 1. Flowchart of the proposed saliency detection algorithm

The flowchart of the proposed saliency detection algorithm is shown in Fig. 1. As [16], our framework is also block-wise and uses the concept of context aware. However, the DCT and superpixels are applied to reduce the computation complexity and further improving the performance of saliency detection.

The image is first partitioned into 8 x 8 blocks. Then, the DCT is applied to each block. We adopt the DCT because it can well separate the low and the high frequency parts and the high frequency part. The high frequency part is always related to noise or tiny details and may worsen the performance of saliency detection.

We preserve only the DC term and the first five AC values for each color channel. The high frequency AC terms are ignored due to the considerations of computation efficiency and reducing the effect of noise. Then, the color features of each block are represented as a $(1+5) \times 3 = 18$-tuple vector.

Then, the color distance between two blocks i and j is defined as the Euclidean distance in the 18-D space $d_{color}(i, j)$, and the spatial distance $d_{position}(i, j)$ is defined as the Euclidean distance between the centers of blocks i and j. If there are K blocks in an image, then the DCT-based saliency score of each block is defined as follows. The term $\exp[-d_{position}(i, j)]$ is applied because the block j should has a smaller effect on computing $S_{DCT}(i)$ if the distance between blocks i and j is larger.



(a) (b) (c)

Fig. 2. (a) The input image. (b) The DCT-based saliency score of the input image. (c) The final saliency map constructed by the proposed algorithm (together with border measuring and boundary scoring).

$$S_{DCT}(i) = \frac{1}{K}\sum_{j=1}^{K}\left\{d_{color}(i,j)\times\exp\left[-d_{position}(i,j)\right]\right\} \tag{1}$$

In Fig. 2(b), the DCT-based saliency score (defined in (1)) of an image is shown. Larger intensity means a larger score. From Fig. 2(b), one can clearly see that the DCT-based saliency score is high in the two strawberries and the edges of the bowl. Therefore, DCT coefficients are indeed good features to conclude whether a region is a salient part of an image.

There is another advantage for using DCT coefficients in saliency detection. Since the DCT has a fast algorithm, the computation loading is much less than that of the original CA approach.

3.2 Superpixel Based Segmentation

In addition to DCT coefficients, we also adopt superpixels for saliency detection. A superpixel is a perceptually meaningful atomic region [17]. It is a combination of several pixels and the color and the intensity are consistent within a superpixel. An example of the superpixel can be seen from Fig. 3(b). Note that the edges of objects match the boundaries of several superpixels.

There are two advantages of using superpixels for saliency detection. First, since the number of superpixels is much less than that of pixels, the computation complexity can be reduced. Moreover, with superpixels, a better image segmentation result can be achieved, which is helpful for constructing a more accurate saliency map.

In the proposed algorithm, the input image into N SLIC superpixels using the method in [17]. Then, a superpixel graph is constructed to replace the rigid structure of the pixel grid. Then, the spectral clustering technique is used to cluster superpixels into C clusters or R non-split regions, as in Fig. 3(c). Finally, a region merging process, which is called boundary-focused region merging, is performed to merge smaller regions into L larger regions, as the example in Fig. 3(d).

For the spectral clustering process, we use the implementation of [18]. Each superpixel is represented as a 5-D vector in *Labxy* space. For the region merging process, we use a proposed boundary-focused region merging algorithm, which is to compute

the difference between adjacent regions and merge the adjacent regions whose difference is smaller than a threshold T. The difference measure between two adjacent regions is not computed by mean colors of regions. Instead, only the colors of the superpixels on the adjacent boundaries are computed.

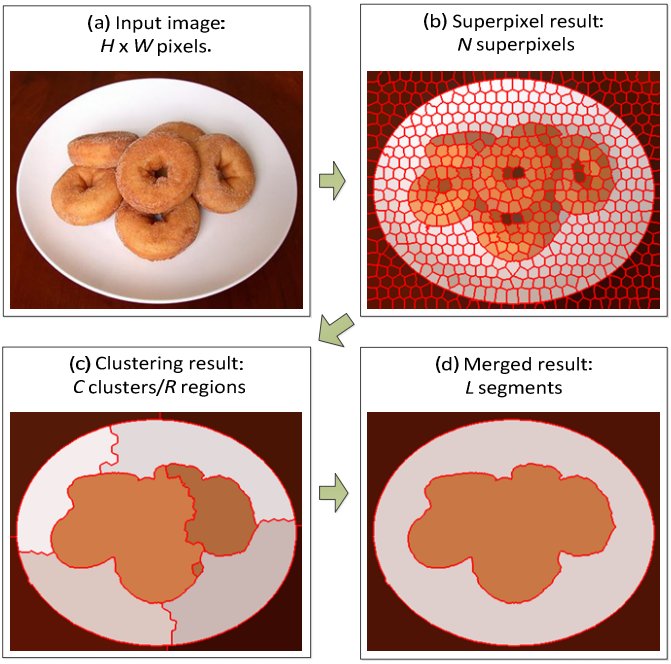


Fig. 3. Using superpixels to segment an image into several regions. (a) Original image. (b) The SLIC superpixels of the input image. (c) The segmentation result after spectral clustering in the 5-D space. (d) The segmentation result after boundary-focused region merging.

For two superpixels k and h belonging to different regions, their distance $E_{k,h}$ is defined as

$$E_{k,h} = \sqrt{(l_k - l_h)^2 + (a_k - a_h)^2 + (b_k - b_h)^2} \tag{2}$$

where (l_k, a_k, b_k) and (l_h, a_h, b_h) are the mean color values of superpixels k and h in the *Lab* color space. Then, the difference measure D between two adjacent regions R_i and R_j is defined as

$$D(R_i, R_j) = \frac{1}{|Adj(i,j)|} \sum_{k,h \in Adj(i,j)} E_{k,h} \tag{3}$$

where $Adj(i, j)$ is the set of distances between the superpixels on the adjacent boundaries of R_i and R_j, and $|Adj(i, j)|$ is the number of $Adj(i, j)$. It means that, when determining whether two regions should be merged, we consider only the color difference

of the superpixels on the adjacent boundaries. After applying superpixel-based segmentation with different clustering numbers, we obtain a coarse scale segmentation result and a fine scale segmentation result. Then, the border measurement is applied to the coarse scale segmentation result and the boundary scoring is applied to the fine scale segmentation result, as in Fig. 1.

3.3 Border Measurement and Boundary Scoring

As most state-of-the-art saliency detection methods, we make an assumption that the regions in the center of an image are more important than those near the image border. Therefore, after performing image segmentation using superpixels, we calculate the value of *border*(*i*) for each region of the coarse scale segmentation result:

$$border(i) = \frac{B_1(i)}{2(H \times W)}$$

where $B_1(i)$ means that in region i there are $B_1(i)$ pixels on the image border. H and W are the height and the width of the image. Then, the *border measurement based saliency value* is defined as:

$$S_{BM}(i) = \exp(-\sqrt{2} \cdot border(i)) . \tag{4}$$

The exponential function is applied because the value of $S_{BM}(i)$ should be smaller if the boundary of a region highly overlaps with the image border.

Then, we calculate boundary scoring for each region of the fine scale segmentation result. First, the DCT-based saliency score in (1) is converted into the following form:

$$S_1(m,n) = S_{DCT}(j) \tag{5}$$

if the pixel (m, n) is in the j^{th} 8×8 block of the image. Then, the boundary scoring of each region is defined as the average value of $S_1[m, n]$ on the region boundary. That is, for region i, the *boundary scoring based saliency value* is determined from

$$S_{BS}(i) = \frac{1}{|B_i|} \sum_{(m,n) \in B_i} S_1(m,n) \tag{6}$$

where B_i denotes the boundary of region i and $|B_i|$ is the number of pixels of B_i. It means that a region is more salient if it has higher DCT-based saliency scores on its boundary.

3.4 Precision-Enhanced Integration

From the processes in Sections 3.1, 3.2, and 3.3, two saliency values are calculated in our framework. One is the border measurement based saliency value, S_{BM} (defined in (4)). The other one is the boundary scoring based saliency value, S_{BS} (defined in (6)).

Then, we use a precision-enhanced integration method to define the final saliency map S from the two saliency values:

$$S = N\left[S_{BS} + S_{BS} \times S_{BM}\right] \tag{7}$$

where N is a normalizing factor used for making the value of S in the range of [0, 1]. Note that, in (7), S_{BS} (related to DCT coefficients) and S_{BM} (related to border measurement) are both used for determining the saliency score and S_{BS} plays a more important role than S_{BM}, since S_{BS} has higher precision when only one saliency value is used. We first choose S_{BS} as a basic map. Then, the product of S_{BS} and S_{BM} is used to enhance the intersection area of two saliency maps. Theoretically, applying the intersection can reduce the area of salient regions but increase accuracy, since the probability that the values of S_{BS} and S_{BM} for a non-salient part are both high is very low.

4 Simulation Results

4.1 Database

In our simulations, we used the publicly available database provided by Achanta *et al*. [11], which consists of 1000 images from the MSRA dataset together with the ground truth for each image. The ground truths are binary masks obtained by drawing the contour of the salient object manually. This database is widely used in saliency detection simulations, since the number of images is sufficiently large and well-defined human-marked ground truths are included.

4.2 Precision, Recall, and *F*-measure

The proposed algorithm is compared with 11 state-of-the-art saliency detection methods: Itti *et al* (IT) [19], fuzzy growing (MZ) [20], graph-based visual saliency (GBV) [21], spectral residual (SR) [22], Achanta *et al* (AC) [23], context-aware (CA) [16], frequency-tuned (FT) [11], Zhai *et al* (LC) [24], histogram contrast (HC), region contrast (RC) (both were proposed by Cheng *et al*. [12]), and the saliency filter (SF) [15]. We use precision and recall to measure the performance where

$$precision = \frac{TP}{N_1}, \qquad recall = \frac{TP}{N_2}, \tag{8}$$

N_1 is the number of pixels in the detected salient region, N_2 is the number of pixels in the salient region of the ground truth, and TP (true positive) is the number of pixels of the intersection of the detected salient region and the salient region in the ground truth. A higher precision rate means that fewer pixels in the non-salient part of the ground truth are misidentified to be in the salient region. A higher recall rate means that more pixels in the saliency part of the ground truth are correctly identified to be in the saliency part by the saliency detection algorithm.

A higher recall rate can be achieved using a lower threshold, but the precision is then reduced, and vice-versa. In order to generate a fair comparison result, we binarized the saliency maps at a threshold T_f where T_f was varied from 0 to 255.

Then, he precision-and-recall curve was generated by computing the precision rate and the recall rate for each T_f.

In Fig. 4, the precision-and-recall curves of the proposed algorithm and 11 existing saliency detection methods are shown. From Fig. 4, it can be seen clearly that our proposed algorithm outperforms other 11 methods. The result shows that the proposed algorithm indeed has a very good performance for detecting the salient regions of images.

Similarly to the works of [11][12][15], we also adopted a weighted harmonic mean measure, which is called the *F*-measure, to compare the performance. It computes the precision and recall rates based on a binarized saliency map and uses the image-dependent adaptive threshold T_a proposed by [11] for the thresholding procedure, which is defined as

$$T_a = \frac{2}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x, y) \tag{9}$$

where *W* and *H* are the width and height of an image, respectively. This value is twice of the mean saliency of the image. The *F*-measure is defined as

$$F_\beta = \frac{(1+\beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}. \tag{10}$$

As in [11][12][15], we set $\beta^2 = 0.3$.

In Fig. 5, we show the precision rates, the recall rates, and the *F*-measures of the proposed algorithm and 11 existing algorithms. We also use the 1000 image database provided by Achanta *et al*. [11]. From Fig. 5, the proposed algorithm still outperforms all of the existing methods when the *F*-measure is compared.

4.3 MAE and MSE

In [15], another comparison method was introduced. It evaluates the mean absolute error (MAE) between the continuous saliency maps *S* (the values range from 0 to 1) and the binary ground truth *GT* (the values are either 0 or 1) and is defined as

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left| S(x, y) - GT(x, y) \right| \tag{11}$$

where *W* and *H* are the width and height of an image, respectively.

Moreover, to emphasize the larger error case, we also used another well-known measurement, the mean square error (MSE), to measure the performance

$$MSE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left[S(x, y) - GT(x, y) \right]^2. \tag{12}$$

In Fig. 6, we show the MAEs and the MSEs of 11 existing saliency detection methods and the proposed algorithm (denoted by OURS). From Fig. 6, one can see that the MAE and the MSE of proposed algorithm are much lower than other methods. It proves that the proposed algorithm has very accurate saliency detection results.

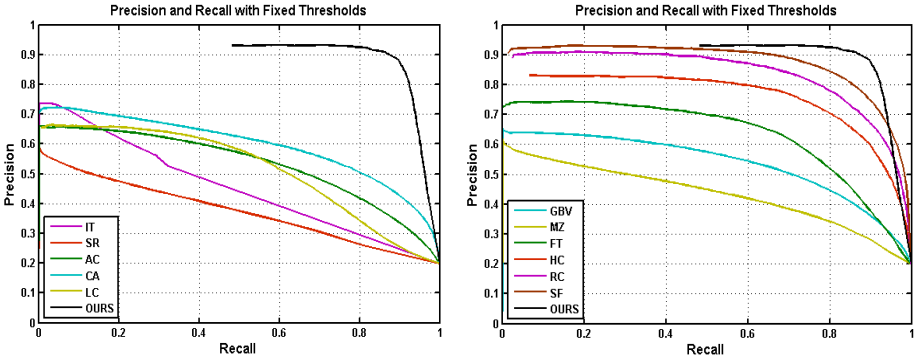


Fig. 4. Precision-and-recall curves of existing saliency detection methods and our proposed algorithm (OURS) when using the database provided by [11]. It can be seen that the proposed algorithm outperforms ALL of the existing methods for saliency detection.

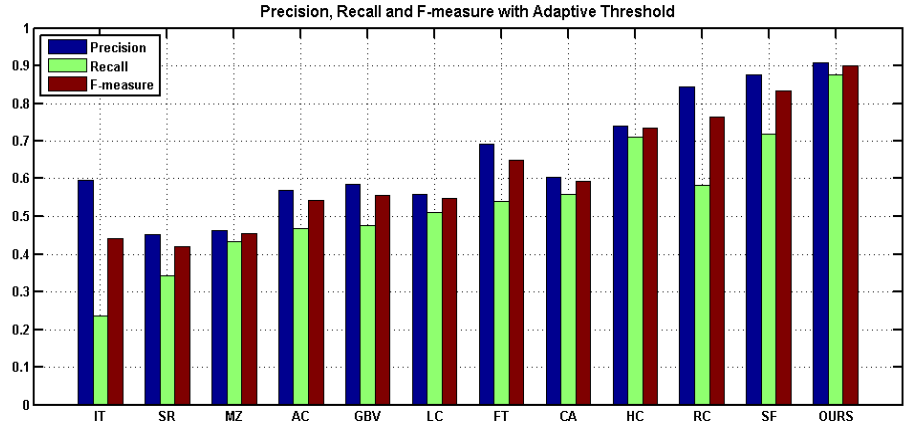


Fig. 5. Precisions, recalls, and F-measures for the existing saliency detection methods and the proposed algorithm (OURS) when using the image-dependent adaptive threshold T_a proposed by [11]

4.4 Analysis

The proposed saliency detection algorithm adopts two important concepts: color contrast and boundary information. With the DCT coefficient method proposed in Section 3.1, the color contrast can be measured in a more precise way. With the segmentation algorithm based on superpixels described in Section 3.2, the boundary information can be extracted in a more accurate way. Therefore, using the proposed techniques of DCT-based color contrast and superpixel-based segmentation, more accurate saliency detection results can be achieved.

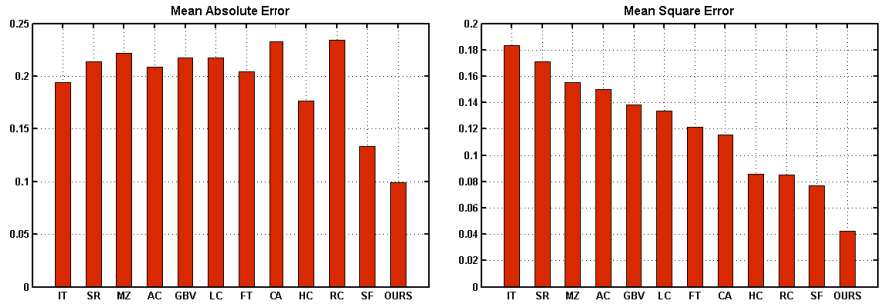
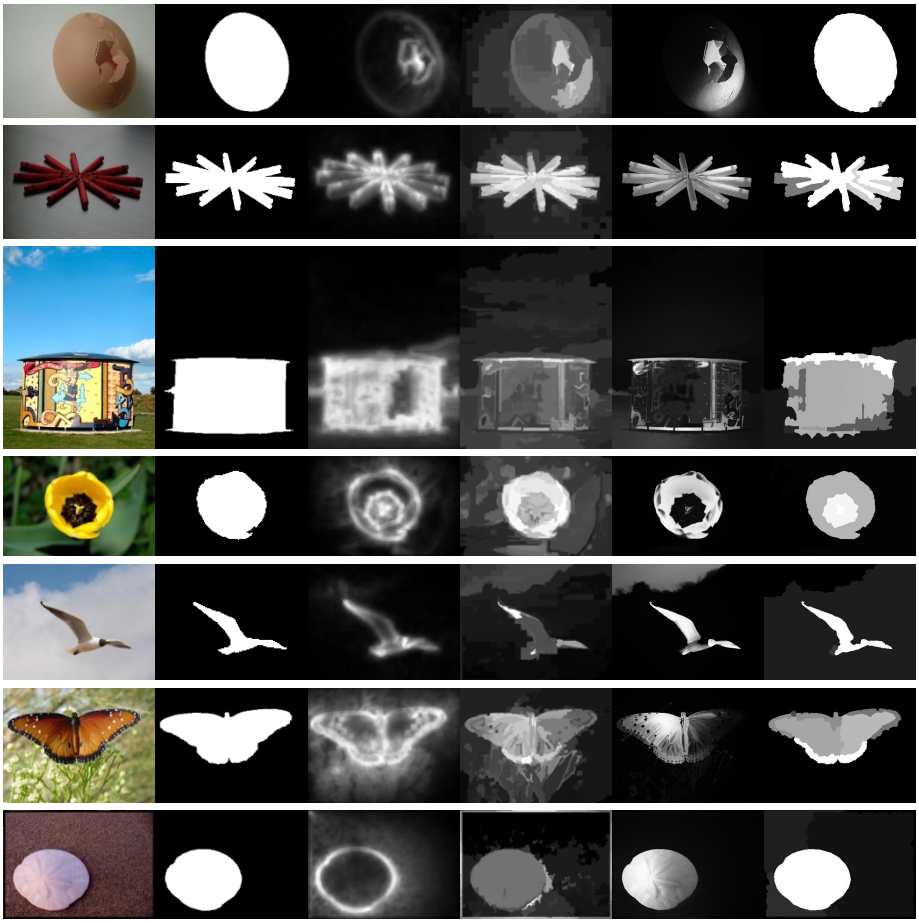


Fig. 6. Left: MAE of the existing saliency detection methods and our proposed algorithm (OURS). Right: MSE of the existing saliency detection methods and our proposed algorithm.



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

Fig. 7. Saliency detection examples for visual comparison. (a) Original images, (b) Ground truths, (c) Goferman *et al*. [16], (d) Cheng *et al*. [12], (e)Perazzi *et al*. [15], and (f) our method.

4.5 Visual Comparison

Visual comparisons of the saliency map detection results are shown in Fig. 7. Here, the methods for comparison are the CA method [16], the RC approach [12] and the SF method [15]. It can be seen that our proposed method performs even better than these state-of-the-art saliency map detection methods.

5 Conclusion

In this paper, a very accurate saliency detection framework is proposed. We adopt two novel techniques, DCT-based color contrast and superpixel-based segmentation. A block-wise saliency map is first generated using the DCT-based color contrast. Then, by employing the superpixel-based segmentation with the border measurement and boundary scoring, we obtain two saliency values with full-resolution. Finally, the precision-enhanced integration method is applied to calculate the saliency score.

We evaluated our proposed approach on the largest publicly available data set with a well-defined ground truth and compared our scheme with 11 state-of-the-art saliency detection methods. Simulation results show that the proposed algorithm achieves the best performance for saliency map generation in terms of the precision rate, the recall rate, the *F*-measure, the MAE, and the MSE.

References

1. Xue, J., Li, C., Zheng, N.: Proto-object Based Rate Control for JPEG2000: An Approach to Content-based Scalability. IEEE Trans. Image Processing 20(4), 1177–1184 (2011)
2. Chen, T., Cheng, M.M., Tan, P., Shamir, A., Hu, S.M.: Sketch2photo: Internet Image Montage. ACM Transactions on Graphics 28(5), article number 124 (2009)
3. Han, J., Ngan, K.N., Li, M., Zhang, H.J.: Unsupervised Extraction of Visual Attention Objects in Color Images. IEEE Transactions on Circuits and Systems for Video Technology 16(1), 141–145 (2006)
4. Chen, O.C., Chen, C.C.: Automatically-determined Region of Interest in JPEG 2000. IEEE Transactions on Multimedia 9(7), 1333–1345 (2007)
5. Rutishauser, U., Walther, D., Koch, C., Perona, P.: Is Bottom-up Attention Useful for Object Recognition? In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 37–44 (2004)
6. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-cut/ Max-flow Algorithms for Energy Minimization in Vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9), 1124–1137 (2004)
7. Avidan, S., Shamir, A.: Seam Carving for Content-aware Image Resizing. ACM Transactions on Graphics 26(3), article number 10 (2007)
8. Wang, Y.S., Tai, C.L., Sorkine, O., Lee, T.Y.: Optimized Scale-and-stretch for Image Resizing. ACM Transactions on Graphics 27(5), article number 118 (2008)
9. Wu, H., Wang, Y.S., Feng, K.C., Wong, T.T., Lee, T.Y., Heng, P.A.: Resizing by Symmetry-summarization. ACM Transactions on Graphics 29(6), article number 159 (2010)
10. Itti, L., Koch, C.: Computational Modeling of Visual Attention. Nature Reviews Neuroscience 2(3), 194–203 (2001)

11. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned Salient Region Detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604 (2009)
12. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global Contrast Based Salient Region Detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–416 (2011)
13. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient Graph-based Image Segmentation. International Journal of Computer Vision 59(2), 167–181 (2004)
14. Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., Li, S.: Automatic Salient Object Segmentation Based on Context and Shape Prior. In: BMVC, vol. 3, p. 7 (2011)
15. Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A.: Saliency Filters: Contrast Based Filtering for Salient Region Detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 733–740 (2012)
16. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware Saliency Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(10), 1915–1926 (2012)
17. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC Superpixels Compared to State-of-the-art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(11), 2274–2282 (2012)
18. Chen, W.Y., Song, Y., Bai, H., Lin, C.J., Chang, E.Y.: Parallel Spectral Clustering in Distributed Systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(3), 568–586 (2011)
19. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)
20. Ma, Y.F., Zhang, H.J.: Contrast-based Image Attention Analysis by Using Fuzzy Growing. In: Proceedings of the 11th ACM International Conference on Multimedia, pp. 374–381 (2003)
21. Harel, J., Koch, C., Perona, P.: Graph-based Visual Saliency. In: Advances in Neural Information Processing Systems, vol. 19, pp. 545–552 (2007)
22. Hou, X., Zhang, L.: Saliency Detection: A Spectral Residual Approach. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
23. Achanta, R., Estrada, F., Wils, P., Süsstrunk, S.: Salient Region Detection and Segmentation. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 66–75. Springer, Heidelberg (2008)
24. Zhai, Y., Shah, M.: Visual Attention Detection in Video Sequences Using Spatiotemporal Cues. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 815–824 (2006)