FROM RARENESS TO COMPACTNESS: CONTRAST-AWARE IMAGE SALIENCY DETECTION

Hsin-Ho Yeh[†] *and Chu-Song Chen*[†‡]

[†] Institute of Information Science, Academia Sinica, Taipei, Taiwan.
[‡] Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
{hhyeh,song}@iis.sinica.edu.tw

ABSTRACT

In this paper, we present a simple but effective method called **C**ontrast-**A**ware **S**aliency (CAS) to detect visual saliency by utilizing two general characteristics: rareness and compactness. In our approach, multiple-salient-spots are used to find initial salient clues, which appear to be rare and unique parts in an image. Then, the salient regions are detected by aggregating the surrounding regions of the spots, which fulfil the compactness nature of salient objects. Experimental results show the proposed CAS performs well in the benchmark dataset.

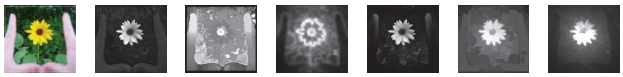
Index Terms— Image saliency, Salient region

1. INTRODUCTION

Visual saliency has been focused for decades since it is fundamental to vision tasks. It is referred to as some image parts attracting most of visual attention when people see it at first glance. In applications, salient regions are considered as foreground, thus detecting saliency mostly separates foreground from background regions [1]. By taking advantage of salient regions, vision tasks such as segmentation [2] and recognition [3] could be benefit from the deletion of cluttered background, easing visual content analysis.

Although methods of detecting salient regions are categorized into top-down [4] and bottom-up approaches [5, 6, 2, 7, 8, 9, 10], bottom-up one interests us more because it detects general regions of interest without any prior knowledge and simply measures visual saliency from **contrast** property. As seeing the picture shown in Fig. 1 (a), for example, most people agree that they are attracted by the yellow flower instead of green leaves or surrounding fingers. It is the visual contrast that arouses more of the human visual attention.

Detecting visual contrast has been discussed for a long time. Achanta et al. [5] present a global-based approach that preserves a reasonable range of frequency to detect the whole salient object instead of object border. It shows an acceptable performance but lacks of the ability to measure the contrast from a local region (i.e., the contrast between the salient object and its surrounding context).



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

Fig. 1. When composing a picture like (a), the photographer tries to guide the audience's visual attention to the yellow flower. In comparison, the saliency results of [5, 2, 9, 11, 12] are reported in (b), (c), (d), (e), (f), respectively. Ours is shown in (g).

To measure local contrast, sliding-window approaches [2, 10, 7] are well studied. Rahtu et al. [2] separate a sliding window into foreground and background regions, and then a Bayesian formulation is employed for saliency detection. Klein and Frintrop [10] measure saliency by KL-Divergence between center and surround parts. Liu et al. [7] present a learning-based framework to detect salient regions. Conditional random field (CRF) is employed to exploit the importance among distinct features such as center-surround histogram. Without knowing the salient object size, however, the sliding-window approaches must resize their window to locate the salient objects. Therefore, similar to object detection, the problem caused by variable object-size will degrade their performance significantly.

To tackle the object-size problem, biology-inspired model and graph-based representation [6, 8] are exploited. In the early work, biology-inspired model [13] tries to simulate the selection mechanism of human visual system (HVS), and it has been realized in graphical model recently. Gopalakrishnan et al. [6] deem that the most salient node must be isolated from others; moreover, it should be tightly connected with the neighboring nodes to express the compactness. Thus, the most salient node is selected by two criteria: global isolation and local compactness. Wang et al. [8] combine biology-inspired model with random walk by proposing the site-entropy-rate (SER). Random walk in [8] is modeled as the process of information transmission to capture the long range relation between adjacent nodes; besides, the center-surround mechanism is exploited as well. As a result, both of

[6] and [8] show a reliable performance.

Pixel-based distinctiveness is an another way to model local contrast. Goferman et al. [9] express the saliency of a node from its surrounding and color-distinctive nodes. Thus, a node's saliency is obtained from the color distances to the other nodes weighted by their inverse spatial distances. To this end, this salient map not only has a good result like humans but can carry the meaning of the scene as well, making it suitable for applications such as image retargeting.

By reviewing the literature, lots of works agree that the objective of saliency is application dependent. Although the topic-specified saliency (or top-down saliency) does well for individual applications [4], it fails in other applications owing to lacking of the generalization ability. On the contrary, without assuming object characteristics, bottom-up approaches have better general detection capabilities for various kinds of images. From this point of view, general-purpose saliency detectors are more promising than topic-specified detectors in most kinds of applications.

We present a novel saliency detection method by employing two general principles: rareness and compactness. Rareness assumes that few salient nodes must be distinctive enough to attract the audience's concentration, while compactness suggests that pixels in an object have similar color to the nearby pixels. In rareness property, multiple-salient-spots hypothesis is employed to discover useful salient clues (nodes). To fulfil the compactness of semantic object, the salient nodes spread their saliency based on color and spatial smoothness to the neighboring nodes. Finally, salient regions can be detected by thresholding the saliency map obtained.

2. PROBLEM FORMULATION AND APPROACH

Given a $n \times n$ image I as input, each pixel in the image is represented by CIELab color space, $I = \{L, a, b\}$. The salient detector f takes the input I and generates its saliency map G. Each pixel in G is proportional to its degree of saliency.

We use patch-based representation: an image I is divided into M non-overlapped patches $\{P_1, ..., P_M\}$. P_i, where $1 \leq i \leq M$, is a $k \times k$ image patch which is centered at the position $P_i^{spa} \in \mathbb{R}^2$. We further denote the color vector formed by the $k \times k$ pixels as $P_i^{cor} \in \mathbb{R}^{k \times k \times 3}$. In the implementation, we set $k = 5$.

2.1. Salient Spot Discovery

Rareness reflects the property of distinctiveness that few pixels must be visually unique to inspire human visual concentration. It can be easily discovered from global and local contrasts. Absolute contrast (Sec. 2.1.1) and bin contrast (Sec. 2.1.2) are to cooperate in the proposed global contrast.

2.1.1. Absolute Contrast

Absolute contrast (AC) preserves the color distinctiveness by measuring the ℓ_2-norm distance between the given patch P_i and the averaged patch from the image. The averaged patch is defined as $mean(I) = \frac{1}{M} \sum_{j=1}^{M} P_j^{cor}$. Thus, absolute contrast is defined as:

$$AC(P_i) = ||P_i^{cor} - mean(I)||_2. \tag{1}$$

2.1.2. Bin Contrast

In order to emphasize the uniqueness of the rare-color bins, their saliency should be larger than the saliency from common-color bins. Thus, we evaluate a bin's frequency to reflect its saliency. In detail, we separate the M patches into B clusters, $H_1, H_2, ..., H_B$, via Normalized Cut [14] in Euclidean space, and $|H_a|$ indicates its population where $1 \leq a \leq B$. The bin contrast for a patch P_i is defined as:

$$BC(P_i) = \exp\left(-\gamma \times \frac{|H_a|}{\sum_{b=1}^{B} |H_b|}\right), \tag{2}$$

for each $P_i^{cor} \in H_a$. We use $\gamma = 1$ to convert the population of a cluster into its contrast degree, and set $B = 2$.

These two contrasts can preserve global saliency well, but they still cannot emphasize local contrasts owing to miss-considering spatial property. Local contrast is thus considered further in the following.

2.1.3. Local Contrast

Local contrast can be exploited as a counter concept of self-similarity. Self-similarity proposed by Shechtman and Irani [15] measures the local structure (LS) likeness across images: For pixel u, the LS measures the similarity of a small patch p_u centered at pixel u with respect to a larger regions R_u: $LS(u) = \exp(\frac{SSD(p_u, p_v)}{\sigma})$, where $v \in R_u$ and SSD denotes the sum of squared differences. Thomas and Vittorio [16] extend this idea into global structure (GS), i.e. R_u is the whole image.

To extend the concept of self-similarity to discover local contrast, we use the term 'distinction' for precisely describing the saliency (or contrast). The distinction between two patches (P_i and P_j) is defined similar to that in [9]:

$$d(P_i, P_j) = \frac{||P_i^{cor} - P_j^{cor}||_2}{1 + \lambda ||P_i^{spa} - P_j^{spa}||_2}, \tag{3}$$

where $\lambda = 1$ is used to balance between the spatial and color information. The local contrast for a patch P_i is defined as:

$$LC(P_i) = \sum_{j=1}^{M} d(P_i, P_j). \tag{4}$$

2.1.4. Multiple Contrast Fusion and Spatial Weighting

The contrasts from local and global visual clues are integrated. We multiply AC, BC, and LC together for fusion them:

$$MC(P_i) = AC(P_i) \times BC(P_i) \times LC(P_i). \quad (5)$$

Besides feature-based contrast, spatial weighting merits consideration according to [17]. A prior spatial weighting is assumed to have the following properties: 1) the center part of an image plays a great importance in catching the visual attention. 2) the visual attention is degraded as the patch is far away from the center part. The spatial weighting for a patch P_i is formulated as follows:

$$PW(P_i) = \exp(-\frac{||P_i^{spa} - \bar{\mathbf{s}}||_2^2}{\varphi}), \quad (6)$$

where $\bar{\mathbf{s}}$ is the 2D position of the image center and $\varphi = 80$ for 400×300 images. The weighted contrast is defined as:

$$MCW(P_i) = MC(P_i) \times PW(P_i). \quad (7)$$

Finally, the salient spots SP are computed by ranking the salient patches according to the MCW criterion (or other weighted contrast criteria discussed in the experiment):

$$SP(MCW) = \{P_i | rank(MCW(P_i)) \leq \alpha \times M, \forall i\}, \quad (8)$$

where $rank(MCW(.))$ ranks the values of MCW in the decreasing order, and $\alpha = 0.05$ is used. After discovering the salient spots by spatial weighting and multiple contrasts fusion, we spread their saliency to the neighbors for the compactness as introduced below.

2.2. Salient Value Transmission

Compactness is another property of saliency focused in this study. It stands for the smooth characteristic of an object either in appearance or space. Two properties are considered: 1) the patches with smooth color should be located in the same object no matter how far they are. 2) the patches nearby the salient spots should be regarded as salient regions as well. To reflect these two characteristics, we define the color smoothness (CS) for each patch P_i to the SP in the following:

$$CS(P_i) = \max_{\forall P_j \in SP} \exp(\frac{-||P_i^{cor} - P_j^{cor}||_2}{\sigma}), \quad (9)$$

and spatial consistency (SC) is further formulated as:

$$SC(P_i) = \max_{\forall P_j \in SP} \exp(\frac{-||P_i^{spa} - P_j^{spa}||_2}{\delta}), \quad (10)$$

where $\sigma = 2$ and $\delta = 100$ in the experiments.

The proposed salient map G, therefore, is achieved by multiplying the color smoothness and spatial consistency: $G(P_i) = CS(P_i) \times SC(P_i), 1 \leq i \leq M$. The salient map G is further normalized into $[0, 1]$.

3. EXPERIMENTAL RESULTS

We evaluate our saliency method in a benchmark dataset from Achanta et al. [5] which contains 1,000 color images, and each one has its pixelwise binary (foreground and background) ground truth. For quantifying the experimental results, we normalize the range of the saliency map into $[0, 255]$ and express it in a binary notation. The pixel is marked as foreground when its salient value is no smaller than the threshold t, and is labelled as background otherwise. For each t varying from 1 to 255, we obtain the averaged precision and recall, and then plot the results into a curve for comparison. The above experimental settings are the same as [5, 11, 12, 10].

In the first experiment, we compare the usefulness of multiple contrasts as shown in Fig. 2. Each contrast is weighted by PW; thus we name contrast AC, BC, and LC as $ACPW$, $BCPW$, and $LCPW$, respectively. The salient spots in Eq. (7) are ranked accordingly. Moreover, MCW is the fused contrasts as in Eq. (6). From the result, we find that AC is more effective than BC and LC since it provides a universal concept to detect the saliency. However, MCW surpasses AC, indicating that the global-based consideration is not enough. Instead, by combining with AC, BC and LC simultaneously, MCW can discover salient regions better.

Later, we compare the performance with the state-of-the art methods as plotted in Fig. 3. The compared performance is reported from their papers or codes. For convenience, we name the compared methods as FT [5], CX [9], RC [12] BITS [10], SS [2], MSSS [11] and CAS (the proposed one), respectively. From the curves, the results show that CAS and RC both have comparable detection performance, which outperform FT, CX, BITS, SS, and MSSS. This result claims that CAS has efficacy similar to RC which measures image saliency from region-level contrast computation. Segmenting an image into regions is a prerequisite for RC. However, we could have no prior knowledge on how to select the parameters for image segmentation to produce better salient maps. By contrast, the general assumptions in CAS make it has less restrictive, and still performs better than the sliding-window approaches (SS [2] and BITS [10]). More saliency-detection results are shown in Fig. 4.

4. CONCLUSION

We have presented a simple but effective saliency detection method by employing two general concepts: rareness and compactness. The experimental results have shown that the proposed method provides reliable detection results owing to multiple contrasts consideration and general salient object assumption. In future, the proposed method has the potential to be extended to video saliency detection by taking motion information into account.

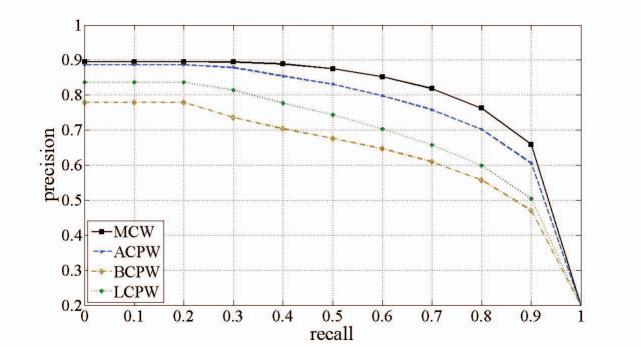


Fig. 2. The compared performance of each individual contrast $ACPW$, $BCPW$, and $LCPW$ as well as the fused contrast MCW.

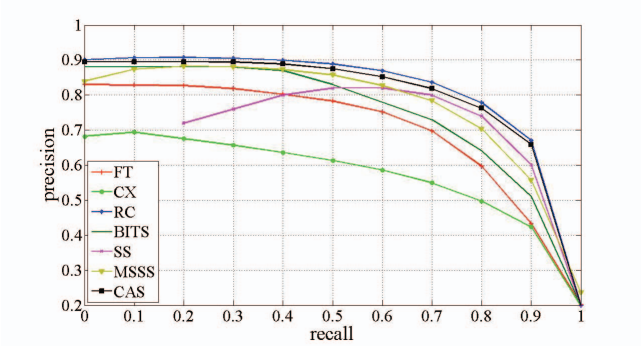


Fig. 3. The compared performance of the methods from FT [5], CX [9], RC [12] BITS [10], SS [2], MSSS [11] and CAS (the proposed one)

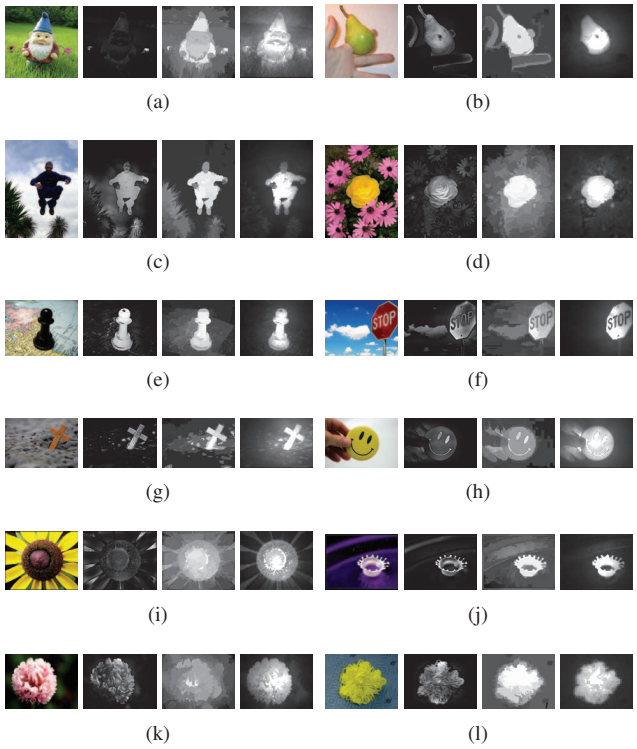


Fig. 4. More saliency maps for visual comparison. In each figure from (a) to (l), the original figure and its saliency maps produced by MSSS [11], RC [12], and CAS (the proposed method) are shown from left to right.

Acknowledgment

This work was supported in part by the National Science Council, Taiwan, under the grant NSC 101-2631-H-001-007.

5. REFERENCES

[1] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *TPAMI*, 2010.

[2] E. Rahtu, J. Kannala, M. Salo, and J. Heikkil, "Segmenting salient objects from images and videos," in *ECCV*, 2010.

[3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition," in *CVPR*, 2004.

[4] D. Gao, S. Han, and N. Vasconcelos, "Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition," *TPAMI*, 2009.

[5] R. Achanta, S. Hemami, F. Estrada, and S. Ssstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009.

[6] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs to model saliency in images," in *CVPR*, 2009.

[7] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *TPAMI*, 2010.

[8] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate," in *CVPR*, 2010.

[9] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *CVPR*, 2010.

[10] D. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *ICCV*, 2011.

[11] R. Achanta and S. Ssstrunk, "Saliency Detection using Maximum Symmetric Surround," in *ICIP*, 2010.

[12] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *CVPR*, 2011.

[13] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *TPAMI*, 1998.

[14] J. Shi and J. Malik, "Normalized cuts and image segmentation," *TPAMI*, 2000.

[15] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *CVPR*, 2007.

[16] D. Thomas and F. Vittorio, "Global and efficient self-similarity for object classification and detection," in *CVPR*, 2010.

[17] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *ICCV*, 2009.