Global contrast based salient region detection

Item Type	Conference Paper
Authors	Cheng, Ming-Ming; Zhang, Guo-Xin; Mitra, Niloy J.; Huang, Xiaolei; Hu, Shi-Min
Citation	Cheng M-M, Zhang G-X, Mitra NJ, Huang X, Hu S-M (2011) Global contrast based salient region detection. CVPR 2011. Available: http://dx.doi.org/10.1109/CVPR.2011.5995344.
Eprint version	Publisher's Version/PDF
DOI	10.1109/CVPR.2011.5995344
Publisher	Institute of Electrical and Electronics Engineers (IEEE)
Journal	CVPR 2011
Rights	(c) 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.
Download date	04/08/2022 20:19:59
Link to Item	http://hdl.handle.net/10754/622089

Global Contrast based Salient Region Detection

Ming-Ming Cheng[1]   Guo-Xin Zhang[1]   Niloy J. Mitra[2]   Xiaolei Huang[3]   Shi-Min Hu[1]
[1] TNList, Tsinghua University   [2] KAUST   [3] Lehigh University
chengmingvictor@gmail.com

Abstract

Reliable estimation of visual saliency allows appropriate processing of images without prior knowledge of their contents, and thus remains an important step in many computer vision tasks including image segmentation, object recognition, and adaptive compression. We propose a regional contrast based saliency extraction algorithm, which simultaneously evaluates global contrast differences and spatial coherence. The proposed algorithm is simple, efficient, and yields full resolution saliency maps. Our algorithm consistently outperformed existing saliency detection methods, yielding higher precision and better recall rates, when evaluated using one of the largest publicly available data sets. We also demonstrate how the extracted saliency map can be used to create high quality segmentation masks for subsequent image processing.

1. Introduction

Humans routinely and effortlessly judge the importance of image regions, and focus attention on important parts. Computationally detecting such salient image regions remains a significant goal, as it allows preferential allocation of computational resources in subsequent image analysis and synthesis. Extracted saliency maps are widely used in many computer vision applications including object-of-interest image segmentation [13, 18], object recognition [25], adaptive compression of images [6], content-aware image editing [28, 33, 30, 9], and image retrieval [4].

Saliency originates from visual uniqueness, unpredictability, rarity, or surprise, and is often attributed to variations in image attributes like color, gradient, edges, and boundaries. Visual saliency, being closely related to how we perceive and process visual stimuli, is investigated by multiple disciplines including cognitive psychology [26, 29], neurobiology [8, 22], and computer vision [17, 2]. Theories of human attention hypothesize that the human vision system only processes parts of an image in detail, while leaving others nearly unprocessed. Early work by Treisman and Gelade [27], Koch and Ullman [19], and subse-

Figure 1. Given input images (top), a global contrast analysis is used to compute high resolution saliency maps (middle), which can be used to produce masks (bottom) around regions of interest.

quent attention theories proposed by Itti, Wolfe and others, suggest two stages of visual attention: fast, pre-attentive, bottom-up, data driven saliency extraction; and slower, task dependent, top-down, goal driven saliency extraction.

We focus on bottom-up data driven saliency detection using image contrast. It is widely believed that human cortical cells may be *hard wired* to preferentially respond to high contrast stimulus in their receptive fields [23]. We propose contrast analysis for extracting high-resolution, full-field saliency maps based on the following observations:

- A global contrast based method, which separates a large-scale object from its surroundings, is preferred over local contrast based methods producing high saliency values at or near object edges.
- Global considerations enable assignment of comparable saliency values to similar image regions, and can uniformly highlight entire objects.
- Saliency of a region depends mainly on its contrast to the nearby regions, while contrasts to distant regions are less significant.
- Saliency maps should be fast and easy to generate to allow processing of large image collections, and facilitate efficient image classification and retrieval.

We propose a *histogram-based contrast method* (HC) to measure saliency. HC-maps assign pixel-wise saliency val-

(a) original   (b) IT[17]   (c) MZ[21]   (d) GB[14]   (e) SR[15]   (f) AC[1]   (g) CA[12]   (h) FT[2]   (i) LC[32]   (j) HC   (k) RC

Figure 2. Saliency maps computed by different state-of-the-art methods (b-i), and with our proposed HC (j) and RC methods (k). Most results highlight edges, or are of low resolution. See also Figure 6 (and our project webpage).

ues based simply on color separation from all other image pixels to produce full resolution saliency maps. We use a histogram-based approach for efficient processing, while employing a smoothing procedure to control quantization artifacts. Note that our algorithm is targeted towards natural scenes, and maybe suboptimal for extracting saliency of highly textured scenes (see Figure 12).

As an improvement over HC-maps, we incorporate spatial relations to produce *region-based contrast* (RC) maps where we first segment the input image into regions, and then assign saliency values to them. The saliency value of a region is now calculated using a global contrast score, measured by the region's contrast and spatial distances to other regions in the image.

We have extensively evaluated our methods on publicly available benchmark data sets, and compared our methods with (eight) state-of-the-art saliency methods [17, 21, 32, 14, 15, 1, 2, 12] as well as with manually produced ground truth annotations[1]. The experiments show significant improvements over previous methods both in precision and recall rates. Overall, compared with HC-maps, RC-maps produce better precision and recall rates, but at the cost of increased computations. Encouragingly, we observe that the saliency cuts extracted using our saliency maps are, in most cases, comparable to manual annotations. We also present application of the extracted saliency maps to segmentation, context aware resizing, and non-photo realistic rendering.

2. Related Work

We focus on relevant literature targeting pre-attentive bottom-up saliency detection, which may be biologically motivated, or purely computational, or involve both aspects. Such methods utilize low-level processing to determine the contrast of image regions to their surroundings, using feature attributes such as intensity, color, and edges [2]. We broadly classify the algorithms into local and global schemes.

Local contrast based methods investigate the rarity of image regions with respect to (small) local neighborhoods. Based on the highly influential biologically inspired *early representation* model introduced by Koch and Ullman [19], Itti et al. [17] define image saliency using central-surrounded differences across multi-scale image features.

Ma and Zhang [21] propose an alternative local contrast analysis for generating saliency maps, which is then extended using a fuzzy growth model. Harel et al. [14] normalize the feature maps of Itti et al., to highlight conspicuous parts and permit combination with other importance maps. Liu et al. [20] find multi-scale contrast by linearly combining contrast in a Gaussian image pyramid. More recently, Goferman et al. [12] simultaneously model local low-level clues, global considerations, visual organization rules, and high-level features to highlight salient objects along with their contexts. Such methods using local contrast tend to produce higher saliency values near edges instead of uniformly highlighting salient objects (see Figure 2).

Global contrast based methods evaluate saliency of an image region using its contrast with respect to the entire image. Zhai and Shah [32] define pixel-level saliency based on a pixel's contrast to all other pixels. However, for efficiency they use only luminance information, thus ignoring distinctiveness clues in other channels. Achanta et al. [2] propose a frequency tuned method that directly defines pixel saliency using a pixel's color difference from the average image color. The elegant approach, however, only considers first order average color, which can be insufficient to analyze complex variations common in natural images. In Figures 6 and 7, we show qualitative and quantitative weaknesses of such approaches. Furthermore, these methods ignore spatial relationships across image parts, which can be critical for reliable and coherent saliency detection (see Section 5).

3. Histogram Based Contrast

Based on the observation from biological vision that the vision system is sensitive to contrast in visual signal, we propose a histogram-based contrast (HC) method to define saliency values for image pixels using color statistics of the input image. Specifically, the saliency of a pixel is defined using its color contrast to all other pixels in the image, i.e., the saliency value of a pixel I_k in image I is defined as,

$$S(I_k) = \sum_{\forall I_i \in I} D(I_k, I_i), \tag{1}$$

where $D(I_k, I_i)$ is the color distance metric between pixels I_k and I_i in the $L^*a^*b^*$ space (see also [32]). Equation 1 can be expanded by pixel order to have the following form,

$$S(I_k) = D(I_k, I_1) + D(I_k, I_2) + \cdots + D(I_k, I_N), \tag{2}$$

---

[1]Results for 1000 images and prototype software are available at the project webpage: http://cg.cs.tsinghua.edu.cn/people/%7Ecmm/saliency/
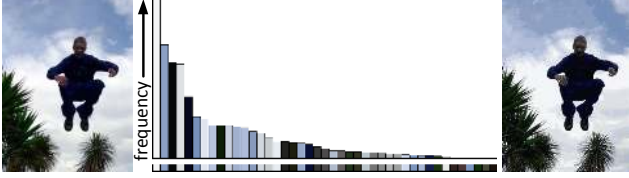
Figure 3. Given an input image (left), we compute its color histogram (middle). Corresponding histogram bin colors are shown in the lower bar. The quantized image (right) uses only 43 histogram bin colors and still retains sufficient visual quality for saliency detection.

where N is the number of pixels in image I. It is easy to see that pixels with the same color value have the same saliency value under this definition, since the measure is oblivious to spatial relations. Hence, rearranging Equation 2 such that the terms with the same color value c_j are grouped together, we get saliency value for each color as,

$$S(I_k) = S(c_l) = \sum_{j=1}^{n} f_j D(c_l, c_j), \qquad (3)$$

where c_l is the color value of pixel I_k, n is the number of distinct pixel colors, and f_j is the probability of pixel color c_j in image I. Note that in order to prevent salient region color statistics from being corrupted by similar colors from other regions, one can develop a similar scheme using varying window masks. However, given the strict efficiency requirement, we take the simple global approach.

Histogram based speed up. Naively evaluating the saliency value for each image pixel using Equation 1 takes $O(N^2)$ time, which is computationally too expensive even for medium sized images. The equivalent representation in Equation 3, however, takes $O(N) + O(n^2)$ time, implying that computational efficiency can be improved to $O(N)$ if $O(n^2) \leq O(N)$. Thus, the key to speed up is to reduce the number of pixel colors in the image. However, the true-color space contains 256^3 possible colors, which is typically larger than the number of image pixels.

Zhai and Shah [32] reduce the number of colors, n, by only using luminance. In this way, $n^2 = 256^2$ (typically $256^2 \ll N$). However, their method has the disadvantage that the distinctiveness of color information is ignored. In this work, we use the full color space instead of luminance only. To reduce the number of colors needed to consider, we first quantize each color channel to have 12 different values, which reduces the number of colors to $12^3 = 1728$. Considering that color in a natural image typically covers only a small portion of the full color space, we further reduce the number of colors by ignoring less frequently occurring colors. By choosing more frequently occurring colors and ensuring these colors cover the colors of more than 95% of the image pixels, we typically are left with around $n = 85$ colors (see Section 5 for experimental details). The colors
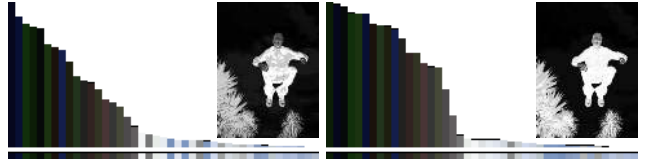

Figure 4. Saliency of each color, normalized to the range $[0, 1]$, before (left) and after (right) color space smoothing. Corresponding saliency maps are shown in the respective insets.

of the remaining pixels, which comprise fewer than 5% of the image pixels, are replaced by the closest colors in the histogram. A typical example of such quantization is shown in Figure 3. Note that again due to efficiency requirements we select the simple histogram based quantization instead of optimizing for an image specific color palette.

Color space smoothing. Although we can efficiently compute color contrast by building a compact color histogram using color quantization and choosing more frequent colors, the quantization itself may introduce artifacts. Some similar colors may be quantized to different values. In order to reduce noisy saliency results caused by such randomness, we use a smoothing procedure to refine the saliency value for each color. We replace the saliency value of each color by the weighted average of the saliency values of similar colors (measured by $L^*a^*b^*$ distance). This is actually a smoothing process in the color feature space. Typically we choose $m = n/4$ nearest colors to refine the saliency value of color c by,

$$S'(c) = \frac{1}{(m-1)T} \sum_{i=1}^{m} (T - D(c, c_i)) S(c_i) \qquad (4)$$

where $T = \sum_{i=1}^{m} D(c, c_i)$ is the sum of distances between color c and its m nearest neighbors c_i, and the normalization factor comes from $\sum_{i=1}^{m} (T - D(c, c_i)) = (m - 1)T$. Note that we use a linearly-varying smoothing weight $(T - D(c, c_i))$ to assign larger weights to colors closer to c in the color feature space. In our experiments, we found that such linearly-varying weights are better than Gaussian weights, which fall off too sharply. Figure 4 shows the typical effect of color space smoothing with the corresponding histograms sorted by decreasing saliency values. Note that similar histogram bins are closer to each other after such smoothing, indicating that similar colors have higher likelihood of being assigned similar saliency values, thus reducing quantization artifacts (see Figure 7).

Implementation details. To quantize the color space into 12^3 different colors, we uniformly divide each color channel into 12 different levels. While the quantization of colors is performed in the RGB color space, we measure color differences in the $L^*a^*b^*$ color space because of its perceptual accuracy. However, we do not perform quantization directly

Figure 5. Image regions generated by Felzenszwalb and Huttenlocher's segmentation method [11] (left), region contrast based segmentation with (left-middle) and without (right-middle) distance weighting. Incorporating the spatial context, we get a high quality saliency cut (right) comparable to human labeled ground truth.

in the $L^*a^*b^*$ color space since not all colors in the range $L^* \in [0, 100]$, and $a^*, b^* \in [-127, 127]$ necessarily correspond to real colors. Experimentally we observed worse quantization artifacts using direct $L^*a^*b^*$ color space quantization. Best results were obtained by quantization in the RGB space while measuring distance in the $L^*a^*b^*$ color space, as opposed to performing both quantization and distance calculation in a single color space, either RGB or $L^*a^*b^*$.

4. Region Based Contrast

Humans pay more attention to those image regions that contrast strongly with their surroundings [10]. Besides contrast, spatial relationships play an important role in human attention. High contrast to its surrounding regions is usually stronger evidence for saliency of a region than high contrast to far-away regions. Since directly introducing spatial relationships when computing pixel-level contrast is computationally expensive, we introduce a contrast analysis method, *region contrast* (RC), so as to integrate spatial relationships into region-level contrast computation. In RC, we first segment the input image into regions, then compute color contrast at the region level, and define the saliency for each region as the weighted sum of the region's contrasts to all other regions in the image. The weights are set according to the spatial distances with farther regions being assigned smaller weights.

Region contrast by sparse histogram comparison. We first segment the input image into regions using a graph-based image segmentation method [11]. Then we build the color histogram for each region as in Section 3. For a region r_k, we compute its saliency value by measuring its color contrast to all other regions in the image,

$$S(r_k) = \sum_{r_k \neq r_i} w(r_i) D_r(r_k, r_i), \quad (5)$$

where $w(r_i)$ is the weight of region r_i and $D_r(\cdot, \cdot)$ is the color distance metric between the two regions. Here we

use the number of pixels in r_i as $w(r_i)$ to emphasize color contrast to bigger regions. The color distance between two regions r_1 and r_2 is defined as,

$$D_r(r_1, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_{1,i}) f(c_{2,j}) D(c_{1,i}, c_{2,j}) \quad (6)$$

where $f(c_{k,i})$ is the probability of the i-th color $c_{k,i}$ among all n_k colors in the k-th region r_k, $k = \{1, 2\}$. Note that we use the probability of a color in the probability density function (i.e. normalized color histogram) of the region as the weight for this color to emphasize more the color differences between dominant colors.

Storing and calculating the regular matrix format histogram for each region is inefficient since each region typically contains a small number of colors in the color histogram of the whole image. Instead, we use a sparse histogram representation for efficient storage and computation.

Spatially weighted region contrast. We further incorporate spatial information by introducing a spatial weighting term in Equation 5 to increase the effects of closer regions and decrease the effects of farther regions. Specifically, for any region r_k, the spatially weighted region contrast based saliency is defined as:

$$S(r_k) = \sum_{r_k \neq r_i} \exp(-D_s(r_k, r_i)/\sigma_s^2) w(r_i) D_r(r_k, r_i) \quad (7)$$

where $D_s(r_k, r_i)$ is the spatial distance between regions r_k and r_i, and σ_s controls the strength of spatial weighting. Larger values of σ_s reduce the effect of spatial weighting so that contrast to farther regions would contribute more to the saliency of the current region. The spatial distance between two regions is defined as the Euclidean distance between their centroids. In our implementation, we use $\sigma_s^2 = 0.4$ with pixel coordinates normalized to $[0, 1]$.

5. Experimental Comparisons

We have evaluated the results of our approach on the publicly available database provided by Achanta et al. [2]. To the best of our knowledge, the database is the largest of its kind, and has ground truth in the form of accurate human-marked labels for salient regions. We compared the proposed global contrast based methods with 8 state-of-the-art saliency detection methods. Following [2], we selected these methods according to: number of citations (IT[17] and SR[15]), recency (GB[14], SR, AC[1], FT[2] and CA[12]), variety (IT is biologically-motivated, MZ[21] is purely computational, GB is hybrid, SR works in the frequency domain, AC and FT output full resolution saliency maps), and being related to our approach (LC[32]).

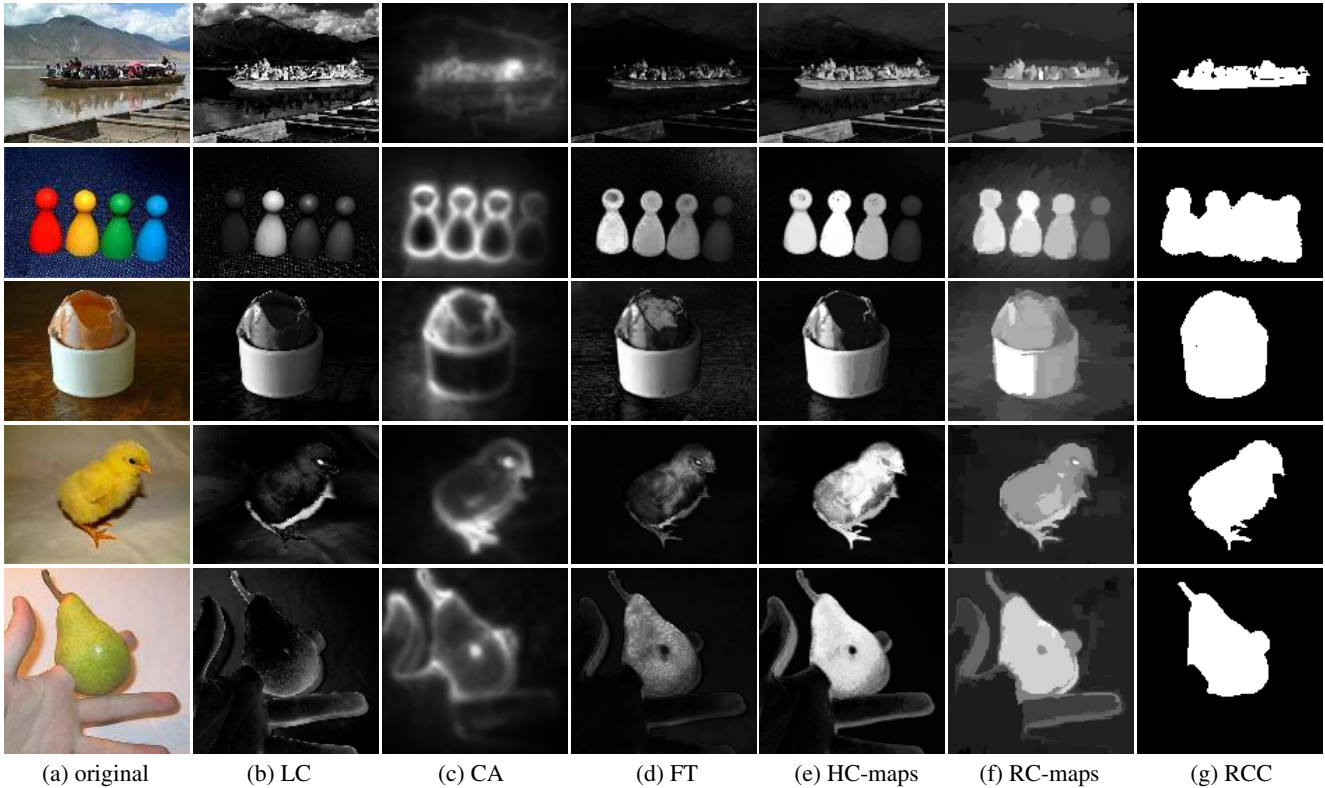| (a) original | (b) LC | (c) CA | (d) FT | (e) HC-maps | (f) RC-maps | (g) RCC |

Figure 6. Visual comparison of saliency maps. (a) original images, saliency maps produced using (b) Zhai and Shah [32], (c) Goferman et al. [12], (d) Achanta et al. [2], (e) our HC and (f) RC methods, and (g) RC-based saliency cut results. Our methods generate uniformly highlighted salient regions. (See our project webpage for all results on the full benchmark dataset.)

We used our methods and the others to compute saliency maps for all the 1000 images in the database. Table 1 compares the average time taken by each method. Our algorithms, HC and RC, are implemented in C++. For the other methods namely IT, GB, SR, FT and CA, we used the authors' implementations, while for LC, we implemented the algorithm in C++ since we could not find the authors' implementation. For typical natural images, our HC method needs $O(N)$ computation time and is sufficiently efficient for real-time applications. In contrast, our RC variant is slower as it requires image segmentation [11], but produces superior quality saliency maps.

In order to comprehensively evaluate the accuracy of our methods for salient object segmentation, we performed two experiments using different objective comparison measures. In the first experiment, to segment salient objects and calculate precision and recall curves [15], we binarized the saliency map using every possible fixed threshold, similar to the fixed thresholding experiment in [2]. In the second experiment, we segment salient objects by iteratively applying GrabCut [24] initialized using thresholded saliency maps, as we will describe later. We also use the obtained saliency maps as importance weighting for content aware image resizing and non-photo realistic rendering.

Segmentation by fixed thresholding. The simplest way to get a binary segmentation of salient objects is to threshold the saliency map with a threshold $T_f \in [0, 255]$. To reliably compare how well various saliency detection methods highlight salient regions in images, we vary the threshold T_f from 0 to 255. Figure 7 shows the resulting precision vs. recall curves. We also present the benefits of adding the color space smoothing and spatial weighting schemes, along with objective comparison with other saliency extraction methods. Visual comparison of saliency maps obtained by the various methods can be seen in Figures 2 and 6.

The precision and recall curves clearly show that our methods outperform the other eight methods. The extremities of the precision vs. recall curve are interesting: At maximum recall where $T_f = 0$, all pixels are retained as positives, i.e., considered to be foreground, so all the methods have the same precision and recall values; precision 0.2 and recall 1.0 at this point indicate that, on average, there are 20% image pixels belonging to the ground truth salient regions. At the other end, the minimum recall values of our methods are higher than those of the other methods, because the saliency maps computed by our methods are smoother and contain more pixels with the saliency value 255.

Saliency cut. We now consider the use of the computed saliency map to assist in salient object segmentation.

Method	IT[17]	MZ[21]	GB[14]	SR[15]	FT[2]	AC[1]	CA[12]	LC[32]	HC	RC
Time(s)	0.611	0.070	1.614	0.064	0.016	0.109	53.1	0.018	0.019	0.253
Code	Matlab	C++	Matlab	Matlab	C++	C++	Matlab	C++	C++	C++

Table 1. Average time taken to compute a saliency map for images in the database by Achanta et al. [2]. Most images in the database (see our project webpage) have resolution 400×300. Algorithms were tested using a Dual Core 2.6 GHz machine with 2GB RAM.
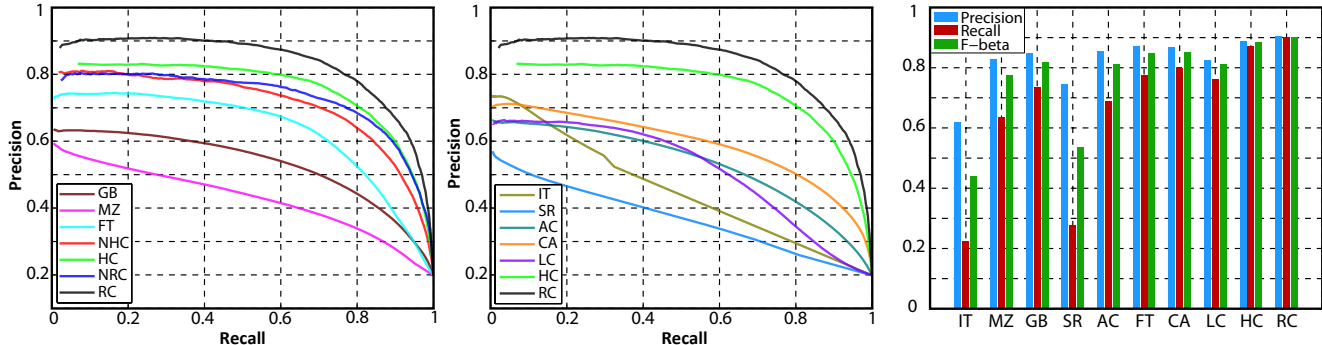
Figure 7. Precision-recall curves for naive thresholding of saliency maps using 1000 publicly available benchmark images. (Left, middle) Different options of our method compared with GB[14], MZ[21], FT[2], IT[17], SR[15], AC[1], CA[12], and LC[32]. NHC denotes a naive version of our HC method with color space smoothing disabled, and NRC denotes our RC method with spatial related weighting disabled. (Right) Precision-recall bars for our saliency cut algorithm using different saliency maps as initialization. Our method RC shows high precision, recall, and F_β values over the 1000-image database. (Please refer to our project webpage for the respective result images.)

Saliency maps have been previously employed for unsupervised object segmentation: Ma and Zhang [21] find rectangular salient regions by fuzzy region growing on their saliency maps. Ko and Nam [18] select salient regions using a support vector machine trained on image segment features, and then cluster these regions to extract salient objects. Han et al. [13] model color, texture, and edge features in a Markov random field framework to grow salient object regions from seed values in the saliency maps. More recently, Achanta et al. [2] average saliency values within image segments produced by mean-shift segmentation, and then find salient objects by identifying image segments that have average saliency above a threshold that is set to be twice the mean saliency value of the entire image.

In our approach, we iteratively apply GrabCut [24] to refine the segmentation result initially obtained by thresholding the saliency map (see Figure 8). Instead of manually selecting a rectangular region to initialize the process, as in
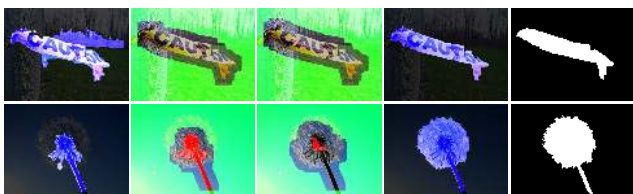

Figure 8. Saliency Cut. (Left to right) Initial segmentation, trimap after first iteration, trimap after second iteration, final segmentation, and manually labeled ground truth. In the segmented images, blue is foreground, gray is background, while in the trimaps, foreground is red, background is green, and unknown regions are left unchanged.

classical GrabCut, we automatically initialize GrabCut using a segmentation obtained by binarizing the saliency map using a fixed threshold; the threshold is chosen empirically to be the threshold that gives 95% recall rate in our fixed thresholding experiments.

Once initialized, we iteratively run GrabCut to improve the saliency cut result. (At most 4 iterations are run in our experiments.) After each iteration, we use dilation and erosion operations on the current segmentation result to get a new trimap for the next GrabCut iteration. As shown in Figure 8, the region outside the dilated region is set to background, the region inside the eroded region is set to foreground, and the remaining areas are set to unknown in the trimap. GrabCut, which by itself is an iterative process using Gaussian mixture models and graph cut, helps to refine salient object regions at each step. Regions closer to an initial salient object region are more likely to be part of that salient object than far-away regions. Thus, our new initialization enables GrabCut to include nearby salient regions, and exclude non-salient regions according to color feature dissimilarity. In the implementation, we set a narrow image-border region (15 pixels wide) to be always in the background in order to avoid slow convergence in the border region.

Figure 8 shows two examples of our saliency cut algorithm. In the flag example, unwanted regions are correctly excluded during GrabCut iterations. In the flower example, our saliency cut method successfully expanded the initial salient regions (obtained directly from the saliency map) and converged to an accurate segmentation result.
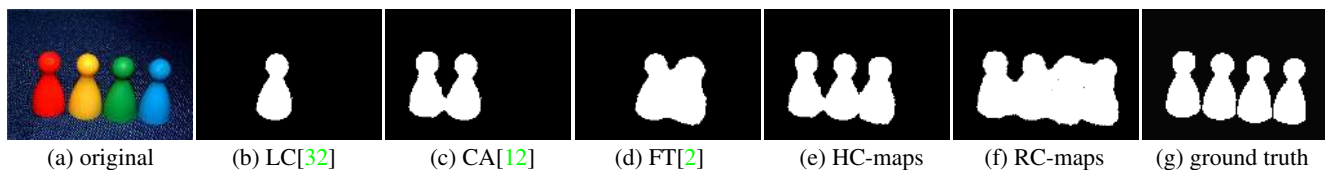
414

(a) original   (b) LC[32]   (c) CA[12]   (d) FT[2]   (e) HC-maps   (f) RC-maps   (g) ground truth

Figure 9. Saliency cut using different saliency maps for initialization. The respective saliency maps are shown in Figure 6.
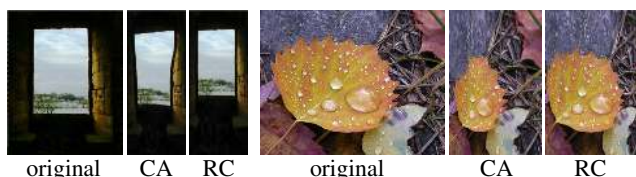

original   CA   RC   original   CA   RC

Figure 10. Comparison of content aware image resizing [33] results using CA[12]saliency maps and our RC saliency maps.

To objectively evaluate our new saliency cut method using our RC-map as initialization, we compare our results with results obtained by coupling iterative GrabCut with initializations from saliency maps computed by other methods. For consistency, we binarize each such saliency map using a threshold that gives 95% recall rate in the corresponding fixed thresholding experiment (see Figure 7). A visual comparison of the results is shown in Figure 9. Average precision, recall, and F-Measure are compared over the entire ground-truth database [2], with the F-Measure defined as:

$$F_\beta = \frac{(1 + \beta^2) Precision \times Recall}{\beta^2 \times Precision + Recall}.$$ (8)

We use $\beta^2 = 0.3$ as in Achanta et al. [2] to weigh precision more than recall. As can be seen from the comparison (see Figures 7-right and 9), saliency cut using our RC and HC saliency maps significantly outperform other methods. Compared with the state-of-the-art results on this database by Achanta et al. (precision = 75%, recall = 83%), we achieved better accuracy (precision = 90%, recall = 90%). (Our demo software is available at our project webpage.)

Content aware image resizing. In image re-targeting, saliency maps are usually used to specify relative importance across image parts (see also [3]). We experimented with using our saliency maps in the image resizing method proposed by Zhang et al. [33][2], which distributes distortion energy to relatively non-salient regions of an image while preserving both global and local image features. Figure 10 compares the resizing results using our RC-maps with the results using CA[12] saliency maps. Our RC saliency maps help produce better resizing results since the saliency values in salient object regions are piece-wise smooth, which is important for energy based resizing methods. CA saliency maps, having higher saliency values at object boundaries, are less suitable for applications like resizing, which require entire salient objects to be uniformly highlighted.

---

[2]The authors' original implementation is publicly available.

Figure 11. (Middle, right) FT[2] and RC saliency maps are used respectively for stylized rendering [16] of an input image (left). Our method produces a better saliency map, see insets, resulting in improved preservation of details, e.g., around the head and the fence regions.
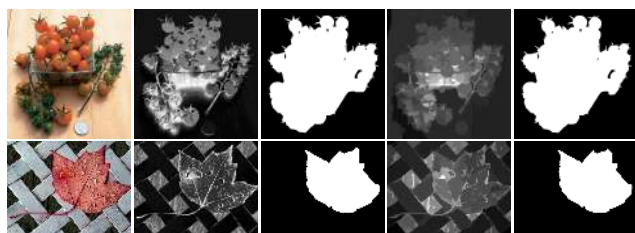

Figure 12. Challenging examples for our histogram based methods involve non-salient regions with similar colors as the salient parts (top), or an image with textured background (bottom). (Left to right) Input image, HC-map, HC saliency cut, RC-map, RC saliency cut.

Non-photorealistic rendering. Artists often abstract images and highlight meaningful parts of an image while masking out unimportant regions [31]. Inspired by this observation, a number of non-photorealistic rendering (NPR) efforts use saliency maps to generate interesting effects [7]. We experimentally compared our work with the most related, state-of-the-art saliency detection algorithm [2] in the context of a recent NPR technique [16] (see Figure 11). Our RC-maps give better saliency masks, which help the NPR method to better preserve details in important image parts and region boundaries, while smoothing out others.

6. Conclusions and Future Work

We presented global contrast based saliency computation methods, namely Histogram based Contrast (HC) and spatial information-enhanced Region based Contrast (RC). While the HC method is fast and generates results with fine details, the RC method generates spatially coherent high quality saliency maps at the cost of reduced computational efficiency. We evaluated our methods on the largest publicly available data set and compared our scheme with eight other state-of-the-art methods. Experiments indicate the

proposed schemes to be superior in terms of both precision and recall, while still being simple and efficient.

In the future, we plan to investigate efficient algorithms that incorporate spatial relationships in saliency computation while preserving fine details in the resulting saliency maps. Also, it is desirable to develop saliency detection algorithms to handle cluttered and textured background, which could introduce artifacts to our global histogram based approach. Finally, it may be beneficial to incorporate high level factors like human faces, symmetry into saliency maps. We believe the proposed saliency maps can be used for efficient object detection [13], reliable image classification, robust image scene analysis [5], leading to improved image retrieval [4].

References

[1] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. In *ICVS*, pages 66–75. Springer, 2008. 410, 412, 414

[2] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604, 2009. 409, 410, 412, 413, 414, 415

[3] R. Achanta and S. Susstrunk. Saliency Detection for Content-aware Image Resizing. In *ICIP*, 2009. 415

[4] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu. Sketch2photo: Internet image montage. *ACM Trans. Graph.*, 28(5):124:1–10, 2009. 409, 416

[5] M.-M. Cheng, F.-L. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Repfinder: Finding approximately repeated scene elements for image editing. *ACM Trans. Graph.*, 29(4):83:1–8, 2010. 416

[6] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 still image coding system: an overview. *IEEE Trans. Consumer Elec.*, 46(4):1103–1127, 2002. 409

[7] D. DeCarlo and A. Santella. Stylization and abstraction of photographs. *ACM Trans. Graph.*, 21(3):769–776, 2002. 415

[8] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995. 409

[9] M. Ding and R.-F. Tong. Content-aware copying and pasting in images. *The Viusal Computer*, 26:721–729, 2010. 409

[10] W. Eihhauser and P. Konig. Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, 17:1089–1097, 2003. 412

[11] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004. 412, 413

[12] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *CVPR*, 2010. 410, 412, 413, 414, 415

[13] J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. *IEEE TCSV*, 16(1):141–145, 2006. 409, 414, 416

[14] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, pages 545–552, 2006. 410, 412, 414

[15] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *CVPR*, pages 1–8, 2007. 410, 412, 413, 414

[16] H. Huang, L. Zhang, and T.-N. Fu. Video painting via motion layer manipulation. *Comput. Graph. Forum*, 29(7):2055–2064, 2010. 415

[17] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI*, 20(11):1254–1259, 1998. 409, 410, 412, 414

[18] B. Ko and J. Nam. Object-of-interest image segmentation based on human attention and semantic region clustering. *J Opt Soc Am*, 23(10):2462, 2006. 409, 414

[19] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurbiology*, 4:219–227, 1985. 409, 410

[20] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, T. X., and S. H.Y. Learning to detect a salient object. *IEEE TPAMI*, 33(2):353–367, 2011. 410

[21] Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia*, pages 374–381, 2003. 410, 412, 414

[22] S. K. Mannan, C. Kennard, and M. Husain. The role of visual salience in directing eye movements in visual object agnosia. *Current biology*, 19(6):247–248, 2009. 409

[23] J. Reynolds and R. Desimone. Interacting roles of attention and visual salience in v4. *Neuron*, 37(5):853–863, 2003. 409

[24] C. Rother, V. Kolmogorov, and A. Blake. "Grabcut"– Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004. 413, 414

[25] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? In *CVPR*, pages 37–44, 2004. 409

[26] H. Teuber. Physiological psychology. *Annual Review of Psychology*, 6(1):267–296, 1955. 409

[27] A. M. Triesman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980. 409

[28] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.*, 27(5):118:1–8, 2008. 409

[29] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, pages 5:1–7, 2004. 409

[30] H. Wu, Y.-S. Wang, K.-C. Feng, T.-T. Wong, T.-Y. Lee, and P.-A. Heng. Resizing by symmetry-summarization. *ACM Trans. Graph.*, 29(6):159:1–9, 2010. 409

[31] S. Zeki. *Inner vision: An exploration of art and the brain*. Oxford University Press, 1999. 415

[32] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, pages 815–824, 2006. 410, 411, 412, 413, 414, 415

[33] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin. A shape-preserving approach to image resizing. *Comput. Graph. Forum*, 28(7):1897–1906, 2009. 409, 415