Low-Level Salient Region Detection Using Multi-Scale Filtering

Gökhan Yildirim
EPFL, Switzerland
gokhan.yildirim@epfl.ch

Sabine Süsstrunk
EPFL, Switzerland
sabine.susstrunk@epfl.ch

Abstract

Various computer visions tasks require a summary of visually important regions in an image. Thus, developing simple yet accurate salient region detection algorithms has become an important research topic. The currently best performing state-of-the-art saliency detection algorithms incorporate image segmentation for abstraction. However, errors introduced in this step of the algorithms are transferred to the final saliency map estimation. In order to avoid this problem, we propose a simple low-level salient region detection algorithm that uses multi-scale filters. We consider each possible combination of filtered image pairs as weak saliency maps and combine them according to their adaptively computed compactness and center prior. Our filter-based method successfully eliminates the texture in the background and gives relatively uniform salient regions for multi-colored objects. In addition, the combination of several multi-scale filters produces a full-resolution saliency output, which preserves object boundaries. We show that our algorithm outperforms the most recent state-of-the-art methods on a database of 1000 images with pixel-precision ground truths.

1 Introduction

When we look at an image or a video, distinctive objects and/or actions in these visual stimuli instantly attract our attention. Visual saliency can thus be defined as the collection of the perceptual attributes that make these distinctive items stand out from their neighborhood and, consequently, be easily detectable.

Visual saliency is a very popular and well-studied topic in the computer vision community. Automatic detection of salient regions in an image provides valuable information for various applications, such as video compression [10], image retargeting [3], video retargeting [24, 23] and object detection [15].

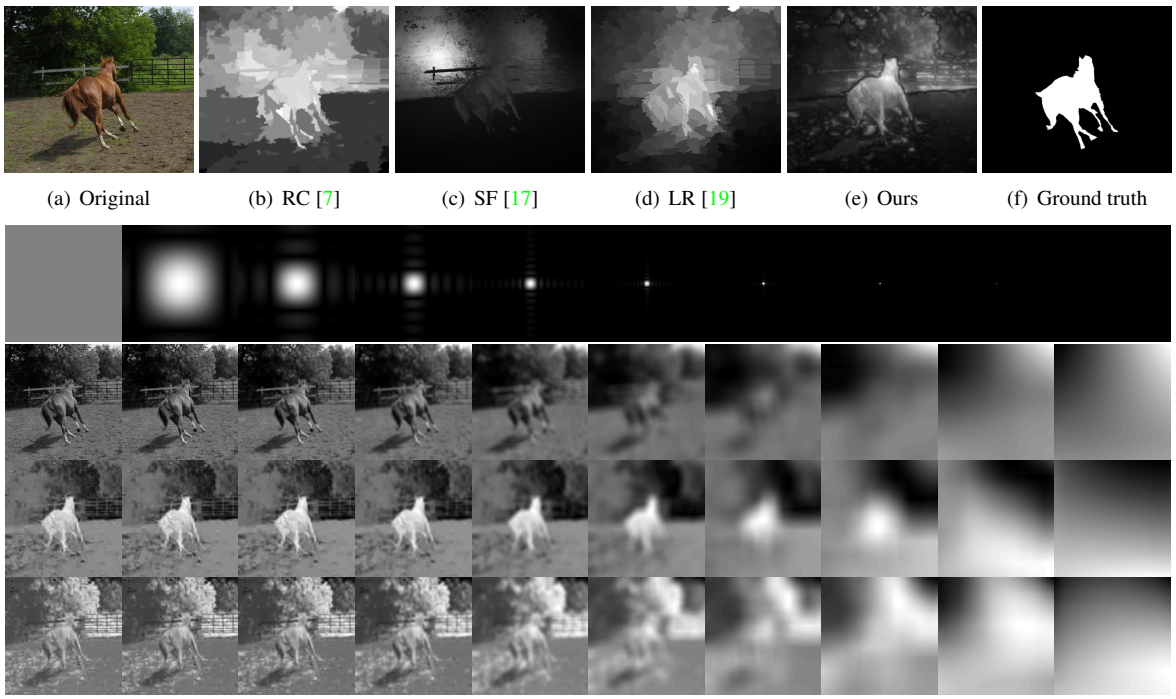
Humans are very efficient and accurate at detecting visually salient regions in images and videos, because we benefit from both high- and low-level saliency cues. Research that attempts to solve high-level saliency incorporates various prior object knowledge, such as face, pedestrian, and vehicle detection [4, 12, 19]. These objects are considered as the most salient components of an image, as their context may override low-level saliency cues. We do not consider high-level saliency here, we only focus on low-level spatial cues via filtering.

Low-level saliency deals with the *contrast* between fundamental properties of an image. The low-level processes in the retina and visual cortex of the human brain is well-investigated. Thus, studies on low-level visual saliency extraction techniques are, in general, inspired by the human visual system (HVS). One of the pioneering investigations in this field was done by Itti et al. [11], where center-to-surround differences in color, intensity, and orientations are combined at different scales to generate a final saliency map. Their algorithm imitates the receptive fields of the early-primate visual system.

The most recent and best-performing low-level saliency detection algorithms incorporate a segmentation step. They rely on the image being abstracted by a segmentation algorithm, which has several drawbacks.

1. Inaccurate segmentation of a highly textured or small object produces errors on the final saliency map.
2. Unnecessary over-segmentation of textured or uniform background pixels introduce false positives.
3. Pair-wise segment comparison in terms of color and distance might cause non-uniform saliency maps.

In order to solve these problems, we propose a salient region detection algorithm using multi-scale filtering instead of segmentation for image abstraction. Our method is a more general version of Achanta et al. [1, 2], where center-to-surround differences are computed using multi-scale filtering by keeping the center filter at a constant size. In our algorithm, we vary both center and surround filter sizes. We then combine multi-scale filter outputs, where every combination is considered as a *weak* saliency map. In order to weigh each weak map, we present two new adaptive sub-modules based on object compactness and center prior. We evaluate the performance of our algorithm using the MSRA-1000 dataset [2], which includes pixel-precision ground truths for 1000 images.



(a) Original (b) RC [7] (c) SF [17] (d) LR [19] (e) Ours (f) Ground truth

Figure 1: In the first row, (a) an example image from MSRA-1000 dataset, estimated saliency maps for (b) RC, (c) SF, (d) LR, and (e) our method are given with the (f) ground truth. Second row shows multi-scale filters in frequency domain. Third, fourth, and fifth rows show filtered L*, a*, and b* channels of the image, respectively.

Finally, we compare our results to the best-performing state-of-the-art algorithms [7, 17, 19]. As our algorithm is not over-segmenting the input image and is filter-based, we obtain a better abstraction and thus final saliency map. The false positives due to a segmented textured background and inaccurately segmented object boundaries can be observed in Figure 1 (first row).

2 Related Work

We summarize here the recent research on salient region detection, which uses low-level, high-level, or combination of features. The studies on salient region detection or visual attention estimation can be grouped under two different goals. The first goal of automatic saliency detection techniques is to predict where humans might look in an image. In order to accomplish that, various eye-fixation datasets are collected through gaze tracking experiments and these datasets are analyzed from a visual saliency point of view. For example, to estimate image regions that might grab the attention of humans, researchers use spectral analyses [18], color co-occurence histograms [13], statistical analysis of saccadic eye movements [20], combination of low- and high-level saliency features [4], local and global patch rarities [5], and machine learning using low, medium-, and high-level features [12].

The second goal is to develop techniques to detect and segment the salient objects on a pixel-level precision. For this purpose, researchers fused additional information related to an object, such as context [21], concavity [14], objectness [6], and composition [8]. Another option to find visually salient objects is to structurally define the foreground and background regions. This can be achieved by using surround prior for background [22], employing global contrast based on histograms [7], exploiting uniqueness and distribution of foreground and background image segments [17], and considering background as a low rank matrix [19]. Our algorithm falls under this category and is compared in the subsequent sections with the best-performing techniques having the same goal.

In Section 1, we summarize the possible problems that the segmentation-based algorithms might encounter (see Figure 1). Consequently, we propose a multi-scale filtering based approach. Even though there are previous salient region detectors based on multi-scale [11, 9], and filtering [1, 2, 25], our algorithm differs from them with an adaptive combination of multi-scale maps and having intermediate-scale saliency detection that are explained in Section 3 and 4.

3 Frequency Domain Analysis

Spatial frequency content and its relation to visual saliency is extensively analyzed in [2]. Here, we present a multi-scale extension of their work providing further analysis on the importance of frequency content in saliency map extraction. We illustrate the analysis using one-dimensional Gaussian filters defined as follows:

$$F_n[x] = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\Big(-\frac{x^2}{2\sigma_n^2}\Big) \tag{1}$$

Here, $F_n[x]$ is a discrete Gaussian filter of size $2^n + 1$, and σ_n is related to the filter size $\sigma_n = 2^{n-1}$. In Figure 2 we illustrate the frequency spectra of 8 filters (an all-pass filter of size one, and 7 filters with $n = 1...7$). The filters in Figure 2 are capable of representing different image frequency intervals starting from the whole spectrum $[0, \pi)$ until $[0, \pi/2^n)$ by approximately halving at each step (see 2^{nd} to 5^{th} rows of Figure 1). We can generate band-pass filters using the Difference-of-Gaussian (DoG) idea in [2], which gives us a very flexible method capable of processing the whole frequency spectra with adjustable filters.

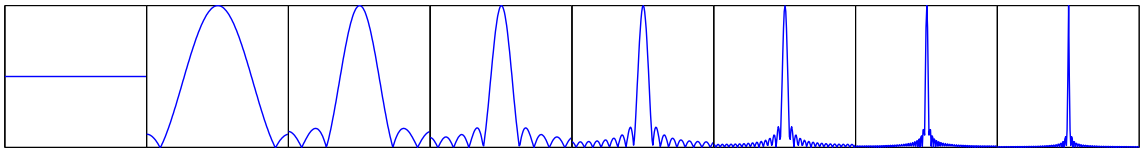
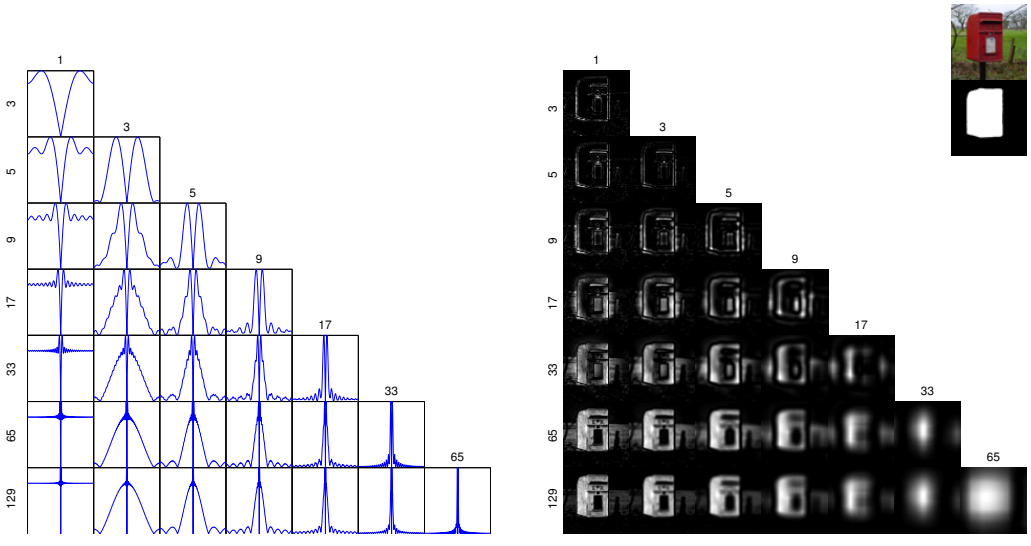


Figure 2: Frequency spectrum of 8 filters (same FFT size is used in the computation, y-axis is magnitude, x-axis is discrete time frequency for $(-\pi, \pi)$)

Frequency spectra of all possible filter differences and a 64×64 image that is processed with corresponding 2D DoG filters are shown in Figure 3. The salient red object is outlined when a small- and a large-scale filter, such as 1 and 129, is combined. This corresponds to subtracting the mean value of an image. The difference of two small-scale filters, such as 1 and 3, acts as an edge detector, which in turn helps us preserve the object boundaries in the estimated saliency map. The combination of two high-scale filters, such as 65 and 129, detects the salient object as a blob and provides uniformity on salient regions. Achanta et al.'s method [1] uses only the first column of difference filters thus lacks the ability of uniformly highlighting the salient regions as shown in the examples in Section 5.



(a) Filter combinations (b) Image abstraction

Figure 3: (a) The frequency content of the 1D filter differences (b) Same frequency content on a 2D image (a* channel of CIELab color space is used due to red-green contrast in the image, see top-right corner). The numbers represent the size of the filters that are combined to get that filter. For example the filter on bottom-left cornered is obtained by taking difference of filters with size 1 and 129.

4 Low-level Saliency Detection

Our model includes three main steps for saliency detection. First, we discuss our multi-scale filtering framework, which constitutes a saliency baseline. We then continue with two adaptive improvements, which are compactness and center prior measure.

4.1 Multi-scale Filtering

In order to decide if a pixel is salient, we can compare its color and position with its surroundings via Gaussian filtering. However, it blurs out the edge of an object, preventing us from extracting a saliency map with a pixel-level accuracy. In addition, a fixed-size filter cannot compare an object larger than the filter size with its surroundings. Thus, we introduce a multi-scale filtering approach.

In order to compute color contrast correctly, we first convert an image into CIELab space. We then filter all the channels of an input image I using a series of 2D Gaussian filters F_i $(i = 0, 1, 2, ...N_F)$ with different sizes. F_0 represents an all-pass filter (i.e. image itself). The remaining filters have a size of $(2^i + 1)$. The σ value of the filters are equal to 2^{i-1}. The number of filters (N_F) should depend on the input image size, because fixed-size F_{N_F} cannot cover whole frequency spectrum $[0, \pi)$ for images larger than F_{N_F}. In addition, a larger image requires a finer resolution in frequency domain, due to the increased amount of detail. F_{N_F} is the smallest filter that is larger than the image (i.e. F_{N_F} has a size of 513, if the input image is 400×300). As the last filter is very large, one should carefully filter image borders. Assuming that the border pixels belong to background regions in general, we replicate the border value of the image for correct filtering. All filtering operations are performed in frequency domain.

Pixel-wise square of differences of every possible filtered image pairs are computed (see Figure 3(b)) for the initial image abstraction. Each pair that uses filters F_i and F_j gives us a weak saliency map $\mathbf{W}_{ij}^C$, where C is either L*, a* or b*. A weak saliency map is an image representing a certain band-pass spatial frequency interval for a channel. The final saliency map is a weighted combination of these weak maps. We introduce two different adaptive measures, compactness K (scalar) and center prior $\mathbf{P}$ (same size as the image). They are explained in Section 4.2 and 4.3, respectively. The flow of this algorithm is as follows:

Initialize saliency map $\mathbf{S} = \mathbf{0}$
for $C = L*, a*, b*$ **do**
  for $i = 1, 2, 3, ...N_F$ **do**
    for $j = i + 1, i + 2, ...N_F$ **do**
      I_i = filter(C, F_i);
      I_j = filter(C, F_j);
      $\mathbf{W}_{ij}^C = (I_i - I_j) \bullet (I_i - I_j)$
      K= compactness($\mathbf{W}_{ij}^C$); (Section 4.2)
      $\mathbf{P}$= centerPrior($\mathbf{W}_{ij}^C$); (Section 4.3)
      $\mathbf{S} = \mathbf{S} + \mathbf{P} \bullet K\mathbf{W}_{ij}^C$;
    end
  end
end

Algorithm 1: Multi-scale filtering algorithm

Here $\bullet$ represents an element-wise multiplication.

4.2 Adaptive Compactness Measure

Our method combines all weak saliency maps to get a final saliency estimation (see Algorithm 1). Naive summation of these filter pairs might cause noisy saliency maps with false positives. In order to avoid that, we introduce a compactness measure, which evaluates the distribution of salient pixels around the image.

In order to compute the compactness, we first normalize a weak saliency map $\mathbf{W}_{ij}^C$ between 0 and 1 and get $\overline{\mathbf{W}}_{ij}^C$. We then calculate the center of mass (μ_x, μ_y) (used in Section 4.3) and spatial distribution (σ_x^2, σ_y^2) of a

weak map alone x and y image dimensions as follows:

$$\mu_x = \frac{1}{|T|}\sum_x\sum_y x\overline{\mathbf{W}}^C_{ij}(x,y)$$
$$\sigma_x^2 = \frac{1}{|T|}\sum_x\sum_y (x-\mu_x)^2\overline{\mathbf{W}}^C_{ij}(x,y) \tag{2}$$

Here, T is the sum of all values in $\overline{\mathbf{W}}^C_{ij}$. Similar equations are used for the computation of μ_y and σ_y^2. These variables measures the position and the compactness of a saliency map along each dimension as illustrated in Figure 4(b).

We also calculate the same variables using $1-\overline{\mathbf{W}}^C_{ij}$ and call them $\tilde{\mu}_x, \tilde{\mu}_y, \tilde{\sigma}_x^2, \tilde{\sigma}_y^2$. These variables represent the compactness of the background. Final compactness value is computed as follows.

$$K = \exp\Big(-k * \Big(\frac{\sigma_x^2\sigma_y^2}{\tilde{\sigma}_x^2, \tilde{\sigma}_y^2}\Big)\Big) \tag{3}$$

Here K is the compactness measure and k is an adjustment parameter. In our experiments, we find $k=4$ gives the best performance. It can be seen from (3) that, at low σ_x^2,σ_y^2 values (compact object) and high $\tilde{\sigma}_x^2$,$\tilde{\sigma}_y^2$ values (distributed background), compactness approaches to 1 and vice versa. An example of the compactness measure computation is given in Figure 4, where non-compact weak maps, such as Figure 4(b) are suppressed thus resulting a better saliency map estimation.

(a) (b) (c) 0.0208 (d) 0.4330 (e) 0.5987 (f) (g) (h)

Figure 4: Example weak maps an their compactness values. (a) Original image (b) the spatial meanings of each term in compactness computation. μ_x and μ_y are related to the center of mass of the object, and σ_x^2 and σ_y^2 are related to the 2D size of the object. (c-e) weak saliency maps with compactness values (f) output without compactness (g) output with compactness (h) ground truth

The compactness idea is investigated in [7] and [17] by combining spatial distribution and color differences on image segment level. However, our measure globally estimates the compactness and uses it as a contribution factor of that weak saliency map.

4.3 Adaptive Center Prior

Humans tend to look at the center of an image in order to understand it [12]. We can benefit from this property and assume a center prior in saliency map computations. Even tough it eliminates false positives, applying a non-adaptive center prior might also eliminates non-centered salient regions, which is not desirable.

Motivated by that, we introduce an adaptive center prior for each weak saliency map using the image compactness statistics we compute in Section 4.2.

$$\mathbf{P}(x,y) = \exp\Big(-\frac{(x-\mu_x)^2}{n*\sigma_x^2} - \frac{(y-\mu_y)^2}{n*\sigma_y^2}\Big) \tag{4}$$

Here, n is an adjustment factor, and $n=12$ is used throughout our experiments. The adaptive center prior $\mathbf{P}$ has the same size with the weak saliency map and multiplies it element-wise.

State-of-the-art methods, such as Shen and Wu [19], also employ center prior. Our method differs from a classical center assumption by taking distribution of salient pixels into account and adaptively shifting the prior mask. In addition, as center prior is computed for each weak saliency map, actual non-centered saliency information is not lost. In Figure 5, an illustration of center prior is given. As same statistics are used for both compactness and center prior measure, their effect on the image are not completely independent from each other. However, compactness eliminates undesired weak saliency maps in a global sense, which works locally for adaptive center prior.

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

Figure 5: Example weak maps an their adaptive center priors. (a) Original image (b) output without prior (c) output with prior (d) ground truth (e-h) two weak saliency maps and their priors

5 Results

We calculate the individual and combined performances of the different steps in our method. In addition, we compare our algorithm with four state-of-the-art methods, Cheng et al. (RC) [7], Perazzi et al. (SF) [17], and Shen and Wu (LR) [19], which are shown to outperform previous saliency detection methods on the MSRA-1000 dataset. We also present the results from Achanta et al. (FT) [2] as it has a similar filtering structure with our algorithm.

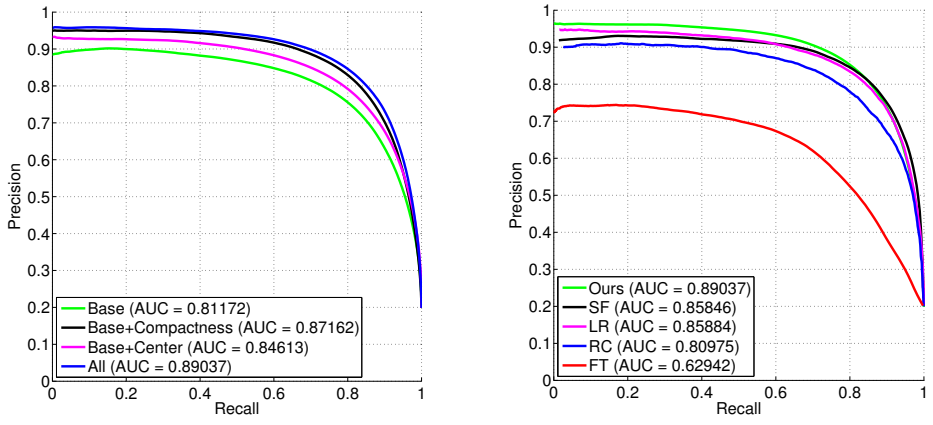
5.1 Contribution of Different Steps

In Section 4, we explain the three main steps of our method. In order to understand the contributions of these steps, we compute precision and recall curves. These calculations are performed by normalizing an output saliency map between 0 and 255, varying the threshold in the same interval and recording the precision-recall values. Figure 6(a) illustrates these curves along with the area-under-the-curve (AUC) values for different combinations of our algorithm steps. Base represents the performance of the multi-scale filtering output.

As we can see from Figure 6(a), our simple multi-scale filtering base performs well even without the adaptive improvements. Compactness step eliminates non-compact weak saliency maps and increases the base performance significantly. Adaptive center prior further refines our performance even though it rely on the same statistics with compactness measure.

5.2 Comparison to Other Methods

Similar to state-of-the-art methods, we evaluate our saliency detection performance calculating precision-recall curves. A comparison with other methods are given in Figure 6(b). Our multi-scale filtering algorithm significantly outperforms Achanta et al.'s filter-based method, FT, [2]. This results indicates that using and combining more filters recover a considerable amount of saliency information. Moreover, our method also outperforms the best-performing state-of-the-art techniques in both precision-recall curves and area-under-the-curve (AUC) values.



(a) (b)

Figure 6: (a) The precision-recall values for different algorithm steps using MSRA-1000 dataset (b) The precision-recall values for the MSRA-1000 dataset

Unlike [7, 17, 19], our accuracy does not depend on the performance of the segmentation method. We can see in Figure 8 some example images that illustrates these problems we indicate in Section 1. For example, in the first row, due to the thin structure of the bicycle, certain image segments do not properly follow the object boundaries, thus incorrectly segmented. In other methods, these errors are then transferred to the final saliency map as false positives or negatives. In our case, multi-scale filtering correctly recovered the structure of the object.

In the second, third and fourth rows of Figure 8, we see three examples, where our algorithm outputs a better saliency map than other methods. The textured structure cause RC and LR to produce unwanted salient segments in the background. In addition, due to multi-colored and textured structure of the foreground object, SF detects a non-uniform saliency map. Our algorithm is able to output a relatively uniform saliency map for this case.

Unnecessary segmentation of the uniform-colored areas, such as the sky and snow, are not properly suppressed in fifth row. Moreover, in RC, segmentation problems due to JPEG compression artifacts is transferred to the final map. In our map, we successfully suppress all uniform background regions.

The sixth and seventh rows illustrates the possible problems of pair-wise image segment comparison. In the other algorithms, only small red patches are marked as salient parts thus resulting a non-global saliency map. Our method produces a uniform salient object, event though object is multi-colored.

In the eighth row, textured areas, that can easily be smoothed by filtering are erroneously detected (RC and LR) or the right-most object that has similar colors with the background is not detected (SF). This object differs from its background not in terms of color but in terms of texture. As our method effectively measures different bands of the frequency spectrum, one of the weak saliency maps separates the right-most object from the background providing an accurate detection. In the last row, our adaptive compactness step overemphasizes a very compact and salient region of the image resulting an inaccurate saliency map.

5.3 Adaptive Thresholding

In addition to constant thresholding, we compare the precision-recall values of our method with the state-of-the-art algorithms estimating an adaptive threshold. As before, in order to find a threshold we use Otsu's method [16] for all methods and compare them in Figure 7. We also combine precision and recall with the following F-measure that is commonly used in saliency detection literature [2, 7, 17, 19].

$$F = (1 + \beta^2)\frac{Precision * Recall}{\beta^2 * Precision + Recall} \tag{5}$$

Here, adjusting parameter $\beta^2 = 0.3$. As we can see from Figure 7 our method slightly has higher F-score than other methods. One advantage of our saliency maps is the precision and recall values are well-balanced using Otsu's threshold.

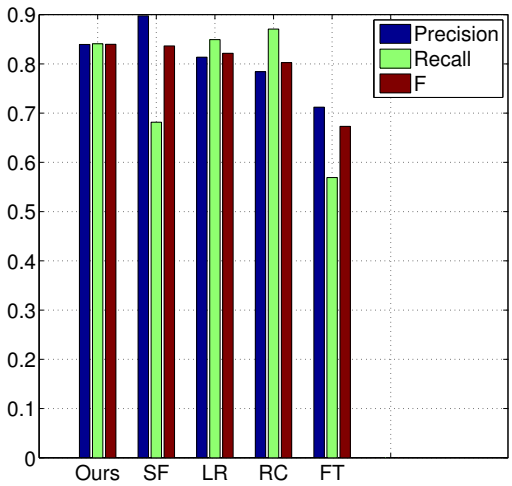
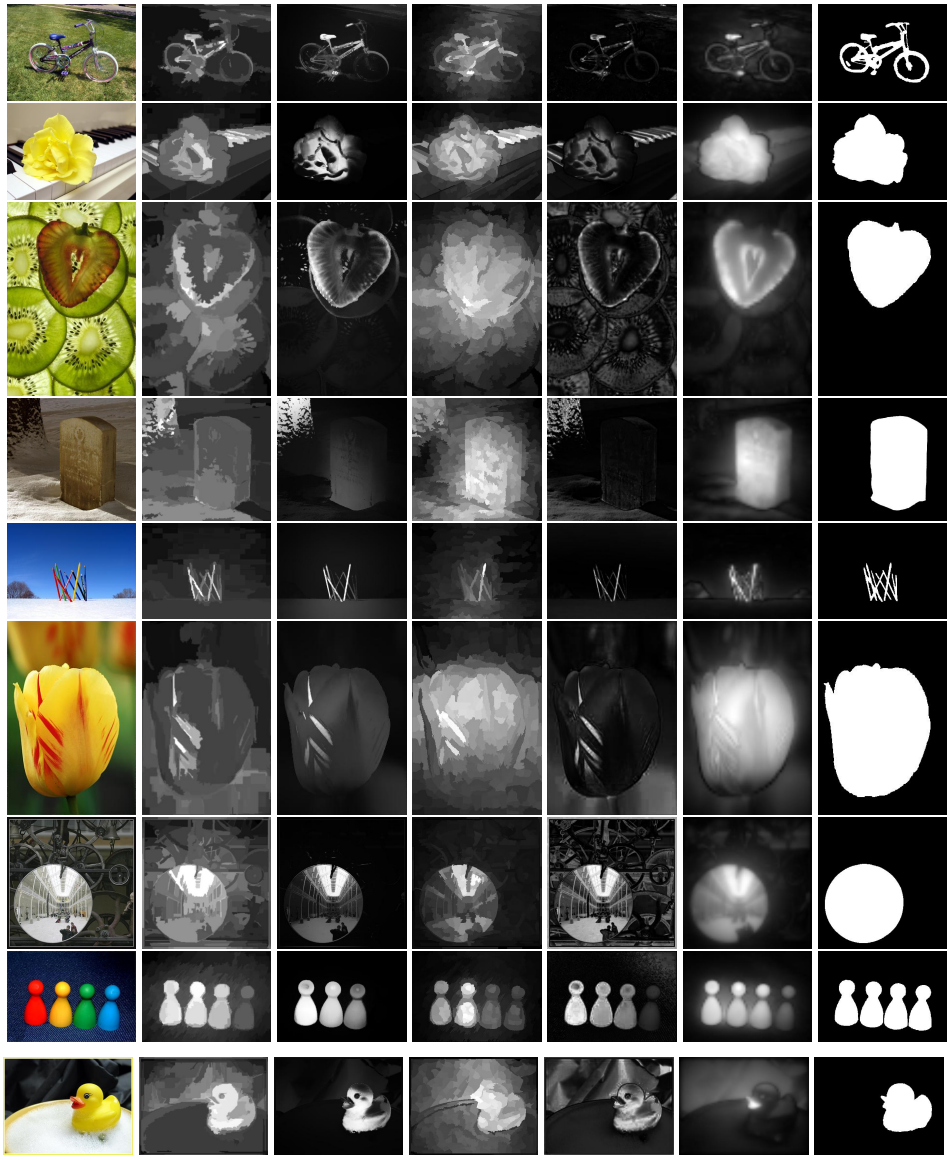


Figure 7: The precision-recall values for adaptive thresholding, from high to low in terms of F-Score.



(a) Original (b) RC [7] (c) SF [17] (d) LR [19] (e) FT [2] (f) Ours (g) Ground truth

Figure 8: Example images from MSRA-1000 dataset and estimated saliency maps for different methods.

6 Conclusion and Future Work

Even though it is an easy task for humans, automatic salient region detection is a challenging problem in computer vision. In this paper, we attempt to solve this problem by introducing a multi-scale filtering approach, which produces weak saliency maps of an image. We then combine these maps using an adaptive compactness measure and center prior and acquire a final pixel-precision saliency map. We indicate the possible problems the segmentation-based methods might have and show that our algorithm outperforms them using varying and adaptive thresholds to compute precision and recall values.

For future work, in order to combine weak saliency maps, a non-heuristic way of combining, such as learning, can be employed. A similar argument can go for the compactness measure. However, to avoid curse of dimensionality, for both of these works, a larger pixel-precision dataset is required. Our method focus only on low-level properties of the image and do not deal with image context. Incorporating semantic information can also be a future research topic to follow.

References

[1] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. In *Proceedings of International Conference on Computer Vision Systems*, pages 66–75, 2008. 1, 2, 3

[2] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned Salient Region Detection. In *Proceedings of IEEE CVPR*, pages 1597 – 1604, 2009. 1, 2, 3, 6, 7, 9

[3] R. Achanta and S. Süsstrunk. Saliency Detection for Content-aware Image Resizing. In *Proceedings of IEEE ICIP*, pages 1001–1004, 2009. 1

[4] A. Borji. Boosting bottom-up and top-down visual features for saliency estimation. In *Proceedings of IEEE CVPR*, pages 438–445, 2012. 1, 2

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

[6] K. Chang, T. Liu, H. Chen, and S. Lai. Fusing generic objectness and visual saliency for salient object detection. In *Proceedings of IEEE ICCV*, pages 914–921, 2011. 2

[7] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In *Proceedings of IEEE CVPR*, pages 409–416, 2011. 2, 5, 6, 7, 9

[8] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun. Salient object detection by composition. In *Proceedings of IEEE ICCV*, pages 1028–1035, 2011. 2

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

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

[11] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on PAMI*, 20(11):1254–1259, 1998. 1, 2

[12] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *Proceedings of IEEE ICCV*, pages 2106–2113, 2009. 1, 2, 5

[13] S. Lu and J. Lim. Saliency modeling from image histograms. In *Proceedings of ECCV*, pages 321–332, 2012. 2

[14] Y. Lu, W. Zhang, H. Lu, and X. Xue. Salient object detection using concavity context. In *Proceedings of IEEE ICCV*, pages 233–240, 2011. 2

[15] A. Oliva, A. Torralba, M. Castelhano, and J. Henderson. Top-down control of visual attention in object detection. In *Proceedings of IEEE ICIP*, volume 1, pages 253–256, 2003. 1

[16] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979. 7

[17] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *Proceedings of IEEE CVPR*, pages 733–740, 2012. 2, 5, 6, 7, 9

[18] B. Schauerte and R. Stiefelhagen. Quaternion-based spectral saliency detection for eye fixation prediction. In *Proceedings of ECCV*, pages 116–129, 2012. 2

[19] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *Proceedings of IEEE CVPR*, pages 853–860, 2012. 1, 2, 5, 6, 7, 9

[20] X. Sun, H. Yao, and R. Ji. What are we looking for: Towards statistical modeling of saccadic eye movements and visual saliency. In *Proceedings of IEEE CVPR*, pages 1552–1559, 2012. 2

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

[22] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *Proceedings of ECCV*, volume 7574, pages 29–42. 2012. 2

[23] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *Proceedings of IEEE ICCV*, pages 1–6, 2007. 1

[24] Y. Xiang and M. S. Kankanhalli. Video retargeting for aesthetic enhancement. In *Proceedings of ACM Multimedia*, pages 919–922, 2010. 1

[25] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *Proceedings of ACM Multimedia*, pages 815–824, 2006. 2