Saliency Detection with Flash and No-flash Image Pairs

Shengfeng He and Rynson W.H. Lau

Department of Computer Science
City University of Hong Kong, Hong Kong, China
shengfeng_he@yahoo.com, rynson.lau@cityu.edu.hk

Abstract. In this paper, we propose a new saliency detection method using a pair of flash and no-flash images. Our approach is inspired by two observations. First, only the foreground objects are significantly brightened by the flash as they are relatively nearer to the camera than the background. Second, the brightness variations introduced by the flash provide hints to surface orientation changes. Accordingly, the first observation is explored to form the background prior to eliminate background distraction. The second observation provides a new orientation cue to compute surface orientation contrast. These photometric cues from the two observations are independent of visual attributes like color, and they provide new and robust distinctiveness to support salient object detection. The second observation further leads to the introduction of new spatial priors to constrain the regions rendered salient to be compact both in the image plane and in 3D space. We have constructed a new flash/no-flash image dataset. Experiments on this dataset show that the proposed method successfully identifies salient objects from various challenging scenes that the state-of-the-art methods usually fail.

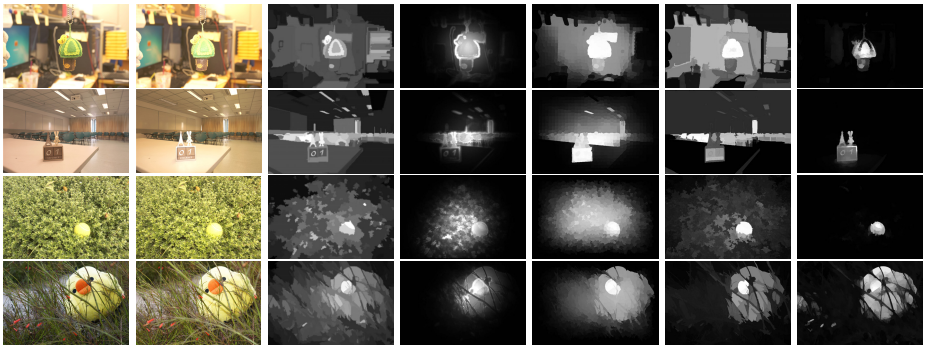
Keywords: Saliency detection, Flash photography, Background elimination, Surface orientation.

1 Introduction

The underlying goal of saliency detection is to locate regions or objects in an image that gain the viewer's visual attention. There is a wide range of computer vision, graphics and multimedia applications of saliency detection, including classification [31], image segmentation [10], image retrieval [6], and content-aware image/video resizing [23].

Numerous studies in psychological science [29,13,25] have shown that the most influential factor to visual saliency in the human visual system is *contrast*. As such, a lot of algorithms that are based on some kind of contrast priors have been proposed to detect salient objects from images. The most widely adopted visual attribute is color contrast [7,26,40]. However, contrast based methods are usually limited to scenes with simple background or with high contrast between foreground and background. In the case where salient objects cannot be clearly distinguished from a complex background or the background has a similar color as the foreground, detecting salient objects become very challenging with existing methods, as shown in Figures 1b to 1e. Background priors [37,41] have been used to tackle this limitation by considering both foreground and background cues in a different way with assumptions. However, these methods fail

D. Fleet et al. (Eds.): ECCV 2014, Part III, LNCS 8691, pp. 110–124, 2014.



(a) Input images **(b)** RC [7] **(c)** PCA [21] **(d)** GMR [41] **(e)** HS [40] **(f)** Ours

Fig. 1. Comparison with the state-of-the-art methods [7,21,41,40] on the output saliency maps. (a) are the input flash and no-flash images. These scenarios contain challenging factors, including similar colors between the salient object and its surrounding, and complex background. All the state-of-the-art methods (b-e) fail in these scenarios. (f) Our method considers the rough depth, surface orientation contrast, color contrast and compactness together to detect salient objects from the foreground (rows 1 and 2) as well as from the background (rows 3 and 4).

when one of the assumptions is invalid, as illustrated in Figure 1d. Orientation contrast [24,30] has also been shown to play an important role in the human visual system. However, this principle is not appropriate for detecting salient objects in practice. This is because orientation contrast (as well as other related cues such as curvedness [36] and pattern [21]) focuses on object boundaries, leading to attenuated object interior when detecting homogeneous regions, as shown in Figure 1c.

In this paper, we propose a new approach to saliency detection using a flash/no-flash image pair. Our approach is based on two observations. First, only the foreground objects are significantly brightened by the flash, as they are nearer to the camera, while the background is less affected by the flash. Second, the brightness variations on object surfaces due to the flash provide hints on surface orientation. These two observations suggest additional cues for saliency detection. First, we can make use of the difference between the flash and no-flash images from the first observation to extract the foreground layer from the image pair. However, background near to the foreground objects (in terms of depth) may not be trivially separated, and not all objects in the foreground layer are necessarily salient. As a complement to the first observation, we estimate the surface orientation contrast and spatial priors from the second observation. These surface orientation contrast and spatial priors help detect and recover salient regions when background prior is invalid. As demonstrated in [9], surface orientation contrast is as effective as 2D orientation contrast in attracting visual attention in the human visual system. However, unlike 2D orientation contrast which focuses on object boundaries, our surface orientation contrast identifies homogeneous salient regions. The spatial priors focus on three types of compactness of the salient objects – regions that are compact in both the image plane and the 3D space tend to be salient.

In order to evaluate the proposed method, we have constructed a dataset of 120 flash/no-flash image pairs. Flash/no-flash photography has been studied for decades, but to the best of our knowledge, this is the first flash/no-flash dataset created for saliency detection. Our experimental results from this dataset show that the proposed method is able to detect salient objects even when a salient object and its surrounding have similar colors or when the background is very complex, which is extremely challenging for the state-of-the-art methods.

2 Related Work

Saliency Detection. Existing bottom-up saliency detection methods commonly utilize low-level processing to determine saliency within a certain context. They are mainly based on some contrast priors. Depending on the extent of the context where saliency is computed, these methods can be roughly categorized into *local methods* and *global methods*. Comprehensive literature review on these saliency detection methods can be found in [4,35]. Here, we briefly review the representative ones.

Local methods compute saliency of an image region with respect to a small neighborhood. An earlier local saliency detection method [14] is based on a biologically-plausible architecture [16]. It uses an image pyramid to compute color and orientation contrasts. Ma and Zhang [20] combine local contrast analysis with a fuzzy growth model. Harel et al. [11] propose a graph-based random walk method using multiple features. As these methods are based on computing local contrast, they are only sensitive to high frequency content like image edges or noise, and they attenuate any homogenous interior regions.

Global methods estimate saliency by considering contrast relations over the entire image. Achanta et al. [1] detect salient regions by computing color deviation from the mean image color on a per-pixel basis. Cheng et al. [7] propose a fast color histogram based method, and compute saliency based on dissimilarity among the histogram bins. To take into account spatial relationships, Perazzi et al. [26] apply two contrast measures based on the uniqueness and spatial distribution of elements. Yan et al. [40] propose a hierarchical framework to reduce the effect of small-scale structures on saliency detection. Despite the demonstrated success, it is still difficult to distinguish salient objects from clustered and textured background by using global contrast alone.

Recently, other than contrast priors, background priors are also used to reduce the distraction from the background. These methods are generally based on assumptions such as having a large and homogeneous background [37] or image boundaries mostly belonging to the background [41]. They become unreliable when these assumptions are not valid. Another method proposes to detect salient objects with the depth cue from stereo images [22]. Although background distraction can be mitigated, its performance strongly depends on the quality of the disparity map. In addition, it fails if the salient object cannot be distinguished in the depth level. Despite many recent improvements, the saliency detection problem on complex background is still highly ill-posed. Our method introduces additional cues from the flash image, and integrates the background, contrast and spatial priors together to address this problem. Our results show that it is able to detect saliency robustly.

Flash Photography. Many applications have adopted flash-based techniques to address different problems in recent years. Eisemann and Durand [8] use joint-bilateral filtering for flash image enhancement. Agrawal et al. [3] propose a gradient projection scheme to remove photographic artifacts introduced by the flash. Petshnigg et al. [27] show that additional information from the flash image can benefit many applications, such as denoising, detail transfer, white balancing, continuous flash, and red-eye correction. Some other applications extract cues by changing the light source position. In multiple flash imaging [28], the shadows casted by the flash are used to extract depth edges for non-photorealistic rendering. Active illumination [18] has also been used for depth recovery by adjusting the light source position. In this paper, we use flash/no-flash images for a different application - saliency detection.

The flash/no-flash idea has also been used for matting [33], foreground extraction [34] and stereo matching [42]. The work most related to ours is foreground extraction [34], in which the rough depth information and a motion compensation model are used to extract foreground objects with some amount of motion. However, this method is limited to very distant background, and the whole foreground layer will be segmented as long as it receives the flash light (even though some of the foreground objects may not necessary be salient). The proposed saliency detection method may extract salient objects from the foreground as well as from close background by taking four complementary cues into account.

3 Flash/No-flash Saliency Detection

The proposed method consists of three complementary priors: background, contrast and spatial priors. The difference between the flash and no-flash images provides reliable information for foreground-background separation. We treat this rough depth information as the background prior. We also use two contrast priors in our approach: surface orientation and color contrasts. Surface orientation contrast is used to differentiate objects with different orientations. Finally, we use the spatial priors to constraint the rendered salient pixels to be compact in both the image plane and the 3D space. These three priors are combined to form the final saliency map.

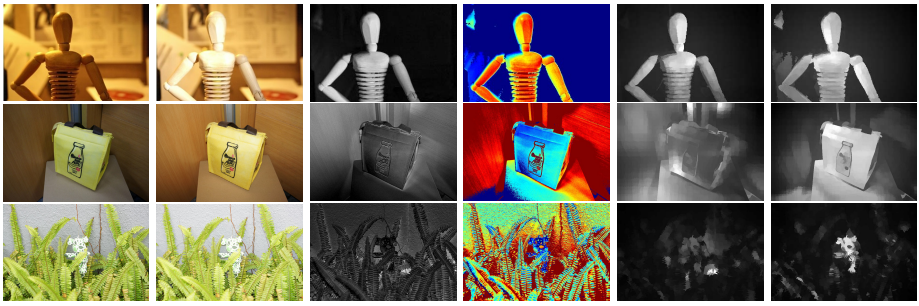
We first model the flash and no-flash images as follows. Considering Lambertian surfaces, the intensity I_n of a pixel p in the no-flash image is determined according to ambient illumination, surface shape, and reflectivity as:

$$I_n(p) = \eta \cdot L_a \cdot U, \tag{1}$$

where η is a proportionality constant between scene radiance and irradiance. L_a is the intensity of the ambient illumination at p, while U is the surface reflectivity at p. The intensity I_f of a pixel p in the flash image is modeled as:

$$I_f(p) = \eta \cdot L_a \cdot U + \eta \cdot L_f \cdot \frac{\cos\theta}{r^2} \cdot U, \tag{2}$$

where L_f is the flash intensity, θ is the angle between the direction of the flash and the surface normal at p. r is the distance from the flash unit to p. The following subsections introduce our difference image, ratio image and three proposed priors.



(a) Input images **(b)** Diff. image **(c)** Ratio image **(d)** Rough depth saliency **(e)** Surf. orien. contrast

Fig. 2. The saliency maps produced by the difference and ratio images: (a) the input images; (b) the difference images; (c) the background-eleminated ratio images, mapped to a color space so that surfaces of similar orientations can be easily seen; (d) the rough depth saliency. Although the result in the first row is good, the other two are not as the salient objects cannot be distinguished in depth level; (e) the surface orientation contrast saliency.

3.1 Saliency from the Difference Image

To cope with the difficulties in detecting saliency in complex scenarios, existing methods usually assume that salient objects are placed in the foreground [37,41]. Although the role of depth cues in the human visual system is still under debate [39], this assumption is practical since people tend to arrange the target object in a photo at a different depth level from the background. Hence, we extract the foreground layer based on this assumption. Our idea to achieve this is by computing a difference image, D, from Eq. (1) and (2) as:

$$D(p) = I_f(p) - I_n(p) = \eta \cdot L_f \cdot \frac{\cos\theta}{r^2} \cdot U. \tag{3}$$

We can see that D is only related to a single light source, the flashlight, and the flash intensity L_f falls off quickly with distance r, due to the inverse square law.

Since we assume that the background scene is further away from the camera than the foreground objects, the appearance change of the background in the flash and no-flash images is expected to be much smaller. The first row of Figure 2 shows such an example. Consequently, the difference image D provides us a very robust cue for foreground extraction. We consider the difference image D as a rough depth map to obtain rough depth saliency.

Similar to previous approaches [21,41], we first employ SLIC super-pixels [2] to segment the input image, and then determine the salient regions. In order to obtain better foreground-background separation, we combine the flash image, the no-flash image and the ratio image (see Section 3.2) into a three-channel image for segmentation. A rough depth saliency value for a region R_i is computed based on its contrast with all other regions in the image as:

$$S_d(R_i) = \sum_{j=1}^{M} w(R_j)\phi(i,j)\|d_i - d_j\|, \tag{4}$$

where M is the total number of regions. d_i and d_j are the mean rough depth values of regions R_i and R_j, respectively. $w(R_j)$ is the number of pixels in R_j. We consider that a larger region contributes higher to saliency than a smaller one. $\phi(i,j)$ is defined as $\exp(-\frac{\|R_i - R_j\|^2}{2\sigma_r^2})$ to control the influenced distance, where $\|R_i - R_j\|^2$ is the L_2 distance between the centers of R_i and R_j, and σ_r is set to 0.45 in this paper. With $\phi(i,j)$, regions that are closer to R_i have higher influence on the computed saliency.

The rough depth can be used to identify foreground objects. However, it only considers the distance and is not sufficient for all cases. The second and the third rows of Figures 2b and 2d show two failure examples – the background is not sufficiently distant or the salient object is placed in close background. We alleviate this limitation by considering contrast and spatial priors.

3.2 Saliency from the Ratio Image

The effectiveness of orientation cues in the image plane has been demonstrated both in psychological science [39] and computer vision [14]. It is natural to ask if the same conclusion holds for surface orientation in a 3D scene. Experiments have been conducted on surface orientation in psychological science [9,12,38], showing that surface orientation is more important than other 3D factors such as depth and shading to visual attention. However, traditional techniques based on analysing a single image are not able to recover surface orientation. We notice that the ratio value obtained from the flash/no-flash images provides cues on surface orientations.

We divide Eq. (2) by Eq. (1) and take the logarithm to obtain the ratio image T as:

$$T(p) = \log \frac{I_f(p)}{I_n(p)} = \log(1 + \frac{L_f}{L_a} \cdot \frac{\cos\theta}{r^2}). \tag{5}$$

To avoid division-by-zero, the ratio image is defined as $T = \log(I_f + \epsilon) - \log(I_n + \epsilon)$ in our implementation, where ϵ is a small positive value. We can see that $T(p)$ is essentially independent of surface reflectivity. Instead, it varies according to depth and surface orientation. Our key finding here is that two neighboring regions with different ratio values indicate that they either have different surface orientations or are at different depth levels, while two neighboring regions with similar ratio values likely belong to the same surface. We note that Eq. (5) may not be accurate for non-Lambertian surfaces, and given two neighboring regions with different ratio values, we are not able to tell if they have different surface orientations or are at different depth levels. Nevertheless, the ratio values can be used to differentiate different surfaces. By considering also the information from the difference image, we may obtain the surface orientation contrast.

Although the properties of the ratio image hold whether the background is distant or not, removing distant background may improve detection performance. Here, we aim at eliminating the influence of distant background, while making sure that the computation of surface orientation saliency would not be affected by close background. To do this,

we determine the first local minimum β of a 128-bin histogram of the difference image D. The histogram is smoothed using a Gaussian kernel. We then apply a threshold of 0.6β to remove pixels that are highly likely belonging to the distant background. Finally, we obtain a new background-eliminated ratio image $\hat{T}$ as:

$$\hat{T}(p) = \begin{cases} T(p) & \text{if } D(p) > 0.6\beta \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

Based on $\hat{T}$, we may compute the surface orientation contrast between R_i and all other regions in the image as:

$$S_s(R_i) = \sum_{j=1}^{M} w(R_j)\phi(i,j)\|\hat{t_i} - \hat{t_j}\|, \tag{7}$$

where $\hat{t_i}$ and $\hat{t_j}$ are the mean background-eliminated ratio values of R_i and R_j, respectively. As shown in Figure 2e, our surface orientation contrast produces much better salient maps than the rough depth in the scenes with close background or with salient objects that cannot be differentiated in depth level. Note that this background-eliminated ratio image is only used in here. When computing the spatial priors in Section 3.4, we use the original ratio image T instead.

3.3 Color Saliency

Although the above background prior and the surface orientation contrast can detect foreground and distinct surfaces, it is not sufficient for all cases. We further consider color contrast as another contrast prior. We note that the flash image and the no-flash image may each have different color contrast. Hence, for each region in the two images, we compute its color contrast with each of the other regions in both two images in order to reduce the background contribution, since the background changes relatively small in both images. In other words, we compute two inter-color contrasts as the sum of L_2 distances in CIE LAB color-space for each pair of regions in the two images:

$$S_n(R_i) = \sum_{j=1}^{M} w(R_j)\phi(i,j)(\|n_i - n_j\|^2 + \|n_i - f_j\|^2), \tag{8}$$

$$S_f(R_i) = \sum_{j=1}^{M} w(R_j)\phi(i,j)(\|f_i - f_j\|^2 + \|f_i - n_j\|^2), \tag{9}$$

where f_i and n_i are the mean color values of R_i in the flash and no-flash images, respectively. Likewise, f_j and n_j are the mean color values of R_j in the flash and no-flash images. The final color contrast is the average of the two inter-color contrasts, i.e., $S_c(R_i) = \frac{S_n(R_i)+S_f(R_i)}{2}$.

3.4 Spatial Priors

Compactness is the main guiding principle used in spatial priors. Previous approaches consider color-spatial compactness [26,32]: generally, colors belonging to the background have high spatial variance, while salient objects are typically more compact.

To extend this idea further, we consider two more types of compactness to describe salient objects. A salient object should be compact both in the image plane and in the 3D space. (As mentioned in Section 3.2, the ratio image T is used to separate different surfaces.) Hence, three types of compactness are used in our approach:

- *Color-spatial compactness*. A region with a low spatially distributed color is considered as more salient than one with high spatially distributed colors.
- *Color-surface compactness*. A region with an average color that mainly appears in the surface that the region belongs to should be considered as salient.
- *Surface-spatial compactness*. A region with low spatially distributed ratio values is considered as more salient than one with high spatially distributed ratio values.

Here, we may treat the ratio value as *Z* dimension information to compute 3D compactness. Although this may not be accurate, regions belonging to the same surface usually have similar ratio values. Hence, we compute the color-spatial compactness as:

$$S_{cs}(R_i) = \sum_{j=1}^{M} \phi(c_i, c_j) \|x_j - \mu_i^{cs}\|^2, \tag{10}$$

where x_j is the center of region R_j. $\phi(c_i, c_j) = \exp(-\frac{\|c_i - c_j\|^2}{2\sigma_c^2})$ describes the similarity of two colors. Note that the color used here is the average color of the flash/no-flash images. We define $\mu_i^{cs} = \sum_{j=1}^{M} \phi(c_i, c_j) x_j$ as the weighted mean position of color c_i to take into account all color contributions. In our implementation, σ_c is set to 20.

The color-surface compactness can be computed by substituting the mean position μ_i^{cs} of Eq. (10) by the mean ratio $\mu_i^{cp} = \sum_{j=1}^{M} \phi(c_i, c_j) t_j$ as:

$$S_{cp}(R_i) = \sum_{j=1}^{M} \phi(c_i, c_j) \|t_j - \mu_i^{cp}\|^2. \tag{11}$$

The surface-spatial compactness is defined similarly as:

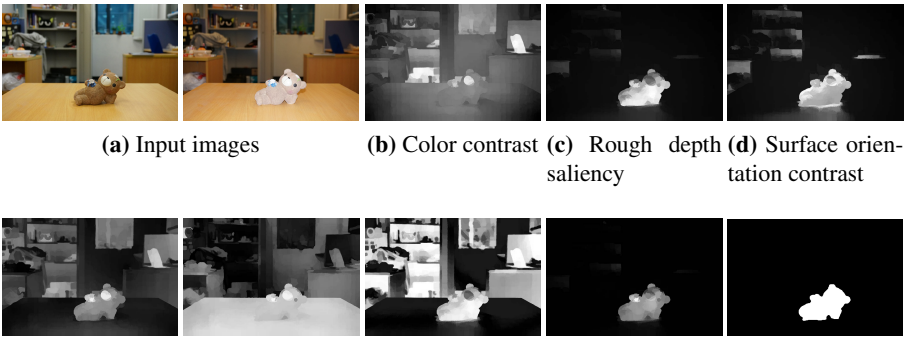
$$S_{ps}(R_i) = \sum_{j=1}^{M} \phi(t_i, t_j) \|x_j - \mu_i^{ps}\|^2, \tag{12}$$

where $\mu_i^{ps} = \sum_{j=1}^{M} \phi(t_i, t_j) x_j$ is the weighted mean position of ratio t_i. $\phi(t_i, t_j) = \exp(-\frac{\|t_i - t_j\|^2}{2\sigma_t^2})$ is the similarity between ratio values t_i and t_j. σ_t is set to 0.2 in all our experiments.

Note that in Eq. (10)-(12), a lower value indicates a higher compactness and hence higher saliency. We use an exponential function to emphasize on small values. The final compactness is obtained by combining the three normalized compactness values as:

$$S_m(R_i) = \exp(-k(S_{cs} + S_{cp} + S_{ps})), \tag{13}$$

where k is set to 0.15 in all the experiments.



(a) Input images **(b)** Color contrast **(c)** Rough depth saliency **(d)** Surface orientation contrast

(e) Color-spatial compactness **(f)** Color-surface compactness **(g)** Surface-spatial compactness **(h)** Final saliency **(i)** Ground Truth

Fig. 3. Outputs of various components used in our approach

3.5 Final Saliency Map

We have now introduced four cues used in our approach, i.e., rough depth, surface orientation, color contrast and spatial compactness. Although each of them has its own advantages and limitations, their advantages largely compliment each other's limitations. As we seek to obtain objects that are salient in all four cues, we multiple the normalized saliency maps as:

$$S_f(R_i) = S_d(R_i) \cdot S_s(R_i) \cdot S_c(R_i) \cdot S_m(R_i). \tag{14}$$

The final saliency map is again normalized to $[0, 1]$.

In order to robustly detect salient regions of different sizes, we use a multi-scale approach to saliency detection. All the saliency maps are computed using four different numbers of super-pixels: 100%, 75%, 50% and 25%. The final saliency map is the average of these four scales.

Figure 3 shows the outputs of individual components of our saliency detection method. Note that none of the fours cues alone suffices to achieve good results.

4 Experiments and Results

This section evaluates the proposed saliency detection method on a new flash/no-flash database of 120 image pairs with manually labeled binary ground truth. We have implemented the proposed method in Matlab using a single core and tested it on a PC with an Intel i5 3.3GHz CPU and 8GB RAM. Our algorithm takes on average 2.1s to process one pair of images of resolution 400×300.

4.1 The Flash/No-flash Image Database

Although flash photography has been studied for decades, there is not a standard dataset for this purpose. In order to evaluate the proposed method, we have constructed a

(a) Successful label **(b)** Salient object mask **(c)** Unsuccessful label

Fig. 4. Two example images of our dataset. (a) is consistently labeled by all users and included in our database. (b) is the ground truth manually segmented by a user. (c) is a failure example. The users have different opinions on what should be the salient objects.

flash/no-flash dataset with ground truth, following the benchmark dataset building procedure [19].

We took 170 photo pairs with a tripod, and asked three separate users to mark the most salient object in each image pair using a rectangle. For images with inconsistent users selections, we remove them from consideration. The most consistent 120 image pairs, according to the overlapping ratio of the rectangles, were chosen to form the database. One user was asked to manually segment the salient object(s) using *Adobe Photoshop* to serve as the ground truth. Figure 4 shows a successful example and an inconsistent labeled example while constructing our database. The flash/no-flash dataset with ground truth can be downloaded from our project website [1].

4.2 Evaluation on the Benchmark

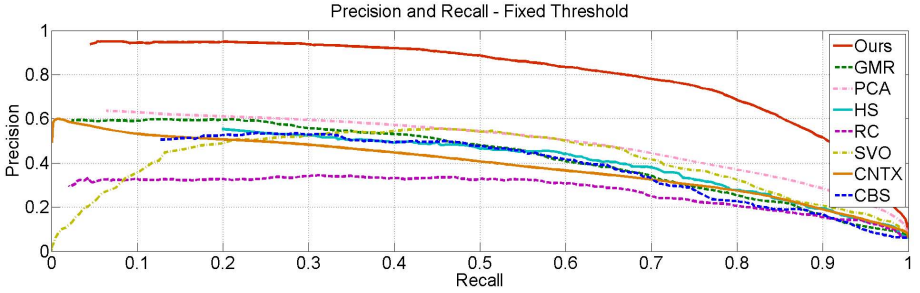
We have compared the proposed method with seven state-of-the-art methods, the top four algorithms (RC [7], SVO [5], CNTX [10], CBS [15]) according to [4], plus three latest algorithms (GMR [41], HS [40], PCA [21]). The implementations provided by the authors were used for fair comparison. Similar to previous work [26,40], we quantitatively evaluate the performances of these methods by measuring the *precision-recall* values. Precision indicates the percentage of the output salient pixels that are correct, while recall indicates the percentage of the ground truth pixels detected. Since we have a pair of two input images in each test, the performances of the seven state-of-the-art methods on these two images may not be consistent. We use the average score obtained from each pair of images.

We evaluate all the methods using the Precision-Recall curves and the F-measure metrics, which is defined as:

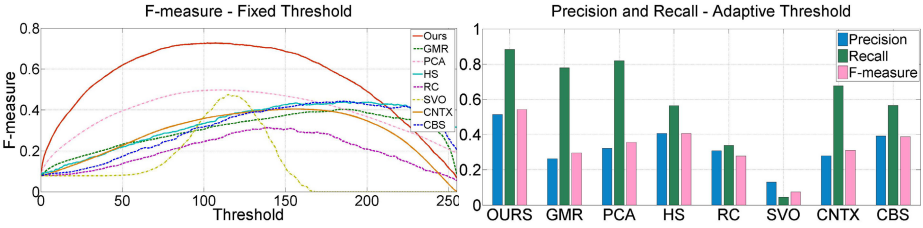
$$F_\beta = \frac{(1+\beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall}, \tag{15}$$

where β^2 is set to 0.3 to emphasize on precision [1].

[1] `http://www.cs.cityu.edu.hk/~shengfehe2/saliency-detection-with-flash-and-no-flash-image-pairs.html`

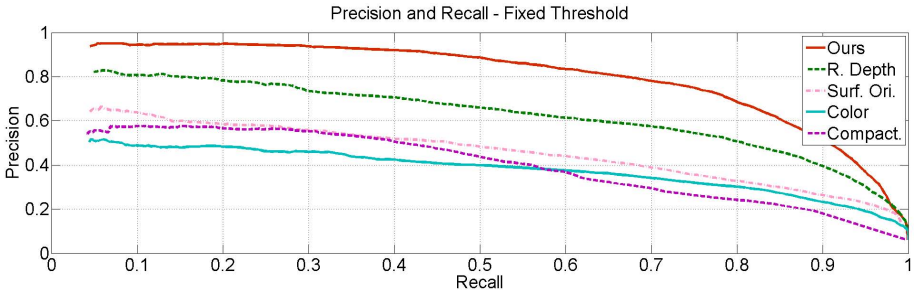


(a)

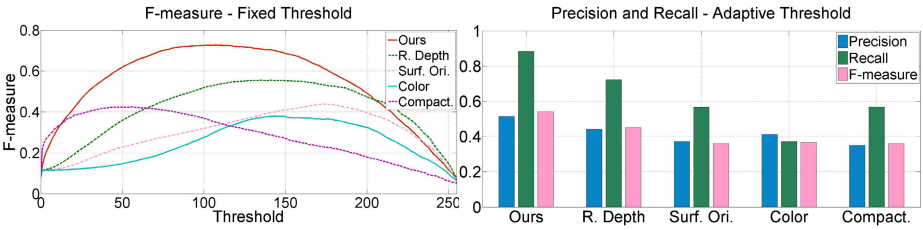


(b) (c)

Fig. 5. Comparison with the state-of-the-art methods. (a-b) are the precision-recall curves and F-measure curves using 255 fixed thresholds. (c) shows the precision, recall and F-measure using an adaptive threshold.

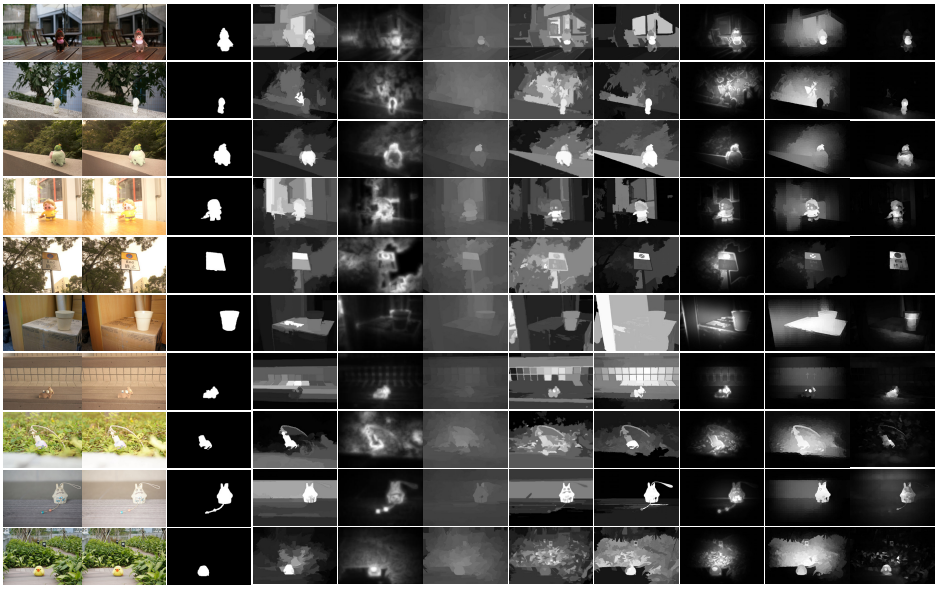


(a)



(b) (c)

Fig. 6. Performance of each component of the proposed method. (a-b) are the precision-recall curves and F-measure curves using 255 fixed thresholds. (c) shows the precision, recall and F-measure using an adaptive threshold.



(a) Input images **(b)** GT **(c)** CBS **(d)** CNTX **(e)** SVO **(f)** RC **(g)** HS **(h)** PCA **(i)** GMR **(j)** Ours

Fig. 7. Comparison with the state-of-the-art methods. The proposed method consistently produces better saliency results in distant as well as close backgrounds.

We conduct two experiments using different threshold criteria. In the first experiment, we binarize the saliency map for every threshold in the range of [0, 255]. Figures 5a and 5b show that the proposed method achieves the best performance in most of the recall rates and thresholds. In the second experiment, we use an image dependent adaptive threshold [1], which is computed as twice the mean value of the saliency map. Figure 5c shows that the proposed method outperforms the state-of-the-art methods in all three aspects. We further analyze the effectiveness of each component. Figure 6 shows that these components complement each other. Thus, the final saliency outperforms individual ones.

Figure 7 shows the qualitative comparison of all the methods. As the seven existing methods produce two results from each input image pair, we choose the result with higher F-measure value using the adaptive threshold. We can see that the proposed method can detect salient objects properly from complex background and in scenes with similar color between salient objects and surrounding regions. Benefiting from four complementary cues, the proposed method still works while the background prior is invalid to certain extent, e.g., the background is not distant enough or the salient objects are not distinct in depth level (rows 6 to 10 in Figure 7).

4.3 Limitations and Discussion

In general, the proposed method suffers from similar limitations as flash/no-flash photography. First, the difference image and the ratio image benefit saliency detection only when the salient objects are reachable by the flash. Second, the proposed method is

only suitable for static scenes. Extending it to cover non-static scenes may be a possible topic for future work. Third, the proposed method may not produce reliable results when applied to scenes with strong specular surfaces.

Despite these limitations, the proposed method can be applied in mobile phones, where the shutter speed is typically fast. We can easily take two consecutive (no-flash followed by flash) images, without the user being aware that two images are taken instead of one. The proposed method may benefit a lot of existing mobile applications that require automatic detection / extraction of salient objects, such as cloud-based object detection/retrieval and object level image manipulation. On the other hand, integrating with the dark flash [17] is another possible topic for future work, which is able to extract salient objects without revealing the detectors.

5 Conclusion

In this paper, we have proposed a new saliency detection approach using flash/no-flash image pairs. The proposed method takes advantages of the additional information obtained from the two image pair to discriminate the salient object from the background and surfaces with different surface orientations. We further introduce the spatial priors to ensure that salient objects are compact in both the image space and the 3D space. We have evaluated the proposed method on a newly constructed flash/no-flash dataset and compared its performance with those of the state-of-the-art methods. Experimental results show that the proposed method outperforms those methods consistently, even in challenging scenarios.

Acknowledgement. We thank Zhe Huang, Chengcheng Dai and Xufang Pang for their help with flash/no-flash data collection. This work was partially supported by a SRG grant from City University of Hong Kong (Project Number: 7002768).

References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: CVPR, pp. 1597–1604 (2009)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE TPAMI, 2274–2282 (2012)
3. Agrawal, A., Raskar, R., Nayar, S., Li, Y.: Removing photography artifacts using gradient projection and flash-exposure sampling. ACM TOG 24(3), 828–835 (2005)
4. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: A benchmark. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 414–429. Springer, Heidelberg (2012)
5. Chang, K., Liu, T., Chen, H., Lai, S.: Fusing generic objectness and visual saliency for salient object detection. In: ICCV (2011)
6. Chen, T., Cheng, M., Tan, P., Shamir, A., Hu, S.: Sketch2photo: Internet image montage. ACM TOG 28(5), 124:1–124:10 (2009)
7. Cheng, M., Zhang, G., Mitra, N., Huang, X., Hu, S.: Global contrast based salient region detection. In: CVPR, pp. 409–416 (2011)

8. Eisemann, E., Durand, F.: Flash photography enhancement via intrinsic relighting. ACM TOG 23(3), 673–678 (2004)
9. Enns, J., Rensink, R.: Sensitivity to three-dimensional orientation in visual search. Psychological Science 1, 323–326 (1990)
10. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. In: CVPR (2010)
11. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: NIPS, pp. 545–552 (2007)
12. He, Z., Nakayama, K.: Visual attention to surfaces in 3-d space. Proc. National Academy of Sciences 92, 11155–11159 (1995)
13. Itti, L., Koch, C.: Computational modelling of visual attention. Nature Reviews Neuroscience 2(3), 194–203 (2001)
14. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE TPAMI 20(11), 1254–1259 (1998)
15. Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N.: Automatic salient object segmentation based on context and shape prior. In: BMVC (2011)
16. Koch, C., Ullman, S.: Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. Human Neurobiology 4, 219–227 (1985)
17. Krishnan, D., Fergus, R.: Dark flash photography. ACM TOG 28(3), 96:1–96:11 (2009)
18. Liao, M., Wang, L., Yang, R., Gong, M.: Light fall-off stereo. In: CVPR, pp. 1–8 (2007)
19. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. IEEE TPAMI 33(2), 353–367 (2011)
20. Ma, Y., Zhang, H.: Contrast-based image attention analysis by using fuzzy growing. ACM Multimedia, 374–381 (2003)
21. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: CVPR (2013)
22. Niu, Y., Geng, Y., Li, X., Liu, F.: Leveraging stereopsis for saliency analysis. In: CVPR (2012)
23. Niu, Y., Liu, F., Li, X., Gleicher, M.: Warp propagation for video resizing. In: CVPR (2010)
24. Nothdurft, H.: Salience from feature contrast: additivity across dimensions. Vision Research 40(10), 1183–1201 (2000)
25. Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision Research 42(1), 107–123 (2002)
26. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: CVPR, pp. 733–740 (2012)
27. Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., Toyama, K.: Digital photography with flash and no-flash image pairs. ACM TOG 23(3), 664–672 (2004)
28. Raskar, R., Tan, K., Feris, R., Yu, J., Turk, M.: Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. ACM TOG 23(3), 679–688 (2004)
29. Reinagel, P., Zador, A.: Natural scene statistics at the centre of gaze. In: Network: Computation in Neural Systems, pp. 341–350 (1999)
30. Reynolds, J., Desimone, R.: Interacting roles of attention and visual salience in V4. Neuron 37(5), 853–863 (2003)
31. Sharma, G., Jurie, F., Schmid, C.: Discriminative spatial saliency for image classification. In: CVPR (2012)
32. Shi, K., Wang, K., Lu, J., Lin, L.: PISA: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors. In: CVPR (2013)
33. Sun, J., Li, Y., Kang, S., Shum, H.: Flash matting. ACM TOG 25(3), 772–778 (2006)
34. Sun, J., Sun, J., Kang, S., Xu, Z., Tang, X., Shum, H.: Flash cut: Foreground extraction with flash and no-flash image pairs. In: CVPR (2007)
35. Toet, A.: Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. IEEE TPAMI 33(11), 2131–2146 (2011)
36. Valenti, R., Sebe, N., Gevers, T.: Image saliency by isocentric curvedness and color. In: ICCV (2009)

37. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012)
38. Wexler, M., Ouarti, N.: Depth affects where we look. Current Biology 18(23), 1872–1876 (2008)
39. Wolfe, J., Horowitz, T.: What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience 5(6), 495–501 (2004)
40. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierachical saliency detection. In: CVPR (2013)
41. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: CVPR (2013)
42. Zhou, C., Troccoli, A., Pulli, K.: Robust stereo with flash and no-flash image pairs. In: CVPR, pp. 342–349 (2012)