Search Guided Saliency

Shijian Lu[1], Byung-Uck Kim[1], Nicolas Lomenie[2], Joo-Hwee Lim[1]
, and Jianfei Cai[3]

[1] Institute for Infocomm Research, A*STAR
1 Fusionopolis Way, #21-01 Connexis, Singapore 138632
[2] The Laboratory of Informatics Paris Descartes, University Paris Descartes
12 Rue de l'Ecole de Mdecine, 75006 Paris, France
[3] School of Computing Engineering, Nanyang Technological University
Nanyang Avenue, Singapore 639798

Abstract. We propose a new type of saliency as inspired by findings from visual search studies - the searching difficulty is correlated with the target-distractor contrast, the distractor homogeneity, as well as the target uniqueness. By putting an image pixel as the target and the surrounding pixels as distractors, a search guided saliency model is designed in accordance with these findings. In particular, three saliency measures in correspondence to the three searching factors are simultaneously computed and integrated by using a series of contextual histograms. The proposed model has been evaluated over three public datasets and experiments show superior prediction of the human fixations when compared to the state-of-the-art models.

Keywords: visual attention, saliency model, visual search

1 Introduction

Visual saliency [27] characterizes the distinct perceptual quality of an object or image region with respect to its surrounding. It helps to serialize the attending of objects in scenes which the human vision system cannot process in parallel due to the tremendous amount of visual information involved. Computational modeling of visual saliency aims to build an attention model that is capable of predicting where people will look at given an image or scene. It has increasingly attracted research interest in recent years due to its importance in both human visual attention study and a wide range of applications in object detection, object segmentation, visual search, etc [8, 21, 24, 29].

Quite a number of saliency models [3] have been reported in recent years that exploit the local contrast and global image contrast. The local contrast based models make use of a center surround difference to compute the contrast of an object or image region with respect to its surrounding [9, 14, 15, 18]. Itti and Koch's model [14, 15] is probably one of the earliest that exploit the center surround difference which is computed using a set of spatial filters. Other approaches have also been proposed that exploit decision-theoretic discrimination [9], object-level

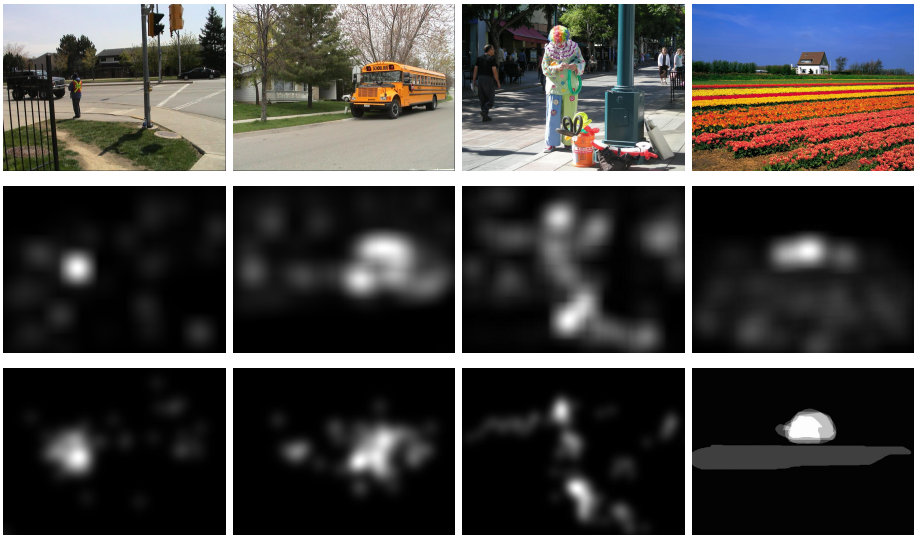


Fig. 1. Illustration of the Search Guided Model: The search guided model is tolerant to image edges and dynamic structures and capable of predicting the human fixations accurately. For the sample images in the first row (from AIM dataset [5], SR dataset [13], and MIT300 dataset [16]), rows 2-3 show the searched guided saliency maps and the corresponding fixational maps, respectively.

segmentation [18], etc. for center surround difference computation. The major limitation of the local contrast based models is that certain global features which are closely related to the perceptual saliency is not captured.

A number of models have been reported to incorporate the global contrast that is often pertaining to the perceptual rarity, uniqueness, unusualness. In particular, image histograms have been extensively used to capture the low-frequency global features. For example, Cheng et al. [6] employ color histograms to capture the global color contrast and combine it with a local region contrast for saliency computation. Lu et al. [19, 20] exploit a 2D co-occurrence histogram that captures both local contrast and global unusualness simultaneously. Contextual information has also been exploited [10, 28, 22] to capture both local and global contrast in different ways. In addition, several frequency space models [1, 11–13] are reported that compute saliency based on global unusual amplitude or phase spectrum of the Fourier transform of an image.

Though local and global contrast has been exploited in various ways, the contrast-based models are often over-responsive to inconspicuous image edges which are associated with the local contrast and often represent certain globally abnormal features. Several learning based models [4, 5, 17, 31, 32, 23] have been proposed to learn the statistics or eye fixation data directly. These learning based models are not sensitive to inconspicuous image edges but the computed saliency often lacks discrimination between salient and inconspicuous objects.

We propose a novel saliency model as inspired by findings from the visual search studies, i.e., searching difficulty is determined by the target-distractor difference, the center uniqueness, as well as the distractor homogeneity as illustrated in Fig. 1. In the proposed model, the three saliency features are computed and integrated by using a series of *contextual histograms* that can be directly determined within a local neighborhood window. The proposed model has a number of novelties. First, it computes and integrates several search-guided saliency features and obtains superior human fixation prediction performance compared with state-of-the-art models. Second, it makes use of contextual histograms and overcomes one typical limitation of many existing models, i.e., high saliency response around inconspicuous image edges or other dynamic structures. Third, it is simple and easy for implementation.

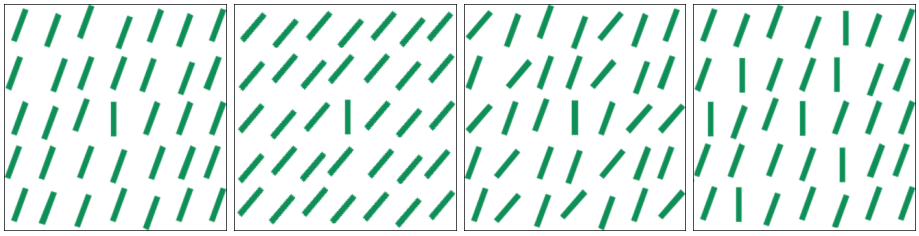


Fig. 2. Visual Search Principle: Compared with the target (the vertical bar at the center) in (a), the target in (b) has a higher pop-out effect due to its higher target-distractor difference, the target in (c) has a lower pop-out effect due to its lower distractor homogeneity (the distractors have the same average slant angle as those in (a)), and the target in (d) has a lower pop-out effect due to its lower uniqueness level.

2 Principle of Search Guided Saliency

The visual search guided saliency is inspired by the feature integration and stimulus similarity theory [26, 7] as illustrated in Fig. 2 - the target will have a shorter searching time and stronger pop-out effect when the target-distractor contrast is stronger, the surrounding distractors have a higher homogeneity level and the target has a high uniqueness level. It integrates three principles from the visual search studies [7, 26, 30] including:

1. A target will have a stronger pop-out effect when it has a larger contrast to the surrounding distractors. As shown in Fig. 2, the target - the vertical bar at the center in 2b has a stronger pop-out effect than that in 2a. This can be further illustrated by simulated images in Fig. 3 where the target pixel at the center in 3b is more salient than the one in 3a due to its stronger contrast to the surrounding distractor pixels.

2. A target will have a stronger pop-out effect when the surrounding distractors have a higher homogeneity level. As shown in Figs 2, the vertical bar at the center in 2c has a lower pop-out effect than that in 2a. This can be further illustrated by simulated images in Fig. 3 where the target pixel in 3c is less salient than the one in 3a due to its lower homogeneity level.
3. A target will have a stronger pop-out effect when it has a higher uniqueness level, i.e., fewer distractors have the same visual properties as the target. This can be shown in Fig. 2 where the vertical bar at the center in 2d has a lower pop-out effect than that in 2a. It can be further illustrated by simulated images in Fig. 3 where the target pixel in 3d is less salient than the one in 3a due to its lower uniqueness level.

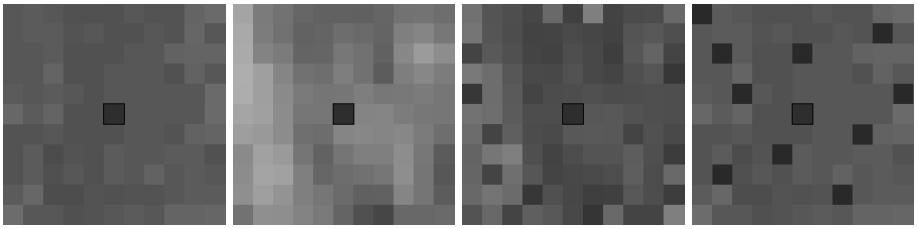


Fig. 3. Visual Search Principle: Compared with the target pixel at the center of the image in (a), the target pixel in (b) has higher saliency due to its higher center-surround contrast, the target pixel in (c) has lower saliency due to its lower surround homogeneity (the distractor pixels have the same mean but much smaller variance than those in (a)), the target pixel in (d) has lower saliency due to its lower uniqueness.

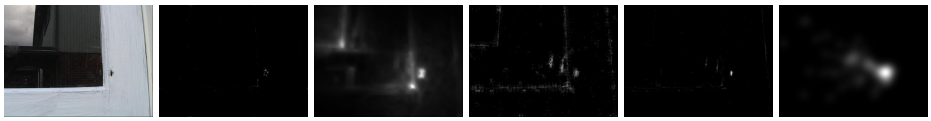
Under the same target-distractor paradigm, the proposed model constructs a contextual histogram based on the "distractor" pixels that surround each "target" pixel at the center. The saliency of the target pixel is computed by integrating three search-inspired saliency measures including the surround homogeneity, the center uniqueness, and the center surround contrast, all of which can be computed from a series of contextual histogram simultaneously. Due to the integration of the three saliency features by using the contextual histograms, the proposed model is tolerant to the image edges and demonstrates better prediction of the human fixations when compared with those local and global contrast based models as illustrated in Fig. 4.

Related models also exploit the image histogram and contextual information to capture the local and global image contrast [10, 12, 19, 20, 2]. The frequency space model in [12] makes use of the global contrast which is tolerant to image edges but often detects only the boundary of salient objects as illustrated in Fig. 4b. The context models [10, 2] integrate the local and global contrast and but are over-responsive to inconspicuous image edges and corners as shown in Fig. 4c. The histogram based models [19, 20] exploit occurrence and co-occurrence of

image intensity and color to capture the local and global contrast concurrently. They are capable of detecting multiple salient objects with a complex background but are also over-responsive to image edges as shown in Fig. 4d.

3 Visual Search Guided Saliency Modeling

This section describes the visual search guided saliency modeling technique. For each "target" pixel, several contextual histograms are first constructed based on the "distractor" pixels that surround the target pixel. A saliency level is then determined by integrating the three saliency measures that are computed from each contextual histogram. The overall saliency is finally determined by integrating the saliency that is computed across multiple contextual histograms and multiple image channels.



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

Fig. 4. Search Guided Saliency: For the sample image in 3a [5], the images in 3b-3e show the saliency that is computed using the spectral residual [12], the context [10], the co-occurrence histogram [20], the search guided principles, and the corresponding fixational map (Gaussian smoothing of eye fixations of 20 subjects), respectively.

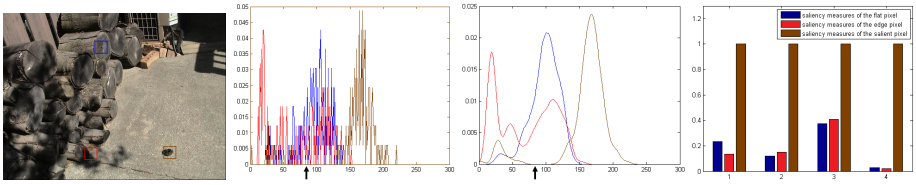
3.1 Contextual Histogram

A contextual histogram is constructed to emulate the target-distractor paradigm. For each target pixel, a contextual histogram H is constructed by using a number of distractor pixels that surround the target pixel at the center. Neighborhoods of different shapes such as circular or square-shaped can be used to pick the distractor pixels. At the same time, neighborhoods of different sizes can be used to capture contexts of different distances to the target pixel (to be described in Section 3.4). Note that the distractor pixels are picked along the neighborhood boundary instead of from within the neighborhood.

Fig. 5b shows the contextual histogram of three typical image pixels as labeled in the image in Fig. 5a, where histogram graphs have the same color as the corresponding labeling neighborhood squares. In particular, the three example pixels have the same intensity (80 as indicated by the arrow) but are picked from a homogeneous image region (blue square), an inconspicuous image edge (red square), and a salient image region (brown square), respectively. The corresponding three contextual histograms are distinctive as illustrated in Fig. 5b. For the pixel in the homogeneous region, its contextual histogram (blue graph)

has a large global peak and its intensity lies close to the global peak. For the pixel along the image edge, its contextual histogram (red graph) usually has two major peaks and its intensity lies somewhere between the two major peaks. For the pixel in the salient image region, its contextual histogram (brown graph) often has a large global peak and its intensity lies far away from the global peak.

The contextual histogram can be smoothed to suppress the undesired saliency response (for the three saliency measures) around the image edge. For the three contextual histograms shown in Fig. 5b, Fig. 5c shows the smoothed contextual histograms by using a Gaussian filter. As Fig. 5c shows, the salient pixel and the one in the homogeneous region still have a small and large histogram values, respectively, after the smoothing. As a comparison, the edge pixel has a much higher value after the smoothing because the smoothing raises its histogram values due to the two major peaks at both sides as illustrated in Fig. 5c. The three saliency measures can be computed from the smoothed contextual histogram as shown in Fig. 5d (to be described in the next subsection).



(a) (b) (c) (d)

Fig. 5. Search Guided Saliency Measures: (a) shows a sample image with three typical pixels from a homogeneous region (square neighborhood of blue color), an image edge (square neighborhood of red color) and a salient region (square neighborhood of brown color). (b) and (c) show the contextual histogram H and the smoothed contextual histogram H' of the three pixels, respectively (with the same coloring as the square neighborhood in Fig. 5a). (d) shows the three saliency measures of the three pixels as well as their integrated saliency.

3.2 Search Guided Saliency Measures

With respect to the three visual search principles, three saliency measures can be defined and computed from the contextual histogram simultaneously. The first saliency measure is center surround difference that has been widely exploited in the literature. It is defined as follows:

$$S_c(x, y) = \|I(x, y) - H_c\| \tag{1}$$

where $I(x, y)$ denotes the intensity of the target pixel at (x, y). H_c is the centroid of the contextual histogram which is equal to the mean of the surrounding distractor pixels that are picked to construct the histogram. Note that the contextual histogram here is the original before smoothing.

Image pixels in a salient region usually have much larger center-surround difference S_c than those in a homogeneous region whose intensity is usually close to the histogram centroid H_c. Edge pixels often have a small S_c because the intensity of edge pixels usually lies between the two major histogram peaks and will be close to the histogram centroid H_c as illustrated in Fig. 5c (red graph). This can be further illustrated in Fig. 5d where the first bar group (with the same labeling colors as in Figs. 5a-5c) shows the S_c of the three pixels which is normalized by their maximum. As Fig. 5d shows, the pixel in a salient region has a much higher S_c than the edge pixel and the pixel in a homogeneous region.

The second saliency measure is the surround homogeneity which is closely correlated with the perceptual visual saliency. Within a smoothed contextual histogram H', the surround homogeneity is mainly demonstrated by a large global peak. This saliency measure is defined as follows:

$$S_h(x,y) = \Big(H'_x - H'\big(I(x,y)\big)\Big)^p \tag{2}$$

where H'_x denotes the global peak of the smoothed contextual histogram H'. Parameter p is a number larger than 1 which controls the weight of this saliency measure.

For image pixels having the same center-surround difference, those with a more homogeneous surrounding should have a larger S_h. In particular, target pixels with a more homogeneous surrounding should have a larger histogram peak H'_x but a smaller $H'(I(x,y))$, where the smaller $H'(I(x,y))$ is largely due to fewer distractor pixels whose intensity lies between the global peak intensity and the target pixel's intensity $I(x,y)$. This can be illustrated in Fig. 5d where the second bar group shows the S_h of the three sample pixels that is normalized by their maximum. It should be noted that image pixels in a homogeneous region usually have a large H'_x but a small S_h because the histogram value of these pixels, i.e., $H'(I(x,y))$, is usually large and close to H'_x. As Fig. 5d shows, the pixel in a salient region has a much higher S_h than the edge pixel and the pixel in the homogeneous region.

The third saliency measure captures the uniqueness level of the target pixel. With a smoothed contextual histogram H', this measure is defined as follows:

$$S_u(x,y) = 1 - H'(I(x,y))^q \tag{3}$$

where $I(x,y)$ denotes the intensity of the target pixel and q is a number lying between 0 and 1 which controls the weight of this saliency measure in the integrated saliency.

Image pixels with a higher saliency level usually have a larger S_u. In particular, a higher center uniqueness level means fewer distractor pixels with the same intensity as the target pixel, i.e., a smaller $H'(I(x,y))$ that leads to a larger S_u. Note that S_u is related to the S_c and S_h as a larger center-surround difference and surround homogeneity usually lead to a higher center uniqueness. On the other hand, S_u captures certain specific saliency information, i.e., the center uniqueness, that is not captured in either S_c or S_h (as illustrated in Figs. 2d

and 3d). This can be illustrated in Fig. 5d where the third bar group shows the S_u of the three sample pixels that is normalized by their maximum. As Fig. 5d shows, the pixel in a salient region has a much higher S_u than the edge pixel and the pixel in a homogeneous region.

3.3 Saliency Modeling

The saliency of an image can be determined by integrating the three saliency measures that are computed for different channel images of different scales. We use the **Lab** color space where channel **L** encodes the image lightness and contrast information and channels **a** and **b** encode the image color information. In addition, each channel image is down-sampled to n image scales to capture contexts of different sizes (to be described in Section 3.4). Note that image values in the three image channels are first mapped to 0∼255 (first subtracted by the minimum intensity, then divided by the maximum value, and finally multiplied by 255) and then rounded to integers for the contextual histogram construction.

With the three saliency measures as defined in Eqs. 1-3 in the previous subsections, the saliency level of a target image pixel is determined as follows:

$$S(x, y) = S_c(x, y) * S_h(x, y) * S_u(x, y) \tag{4}$$

where $S_c(x, y), S_h(x, y)$, and $S_u(x, y)$ denote the three saliency measures that are computed for the target image pixel at (x, y), respectively. A multiplication strategy is adopted because all the three saliency measures change in the same direction as the overall perceptual saliency. For the three sample pixels labeled in Fig. 5a, the fourth bar group in Fig. 5d shows the corresponding integrated saliency, where the pixel from a salient region has much higher saliency compared with the other two sample pixels.

The overall saliency of a target pixel can be finally computed as follows:

$$S_o(x, y) = \sum_{Lab} \max\big(S_1(x, y), \cdots, S_n(x, y)\big) \tag{5}$$

where $S_1(x, y), \cdots, S_n(x, y)$ refer to the integrated saliency in Equation 4 that is computed for one image channel of different scales. Two strategies are adopted to integrate the computed saliency. First, max-pooling is employed to take the maximum of the saliency that is computed across n image scales, i.e., $S_1(x, y), \cdots, S_n(x, y)$. Note that saliency computed at different image scales is first scaled back to the original image scale before the max-pooling. Second, average-pooling is implemented to determine the overall saliency level by averaging the saliency that is computed over the three image channels.

3.4 Discussion

The proposed model involves several parameters. In particular, p and q are used to control the weights of the surround homogeneity and the center uniqueness

as described in Section 3.2. Evaluation over the eye fixational map shows that saliency can be detected properly when p and q are set around 0.05-0.2 and 2-4, respectively. Neighborhoods of different shapes and sizes can be set to pick the distractor pixels as described in Section 3.1. In our implementation, a square-shaped neighborhood is used and the neighborhood radius is set at 10 pixels. In addition, each channel image is down-sampled to n image scales for saliency computation as described in Section 3.3. In our implementation, five image scales are used where each image is down-sampled to 0.6, 0.5, 0.4, 0.3, and 0.2 of the original image scale. Last, the contextual histogram is smoothed by using a Gaussian filter as described in Section 3.1. The width of the filter window can be around 20-40 based on the humans' perceptible visual contrast.

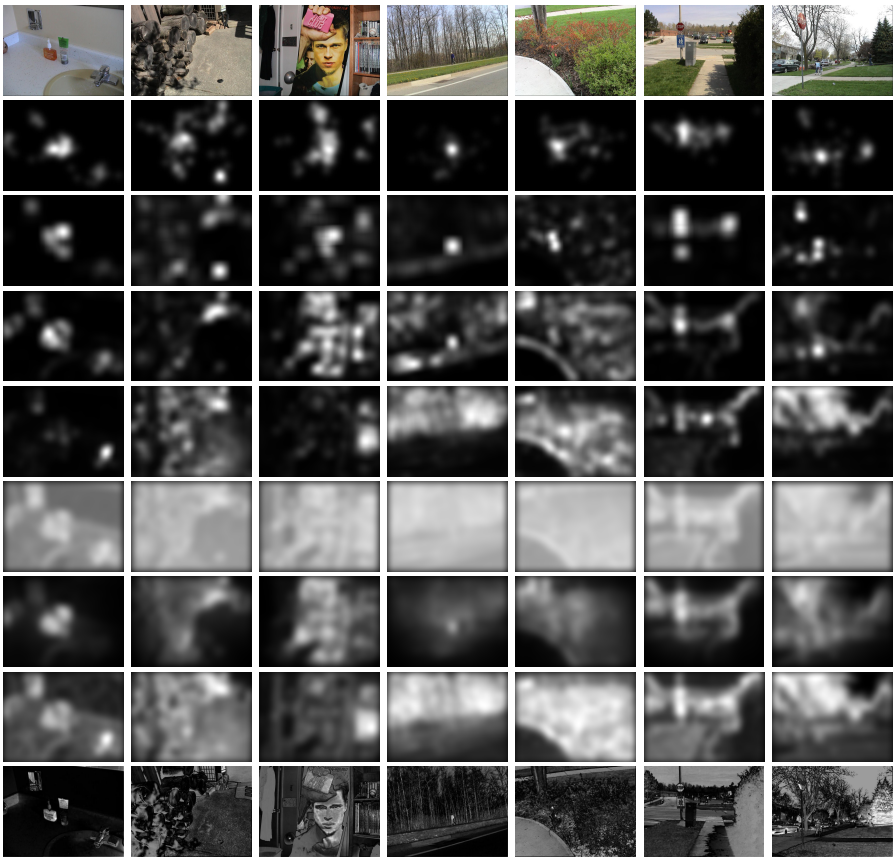


Fig. 6. Comparison of the search guided model with six state-of-the-art models over the AIM dataset: For the sample images in the first row, row 2 shows the corresponding fixational maps as described in Section 4, row 3-9 show the corresponding saliency maps by the search guided model and the six state-of-the-art models in [20, 12, 5, 10, 31, 1], respectively.

The three saliency measures help to suppress the response at the inconspicuous edges and other dynamic structures such as tree branches effectively. First, edge pixels usually have a small S_c because their intensity lying between the two major histogram peaks is often close to the H_c, the mean of the distractor pixels that are counted into the contextual histogram. Second, edge pixels usually have a small S_h because they have a smaller global histogram peak H'_x but a larger $H'(I(x,y))$, largely due to the smoothing of the contextual histogram with two major peaks at the left and right sides. Third, edge pixels usually have a small S_u because a small q (e.g. $q = 0.1$) will left $H'(I(x,y))^q$ to be close to 1 even when $H'(I(x,y))$ is very small by itself.

4 Results

The proposed model has been evaluated over three public datasets including the SR dataset [13], the AIM dataset [5], and the MIT300 dataset [16]. The SR dataset consists of 62 static images and for each image, salient regions are manually labeled by four subjects which are further averaged to form a hit map as illustrated by the first three graphs in the second row of Fig. 7. The AIM dataset includes 120 static images and the corresponding fixational maps as illustrated in the second row in Fig. 6, which are created by Gaussian smoothing of the eye fixations that are collected from 20 subjects for each image. The MIT300 dataset consists of 300 natural images that capture different scenes such as humans, buildings, flowers, etc. For each image, fixations of 39 subjects are collected in similar way as the AIM dataset with which a fixational map is computed as illustrated in the last three graphs in the second row of Fig. 7.

The proposed model is compared with six state-of-the-art models including the context model [10], the signature model [12], the frequency tuned (FT) model [1], the AIM model [5], the SUN model [31], and the CCH model [20]. The implementations of the state-of-the-art models are downloaded from the authors' websites. For the search guided model, parameters p and q are set at 0.1 and 3, and the window width of the histogram filter is set at 30. Fig. 6 show several images of the AIM dataset in the first row, the corresponding fixational maps in the second row, and the saliency maps that are computed by using the search guided model and the six state-of-the-art models [20, 12, 5, 10, 31, 1] in 3-9 rows, respectively. Fig. 7 show the saliency maps of the search guided model and the six compared models for the SR dataset (the first 3 images) and the MIT300 dataset (the last 3 images). As Figs. 6 and 7 show, the search guided model predicts the human fixations accurately.

In particular, the contrast-based models [20, 12, 10]are often over-responsive to the inconspicuous image edges as illustrated in rows 4, 5, and 7. As a comparison, the search guided models helps to suppress such "false alarms" effectively as shown in row 3. For example, most contrast-based models are over-responsive to the inconspicuous image edges and dynamic tree branches and grasses as shown in the second, fourth, fifth and seventh images in Fig. 6 where the search guided model has little responses as shown in row 3. The learning based models [31,

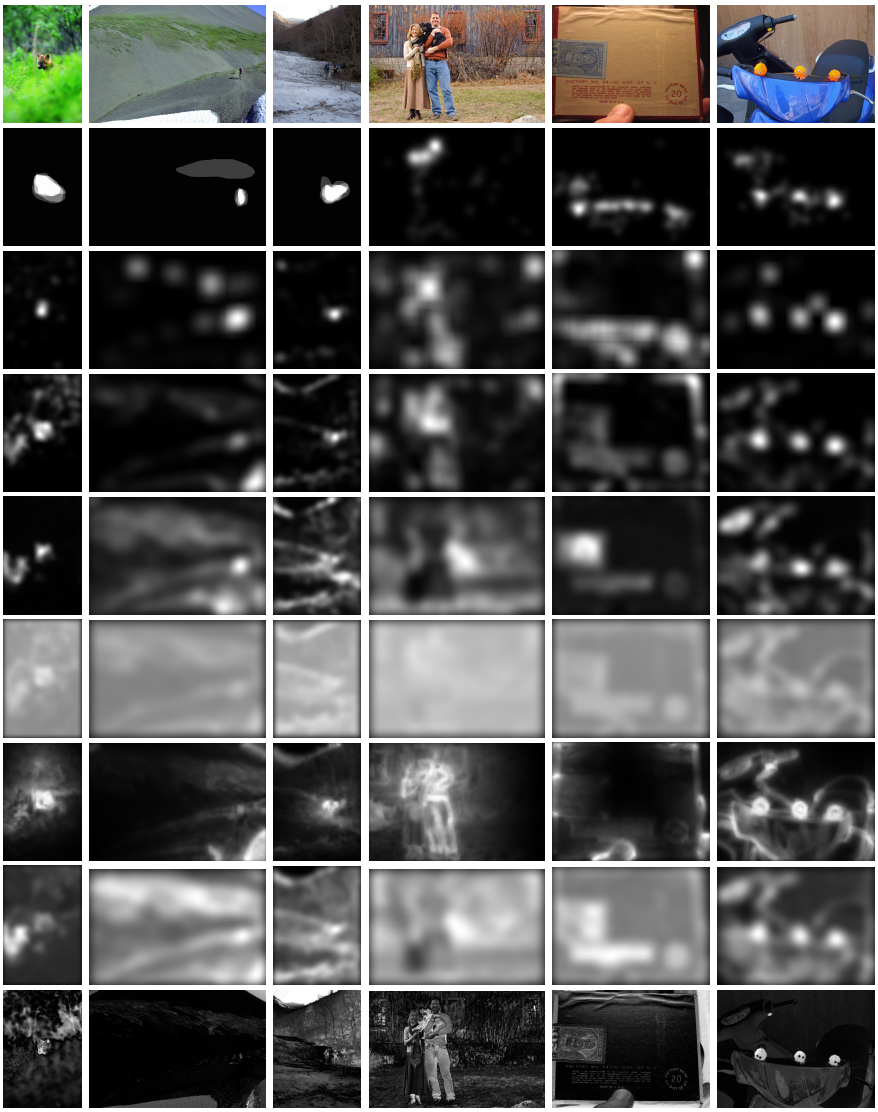


Fig. 7. Comparison of the search guided model with six state-of-the-art models over the SR and MIT300 datasets: For the sample images in the first row, row 2 shows the corresponding fixational maps as described in Section 4, row 3-9 show the corresponding saliency maps by the search guided model and the six state-of-the-art models in [20, 12, 5, 10, 31, 1], respectively.

5] are instead blurry where salient and non-salient regions both have certain saliency as illustrated in rows 6 and 8. In addition, the search guided model is capable of detecting salient objects or image regions of small scale such as the dark object in the second image and the red flowers in the fifth image in Fig. 6, largely due to the incorporated context homogeneity information. As a comparison, the contrast-based models often fail to detect such salient objects because other inconspicuous objects often have much higher contrast.

Table 1. sAUC of the search guided model and six compared state-of-the-art models based on the AIM dataset [5] (CSC: center-surround contrast in Eq. 1; SH: surround homogeneity in Eq. 2; CU: center uniqueness in Eq. 3).

Models	Shuffled AUC	
	AIM dataset	SR dataset
Search Guided Model	**0.7311**	0.7224
SH+CU Model	0.7217	N.A.
SH Model	0.7039	N.A.
CU Model	0.6942	N.A.
CSC Model	0.6899	N.A.
Co-Occurrence Model in [20]	0.7221	0.7291
Signature model [12]	0.7147	0.6881
AIM model [5]	0.6990	0.7149
Context Model in [10]	0.6958	**0.7458**
SUN model [31]	0.6813	0.6668
FT model [1]	0.5885	0.6108

Quantitative experiments have also been conducted based on the AIM dataset. The MIT300 dataset is not evaluated as only nine images have fixational maps available (used for comparison of different models on the authors' website) whereas fixational maps of the rest images is not available. The performance is evaluated through the analysis of the receiver operating characteristic (ROC) and the corresponding shuffled area under the ROC curve (sAUC). For the saliency computed by different models, 25 rounds of Gaussian smoothing are implemented by changing the smoothing window size from 0.01 to 0.13 of the image width with an increase step of 0.005 as described in [12]. In addition, the ROC computation procedure in [25] is adopted which compensates for center-bias that commonly exists within the human fixations.

Table 1 shows the sAUC of the search guided models and the six compared models. With the three saliency measures, five sAUCs are computed where the "Search Guided Model" integrates all three saliency measures, the "SH+CU Model" integrates the surround homogeneity and the center uniqueness, the "CSC Model" uses the center-surround contrast alone, the "SH Model" uses the surround homogeneity alone, and the "CU Model" uses the center uniqueness alone. As For the AIM dataset, the "CSC Model" does not model the vi-

sual saliency well, with a sAUC at 0.6899. The "SH+CU Model" integrates the novel surround homogeneity and center uniqueness, which greatly outperforms the "CSC Model". In addition, the "SH Model" clearly outperforms the "CU Model", meaning that the surround homogeneity plays a heavier role in perceptual saliency compared with the center uniqueness. Furthermore, the "Search Guided Model" obtains a sAUC of 0.7311 which outperforms all sub-component models as well as the six contrast-based models. For the SR dataset, the search guided model obtains a sAUC of 72.24% which is close to that of the AIM dataset (sAUC of the sub-component models are not computed). Note that the context based model [10] obtains a clearly higher sAUC, largely due to a face detector it incorporates that helps to predict the high saliency of human and animal faces within a number of images of the SR dataset.

The proposed search guided model exploits only the low-level features. On the other hand, the human eyes are often attracted by familiar objects with semantic meaning such as human bodies, animals, human faces, vehicles, texts in scenes, etc. Relevant visual search models such as face detector, text detector, vehicle detector, etc. will be investigated and combined with the search guided model for better prediction of the human fixations.

5 Conclusions

This paper presents a novel saliency model that is inspired by visual search studies. Three saliency measures including the widely used center-surround contrast, the surround homogeneity, and the center uniqueness are defined and integrated for saliency modeling. A series of contextual histograms are constructed for each image pixel from which all the three saliency measures can be computed simultaneously. Experiments over three widely used public benchmarking datasets show that the proposed model predicts the human fixations accurately.

References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. IEEE CVPR p. 15971604 (2009)
2. Borji, A., Itti, L.: Exploiting local and global patch rarities for saliency detection. IEEE CVPR pp. 478–485 (2012)
3. Borji, A., Sihite, D., Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. IEEE TIP 22(1), 55–69 (2012)
4. Bruce, N., Tsotsos, J.: Saliency based on information maximization. NIPS pp. 155–162 (2006)
5. Bruce, N., Tsotsos, J.: Saliency, attention, and visual search: An information theoretic approach. Journal of Vision 9(3), 1–24 (2009)
6. Cheng, M., Zhang, G., Mitra, N., Huang, X., Hu, S.: Global contrast based salient region detection. IEEE CVPR pp. 409–16 (2011)
7. Duncan, J., Humphreys, G.: Visual search and stimulus similarity. Psychological Review 96(3), 433–458 (1989)

8. Feng, J., Wei, Y., Tao, L., Zhang, C., Sun, J.: Salient object detection by composition. IEEE ICCV pp. 1028–1035 (2011)
9. Gao, D., Vasconcelos, N.: Bottom-up saliency is a discriminant process. IEEE ICCV pp. 1–6 (2007)
10. Goferman, S., Zelnik-Manor, Tal, A.: Context-aware saliency detection. IEEE CVPR pp. 2376–2383 (2010)
11. Guo, C., Ma, Q., Zhang, L.: Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. IEEE CVPR pp. 1–8 (2008)
12. Hou, X., Harel, J., Koch, C.: Image signature: Highlighting sparse salient regions. IEEE TPAMI 34(1), 194–201 (2012)
13. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. IEEE CVPR pp. 1–8 (2007)
14. Itti, L., Koch, C.: Computational modeling of visual attention. Nature Reviews Neuroscience 2(3), 194–203 (2001)
15. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE TPAMI 20(11), 1254–1259 (1998)
16. Judd, T., Durand, F., Torralba, A.: A benchmark of computational models of saliency to predict human fixations. MIT Computer Science and Artificial Intelligence Laboratory Technical Report pp. MIT–CSAIL–TR–2012–001 (2012)
17. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. IEEE ICCV pp. 2106–2113 (2009)
18. Liu, T., Sun, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. IEEE CVPR pp. 1–8 (2007)
19. Lu, S., Lim, J.H.: Saliency modeling from image histograms. ECCV pp. 321–332 (2012)
20. Lu, S., Tan, C., Lim, J.: Robust and efficient saliency modeling from image co-occurrence histograms. IEEE TPAMI 36(1), 195–201 (2013)
21. Marchesotti, L., Cifarelli, C., Csurka, G.: A framework for visual saliency detection with applications to image thumbnailing. IEEE ICCV pp. 2232–2239 (2009)
22. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? IEEE CVPR pp. 1139–1146 (2013)
23. Rudoy, D., Goldman, D., Shechtman, E., Zelnik-Manor, L.: Learning video saliency from human gaze using candidate selection. IEEE CVPR pp. 1147–1154 (2013)
24. Sharma1, G., Jurie, F., Schmid, C.: Discriminative spatial saliency for image classification. IEEE CVPR pp. 3506–3513 (2012)
25. Tatler, B., Baddeley, R., Gilchrist, I.: Visual correlates of fixation selection: Effects of scale and time. Vision Research 45(5), 643659 (2005)
26. Treisman, A., Gelade, G.: A feature-integration theory of attention. Cognitive Psychological 12(1), 97–136 (1980)
27. Tsotsos, J.: Analyzing vision at the complexity level. Behavioral and Brain Science 13(3), 423–445 (1990)
28. Wang, L., Xue, J., Zheng, N., Hua, G.: Automatic salient object extraction with contextual cue. IEEE ICCV pp. 105–112 (2011)
29. Wang, P., Wang, J., Zeng, G., Feng, J., Zha, H., Li, S.: Salient object detection for searched web images via global saliency. IEEE CVPR pp. 3194–3201 (2012)
30. Wolfe, J.: Guided search 2.0 a revised model of visual search. Psychonomic Bulletin and Review 1(2), 202–238 (1994)
31. Zhang, L., Tong, M.H., Marks, T.K., Cottrell, G.W.: Sun : A bayesian framework for saliency using natural statistics. Journal of Vision 8(7), 1–20 (2008)
32. Zhao, Q., Koch, C.: Learning a saliency map using fixated locations in natural scenes. Journal of Vision 3(9), 1–15 (2011)