LETTER

Isophote Based Center-Surround Contrast Computation for Image Saliency Detection

Yuelong CHUANG[†], ***Student Member***, **Ling CHEN**[†a)], **Gencai CHEN**[†], ***and*** **John WOODWARD**[††], ***Nonmembers***

SUMMARY In this paper, we introduce a biologically-motivated model to detect image saliency. The model employs an isophote based operator to detect potential structure and global saliency information related to each pixel, which are then combined with integral image to build up final saliency maps. We show that the proposed model outperforms seven state-of-the-art saliency detectors in experimental studies.

key words: *image saliency, isophote, center-surround contrast*

1. Introduction

Image saliency can be defined as local regions that can be easily differentiated from their surroundings, these differentiators being color, orientation and intensity [1], [2]. The key issues in detecting image saliency include the following: 1) how to determine position and size for both center regions and neighboring regions, and 2) how to compute differences between center regions and adjacent regions for the three feature channels.

Itti et al. [1] solve the problems by building image pyramids and subtracting different image pyramids to determine center-surround contrast. Frintrop et al. [3] build image pyramids by integral filters [4], and compute the center-surround contrast by exhaustively searching at each pyramid. However, both approaches have a problem: different methods are employed to compute saliency information for each feature channel, which results in the problem that the fusion of the different feature channels with non-comparable properties is somewhat arbitrary (For the different feature channel maps, normalization has to be done to make the maps comparable. However, normalizing maps to a fixed range could remove important information about the magnitude of the maps).Instead of building image pyramids, Achanta et al. [5] build up the saliency map by filtering the original image in a raster scan fashion. However, determining reasonable values for both center and surrounding regions is rather hard to achieve, and the performance is easily affected by the type of background in the image.

In recent years, several models have been proposed to compute image saliency with mathematical methods [6]–[8]. Achanta et al. [6] propose a frequency tuned method to compute pixel saliency directly. Since the model only considers first-order average color, it could be insufficient to analyze complicated situations. Valenti et al. [7] adopt an isocentric feature approach to represent image saliency in a global manner. However, the performance is seriously influenced by complicated backgrounds. Hou and Zhang [8] propose a novel model to detect image saliency by exploring spectral components in an image. The model is very fast, but since it is based on global considerations, detailed information about salient objects could be overlooked.

To solve these two problems, we propose an isophote based model to detect the salient pixels in images. In the proposed model, an isophote based operator is employed to capture potential structure and global saliency information related to each pixel. The potential structure is used to determine center and surrounding regions that are then combined with global saliency to determine the final saliency information. Moreover, the integral image is employed to compute center-surround contrast, which is conducive to the fusion of all feature channel maps.

2. The Framework

2.1 Isophote Based Operator

Isophotes are contour lines connecting points of equal luminance. An image can be fully described in terms of its isophotes because: 1) isophotes do not intersect each other, and 2) the shape of each isophote is independent of changes in contrast and brightness [7]. Moreover, it has been observed that for highly curved isophotes their osculating circles tend to concentrate on small regions around an object's corners, while for minimally curved isophotes their osculating circles tend to concentrate on large regions around an object's centers, which means that the curvature of isophotes could indicate an object's scale (Fig. 1 (a)). Due to these properties, we use an isophote based operator to detect each pixel's center and surrounding regions.

Given an input $L(x, y)$, where x, y are the Cartesian coordinates in the image plane, an isophote is defined as $L(x, y) = L_v$ (where L_v stands for a particular value of luminance). The curvature c of L_v is obtained by $c = \frac{y''}{(1+y'^2)^{\frac{3}{2}}}$, where $y' = \frac{dy}{dx}$ and $y'' = \frac{d^2y}{dx^2}$. The first derivative of y with respect to x is arrived at by implicit differentiation of the isophote:

Manuscript received May 15, 2013.

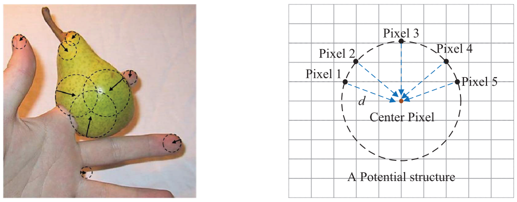
Manuscript revised August 29, 2013.

[†] The authors are with the College of Computer Science, Zhejiang University, P.R. China.

[††] The author is with the Computing Science and Mathematics, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK.

a) E-mail: lingchen@zju.edu.cn

DOI: 10.1587/transinf.E97.D.160



(a) (b)

Fig. 1 An example of detecting potential structures by isophote based operator (the saliency map and segmentation result of the image are shown in fifth column of Fig. 7). In (a), only the potential structures related to the edge of the image are shown. It is because most of the potential structures within the smooth surface of the image could be filtered out by the center-surround contrast computation demonstrated in Sect. 2.2.

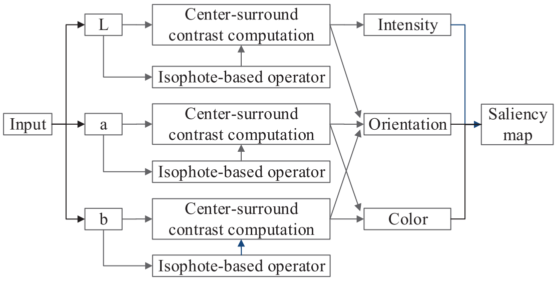


Fig. 2 The framework to build a saliency map.

$$\frac{dy}{dx} = -\frac{L_x}{L_y} = -L_x L_y^{-1} \tag{1}$$

where L_x and L_y are the first derivatives of L_v with respect to x and y respectively. The second derivative is:

$$\frac{d^2y}{dx^2} = -L_y^{-1}L_{xx} - L_y^{-1}L_{xy}\frac{dy}{dx} + L_xL_y^{-2}L_{yx} + L_xL_y^{-2}L_{yy}\frac{dy}{dx} \tag{2}$$

where L_{xx}, L_{xy}, L_{yx}, and L_{yy} are the second partials in x and y. Substituting (1) in (2):

$$\frac{d^2y}{dx^2} = \frac{-L_y^2L_{xx} + 2L_xL_yL_{xy} - L_x^2L_{yy}}{L_y^3} \tag{3}$$

According to Eq. (1) and Eq. (3), the curvature c is obtained by:

$$c = \frac{-L_y^2L_{xx} + 2L_xL_yL_{xy} - L_x^2L_{yy}}{(L_x^2 + L_y^2)^{\frac{3}{2}}} \tag{4}$$

We are interested in the osculating circle, and use it to represent objects' scale information. Knowing that the curvature is the reciprocal of the radius of the osculating circle, and the sign of the isophote curvature depends on the intensity of the outer side of the curve [7]. Thus, the final formulation of the radius is obtained by multiplying the gradient with the inverse of the isophote curvature c:

$$\begin{aligned} d(x,y) &= \frac{\{L_x, L_y\}}{\sqrt{L_x^2 + L_y^2}} \times \frac{(L_x^2 + L_y^2)^{\frac{3}{2}}}{-L_y^2L_{xx} + 2L_xL_yL_{xy} - L_x^2L_{yy}} \\ &= \frac{\{L_x, L_y\}(L_x^2 + L_y^2)}{-L_y^2L_{xx} + 2L_xL_yL_{xy} - L_x^2L_{yy}} \end{aligned} \tag{5}$$

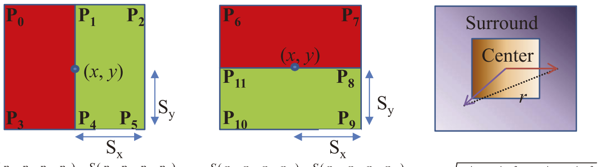
where d stands for a displacement vector to the estimated position of a potential structure. Once d is obtained, a potential structure is available. Potential structures are distributed in the different parts of an image. We focus on the potential structures that are different from their surroundings. Thus, potential structures are used as center regions for computing center-surround contrast. As for the surrounding region, d can be used as a cue to set its size: the radius of surrounding region can be set by a constant multiplicative factor $k(k > 1) : d_{\mathrm{outer}} = k \times d$, and the value of k is set based on experiments. After determining the center and surrounding regions, the center-surround contrast can be computed, which is used to represent corresponding center pixel's saliency information. Since center-surround contrast computation is a local operation, the saliency information obtained by it is called *local saliency*.

Another point which should be mentioned is the number of pixels that belong to a potential structure (having the same displacement to a center pixel) (Fig. 1 (b)). The larger the number is, the more significant the corresponding center pixel is. Compared to the center-surround contrast computation, the number of pixels belonging to it is computed by the isophote based operator in a global manner. Therefore, we call it *global saliency* S_{g}. Valenti et al. [7] directly use this global saliency as image saliency. However, this strategy is liable to be affected by complicated backgrounds and especially for images of nature (Fig. 4). In contrast to [7], a pixel's saliency is determined by the combination of its global and local saliency information, which could contribute to background reduction.

2.2 Center-Surround Contrast Computation and Final Saliency Map Construction

In the previous section we introduce an isophote based operator to extract potential structure and global saliency information related to each pixel. In this section, an integral image based operator is employed to compute center-surround contrast for all three feature channels that are then fused into a final saliency map. The overall process proposed to detect image saliency is shown in Fig. 2: an image is first converted into three sub-images based on CIELAB color space; the isophote based operator is then employed to determine potential structures and global saliency information for each sub-image. These two steps are described in the following paragraphs. The potential structures are used to compute center-surround contrast which is then combined with global saliency to build a final saliency map.

1) Given an image I, we first convert it into sub-images based on CIELAB color space: $I \rightarrow \{\mathrm{L, a, b}\}$. The color space has the dimension 'L' for luminosity, 'a' for the variation from red to green, and 'b' for the variation from blue to yellow. The reason we chose CIELAB color space is its perceptually uniformity, which means a change in a color value



$$I_x^{'} = \frac{S(p_0, p_1, p_4, p_3) - S(p_1, p_2, p_5, p_4)}{2S_xS_y} \quad I_y^{'} = \frac{S(p_6, p_7, p_8, p_{11}) - S(p_{11}, p_8, p_9, p_{10})}{2S_xS_y} \quad R_{ori} = \sqrt{(I_x^{'c} - I_x^{'s})^2 + (I_y^{'c} - I_y^{'s})^2}$$

Fig. 3 Orientation center-surround contrast computation. $S(.)$ represents mean within a region, and r is the orientation contrast between center and surrounding regions.

is perceived as approximately a change of the same amount in the human visual perception system.

2) For each sub-image $I_{\text{sub\_image}} \in$ {L, a, b}, the isophote based operator is adopted to extract potential structure and global saliency information S_{g} related to each pixel respectively. The potential structure is then used to determine the center and surrounding regions for each pixel as mentioned in Sect. 2.1. For the intensity and color feature channels, an integral image mechanism is directly employed to compute center-surround contrast. To extract the most salient pixels, the global saliency information S_{g} is used as a weighting for the center-surround contrast computation:

$$R_{\text{Int}}(x,y) = S_{\text{g}}^{\text{L}}(x,y) \cdot |r_{\text{center}}^{\text{L}}(x,y) - r_{\text{surround}}^{\text{L}}(x,y)| \tag{6}$$

$$R_{\text{Col}}(x,y) = \frac{1}{2} \sum_{\eta=\{\text{a,b}\}} S_{\text{g}}^{\eta}(x,y) \cdot |r_{\text{center}}^{\eta}(x,y) - r_{\text{surround}}^{\eta}(x,y)|$$

where R_{Int} and R_{Col} represent intensity and color feature channel maps respectively; (x, y) is a Cartesian coordinate in the image plane; S_{g} is the global saliency information computed by isophote based operator at the position (x, y); and r_{center} and r_{surround} represent the average of center and surround regions around the position (x, y) within $I_{\text{sub\_image}} \in$ {L, a, b} respectively. The orientation computation method proposed by [9] is based on an integral image strategy, and it is naturally integrated into our model to compute orientation center-surround contrast. Because all three feature channel maps can be computed by the integral image based operator, it is conducive to the fusion of all three feature channel maps. Instead of computing angle and magnitude [9], we compute orientation contrast between the center and surrounding regions by the Euclidean distance between their orientation vectors. The orientation feature map is obtained by:

$$R_{\text{Ori}}(x,y) = \frac{1}{3} \sum_{\eta=\{\text{L,a,b}\}} S_{\text{g}}^{\eta}(x,y) \cdot r^{\eta}(x,y) \tag{7}$$

where R_{Ori} is the orientation feature map; r stands for the orientation center-surround contrast at position (x, y), and its meaning is illustrated in Fig. 3. The final saliency map is a linear combination of all three feature maps:

$$S_{\text{Final}} = \frac{1}{3}(R_{\text{Int}} + R_{\text{Col}} + R_{\text{Ori}}) \tag{8}$$

Figure 4 shows examples of the combination of global and local saliency information for building the final saliency map. As illuminated, the combination of global and local saliency information can effectively detect the most salient pixels.

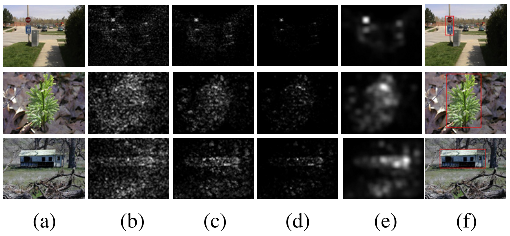


Fig. 4 Examples of the combination of global and local saliency information: (a) inputs, (b) global saliency maps, (c) local saliency maps, (d) final saliency maps combined by global and local saliency information, (e) smoothed saliency maps with Gaussian filter, and (f) ground truths.

3. Experiments

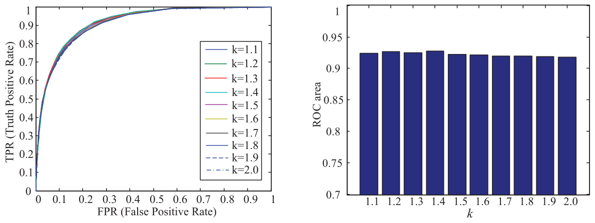
3.1 Database and Parameter Setting

We evaluated the proposed model on the publicly available database provided by Achanta et al. [6], which includes 1000 images, and has ground truths in the form of accurate human-marked labels for saliency regions. The parameter k is used to determine the size of outer regions for computing center-surround contrast. As previously mentioned, a potential structure extracted by an isophote based operator would only belong to part of an object, thus it is not necessary to set an overly large size to outer region. First of all, the range of values of k was limited to [1.1, 2.0]. For 100 images randomly selected from the database, values of k in steps of 0.1 were used to build saliency maps. The measurement used in [6] was employed to evaluate the performance. The threshold value of what was varied from 0 to 255. For each value, a binary mask was obtained that was used to compute the True Positive Rate (TPR) and the False Positive Rate (FPR) against the ground truth. The resulting Receiver Operating Characteristic (ROC) is shown in Fig. 5. As illustrated, all values of k delivered similar performance. Based on the experiments, k was arbitrarily set to 1.4 in the following experiments.

3.2 Performance Evaluation

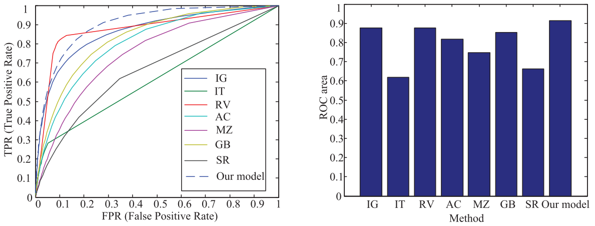
We compared the proposed model to seven state-of-the-art saliency detection models, which include IT [1], AC [5], IG [6], RV [7], SR [8], MZ [10], and GB [11]. Following [6], two measurements were employed to evaluate the performance: segmentation by fixed thresholding and segmentation by adaptive thresholding, which we now describe.

Segmentation by fixed thresholding. This measure has been used in Sect. 3.1, and the comparison is shown in Fig. 6. The figure clearly shows that the proposed model outperforms the other seven models. It is interesting to note that the model, RV, proposed by Valenti et al. [7] shows very



(a) (b)

Fig. 5 The comparison of different saliency maps by using different values of multiplicative factor k: (a) is ROC curvature, and (b) is ROC areas.



(a) (b)

Fig. 6 The comparison of segmentation by fixed thresholding: (a) is ROC curvature, and (b) is ROC area.

Table 1 Precision-Recall-F-measure results for segmentation by adaptive thresholding.

	F-measure	Recall	Precision
IG	0.6417	0.7950	0.6213
IT	0.6017	0.6622	0.6253
RV	0.6334	0.8513	0.5870
AC	0.5681	0.7373	0.5428
MZ	0.5249	0.6964	0.4950
GB	0.5716	0.7675	0.5328
SR	0.4900	0.6558	0.4633
Our model	**0.6967**	**0.8566**	**0.6654**

high accuracy for a very low FPR, but the growth of accuracy is slower than our model (Fig. 6 (a)). This might be because that the RV is easily affected by complicated backgrounds.

Segmentation by adaptive thresholding. Achanta et al. [6] introduce an adaptive threshold method that has twice the mean saliency of an input saliency map. The images are segmented using a mean-shift segmentation algorithm, and retain only those regions whose average saliency is greater than the threshold. We replaced the mean-shift segmentation algorithm by a graph-based segmentation algorithm [12]. Compared to the mean-shift algorithm, the graph-based segmentation algorithm can provide larger and more uniform segmentations. After segmentation, the Precision, Recall and F-measure (Eq. 9) are obtained over the ground truth.

$$F = \frac{(1+\alpha)Precision \times Recall}{\alpha \times Precision + Recall} \tag{9}$$

where α is set to 0.5. The results are listed in Table 1. RV [7] shows a high recall but low precision, indicating that the foreground regions obtained by RV include many background pixels, which is consistent with the results in the above measurement. Among all models, the proposed model shows the highest performance. Figure 7 shows examples of saliency and segmentation maps constructed by our model.

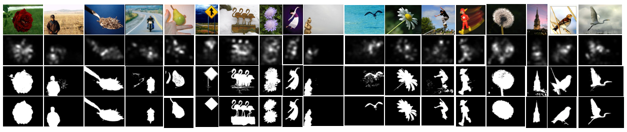


Fig. 7 Examples of saliency and segmentation maps constructed by our model. From top to down: input images, saliency maps, segmentation results and ground truth.

4. Conclusions

In this paper, we propose an isophote based operator to detect the salient pixels in images. The operator guarantees that most of the significant pixels can be detect, and the combination of global and local saliency information can filter out most of the background pixels. Additionally, the integral image mechanism ensures a consistent computation for all three feature channels, which is conducive to the fusion of all feature channel maps. The experiments have shown that the proposed model has reliable performance for a wide range of images.

Acknowledgements

This research is supported by the Ministry of Industry and Information Technology of China (No.2010ZX01042-002-003-001), the Natural Science Foundation of China (No. 60703040), Science and Technology Department of Zhejiang Province (Nos. 2007C13019, 2011C13042).

References

[1] L. Itti, C. Koch, and E. Neibur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.11, pp.1254–1259, 1998.

[2] A. Kimura, R. Yonetani, and T. Hirayama, "Computational models of human visual attention and their implementations: A Survey," Proc. IEICE Trans. Inf. & Syst., vol.E96-D, no.3, pp.562–578, March 2013.

[3] S. Frintrop, M. Klodt, and E. Rome, "A real-time visual attention system using integral image," Proc. ICVS, 2007.

[4] P. Viola and M. Jones, "Robust real-time object detection," Proc. IJCV, vol.57, no.2, pp.137–154, 2004.

[5] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," Proc. ICCVS, pp.66–75, 2008.

[6] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," Proc. CVPR, pp.1957–1604, 2009.

[7] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," Proc. ICCV, pp.2185–2192, 2009.

[8] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," Proc. CVPR, pp.1–8, 2007.

[9] D.A. Klein and A.B. Cremers, "Boosting scalable gradient features for adaptive real-time tracking," Proc. ICRA, pp.4411–4416, 2011.

[10] Y.F. Ma and H.J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," Proc. ACM MM, pp.374–381, 2003.

[11] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," Proc. NIPS, pp.545–552, 2007.

[12] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient graph-based image segmentation," Proc. IJCV, vol.59, no.2, pp.167–181, 2004.